



Firm location, interaction, and local characteristics: A case study for Madrid's electronics sector*

Miguel Gómez-Antonio¹, Stuart Sweeney²

¹ Department of Public Finance and Tax System, Complutense University of Madrid, Somosaguas Campus, 28223 Pozuelo de Alarcón, Madrid, Spain (e-mail: mgomezan@ucm.es)

² Department of Geography, University of California, Santa Barbara, CA 93106-4060; Institute for Social, Behavioral, and Economic Research, UC Santa Barbara, USA (e-mail: stuart.sweeney@ucsb.edu)

Received: 30 September 2015 / Accepted: 22 September 2016

Abstract. We use geo-referenced establishment data to estimate parameters of a Gibbs model. The statistical model is used to decompose the conditional intensity of the spatial point process into trend and interaction components. The trend captures covariates related to firms' costs, accessibility to public transport infrastructures, to technical universities and to cultural and recreational facilities. The ability to specify a Geyer interaction component captures the existence of additional spillovers providing a deeper insight into inter-establishment spatial dynamics. The results challenge some of the outcomes of the inter-urban industrial location literature, confirming that spatial aggregation compromises results in studies of business location. Firms' location decisions are dominated by site costs, with transportation costs being much less of a consideration; universities geographical knowledge spillovers are confirmed for large establishments; and amenities are unlikely to be important location factors.

JEL classification: C49, C51, H41, H54, R38

Key words: Industrial location, Gibbs models, agglomeration economies, local public goods

1 Introduction

Empirical research on intra-metropolitan firm location has been hindered by the lack of appropriate models and disaggregated data. Most prior work in the industrial location literature uses areal measures of location based on administrative boundaries which are exogenously and arbitrarily defined. In this paper we use the geo-coded point location of the firm as the outcome variable and covariates and firm interaction measures are defined on continuous geographic

* We would like to thank the participants in the 6th Jean Paenlick workshop and three anonymous referees for their valuable comments and suggestions. Miguel Gómez-Antonio acknowledges financial support from the ECO2015-65758-P research funding and Stuart Sweeney acknowledges support from National Science Foundation award BCS-0454993 and from the ECO2015-65758-P research funding.

space. As a consequence results do not embed arbitrary scale-effects that may be present in models based on areal data.

The literature using firm geo-referenced data dates back to Barff (1987). This seminal paper measures the absolute concentration of manufacturing firms assuming that the firm's distribution is stationary in the study area. To circumvent this assumption, several distance-based approaches have been empirically tested in the presence of spatial inhomogeneity.¹ These approaches can be broadly grouped into four categories: the D function (Diggle and Chetwynd 1991), the Duranton and Overman (D-O) approach (Duranton and Overman 2005), the M function (Marcon and Puech 2010), and the inhomogeneous K function (Baddeley et al. 2000). All of these methods are grounded in a case-control strategy that consists of selecting a group of controls that account for the observed inhomogeneity, and comparing its spatial distribution with one of the selected cases. These approaches are very accurate at detecting and identifying the relevant distance for the cluster's existence. However, they are purely descriptive and are silent about its determinants. Gibbs models for point processes (or Markov point processes) provide a regression framework that constitutes a fruitful approach to empirically exploring a cluster's determinants. We model firm location decisions as a spatial point process that depends on local environmental characteristics, such as proximity to local public goods, and on the existence of interaction among firms.

A sound economic model of location decisions based on point-referenced firm data can improve our knowledge in several ways. The main advantage of a Gibbs model specification in analysing industrial location choices is that they have the potential to distinguish spatial concentrations due to exogenously given endowments (land-use regulations, proximity to airports, to major infrastructures or to public amenities) from agglomeration of essentially footloose firms attracted to each other by various types of externalities. Co-location may occur without linkages or interaction between proximate firms. A sector's propensity to agglomerate may be due to the spatial concentration of immobile resources that are important to that sector, or it may be determined by pecuniary and/or technological spillovers. First nature geography echoes comparative advantages provided by certain locations creating incentives for firms and industries to cluster around focal points of interest. As a consequence of intense interactions between producers at the same location, urbanization and localization economies eventually arise and generate additional benefits derived from second nature geography (Jacobs 1969). Gibbs models separate the factors affecting the location choice into those that are related to the environment and those related to true interaction between firms. Assessing the significance and magnitude of certain types of local public goods on firm location choices can also be facilitated with Gibbs models. Spatial aspects of local public goods are seldom considered in the existing literature because of the analytical complexity involved in measuring their impact. Most of the studies in which the implications of spatial variation in public service provision are considered rely on the assumption that public services are smoothly 'diffused' over exogenous jurisdictional boundaries, rather than being provided at just a few discrete locations in an urban area. However, the distribution of public activities can have a substantial impact on private sector locational choices, as benefits from public services may not accrue by the same amount to all residents in a region. Additionally, there have been conceptual problems in the literature associated with the use of expenditure data as a measure of public services because variation in expenditure does not necessarily imply variation in the quality of services. Gibbs models solve this problem by leveraging information on the point locations of establishments and public goods to reveal how local public goods affect the urban structure, and more specifically, the location of firms.

¹ Economic activities are not homogeneously distributed because of rivers, mountains, or specific town regulations (parks, pure residential zones, etc.).

Details related to Gibbs model statistical interpretation in applications to firm location are provided in Sweeney and Gómez-Antonio (2015). It demonstrates that Gibb's models can be fruitfully employed to disentangle trend from interaction in industrial location modelling. It provides considerable depth on the theory of Gibb's models, relate them to prior work on industrial location and industry clustering, and provide economic intuition behind alternative interaction specifications. It includes an empirical application, but it is a minimal model specification used in a pedagogic capacity to provide an example of alternative model specification diagnostics and to give a sense of model interpretation. In this paper we provide a detailed description of the geographic concentration of Madrid's electronic manufacturing industries. We estimate a fully specified model of locational factors in the electronics sector after accounting for all of the variables previously used in the literature and emphasizing the role of local public goods in locational choices.

The aim of the paper is twofold. First, we test if the results obtained from area-based models are compromised once we implement techniques that do not rely on administrative boundary datasets. Second, we demonstrate the potential of Gibb's model for detecting certain phenomena that cannot be detected under traditional methodologies. Gibb's models may be able to detect and quantify certain type of knowledge spillovers. The geographical scope of agglomeration effects is typically unknown and is unlikely to align with the scope imposed by administrative boundaries. The strength of different agglomeration mechanisms may differ at different geographical scales. Knowledge spillovers might operate via face to face contacts, so its geographical scope could be very limited and even the municipality may not be an appropriate geographical unit to capture these effects (Jofre-Montseny et al. 2011). There is evidence suggesting that some agglomeration effects like knowledge spillovers decay very sharply with distance making it difficult to test for their existence. For analysing the role that knowledge spillovers have on firms' location we need tools that are not constrained by the arbitrary delimitation of space. The effects of knowledge spillovers might be limited to a certain zone inside the urban area, the municipality or the city administrative boundary and when using area based data it is not possible to measure and test for their existence.

In this paper we use Gibbs models to analyse the distribution of firms in the electronics industry within the greater Madrid area. We framework a theoretical model that includes the main potentially important factors described in the literature to affect manufacturing location decisions. Our analysis takes the Madrid electronics industry as its focus for several reasons. First, by concentrating on a single narrowly defined industry the problem of unobserved heterogeneity is reduced. Second, regional growth theories predict that clustering will be particularly strong among high-tech or knowledge-intensive industries. Third, it is a strategic industry in the region. And fourth, several studies using different methodologies conclude that there is a high degree of inter-firm interaction.

The following section briefly describes the proposed modelling framework. Section 3 introduces the theoretical model. Section 4 describes the data and the process utilised to construct the covariates, and Section 5 discusses the main results obtained. Concluding remarks are presented in Section 6.

2 Modelling firm location

Since Carlton (1983), it has become standard practice to estimate firm location choices through the conditional logit model. Alternatively, Papke (1991) suggested that location choice could be represented by a region-level count model, such that estimation is based on maximum likelihood with an assumed Poisson distribution. The basic econometric framework for empirical studies has essentially remained unaltered over the past four decades with different versions of either discrete choice or count data models. In all location choice models the unit of analysis is an administrative metropolitan area or county, for the US, or region or municipality (NUTS 2 or 3)

for Europe.² Nevertheless spatial point processes can be used directly, to model and analyse data which take the form of a spatial point pattern, such as firm locations derived from geo-referenced addresses. The spatial distribution of firms is then considered to have arisen from a stochastic point process, with the observed distribution of firms being a single realization.

Below we work with Gibbs (Markov) processes, X , that can be expressed as exponential family densities and allow for separate estimation of effect sizes on components of the trend ('first-order effects') and a specific representation of the interaction ('second-order effects'). Details on the Gibbs process formulation can be found in Møller and Waagepetersen (2007) with extensions and interpretation for the firm location choice in Sweeney and Gómez-Antonio (2015).

A general expression of an in-homogenous Gibbs point process density (with respect to Poisson unit density) is:

$$f(x) = a \prod_{y \subseteq x} b(y) \gamma(y) \quad (1)$$

with normalizing constant a , localized shift $b(y)$ measuring the density at location y , and the function $\gamma(y)$ measuring interaction among points (in pairs, triples, or higher orders); the product is over all subsets of x .

The trend component $b(y)$ depends only on the spatial location u , and reflects spatial inhomogeneity that affects the location decision of firms. While y reflects the points in the observation window where a firm is observed, u echoes all possible locations in the observation window $b(y)$ captures aspects related to the natural or built environments that are likely to impact a firm's profit function, such as accessibility, proximity to raw materials, proximity to infrastructure, etc.

$$b(y) = a + b_1 u_1 + b_2 u_2 + \dots + b_n u_n = z(u) b^T \quad (2)$$

There are several alternative specifications of the interaction $\gamma(y)$ that might be applied to firm location modelling. Sweeney and Gómez-Antonio (2015) evaluates several, provide economic intuition underlying each, and found that Geyer saturation interaction yielded the best diagnostics. Geyer interaction is specified as:

$$\gamma(y) = \gamma^{\min\{d, N_X(u)\}} \quad (3)$$

The interaction term is a function of the number of R close neighbours $N_X(u)$, but is bounded above by the scalar distance, d , so that the contribution of any point to total interaction is restricted. The saturation term, d , limits the amount of localization benefit that a firm can accrue from having other firms located nearby. In practice, this means that diminishing returns arrive all at once after a ceiling is reached. If the saturation parameter is 3, then the model is stipulating that there are localization benefits related to 1, 2 and 3 'close neighbours' with marginal return proportional to γ , but after the saturation threshold any additional firms locating within r offer no additional benefit. Without that restriction the process simplifies to a Strauss process which is known to yield only 'single ball' clusters when $\gamma > 1$ (Møller and Waagepetersen 2007). But with that constraint the resulting interaction is capable of generating moderate positive interaction that is characteristic of firm location patterns.³

² An excellent review on the administrative area based models literature can be found in Arauzo et al. (2010).

³ In the spatial statistics literature, Gibbs models have typically been employed to model processes of inhibition between events, while other types of model such as Cox models or cluster Poisson models are supposedly more appropriate to model attraction (clusters). Nevertheless the types of cluster process that might be found in other fields such as biology or physics are very different to the cluster structure that might present the spatial distribution of firms. In those fields cluster structure is characterized by a large concentration of events very close to one another, while in the establishments' patterns we assume a moderate level of clusters; Gibbs models are consequently more suitable. Another reason to estimate a Gibbs model is that we also need to account for spatial inhomogeneity, which is not considered in Poisson cluster type models.

The Geyer interaction term is central to the analysis because, when significant, it confirms the existence of attraction among proximate firms in the sector. If firm-to-firm interaction is not relevant in the industry, the estimated γ coefficient in (3) will be zero, implying that having another firm in the neighbourhood is not a relevant aspect in firm's location choice. The firm-to-firm interaction structure relates to theory as implementing a specific functional form of proclivity for firms to co-locate, after controlling for features modelled in the trend. Firm's interaction could be asymmetric depending on firm's size. To determine and gain deeper insight about the firm's interaction, a natural extension of the Gibb's model approach is to include marks in the analysis such as the size of the firm or the sector to which the firm belongs (see Högmänder and Särkkä 1999).

The normalizing constant, a , in (1) makes it difficult to work directly with the density, and the model is made tractable by working instead with the Papangelou conditional intensity function:

$$\lambda(u, X) = b(u)\gamma^{\min\{d, N_X(u)\}} \quad (4)$$

Under this reformulation, and looking toward the model interpretation, $\lambda(u, X)du$ is the probability of observing a point u of the process in a small neighbourhood du of u , conditional upon the rest of the process X .

2.1 Estimation

Because of the exponential form of Gibbs model, standard software implementations for generalized linear (additive) models can be used to estimate parameters of the conditional intensity function. Expanding the right hand side in log-linear form, the model is decoupled into two components:

$$\lambda(u, x) = \exp\{\phi^T b(u) + \vartheta^T S(u, x)\} \quad (5)$$

On the right hand side, the first term is the spatial trend component and the second term is the interaction component. Canonical parameters $\theta = \langle \phi, \vartheta \rangle$ are the focus of estimation. Note that the trend component depends only on the spatial location u , and reflects spatial in-homogeneity that affects the location decision of firms. History might affect certain locations' population density, public service endowments and geographical characteristics also play an important role for certain industries. For instance, the proximity to natural resources will be important for those firms in the aquaculture industry that need to be close to a river. Thus the factors affecting inhomogeneity might depend on the analysed industry. Two groups of covariates are introduced in this part of the model, either related to firms' costs (land prices, population density, distance from distribution hubs, distance from CBD, taxation variables), or to the distance from certain local public goods (public transportation infrastructures, geographical knowledge spillovers and cultural and recreational facilities).

Estimation of the canonical parameters is based on the Huang and Ogata (1999) maximum likelihood as encoded in the R package *Spatstat* (Baddeley and Turner 2005). That method also provides an estimate of the asymptotic variance-covariance matrix for the canonical parameters. Standard errors and significance tests for those parameters are based on that covariance estimate. The interaction component contains two irregular parameters, the range of interaction, R , and the threshold capping interaction at d . Profile likelihood is used to estimate the irregular parameters.⁴

⁴ When estimating the model with a Geyer interaction specification, two irregular parameters need to be estimated: the saturation threshold, and the interaction radius of influence. We use a small set of integer values for the saturation parameter (1 to 8) and interaction radius (2 to 3) and select the combination of these that maximizes the profile pseudo likelihood.

2.2 Diagnostics

In addition to familiar specification tests of individual canonical parameters or groups of parameters using the estimated covariance, the complexity introduced with separate trend and interaction components requires a more varied set of diagnostic and specification tools. Baddeley et al. (2005, 2012) have extended diagnostics designed for GLMs to the point process setting. That translation is made possible by defining (pseudo-) innovations and (pseudo-) residuals for spatial point processes; accomplished essentially by defining the difference in observed points and expected points in subspace balls of the study region. In practice those balls are defined as grids in the study region.

Diagnostic measures and tests can be defined to isolate the adequacy of fit in the trend, the interaction, or the overall model goodness of fit like the Akaike information criteria (AIC).⁵ The diagnostics for trend include lurking variable plots and contour plots of smoothed residuals. Lurking variable plots indicate the cumulative error, measured as exponential energy weights in this case, over the domain of a covariate in the trend term, $b(u)$. The lurking variable plot investigates the presence of spatial trends in point processes. The residuals are plotted against a spatial covariate which must be spatially continuous in the study region. We evaluate the residual measure on each sub-region, yielding a cumulative residual function that should be approximately zero if the fitted model is correct. Any systematic pattern in the lurking variable plot suggests an appropriate modification of the model to better account for that spatial covariate. The plots contain the expected value under the ‘null’ (which predicts that the fitted trend is correct) and the point-wise confidence envelopes (± 2 standard deviations). When the cumulative Pearson residuals are larger than $+2$ sigma at a certain distance, this diagnostic suggests that the null model underestimates the number of firms at this scale. In other words, there are more firms occurring at those spatial locations than the null model predicts. If the curve is below the -2 sigma envelopes, then the null model overestimates the intensity of firms at those locations. Smoothed residuals are used to identify areas in the overall fitted trend that either over or under predict; a well-fitting model would have a spatially mixed pattern of slight under and over fitted akin to points around a trend line in a scatter plot.

The fit and validation of the interaction term is assessed using the QQ-plot and the G-compensator. The QQ-plot compares quintiles of the smoothed residuals to expected quintiles of residuals under the fitted model. If the interaction component is captured by the model then the observed and expected quintiles will be within the error bounds provided in the plots. The G-compensator is an analog of a score test (Baddeley et al. 2005). It compares a nearest-neighbour distribution measure, the G-function, of the observed data and the expected value under the fitted model. Again, error bounds are provided and if the fitted interaction term is well-matched to the underlying process of the data, the resulting G-compensator measure will fall within the error bounds.

3 The empirical model

The estimated Gibbs model is formally derived from a representative firm’s random profit maximization framework, where location choices are made in two sequential stages. The first stage consists of selecting a broad region focusing on wages, market potential, taxes, climate and other features that may have significant interregional variation, but are similar everywhere

⁵ All of the methods we use are available in the R software using the *Spatstat* package. Baddeley et al. (2005, 2012) contain details of the theory underlying the use of pseudo-residuals.

within the region. The second stage considers locational factors that vary at the micro-geographic level. Our empirical model focuses on the second stage of the decision.

The intra-metropolitan location decision is one of cost minimization. We assume that differences in intra-metropolitan location affect firms' profits only through variation in their costs between locations.⁶ The cost function per unit of output has been represented in the literature as the function⁷:

$$C = F(PL, PN, S, t, G, AG) \quad (6)$$

Where PL and PN are effective input prices of industrial land and labour,⁸ S is a vector of general site characteristics, t is the effective property tax rate, AG captures the existence of agglomeration economies and G represents public services in the area.

The process of a firm selecting the location site is equivalent to modelling the conditional intensity of the Gibbs process defined in Section 2. Remember that the conditional intensity function models the probability of observing a firm in a small neighbourhood du of u , conditional upon knowing the location of the rest of the firms. The Papangelou conditional intensity of the process is then defined as:

$$\ln(\lambda(u, x)) = \alpha_1 PL + \theta' S + \alpha_2 AG + \alpha_3 t + \beta' G + \alpha_4 Interaction \quad (7)$$

The probability of a firm selecting a particular location, u , depends on land prices, on the spatial concentration of immobile resources that are important to the sector, on the existence of agglomeration forces, on the role of government at the site through taxes and local public good endowments, and on the existence of others firms with whom the firm could interact.

The expected sign of industrial land price (PL) is negative; however, high-tech firms might prefer to locate in areas where land price is high, as land price might capitalize the value of local public services. Two covariates related to general site characteristics (S) were included: the covariate distance to CBD that captures the importance of proximity to the central city markets for labour and goods,⁹ and the distance to logistic infrastructures, testing the importance of this type of infrastructure in the location choice.

The model specification allows for capturing sources of agglomeration economies (AG) both in the trend component and in the interaction component. The covariate population density evaluates the importance of access to a larger labour and output market seizing the importance of urbanization economies, hence a positive sign is expected. We could develop additional covariates in the model to capture and detangle these agglomeration forces but the focus of the paper is not analysing the different sources of Marshallian agglomeration economies, but is to test for the existence of interaction among firms.

Governments can attract business by manipulating fiscal policies (t) and public services (G) providing a profitable environment for firms. The location literature establishes that fiscal policy constitutes an important covariate for location choices. In the US economy, state and local taxes

⁶ Consumer demand for output is assumed not to vary within intra-metropolitan locations.

⁷ These covariates are in line with the literature (see Brühlhart et al. 2012). The distinctive feature of our empirical model lies on the estimation technique. As we treat space as continuous municipality fixed effects can be ignored and single sector industry fixed effects are not needed due to the existence of only one sector in the analysis. We replace public sector spending covariates by distance to local public goods and eliminate some variables that seek to control for the size of the municipalities that are redundant.

⁸ Wage rates figures could not be included as they were not available for individual communities of metropolitan areas. Even if computed, they would be more a measure of the employment mix in the area than a measure of the wage rate that a firm in a particular industry must pay. The same problem was present for other covariates such as population density, industrial land price, and crime. Industrial land prices were available disaggregated by street and block, but they are not geo-referenced and the process of geo-referencing them and preparing them would constitute a highly time consuming and laborious task that goes beyond the scope of this paper.

⁹ It will allow the hypothesis of the mono-centric or poly-centric central business districts to be checked.

have an important effect on business location, particularly within metropolitan areas where business property taxes can vary substantially between jurisdictions. Accordingly, in the theoretical model local governments can attract business by manipulating fiscal policies and public services. However, a significant relationship between taxation variables and industrial location is not expected for Spain as main taxes are centralized in the national government. Although local taxes can differ among local governments they do not represent an important tax burden for firms. Therefore, local governments can only provide a profitable environment for firms through public services.¹⁰ Nevertheless, due to the novelty of the Gibbs model approach we considered more pedagogical including the fiscal policy covariate and hypothesize that it should not be significant rather than removing it.¹¹ From all the local taxes levied on firms, we decided to include in the analysis the property tax (IBI) because it is the only one that presents spatial variation within the Madrid metropolitan area.

Government services in the area are related to the consumption of local public goods which are specific to a given geographical area. To derive utility from the public good, a firm must locate within sufficient proximity to enjoy the benefits since utility decreases with distance. The role of three local public goods, where distance is an important component in their consumption are included in our model: transportation access, access to knowledge spillovers derived from technical universities and access to recreational or cultural public goods.

Access to transportation is a local public good that represents an amenity for both firms and workers. The effect of public transport infrastructures has been extensively studied and it differs across manufacturing industries, indicating that accessibility requirements may vary with technology and/or demand. The covariates measuring access to main roads and proximity to the airport capture the benefits that firms enjoy in terms of reduced time and transportation cost due to congestion. On the other hand, proximity to public transportation systems such as the metro, measured as distance to local train station, is a local public good that can be considered a worker amenity. Access to these transportation systems represents a reduction in travel time, and an increase in comfort, due to their safety and frequency of service. Traffic congestion represents one of the main negative externalities in big cities, with Madrid being no exception.

The public good nature of knowledge arises from non-proprietary university research. The role of these institutions centres on the existence of knowledge or technology spillover effects and the extent to which their diffusion is facilitated by geographical proximity. Proximity facilitates the diffusion of novel knowledge arising from university research, determines a lower cost of acquiring technological spillovers, easier access to faculty consultants and student interns, and it also enables the continuing education of employees. Of the Marshallian externalities that motivate the literature on agglomeration economies, knowledge spillovers have proven to be the hardest to verify empirically. There is mounting evidence for the attenuation of human capital spillovers at small spatial scales (Carlino et al. 2011; Wang et al. 2012). Gibbs models constitute an appropriate tool to detect knowledge spillovers arising from the existence of technical universities in the area.

Firms employing skilled professionals are supposed to be amenity oriented when selecting sites within metropolitan areas. The role of cities as consumption centres started drawing attention recently. The 'consumer city' literature emphasizes urban amenities as the centripetal force attracting workers into cities. Restaurants and theatres and an attractive mix of social partners are local goods. In order to attract highly skilled workers, cities must cultivate places that offer the qualities that people want in a place to live and work (Glaeser et al. 2001; Currid and

¹⁰ The size of Spanish municipal governments is moderate, with municipal budgets representing 15 per cent of total public spending. Local taxes represent a third of local budgets, inter-governmental grants represent another third, and the remainder comprises user charges.

¹¹ We would like to see these models replicated for metropolitan areas of federal countries where main taxes are decentralized and data are available.

Williams 2010; Lee 2010). High-tech firms directly take into account the locational preferences of their potential professional employees, despite other locational influences arising from organizational purposes. The quality of places attracts creative people, and their presence attracts high-tech and cultural industries. This type of clustering pattern has been documented in the analysis of the more formal aspects of industry production and firm location (Molotch 1996; Currid and Connelly 2008). We test if urban amenities drive intra urban location choice. 'Cultural and recreational opportunities' is considered a local public good and three indicators are included in the model: a district's green area extent; distance from the nearest movie theatre¹²; and violent crime, considered as one disamenity that influences firm location when evaluated at the worksite itself.

To conclude, as stated in previous paragraphs, the ability to specify a mathematical Geyer interaction component is the main strength of a Gibbs model specification. If we assume that the first order effect is specified approximately correctly – in the sense that in all econometric modelling we strive to include the important covariates – the Gibbs model approach helps to solve the simultaneity problem, isolating in the interaction component the advantages firms obtain from locating close to firms within the same industry. Empirical evidence of the positive effect that agglomeration economies have on location decisions are numerous, however none of the approaches implemented so far have attempted to determine if conditional on the covariates related with the site environmental characteristics there exists significant interaction between the firms. The interaction term $\gamma(y)$ might capture advantages related to the use of the same type of workers, to have a customer–supplier relationship, and to the use of the same new technologies. Given that electronics is a high-tech sector we assume that the benefits of proximity are predominantly driven by knowledge spillovers. Nevertheless in other sectors the interaction term might capture other agglomeration sources like labour pooling or input sharing. Additional research will be needed to specify which source of agglomeration economies is behind the interaction, but what we confirm is that when the parameter γ is significant it captures true interaction between firms.

The reduced form to be estimated is the following:

$$\ln(\lambda(u, x)) = \alpha_0 + \alpha_1 PL + \alpha_{2,3} S + \alpha_4 AG + \alpha_5 t + \alpha_{6-12} GTransp + \alpha_{13} Gknow + \alpha_{14-16} Glife + \alpha_{17} Interaction \quad (8)$$

4 Data

The methods proposed here are intended to provide estimates of parametric models of spatial stochastic processes that are consistent with currently observed patterns of industry. Our modelling framework relies on cross-sectional analysis as a reduced form, abstracting from the dynamics, to test for significant association between covariates and firms' locations. We can test the magnitude and the sign (direction) of the covariates but we cannot say anything about causality.¹³ The data sample covers the stock of established firms because when we considered estimating a firm birth model several motivations deterred us from using flow data. First, we had to define the appropriate time intervals to determine what constitutes a new firm to introduce arbitrariness in the analysis. When we narrowed the time interval to decades, the sample of firms was notably reduced, which imposed important data constraints. In addition to estimating the model for different sub-periods, we also needed to construct distance

¹² Theatres are not confined to be within the agglomeration, and are not equally scattered within it, so these might act as a good proxy for the cultural activity of the area.

¹³ There are no other competing methods that exist that could provide a causal answer.

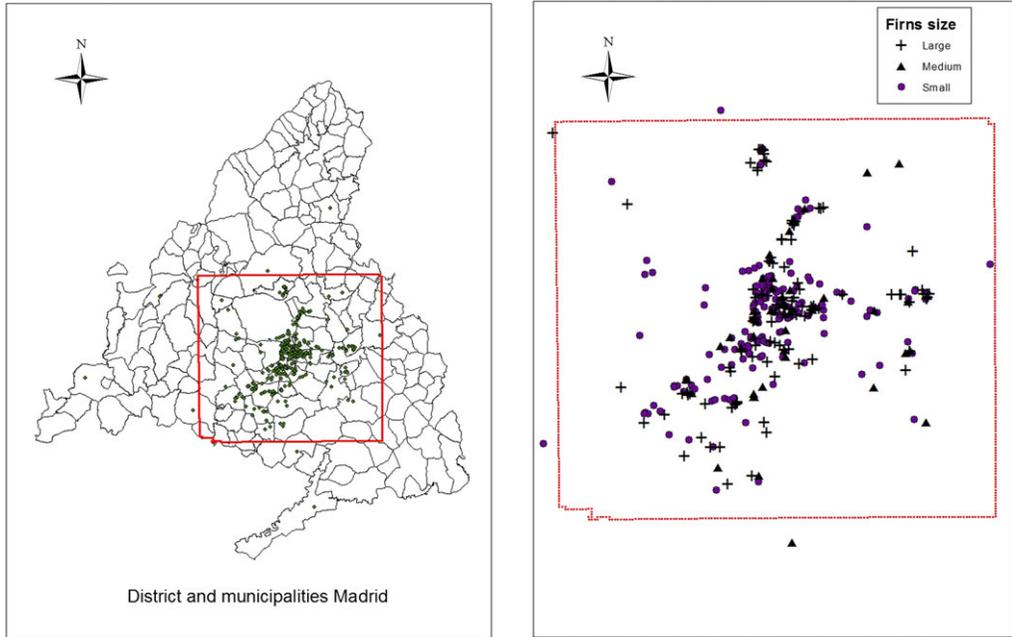


Fig. 1. Districts and municipalities Madrid, observation window

covariates for every sub-period, which causes model estimation to be very data-intensive. Because facilities' location (roads, train, subway network, etc.) changes over time.

The data we use to define our outcome and most of our covariates are from the Madrid Regional Statistic Institute. Our data includes information on establishments in the electronics industry (latitude/longitude coordinates based on geo-coded street addresses and the number of employees) and several vector-representation GIS files containing region boundaries as well as road and rail networks. All data are from the year 2002 and the industry definition is based on the two digits National System of Economic Activities (CNAE93) classification: industry 32 'Manufacture of electronic equipment, manufacturing equipment and radio, television and communication devices'.

A depiction of the industry locations and the region boundary are provided in Figure 1. To ease the computational burden of model fitting we use a simplified regional boundary that captures the greatest density of establishments surrounding Madrid's capital.

There is no agreement between countries and industries as to whether there are marked differences in location patterns between small and large plants.¹⁴ Location factors depend on the type of industry and on the size of the firm. Whether a firm's size affects its location preferences or site characteristics is assessed by dividing the sample into large, medium and small firms according to the following classes: 1–4 employees, 5–9 employees, and 20 or more employees. According to the history of the sector and the number of firms in each subsample this seems an appropriate delimitation. The origin of many firms in the sector is a consequence of employment reductions in large firms that led to a surge of new start-ups owned by highly skilled and specialized labour that complemented the activities of the larger firms. Additionally, after the

¹⁴ In the literature some papers find localization economies associated with smaller establishments. Glaeser and Kerr (2009) and Rosenthal and Strange (2001) consider the idea that agglomeration economies are stronger when there are many small firms. Others report a relationship resembling an inverted "U" (Sweeney and Feser 1998), and many studies suggest the size-agglomeration relationship depends on characteristics of an industry or methodology (Duranton and Overman 2008).

liberalization of the European communication space in 1988 entry costs lowered and the number of smaller suppliers in the market increased.

Several covariates are defined as measuring the shortest distance from a point, u , to a feature of the built environment. The covariates involving distances to facilities included in the model are: *distance to city centre*, *to different types of roads (freeways, ring roads, and radial roads)*, *to subway and train stations*, *to airport*, *to distribution hubs*, *to technical universities*, *to theatres*, and *to green areas*. These covariates are raised to a power to increase the rapidity in which the effect diminishes. We assumed that once a point is at a distance larger than a certain threshold the benefit of proximity to a facility is very low. We hypothesize that it is sensible to assume that once the distance from certain facilities is larger than 10 kilometres, the benefit from proximity drops substantially. After graphing different power transformations, we found that raising the functions to 2.5 obtained the desired assumed effect.

Gibbs model specifications treat geographical space as continuous and covariates need to be defined in any point inside the observation window. Two different procedures are followed in order to assign values to covariates that are not point-referenced to specific sites. For the covariate *population density*, each of the municipalities is randomly assigned as many points as the value of the covariate in the municipality.¹⁵ Then a kernel smoother is implemented to make the covariate continuous in space. This procedure biases the results when the covariate is not divided by area, favouring the smaller municipalities. For example, with this procedure of geo-referencing two municipalities with the same population but different areas would take different values when they should be equal. To avoid this bias, a different geo-referencing method is followed for the covariates *industrial land price*, *property tax*, and *surface of green area*.¹⁶ A grid of 150 by 150 points (points are 300 metres apart) is overlaid to the observation window and then each intersection point in this grid, takes the value of the covariate in the municipality. Finally, a kernel smoother is implemented to make the covariate continuous in space.

The *crime rate* is provided by the International Security Study Group (GESI) from the Spanish Home Office. The number of arrests made by the National Police (Policía Nacional y Guardia Civil) per 10,000 inhabitants is utilized. In order to obtain a homogeneous series, since this rate is not available for Madrid municipal districts, we produce these data based on local police arrests by district. The districts' population figures are obtained from the database of the Town Hall in Madrid, and the municipalities' populations from the Madrid Regional Statistics Institute.¹⁷ These are the only two covariates in the model— *population density* and *crime index*— that cannot be measured at a continuous level. Instead these two covariates have a district based value due to lack of data at a micro-geographical level.

The *area* of the municipalities is provided by the *Statistical Yearbook of the Region of Madrid 1985 to 2011*. And district areas are provided by the *Statistical Yearbook of the City of Madrid*.

The covariate *industrial land price* is the minimum value for the purposes of calculating the tax base of the Inheritance and Gift Tax. *Industrial land price* is measured for sets of streets in each municipality. It is measured in euro per square metre and captures the real value of the property when industrial land is transferred. The estimation method analyses the real estate market on the date of accrual of the tax and calculates mean values for different areas from a

¹⁵ Due to the extent of the municipality of Madrid we used the data of the 21 districts that is the lowest level of aggregation, while the rest of the municipalities only have one district.

¹⁶ Potential readers interested in implementing Gibbs approach to other countries should not dismiss including the fiscal covariate in the analysis for technical reasons. This was an additional motivation to maintain the taxation covariate in the analysis; showing that the Gibbs model approach requires certain transformations on covariates measured on an area based scale.

sufficient number of ‘witnesses’, consisting of properties of the same type and characteristics to the land being transferred.¹⁷

The *property tax* is common for all establishments in each municipality but only those firms which own the facilities where they locate in are required to pay this tax. The nominal property tax rate varies between 0.4 per cent and 1.1 per cent. We utilize the effective tax rate measured by *per capita* payable tax (*CI_IBI*), which also controls for the nominal value of the assets. Although this value is updated periodically on average it is around 50 per cent of the market price of the assets and there are large differences between municipalities.¹⁸

Finally, the Green Urban Areas and Sport and Leisure Facilities are provided by CORINE Land Cover Project developed by the European Environment Agency included in the Land Core Monitoring System of the Global Monitoring for Environment and Security.¹⁹ A map of the municipalities and district boundaries was overlaid in order to assign values to these polygons, and a grid of 150 by 150 points was overlaid as explained above.

5 Results

As a first step we use several diagnostics to assess the model fit both in terms of the trend and interaction components. A diagnostic of fit for significant covariates in the trend, lurking variable plots, are provided in Figures 2–4.²⁰ As observed, the true spatial trend can be approximated by the specified trend. The empirical plot is close to its expected value assuming that the model is correct. The graphs do not exceed the point-wise two standard-deviation error limits calculated for the inhomogeneous Poisson process (dotted lines). When plotting the residuals of the model fitted to all the establishments’ sample, the empirical line exceeded the envelopes for many of the covariates, illustrating the need to fit different models for each of the three establishment employment size classes. Overall model fit in the trend is assessed using a plot of smoothed residuals (see Figure 5). The smoothed residuals diagnostic presents a flat surface with small deviations from zero in all the models suggesting a good overall fit. Assessment of the interaction component is based on the QQ-plot (Figure 6) and the G-compensator diagnostic (Figure 7). The empirical distribution of the smoothed residuals in the QQ-plot lies inside the envelopes (estimated by Monte Carlo samplings) of the expected empirical quintiles obtained from simulations of the model with a Geyer interaction specification. In the G-compensator diagnostic the standardized residuals lie inside the envelopes showing positive values, suggesting that the data are slightly more clustered than the model. Overall the presented diagnostics indicate that the model with the Geyer saturation interaction term correctly captures the dependence on the covariates and the interaction between the establishments. Coefficients estimates for the model defined in equation 8 are shown in Table 1.

One of the main purposes of the paper is to show that results might be different depending on the scope of the analysis because the modifiable area unit problem (MAUP) might be present. When possible we compare our results to demonstrate that some results differ depending on the

¹⁷ In some of Madrid’s districts there is no industrial land use. To assign land price values we followed two different procedures. First, we kernel smoothed industrial land prices so that the districts with no industrial land prices are assigned a value equal to the weighted average of industrial land prices in neighbouring districts. Second, we used real commercial land price data for those districts where industrial land was not present. The results under both alternatives were very similar.

¹⁸ The property tax is the highest local tax in terms of revenue. The business sector is charged a number of municipal taxes and fees. This list includes a local business tax, a property tax, a tax on vehicles, a tax on building activities, and a tax on the sale of land and buildings. The local business tax is the main local tax burden borne by the business sector, but since 2003 firms have a tax allowance in the region.

¹⁹ Land use 141 and 142, respectively.

²⁰ The complete set of lurking variable plots is available from the authors. The full set is not provided in the paper due to space constraints and to maintain visual clarity in the published figure.

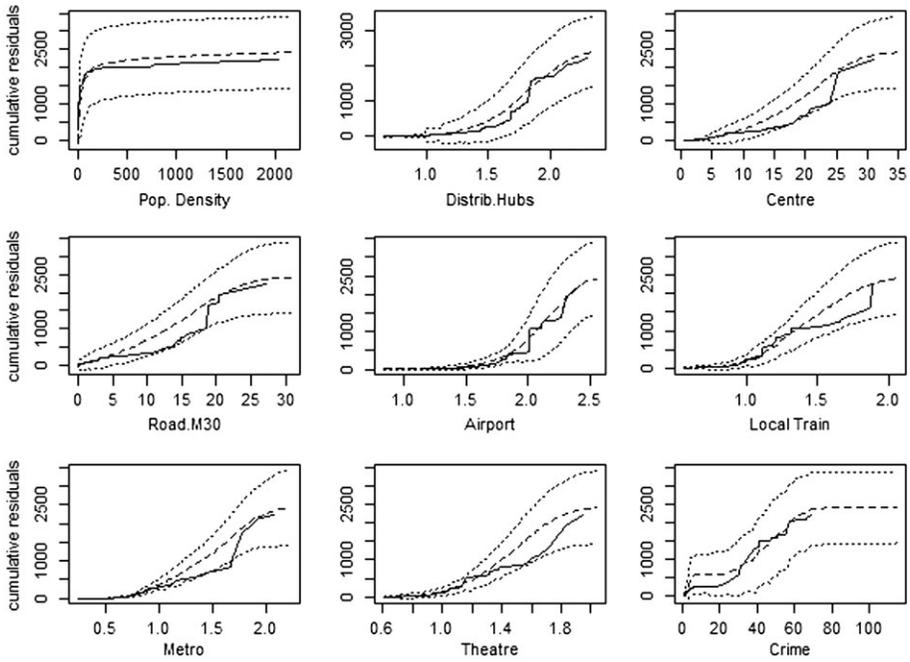


Fig. 2. Lurking variable plot: small establishments

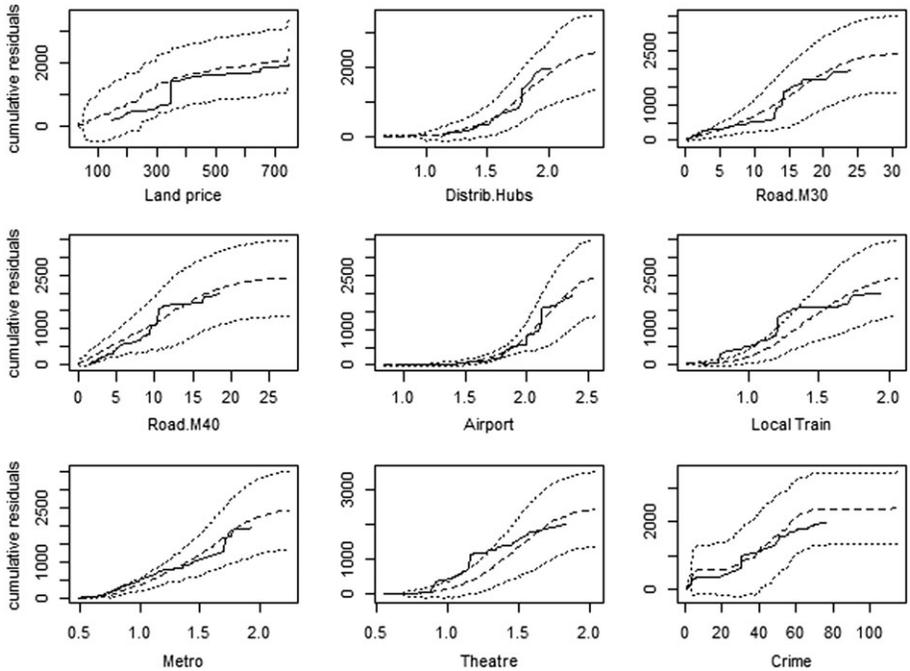


Fig. 3. Lurking variable plot: medium establishments

scope of the analysis. Area-based models could be misguided and Gibbs models constitute the appropriate tool to complement the analysis of intra-urban location. Results differ between those in which the model is estimated for the whole sample and those obtained when firm's size is

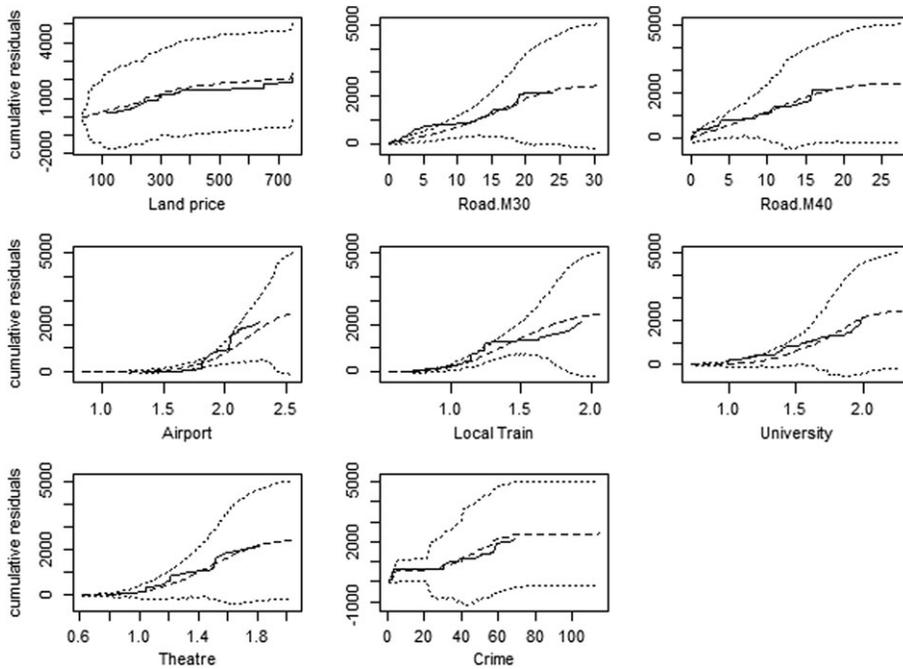


Fig. 4. Lurking variable plot: large establishments

considered in the analysis. Small firms, due to the risk associated with their size and their inability to internalize certain necessary stages of production, tend to locate in dense areas, close to markets and support services. Population density has positive effects on productivity reaching conventional significance levels suggesting that urbanization economies usually outweigh congestion and other diseconomies of high population density. The covariate distance to the centre is significant at the 95 per cent confidence level only in the model for the small firm's subsample, but with a positive sign. This indicates that in Madrid, a polycentric structure is more prevalent than a monocentric one, reflecting the overwhelmingly decentralized nature of the industry. Recent models have acknowledged the polycentric structure of many cities in the world and attempted to explain the emergence of more complex patterns through the interplay of various forces of agglomeration and dispersion (Lucas and Rossi-Hansberg 2002; Ahlfeldt and Wendland 2013). Our result contradicts those obtained by Arauzo and Viladecans (2009) and Alañón et al. (2007) who show that Spanish manufacturing establishments in high-tech industries prefer to locate close to the centre of the metropolitan area. These studies are interurban area/region analyses where different urban areas or cities are considered, so the results are interpreted in terms of the average urban area. The results for intra-industrial location analyses are scarce and inconclusive. Wu (1999) determines that the traditional centre has seen a declining attraction for new firms in the metropolitan area of Guangzhou in China. Shukla and Waddell (1991) examines firm location in the context of poly-centricity, finding that the distance to the CBD (Dallas CBD) cannot provide a singularly powerful explanation of firm location, with the exception of the finance, insurance and real estate industries.

The location decisions of firms in the electronics industry are dominated by site costs. For the small size sample, industrial land price is only significant at the 95 per cent level indicating that the importance of benefits from urbanization economies could dominate the price effect. As firm size increases, however, industrial land price becomes significant at the 99 per cent level in the location choice. The analysis by Figueiredo et al. (2002) of the whole manufacture for the

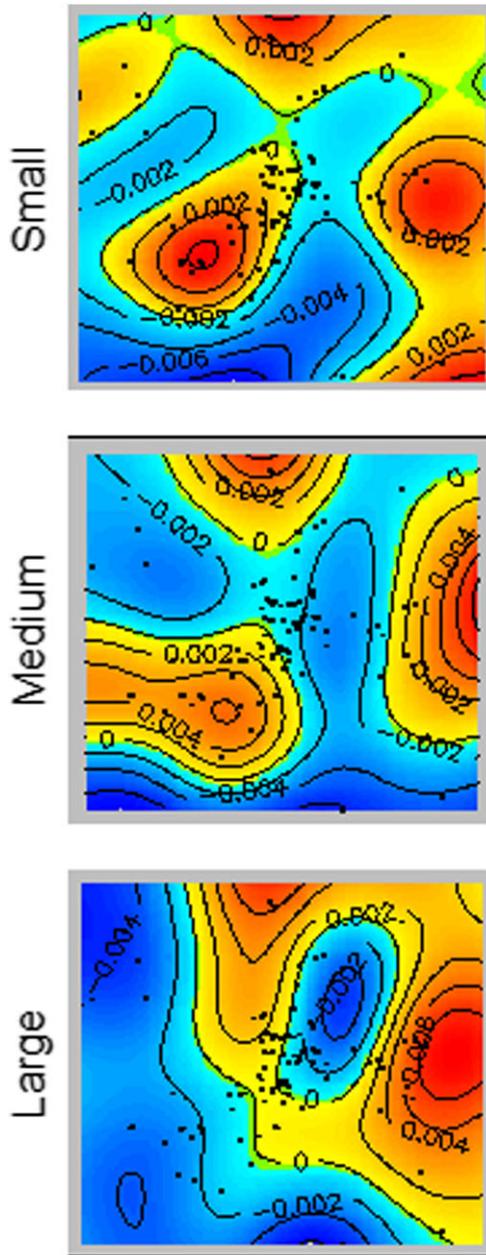


Fig. 5. Smoothed residuals

Portuguese districts, or that of Cheng and Stough (2006) of the location of Japanese FDI plants in Chinese cities obtained similar results.

There is not significant evidence of a higher concentration of firms surrounding the logistic infrastructures in the subsets of data defined by employment size class. Firms do not reveal any preference for being in close proximity to a distribution hub, probably because the electronics industry neither uses these logistics centres to ship their products nor requires large logistics infrastructures.

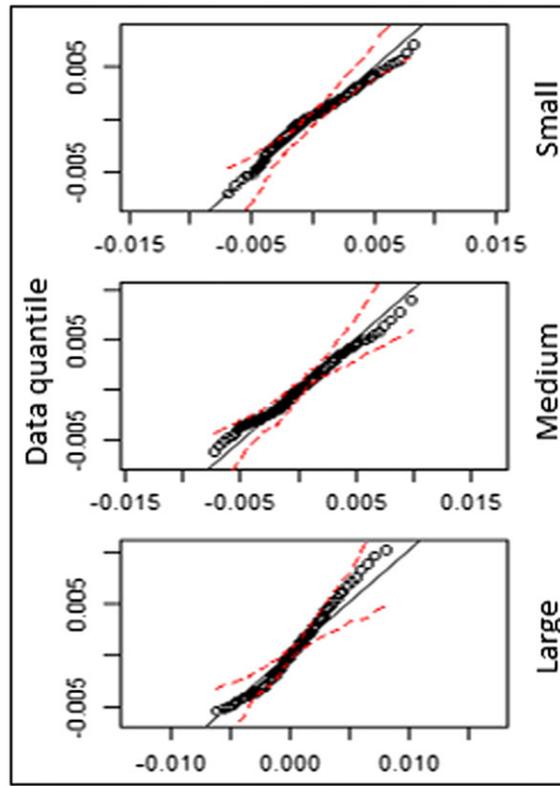


Fig. 6. QQ-plot

As expected, results indicate that the estimated parameter of the fiscal variable is not significantly different from zero. Moreover, when the fiscal covariate was dropped from the analysis the results were unaffected. Only the model for the small firms subsample finds a significant effect at the 90 per cent confidence for the taxation variable. The property tax (IBI) in Spain represents a very small percentage of the tax burden faced by establishments in the analysed sample. Earlier studies were not conclusive on the effects of taxation on industrial location either found a non-significant effect of tax levels on location decisions or a moderately negative effect in the United States. On the contrary, Gabe and Bell (2004) have argued that high-tax locations remain attractive as long as they spend large sums of money on the provision of public goods and services. Recent papers have proposed that the effect of taxes on a firm's location needs to be considered in relation to the existence of agglomeration economies. Brülhart et al. (2012) finds that corporate taxes deter firm births less in more spatially concentrated sectors. Hence, agglomeration economies do appear to attenuate the impact of tax differentials on firms' location choices. Jofre-Montseny and Sole (2012) suggest agglomeration economies can either increase or reduce the intensity of tax competition, depending on the specific circumstances. Gibbs model constitute a fruitful approach to contribute to the debate in economies where taxes are decentralized or where local taxes represent an important tax burden for companies.

Regarding the preference for local public goods, the results support the following conclusions. Transportation costs are much less a consideration in location decisions in the electronics industry. At the 99 per cent level only ring roads are statistically significant for the large firms' model. The almost complete concentration of the telecommunication equipment manufacturing industry in Madrid probably contributed to the development of large establishments for which a location on

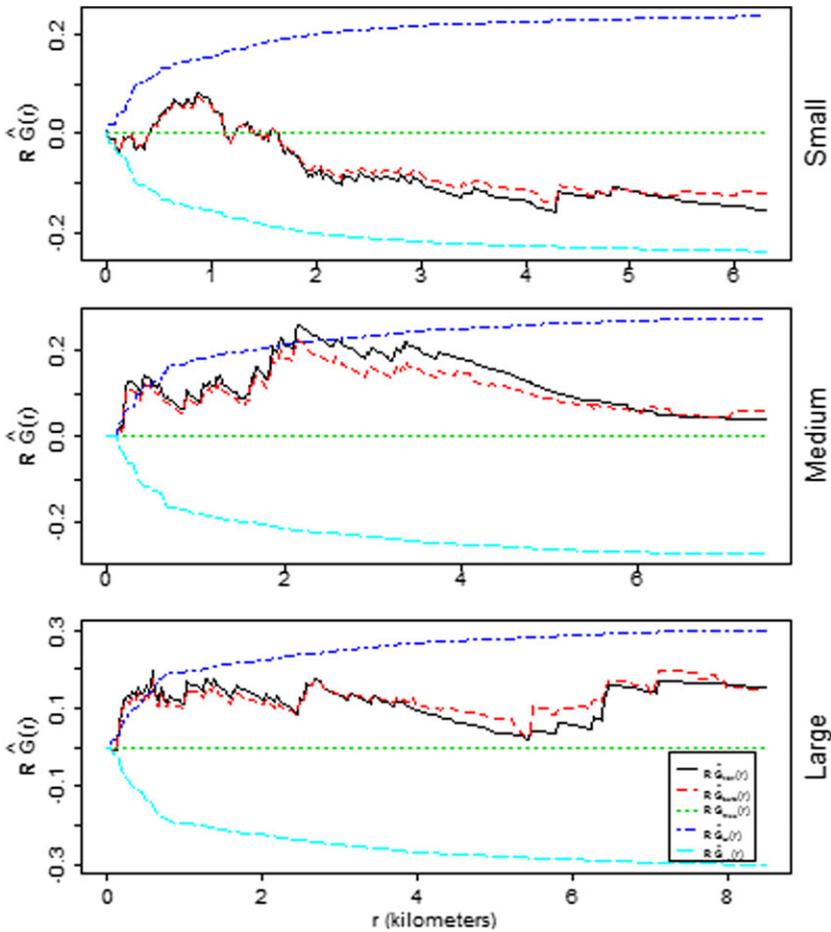


Fig. 7. G-compensator

the periphery of the city and near to the ring road M30 is more attractive. The ring road M40 coefficient estimate has the opposite sign from what we expected. This result could be a consequence of the shape of this particular road, which extends far beyond the west surrounding the metropolitan area. The location of large firms is to some extent conditioned by land zoning restrictions, with most firms located in the eastern and northern municipal districts, and to a lesser extent in the southern districts. Ground transportation is expected to play a more significant role in industries with heavy or bulky products – those industries in which transportation costs are a large percentage of the total product cost – than in industries whose products are lighter and easier to transport.

Firms prefer sites that are in proximity to train stations or metro transportation systems. Even large firms that locate in sites far from the city which might not be covered by the metro network are always in proximity to ‘local train stations’. There is evidence of a significant preference for locations in proximity to Madrid International Airport and this is particularly the case for larger firms. Strauss-Kahn and Vives (2005) find the existence of airport facilities to be a significant location factor in analysing the location of headquarters in US counties. Shukla and Waddell (1991) also find a significant effect in distance to airport in the analysis of Dallas urban area for the whole manufacturing industry.

We tested if there is significant association between technical universities and electronics firms. One of the most inspiring results is what we take as evidence of the existence of

Table 1. Models estimation

Variable	All	Small	Medium	Large
Intercept	1.681 (1225)	1.187 (1849)	-1.879 (1962)	8.351*** (2723)
Industrial Land	-0.001*** (0000)	-0.001* (0001)	-0.002*** (0001)	-0.002*** (0001)
Population density	0.001 (0000)	0.002*** (0001)	-0.000 (0001)	-0.001 (0001)
Distribution Hubs	0.328 (0268)	0.491 (0561)	0.998* (0510)	-0.087 (0539)
Property Tax	-0.002 (0001)	-0.004* (0002)	-0.003 (0002)	0.000 (0002)
Distance to Centre	0.102 (0065)	0.254*** (0105)	-0.018 (0104)	0.166 (0149)
Distance to road R	0.005 (0021)	0.064* (0036)	-0.062 (0040)	-0.012 (0040)
Distance to road A	-0.043 (0036)	-0.035 (0060)	-0.074 (0059)	0.004 (0059)
Distance to road M30	-0.143*** (0059)	-0.150* (0084)	-0.091 (0085)	-0.472*** (0146)
Distance to road M40	0.022 (0041)	-0.111 (0073)	0.039 (0071)	0.288*** (0106)
Distance to airport	-0.887*** (0331)	-1.054 (0687)	-0.881 (0598)	-2.331*** (0702)
Distance to Local Train	-0.873*** (0292)	-0.747 (0466)	-0.746* (0400)	-2.135*** (0534)
Distance to Subway	-1.242*** (0307)	-3.323*** (0552)	-1.858*** (0465)	0.227 (0607)
Distance to University	0.059 (0422)	0.904 (0644)	0.112 (0584)	-1.548*** (0729)
Distance to Theatre	-0.347 (0331)	-0.704 (0544)	0.085 (0478)	-1.231*** (0571)
Crime Index	0.013*** (0004)	0.017*** (0008)	0.016*** (0006)	0.017*** (0007)
% Green area	-0.051* (0030)	-0.115 (0126)	-0.038 (0029)	-0.064 (0043)
Interaction	0.539*** (0054)	0.285*** (0085)	0.444*** (0063)	1.276*** (0233)

Notes: Standard Errors are reported in parentheses below the coefficients. ***, **, * Indicate significance at the 1, 5, 10 per cent levels, respectively.

geographical knowledge spillovers derived from technical universities. Larger firms present a higher probability of locating in the surroundings of technical universities. If productivity effects arising out of area universities are ‘universal’, that is, if they are present on a metropolitan-wide basis without significantly varying across locations, then the exact location of universities within urban areas may not influence the intra-urban geography of high-tech companies. Human capital spillovers attenuate sharply with distance and papers using administrative areas as units of observation might have difficulties in recognizing the existence of spillovers when they emerge inside the administrative boundaries. The Gibbs model constitutes an appropriate approach to detect them. Wang et al. (2006, 2012) conclude that the first phase of high and new technology industry clusters often happens close to universities named as the first ‘Cambridge phenomenon’. Proximity might be important if in order to transfer new scientific findings into marketable products, face-to-face interaction and hands-on participation are required. Examples of this phenomenon are Silicon Valley and Route 128 in Boston which owe their status to their close ties to research institutions at Stanford and MIT. The concentration of R&D firms in Irvine in Greater Los Angeles may be associated with benefits stemming from the California Institute of Technology and the University of California at Irvine campuses (Jürgen et al. 2004). Audretsch and Lehmann (2005) find that the number of knowledge-based start-ups clustered around German universities is positively influenced by the knowledge output of the respective university and the innovative capacity of the region. Proximity to relevant academic research is associated with higher productivity among measuring and controlling device manufacturers and may influence other technology-intensive industries (Drucker and Feser 2012).

Cultural and recreational opportunities are not important location factors in the electronics industry for selecting a site within a metropolitan area. These sorts of amenities are hard to measure because amenities and agglomeration are highly spatially correlated. It is difficult to distinguish whether high-tech location behaviour is truly agglomerative or amenity oriented. Results indicate that only large firms present a higher probability of locating in proximity to a theatre, and the covariate green surface area in the district does not significantly differ from zero

in any of the models. These results suggest that in intra-metropolitan analysis we should be prudent about the relationship between ‘quality of life’ and the location of high-tech industries.²¹ The results reveal that the ‘consumer city’ hypothesis might be relevant in attracting high skill labour to a certain city or metropolitan area. These amenities could be relevant during the first stage of location, the coarse-scale choice among alternative regions, urban areas, cities or municipalities. However, once the coarse-scale choice has been made these services might not be relevant at the intra-metropolitan scale. As the factors attracting this labour profile are scattered around the city, individuals can live in a jurisdiction with pleasant surroundings while working in an area without these characteristics. Employees can still enjoy the benefits of amenities even if they are not in close proximity to them. The high rate of commuting that characterizes the greater Madrid area determines that a firm may derive all possible benefits from amenities in the form of spillovers. Individuals can live in a jurisdiction with pleasant surroundings while working in an area with none of these characteristics. The covariate crime index is always significantly different from zero but with an unexpected positive sign. Crime may present an error-ridden proxy of exposure to crime, as different neighbourhoods within cities may be subject to wide variation in crime rates. If crime figures were available at the neighbourhood or at the street and block level of aggregation, the model would be able to test for the validity of assumption 1. Gibbs specification models are very suitable for testing this hypothesis when geocoded data are available. We can conjecture that ‘Quality of life’ amenities are not likely to be a decisive location factor for firms selecting a site within a metropolitan area. These amenities could be relevant during the first stage of location, in choosing which urban area, city, or municipality to locate. But once this decision has been made these services might not be relevant at the intra-metropolitan scale. There can be a mismatch between the location of one’s job and the location of one’s residence as a firm may derive all possible benefits in the form of spillovers. Residents commute great distances because their residence environment offers amenities they consider to be more important than commuting costs.

To conclude, the interaction component is always significant. The estimated parameter $\gamma > 1$ indicates that the probability of observing a firm is higher if there is another firm in its proximity. Once the covariates that affect the first order intensity of the process are taken into account, the interaction component is significantly different from zero at 99 per cent confidence level in all the estimated models. Recall that the interaction effect can be interpreted as the estimated degree of establishment-to-establishment attractiveness. The narrow interpretation is that conditional on other covariate values in the trend, the probability of observing a firm at a location is higher if another firm is located nearby. The models also allow us to recover a different radius of

²¹ According to the literature, amenities have been proxied by several covariates, including *human capital*, the existence of *good schools in the area*, *toxic emissions levels*, *landfill waste levels*, *per capita recreation expenditure*, *per capita local public expenditure*, *GDP per capita*, *acreage of state parks*, *density of amusement for employees*, *crime index*, *distance to malls*, *green surface area in the district*, and *distance to sport and leisure facilities*. Unfortunately, data on *good schools in the area*, *toxic emissions levels*, *landfill waste levels* and *GDP per capita* were not available for our case study. On the other hand, expenditure level covariates like *per capita recreation expenditure* and *local public expenditure* can be biased due to differences in the provision costs. We replaced these expenditure variables by covariates that measure near distance to certain facilities since they are more accurate at capturing the amenities in the area. We used ArcGis to calculate *the nearest distance to a police station*, *the nearest distance to a hospital*, *the nearest distance to a mall*, *the nearest distance to a recycling center*, *the nearest distance to a sport facility*, and *the nearest distance to a green area*. For our case study the covariates *nearest distance to hospitals* and *nearest distance to a police station* were not a good proxy to capture the amenities in the area. They were always significant since the location of these facilities was planned based on a maximization algorithm ensuring that any point in the district is at a threshold distance to them. The parameters of the covariates *nearest distance to a recycling center*, *nearest distance to a public sport facility*, and *the nearest distance to a green area* were neither found significantly different from zero. Finally, *Human capital* is an area based measure, to include it in the model we overlaid to the municipality boundaries a grid of 150 by 150 points, separated 300 metres apart. We assigned to each intersection point in the grid human’s capital municipality value. Then a kernel smoother is implemented to make the covariate continuous in space. Its estimated parameter did not show any significant effect on firms’ location choices.

interaction based on profile likelihood. More broadly, the estimated interaction effects and radii can be interpreted as evidence of the strength and scope of localization economies in the electronics sector. The range of interaction is estimated to be quite small (14 kilometres or less) and the benefits of co-location are implied to accrue when there are 2 to 4 neighbouring firms, depending on employment size.²² Therefore, once the firm is at a distance larger than 14 kilometres there is no interaction effect. If our models are correct and the range of interaction is 14 km or less, you could not then expect to detect interaction with most area based models because the units of measure (the areas) resolve at a much coarser scale.

Arbia (2001) and Arbia et al. (2015) analysed firm's interaction estimating models based on georeferenced data. However results cannot be straightforwardly compared as both papers introduce the interaction term as an additional covariate modelling the first order concentration (trend). Arbia (2001) estimates an extension of Krugman's (1991) theoretical model of firm demography for the San Marino Republic. The intensity of the firms' birth pattern is estimated based on the covariates' distance from the main roads and on their interaction with previously existing firms. Subsequently, Arbia et al. (2015) estimate three sub-processes (birth, growth and survival) of firm demography to test the relative importance of competitive and co-operative spatial interactions in determining the spatial distribution of convenience food stores in Trento. The spatial firms' birth model is the only one based on georeferenced data and the intensity of the spatial point process is estimated as a function of existing convenience stores and the location of large supermarkets. Sweeney and Gómez-Antonio (2015) is the only paper to our knowledge that explicitly fits Gibbs models to point pattern data and incorporates both spatial inhomogeneity and inter-point interactions. It provides a simple empirical model for Madrid's electronics industry where the interaction parameters coincide.

The interaction result is in line with previous studies of the industry using different methodologies. Rama and Calatrava (2002) determined a marked clustering of producers in the region, embedded in an intense web of subcontracting relationships, with the majority of establishments in Madrid obtaining a significant proportion of their production inputs, such as raw materials, parts and equipment, locally. Holl and Rama (2009) find that co-location implies strong interactions between firms in the electronics industry in Madrid. They determine that subcontracting linkages are among the most localized relations, suggesting that they involve close contacts for which proximity is desirable. It would take additional qualitative or quantitative analysis to attribute the localization to specific forms of co-operation among firms, such as joint production, subcontracting, use of joint inputs or machinery, shared purchases, use of the same marketing campaigns, aftersales services and R&D activities, etc. Possible explanations for the interaction result are related to geographic and professional proximity, as well as to the similar origins of entrepreneurs (Telefonica or Politécnica University), which encouraged the development of stable outsourcing relationships that prompted inter-firm collaboration as a way of minimizing capital risks (Rama et al. 2003).

Different industries agglomerate for different reasons. In some industries knowledge spillovers are essential while other industries are instead driven by labour market pooling or input sharing. High-technology sectors show stronger evidence of knowledge spillovers, while lower-technology industries instead show stronger evidence of input sharing and labour pooling (Faggio et al. 2014). Although this second level of analysis is not discussed in this paper, Gibbs models provide a major step forward in isolating and characterizing the nature of spatial co-location. Under this approach we can identify in which sectors the source of cluster has its origin in true interaction between firms or is just a consequence of environmental characteristics.

²² For the small and the large firms model the interaction radius is 1 km while for the medium is larger 1.4 km. The saturation parameter is 4 for the small firms sample, 3 for the large and 2 for the medium.

6 Conclusions

This paper improves the approaches used to date in measuring geographic concentration. Our improvements allow the disentanglement of first and second order effects, allowing a deeper understanding and quantification of different types of spillovers effects. Our results confirm the validity of the employed methodology for the analysis of intra-urban industrial location.

The intra-metropolitan distribution of electronics industry mainly follows a dichotomous pattern. Different covariates play different roles in explaining the trend component for each firm size class. The location patterns of small firms are mainly linked to current population and job distribution, but those of large firms are less affected by these variables. In the greater Madrid area a polycentric structure is more prevalent than a monocentric one, where location decisions of firms in the electronics industry are dominated by site costs, and distribution hubs is not a factor considered by firms in deciding their location site. Regarding the local public goods analysed, transportation costs are much less a consideration in location decisions in the electronics industry, but firms show a preference for proximity to train and metro stations, or to the airport.

The results obtained challenge some of the outcomes of the intra-urban industrial location literature. The existence of geographical knowledge spillovers derived from technical universities is confirmed for large establishments; amenities are not likely to be important location factors within a metropolitan area; and crime rate results are unexpected. All these results emphasize the need for different approaches in order to identify the factors that affect intra-urban location. Spatial aggregation compromises the validity of results in studies of business location. Intra-urban location decision processes might be rather different to inter-city or inter-regional location decisions. Firms consider different factors of localization at different scales, so results should not be compared between different spatial units of observation.

Interaction is significant in all models indicating the presence of sources of spillovers even after other environmental sources of clustering are taken into account. In terms of the processes that led to interaction, only historical analysis of particular cases could really speak to origins (think of the many historical accounts and actors involved in the emergence of the Silicon Valley). The methods proposed here are instead intended to provide estimates of parametric models of spatial stochastic processes that are consistent with currently observed patterns of industry.

We have tested if there is any significant association between the covariates and the electronics firms' locations but cannot say anything about causality. Therefore we need to be cautious in making policy recommendations. The main policy implication that could be derived from the results obtained in this paper is the role that universities and research centres might play in attracting high tech firms to the area. The returns of investments in these institutions are twofold. First, it increases the demand for high skilled labour, establishing conditions that can foster the creation of startup firms. Second, it may attract high tech industries to the area. We proved that once trend effects are discounted for, high tech firms value the existence of other firms in proximity when determining its location. Therefore, public investments in technical universities and research centres have a multiplicative effect as firms attract other firms. These institutions can act as attraction poles of employment in the region. It has been established in the literature that there is a clear linkage between universities and the birth of high tech firms that may be created as spin outs from these institutions. Given that transport infrastructure allows easy commuting inside the Madrid urban area, we postulate that the existence of these institutions will determine a multiplier effect that will benefit economic growth in the urban area. On the other hand, investing in better amenities in the area may not necessarily attract new firms because workers can commute from different jurisdictions. The factors in attracting high skilled labour are metropolitan area-related, meaning that employees may enjoy living in a jurisdiction with a variety of pleasant amenities while working in areas where such amenities are absent.

The results of this paper suggest that Gibbs models constitute a fruitful approach to empirically explore the sources behind cluster configurations of firms at a variety of spatial scales. Gibbs models are sufficiently flexible to allow for multiple mechanisms, both for the purposes of identification and to avoid confounding results. While the approach does not solve all of the challenges inherent in empirical analysis of localization, the models yield a far richer set of results than prior methods, and move from the simple hypothesis testing to complete model specification and validation that forms the basis for most empirical research in economics and regional science. The specific advantage is that Gibbs models provide a regression framework that takes point-referenced data as unit of observation.

References

- Ahlfeldt GM, Wendland N (2013) How polycentric is a monocentric city? Centers, spillovers and hysteresis. *Journal of Economic Geography* 13: 53–83
- Alañón A, Arauzo JM, Myro R (2007) Accessibility, agglomeration and location In: Arauzo JM, Manjón, M (eds) *Entrepreneurship, industrial location and economic growth*. Edward Elgar, Cheltenham
- Arauzo JM, Viladecans E (2009) Industrial location at the intra-metropolitan level: The role of agglomeration economies. *Regional Studies* 43: 545–558
- Arauzo JM, Liviano-Solis D, Manjón-Antolín M (2010) Empirical studies in industrial location: An assessment of their methods and results. *Journal of Regional Science* 50: 685–711
- Arbia G (2001) Modelling the geography of economic activities in a continuous space. *Papers in Regional Science* 80: 411–424
- Arbia G, Cella P, Espa G, Giuliani D (2015) A micro spatial analysis of firm demography: The case of food stores in the area of Trento (Italy). *Empirical Economics* 48: 923–937
- Audretsch D, Lehmann E (2005) Does the knowledge spillover theory of entrepreneurship hold for regions? *Research Policy* 34: 1191–1202
- Baddeley A, Møller J, Waagepetersen R (2000) Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54: 329–350
- Baddeley AJ, Rubak E, Møller J (2012) Score, pseudo-score and residual diagnostics for Goodness of fit of spatial point process models. *Statistical Science* 26, 4: 613–646
- Baddeley AJ, Turner R (2005) Spatstat: an R package for analysing spatial point patterns. *Journal of Statistical Software* 12: 1–42
- Baddeley AJ, Turner R, Møller J, Hazelton M (2005) Residuals for spatial point processes using the Papangelou conditional intensity. *Journal of the Royal Statistical Society, series B*, 67: 617–666
- Barff R (1987) Industrial clustering and the organization of production: A point pattern analysis of manufacturing in Cincinnati, Ohio. *Annals of the Association of American Geographers* 77: 89–103
- Brühlhart M, Jametti M, Schmidheiny K (2012) Do agglomeration economies reduce the sensitivity of firm location to tax differentials. *The Economic Journal* 122: 1069–1093
- Carlino GA, Carr JK, Hunt RM, Smith T (2011) The agglomeration of R&D labs. Working Paper Research Department, Federal Reserve Bank of Philadelphia
- Carlton DW (1983) The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics* 65: 440–449
- Cheng S, Stough RR (2006) Location decisions of Japanese new manufacturing plants in China: A discrete-choice analysis. *Annals of Regional Science* 40: 369–387
- Currid E, Connelly J (2008) Patterns of knowledge: The geography of advanced services and the case of art and culture. *Annals of the Association of American Geography* 98: 414–434
- Currid E, Williams S (2010) The geography of buzz: art, culture and the social milieu in Los Angeles and New York. *Journal of Economic Geography* 10: 423–451
- Diggle PJ, Chetwynd AG (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* 47: 1155–1163
- Drucker J, Feser E, (2012) Regional industrial structure and agglomeration economies: An analysis of productivity in three manufacturing industries. *Regional Science and Urban Economics* 42: 1–14
- Durantón G, Overman H (2005) Testing for localization using micro-geographic data. *Review of Economic Studies* 72: 1077–1106
- Durantón G, Overman H (2008) Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science* 48: 213–243

- Faggio J, Olmo S, Strange WC (2014) Heterogeneous agglomeration. Discussion Paper Spatial Economics Research Centre 152, London School of Economics.
- Figueiredo O, Guimaraes P, Woodward D (2002) Home-field advantage: Location decisions of Portuguese entrepreneurs. *Journal of Urban Economics* 52: 341–361
- Gabe T, Bell K P (2004) Tradeoffs between local taxes and government spending as determinants of business location. *Journal of Regional Science* 44: 21–41
- Glaeser E, Kolko J, Saiz A (2001) Consumer city. *Journal of Economic Geography* 1: 27–50
- Glaeser EL, Kerr WR (2009) Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics and Management Strategy* 18: 623–663
- Högmander H, Särkkä A (1999) Multitype spatial point patterns with hierarchical interactions. *Biometrics* 55: 1051–1058
- Holl A, Rama R (2009) The spatial patterns of networks, hierarchies and subsidiaries. *European Planning Studies* 17: 1261–1281
- Huang F, Ogata Y (1999) Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics* 8: 510–530
- Jacobs J (1969) *The economy of cities*. Random House, New York
- Jofre-Montseny J, Marín-López R, Viladecans-Marshall E (2011) The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics* 70: 61–74
- Jofre-Montseny J, Sole A (2012) Which communities should be afraid of mobility? The effects of agglomeration economies on the sensitivity of employment location to local taxes. *Regional Science and Urban Economics* 42: 257–268
- Jürgen E, Gottschalk S, Rammer C (2004) Location decisions of spin-offs from public research institutions. *Industry and Innovation* 11: 207–223
- Krugman P (1991) Increasing returns and economic geography. *Journal of Political Economy* 99: 483–499
- Lee S (2010) Ability sorting and consumer city. *Journal of Urban Economics* 68: 20–33
- Lucas REJ, Rossi-Hansberg E (2002) On the internal structure of cities. *Econometrica* 70: 1445–1476
- Marcon E, Puech F (2010) Measures of the geographic concentration of industries: Improving distance-based methods. *Journal of Economic Geography* 10: 745–762
- Møller J, Waagepetersen RP (2007) Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics* 34: 643–711
- Molotch H (1996) LA as design product: How art works in a regional economy. In: Scott AJ, Soja EW (eds) *The city: Los Angeles and urban theory at the end of the twentieth century*. University of California Press, Los Angeles, CA
- Papke LE (1991) Interstate business tax differentials and new firm location: Evidence from panel data. *Journal of Public Economics* 45: 47–68
- Rama R, Calatrava A (2002) The advantages of clustering: The case of Spanish electronics subcontractors. *International Journal of Technology Management* 24: 181–229
- Rama R, Ferguson D, Melero A (2003) Subcontracting networks in industrial districts: The electronics industries of Madrid. *Regional Studies* 37: 71–88
- Rosenthal SS, Strange WC (2001) The determinants of agglomeration *Journal of Urban Economics* 50: 191–229
- Shukla V, Waddell P (1991) Firm location and land use in discrete urban space: A study of the spatial structure of Dallas-Fort Worth. *Regional Science and Urban Economics* 21: 225–253
- Statistical Yearbook of the Region of Madrid (1985–2011) Instituto de Estadística, Comunidad de Madrid
- Statistical Yearbook of the City of Madrid (2011) Anuario Estadístico Municipal, Ayuntamiento de Madrid
- Strauss-Kahn V, Vives X (2005) Why and where do headquarters move? Discussion Paper 5070, Centre for Economic Policy Research
- Sweeney S, Feser E (1998) Plant size and clustering of manufacturing activity. *Geographical Analysis* 30: 45–64
- Sweeney S, Gómez-Antonio M (2015) Localization and industry clustering econometrics: An assessment of Gibbs models for spatial point processes. *Journal of Regional Science* 56: 257–287
- Wang Z, Liu X, Mao KJ (2012) Industry cluster: Spatial density and optimal scale. *Annals of Regional Science* 49: 719–731
- Wang Z, Zhao JY, Liu X, Mao KJ, Xie XL (2006) An analysis to the evolution law of high-tech industry in space and its factors. *Studies in Science of Science* 24: 227–232
- Wu F (1999) Intrametropolitan FDI firm location in Guangzhou, China: A Poisson and negative binomial analysis. *The Annals of Regional Science* 33: 535–555