INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Non-linear externalities in firm localization

Giulio Bottazzi °
Ugo Gragnolati *
Fabio Vanni °

° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
* University of Paris I Pantheon-Sorbonne, France

2015/28 November 2015
ISSN(ONLINE) 2284-0400

Non-linear externalities in firm localization

Giulio Bottazzi *[1], Ugo M. Gragnolati † [2], and Fabio Vanni ‡ [3]

[1]Scuola Superiore Sant'Anna , Institute of Economics , Pisa, Italy.
[2]Université Paris I Panthéon-Sorbonne , Centre d'Économie de la Sorbonne , Paris, France.
[3]Scuola Superiore Sant'Anna , Institute of Economics , Pisa, Italy.

November 17, 2015

Abstract

This paper presents a model of firm localization allowing for non-linear, quadratic externalities. The model and its numerical estimation procedure manage to disentangle localization externalities from the intrinsic advantages of regions. Moreover, the introduction of a quadratic term can accommodate both more-than-linear positive feedbacks as well as congestion effects. Indeed, if the quadratic term is sufficiently negative, one location can reach the point in which the addition of an extra firm decreases the probability for that same location to further attract other firms. In this sense, the present model does not assume a priori that the localization choices of firms are characterized by positive interdependencies. Rather, the methodology allows to estimate whether or not this is actually the case.

JEL codes: C12, C16, C51, R30.
Keywords: Firm localization, Externalities, Non-linearities.

*E-mail: giulio.bottazzi@sssup.it; corresponding author.
†E-mail: ugo.gragnolati@univ-paris1.fr
‡E-mail: fabio.vanni@sssup.it

1 Introduction

The uneven spatial distribution of some scarcely mobile factors of production may drive also the firms that use such factors to mimic their spatial distribution. For example, extractive industries are clustered around mines, gushers, or gas fields, while being virtually absent where these natural resources are unavailable. Many other industries, however, are not so clearly bound to immobile factors of production; nonetheless, they are strongly concentrated in space. In these cases, clustering is generally explained through the self-reinforcing dynamics stemming from different types of interdependencies in the localization choices of firms. In particular, these interdependencies can be categorized according to the role that market mechanisms play in their unfolding (see Scitovsky, 1954). When the choice of a firm enters the production problem of other firms by affecting the prices thereby involved, then the interdependence among firms is classified as a pecuniary externality. Otherwise, when the choice of a firm affects directly and exclusively the factor inputs of other firms, then the interdependence is characterized as technological externality. There is a strictly practical reason for which this taxonomy is of interest especially when the interdependencies at stake are spatially bounded. Namely, if such interdependencies stand at the root of the self-reinforcing dynamics that allow the formation of spatial clusters, then policy makers may well be interested in unleashing localized externalities so a to promote regional development. Yet, the policies that may be thought to foster pecuniary externalities are very different from those that could target the formation of technological externalities.

Pecuniary externalities can lead to self-reinforcing dynamics an thus to spatial agglomeration through a process of cumulative causation based on the accumulation of local demand (see Myrdal, 1957). As an example, take two perfectly symmetric regions. If one firm moves from one region to the other, local wages decrease in the region of departure and increase in the region of destination, thus leading workers to move in the same direction as the firm. In turn, workers are also customers, so that consumer demand rises in the destination region. In this sense, the localization choice of a single firm ends up affecting all other local firms through market demand, thus constituting a pecuniary externality. This mechanism may keep on attracting additional firms as long as the gain from an expansion in local demand outstrips the loss associated to fiercer local competition. Hence, under suitable transportation costs and imperfectly competitive markets, the initial relocation of one firm calls for other firms to move in the same direction and gives rise to spatial concentration (see Krugman, 1991a). In this framework, the geography of production can be characterized in terms of core and periphery. This is especially the case since the attractive pull of demand acts *across* sectors, thus suggesting that the resulting agglomeration should look like a diversified city rather than as a specialized cluster.

Yet, specialized clusters do exist in areas that are neither particularly populated nor especially well-connected to markets. Only a limited part of these agglomerations can be explained through the dependence of one sector on some immobile factor inputs, as the contribution of broadly-meant natural advantages to spatial concentration is empirically modest across sectors (see Ellison and Glaeser, 1999). These specialized clusters could then stem from pecuniary externalities that take place with the co-localization of firms belonging to vertically-integrated industries, as discussed in Venables (1996). Also in this case, however, the empirical evidence suggests that the co-agglomeration between industries with strong upstream-downstream ties is limited (see Ellison and Glaeser, 1997). Therefore, something else is likely to be at play *within* sectors.

Sector-specific technological externalities represent an alternative source of self-reinforcing dynamics that may contribute to explain the riddle of the spatial distribution of firms. A typical example of how these untraded interdependencies may unfold locally is represented by the case of knowledge spillovers. That is, some of the private productive knowledge incorporated in firms spills into a common knowledge pool whose availability is bounded in space, due to the fact that knowledge is at least partly tacit and thus not perfectly transferable. In these circumstances, firms have an incentive to localize where a larger share of other firms are already settled, so as to benefit from the vaster knowledge pool that is available only locally (see Marshall, 1890, Book IV, Ch.X). What makes this mechanism sector-specific is that firms are interested in acquiring knowledge that is related to their own production process, so that they will co-localize with similar—rather than with generic—firms. In this sense, localized technological externalities stemming from knowledge spillovers can make sense of specialized clusters.

On the previous special issue of this JOURNAL, Bottazzi and Gragnolati (2015) provide a methodology to disentangle and measure the determinants of the spatial distribution of firms. Their work focuses in particular on singling out the strength of sector-specific, localized, technological externalities so as to compare them with the pull of local demand, local sectoral variety, and infrastructural advantages. The analysis is based on Italian plant data for a variety of manufacturing and service sectors as observed in year 2001 at the scale of commuting zones. Their main result is that the pull of localized technological externalities and population size are by far the most important determinants of firm localization, and these two drivers are comparable to each other in terms of their magnitudes. In fact, localized technological externalities are found to have an even stronger effect than population size across most sectors. This result contrasts to some extent with the limited weight that technological externalities have received, for instance, in the context of the earlier core-periphery models. For example, Krugman (1991b, p.54 and pp-61–62) states:

> [W]hile I am sure that true technological spillovers play an important role in the localization of some industries, one should not assume that this is the typical reason—even in the high technology industries themselves [p.54].
>
> [...] An accident led to the establishment of the industry in a particular location, and thereafter cumulative processes took over. [...] What the historical record shows us are two things. First, such cumulative processes are pervasive [...]. And second, Marshall's first two reasons for localizations, labor pooling and the supply of specialized inputs, play a large role *even when pure technological externalities seem unlikely* [italics added] [pp.61–62].

By contrast, the results discussed in Bottazzi and Gragnolati (2015) suggest that technological externalities can hardly be considered as "unlikely". In fact, they should be regarded as the rule in the economy, rather than as an exotic exception concerning only high-tech sectors.

There is, however, at least one reason to suspect that such a conclusion may overestimate the strength of technological externalities in determining firm localization. Namely, Bottazzi and Gragnolati (2015) base their estimates on a linear specification of externalities. In their framework, the individual advantage for a firm to choose a location increases proportionally to the number of other firms in the sector that are already placed in the same location. Externalities, however, may well be non-linear. In its "weaker" declination, non-linearity would still preserve monotonicity. In the case of localized knowledge spillovers, for instance, the advantages from having an additional neighbor may increase up to a critical threshold and then keep stable. In this case, externalities are non-decreasing and thus monotone, but they are non-linear. In a "stronger" declination of non-linearity, instead, monotonicity is lost. For example, the advantages of having an additional neighbor may increase up to a critical threshold and decrease beyond it, for instance because firms may incur into growing search costs to select and exploit the relevant pieces of productive knowledge. In this case, firm localization is subject to congestion and externalities are non-monotone.

The present work extends the localization model originally presented in Bottazzi et al. (2007) by allowing for quadratic externalities. In carrying out such an extension, two key characteristics of the original framework are entirely preserved. First, it remains possible to clearly disentangle localized externalities from the effect of other location-specific variables. Second, the localization choices of firms converge in the long-run toward a stationary distribution which is also ergodic. In this sense, the model prescribes the same stochastic equilibrium regardless of the initial distribution of firms across regions. On the other hand, the introduction of non-linear externalities prevents us from deriving the equilibrium distribution in closed form. Hence, we deploy numerical simulations to derive the equilibrium distribution. of firms across regions. Such a theoretical distribution is then compared to the observed one through χ^2 minimization, so as to estimate the parameters of the model. The main objective of the estimation is to test whether the quadratic externality coefficient is statistically different from zero. If not, localized externalities would result to be sufficiently well-approximated by a linear form; otherwise, they are better approximated by a quadratic shape. Notably, depending on the sign and magnitude of the estimated quadratic coefficient, the model allows to accommodate both monotone and non-monotone shapes of externalities. If the quadratic term is estimated to be statistically different from zero and sufficiently negative, a region can reach the point in which the addition of an extra firm decreases the probability for that same region to further attract other firms.

2 Model

The localization choices of firms depend on the intrinsic features of regions as well as on the distribution of other firms across regions. A class of stochastic models which is naturally suited to capture both dependencies is that of "generalized" Polya urn schemes. These models have been applied, for instance, to the description of phenomena like technological adoption and diffusion (see the early contributions in Arthur et al. (1987) and Dosi et al. (1994)). In those applications, however, the "fixed effects" provided by the intrinsic features of the objects of choice tend to progressively disappear, be it in time or space. They are as such less useful to describe cases of persistent fluctuations. Moreover, they often present non-ergodic dynamics, which imply a substantial impossibility to estimate their parameters from the data. For these reasons, Bottazzi and Secchi (2007) and Bottazzi et al. (2008) modified the original urn scheme framework toward a Markov process known in the physics literature as Ehrenfest-Brillouin (E-B) model. The E-B model can be obtained under rather general conditions from discrete utility theory, assuming a linear externality effect very similar to the one adopted in Brock and Durlauf (2001a,b). The equilibrium distribution it entails is in fact equivalent to a finite time Polya process; but, being ergodic, it can be directly estimated from the data. Notably, ergodicity allows to overcome the concerns on identification presented in Blume et al. (2011). In this sense, the present paper introduces an extension of the E-B model to include quadratic externalities while still allowing to estimate them in high dimensional problems.

A population of N firms has to choose among L regions, and each generic firm i does so by maximizing individual utility, u_i. Labeling as n_l the number of firms that have chosen region l and as $\boldsymbol{n} = (n_1, \ldots, n_L)$ the corresponding occupancy vector, the utility that firm i associates to region l depends on the choices made by other firms via the function $u_{i,l}(\boldsymbol{n})$. In this sense, the choices of firms are interdependent.

If $u_{i,l}(\boldsymbol{n})$ were deterministic, an equilibrium would correspond to an occupancy vector $\boldsymbol{n}^*$ toward which individual choices converge from any sufficiently small perturbation $\boldsymbol{n}^* + \boldsymbol{\epsilon}$. In the present case, however, the utility function is stochastic due to an heterogeneous component in the preferences of firms. In line with other strands of literature, this heterogeneous component is regarded as an unobservable which enters the utility function in the form of a stochastic term (see Anderson et al., 1989, McFadden, 1984, Thurstone, 1927). Hence, individual utility is given by the stochastic function $u_{i,l} = g_l(\boldsymbol{n}) + \varepsilon_{i,l}$, where the terms $\varepsilon_{i,l}$ are randomly and independently extracted from the common distribution $F(\varepsilon)$. Thus, the overall attractiveness of region l depends partly on the occupancy n through the common (to all firms) component $g_l(\boldsymbol{n})$, and partly on the idiosyncratic component $\varepsilon_{i,l}$. Under broad assumptions on the shape of the distribution $F(\varepsilon)$, Bottazzi and Secchi (2007) show that the probability p_l for region l to be chosen is independent from the idiosyncratic term $\varepsilon_{i,l}$. The proof follows either from Yellott (1977, Th.6 §4) and Luce (1959, Axiom 2.1) or from Raouf Jaibi and ten Raa (1997). Especially in the latter case, it suffices that $F(\varepsilon)$ has an upper tail decaying faster than the exponential distribution (like for the Gaussian case) for p_l to be independent from $\varepsilon_{i,l}$. As long as the shape of $F(\varepsilon)$ meets these minimal requirements, it can be shown that region l is chosen by a random member of the population with probability

$$p_l = \frac{g_l}{\sum_{j=1}^{L} g_l} \; . \tag{1}$$

Therefore, the individual decision process is entirely characterized, in probability, by the vector of common utility components $\boldsymbol{g} = (g_1, \ldots, g_L)$.

Notice that, in this stochastic context, the notion of equilibrium differs substantively from its deterministic counterpart. In the latter instance, equilibrium corresponds to a situation in which no single firm has any incentive to switch from one region to another. In this sense, a deterministic equilibrium does not contemplate any turbulence in the individual choices of firms. In the stochastic case, conversely, equilibrium is achieved if the *distribution* of firms across regions is stationary. Thus, equilibrium does not rule out the possibility of individual turbulence: if the number of firms switching from region l to region m is offset by an equal number of firms moving in the opposite direction, there would be individual turbulence underneath a stationary distribution.

When $g_l(\boldsymbol{n})$ is linear in n_l, this model entails an ergodic distribution characterized by a Pólya shape (see Bottazzi and Secchi, 2007). In this case, the existence of a stochastic equilibrium can be demonstrated either by proving the attainment of a detailed balance condition, or by studying the convergence of sequential choices (see Scalas et al., 2006). But when $g_l(\boldsymbol{n})$ is not linear, we are unable to determine analytically the detailed balance condition. Therefore, we turn to the second approach by assuming a sequence of choices to then study numerically their convergence properties.

Let us assume that the common utility component g_l depends quadratically on the number of firms n_l that chose region l. Formally,

$$g_l = a_l + bn_l + cn_l^2 \ , \tag{2}$$

where $\boldsymbol{a} = (a_1, \ldots, a_L)$ is and l dimensional vector assumed with positive components and $b \geq 0$ and c are scalars. According to equation (2), the part of utility that is common to all firms has two distinct components. On the one hand, the term n_l establishes an interdependence between the choice of firm i and those made by other firms, thus capturing externalities. On the other hand, the term a_l is meant to capture the intrinsic features of region l.

The sequence of individual choices is structured so that, at each time step, one firm is selected at random to revise its choice. In general, also other non-random rules could be adopted to select who is called to operate a revision, thus affecting the dynamics of the model. In the present case, however, the aim is to keep the structure of selection as agnostic as possible precisely by attributing to all firms an equal probability to be selected for choice revision. Under this premise, the evolution of the system from configuration $\boldsymbol{n}$ at time t to configuration $\boldsymbol{n}'$ with $n'_m = n_m - 1$, $n'_l = n_l + 1$ and $n'_k = n_k$ for any $n \neq l, m$ at $t+1$ is defined in terms of the transition probability $P\{\boldsymbol{n}'_{t+1}|\boldsymbol{n}_t\}$. Such probability corresponds to the intersection between the event "Firm revises its previous choice of region m" (event B) and the event "Firm chooses region l" (event A). That is, the conditional probability reads

$$\begin{aligned} P\{\boldsymbol{n}'_{t+1}|\boldsymbol{n}_t\} =& \Pr\{\text{Firm revises its choice } m\}\cdot \\ & \cdot \Pr\{\text{Firm chooses region } l \mid \text{Firm revise its choice } m\} \ . \end{aligned}$$

Since the firm that is called to revise its current choice m is selected at random, it follows that $\Pr\{B\} = n_m/N$. This probability must be then multiplied by the probability p_l to select region l (as given by equation 2) conditional to the fact that the firm is no longer among those opting for m, that is without considering self-interaction:

$$P\{\boldsymbol{n}'|\boldsymbol{n}\} = \frac{n_m}{N} \frac{a_l + b(n_{l,t} - \delta_{l,m}) + c(n_{l,t} - \delta_{l,m})^2}{\sum_{l=1}^{L} \{a_l + b(n_l - \delta_{l,m}) + c(n_l - \delta_{l,m})^2\}} \tag{3}$$

where the Kronecker term $\delta_{l,m}$ is 1 if $l = m$ and 0 otherwise. Hence, the term $\delta_{l,m}$ is what makes the expression above a *conditional* probability. With the previous assumptions one has the following:

Proposition 2.1. *Assume that a_l, b and c are such that for any occupancy vector $\boldsymbol{n}$ the probability of choosing any location is positive, that is $p_l(\boldsymbol{n}) > 0, \forall l$. Then, the Markov chain is irreducible.*

Proof. For any system configuration $\boldsymbol{n}$, let m be a region such that $n_m > 0$ and consider the configuration $\boldsymbol{n}' = \boldsymbol{n} + \delta_l - \delta_m$ obtained by adding one firm among those who chose l while removing one firm from those who chose m. Then, the one-step probability $P\{\boldsymbol{n}'|\boldsymbol{n}\}$ to move from configuration $\boldsymbol{n}$ to configuration $\boldsymbol{n}'$ is given by equation 3. Since by hypothesis $P\{\boldsymbol{n}'|\boldsymbol{n}\} > 0$, any configuration can be reached from any other configuration in a finite number of steps. The statement follows. □

Since the Markov chain associated to the non linear model is irreducible, it is also ergodic. Namely, there exists a unique distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, b, c)$ which depends on the set of parameters $(\boldsymbol{a}, b, c)$, such that for any realization $\{\boldsymbol{n}_t\}$ of the process and irrespectively of the initial configuration $\boldsymbol{n}_0$ it is

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \delta_{\boldsymbol{n}_t,\boldsymbol{n}} = \pi(\boldsymbol{n}) \ . \tag{4}$$

Equation 4 is precisely the method adopted here to compute the distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, b, c)$. Notice that differently from the analytical solution of the model, the numerical approach does not require that the Markov chain is reversible and does not exploit the detailed balance condition (see Bottazzi and Secchi, 2007, for further details).

Notably, Proposition 2.1 implies a careful investigation of the region of the parameter space in which the positivity condition of the transition probability is fulfilled. This is indeed the only region in which it is guaranteed that numerical time averages converge asymptotically to the invariant distribution of the process. In this respect, it is important to clarify why the parameter $b \geq 0$ can be safely assumed to be non-negative in the context of the present analysis. As a matter of fact, the model will be used here to test the presence of quadratic externalities in firm localization. Such externalities, however, have already been estimated to be strongly positive under a linear specification. In particular, Bottazzi and Gragnolati (2015) have done so using exactly the same data that will be used in Section4. Crucially, their analysis compares two alternative nested models which differ from each other only for the presence of the linear externality term. In their procedure of model selection, the model with linear externalities turns out to systematically outperform the model without externalities, even after penalizing parametric numerosity. For this reason, it is safe to assume $b \geq 0$ in the specific context of the present work. More attention instead needs to be placed of the behavior of the model for varying values of c.

[Figure 1 about here.]

To this purpose, Figure 1 shows the behavior of p_l as a function of n_l for varying values of c. Notwithstanding the constraints imposed on c to guarantee $p_l \geq 0$, a wide range of behaviors can still be observed. Clearly, when $c = 0$ like in Figure 1a, the probability p_l grows linearly in n_l according to equation (2), thus delineating the standard Polya model exposed in Bottazzi and Secchi (2007). By converse, as c moves away from 0 like in Figures 1b–1e, the probability p_l becomes a non-linear function of n_l. At least two further aspects of Figure 1 are worth pointing out. First, having $c < 0$ is not a sufficient condition for $p_l(n_l)$ to be non-monotone. In fact, there exist negative values of c for which p_l is increasing in n_l, as shown in Figures 1d-1e. Second, the magnitude and sign of the derivative $\partial p_l/\partial n_l$ will depend on n_l. These two aspects imply that it is not immediate to determine whether individual choices are affected by a congestion effect. Defining congestion as $\partial p_l/\partial n_l < 0$, such condition occurs only for sufficiently negative values of c and for sufficiently high values of n_l, as shown in Figure 1f. In this sense, if occurring at all, congestion is not a general condition affecting all regions, but rather a condition that would be specific to some regions according to their actual occupancy.

3 Estimation and inferential analysis

Under the assumption that the model described in Section 2 represents the true data generating process, its estimation consists in selecting the set of parameters $(\boldsymbol{a}, b, c)$ that generate the equilibrium occupancy vector $\boldsymbol{n}'(\boldsymbol{a}, b, c)$ which is most statistically "compatible" with the observed occupancy vector $\boldsymbol{n}_o$. Given this general intuition, the following exposition will make explicit reference only to the parameter c while assuming that $\boldsymbol{a}$ and b are known, so as to simplify the notation. Hence, the generic equilibrium occupancy vector that is predicted by the model will be labeled as $\boldsymbol{n}'(c)$.

[Figure 2 about here.]

To find the particular value of c that generates the occupancy vector $\boldsymbol{n}'(c)$ which is closest to the observation $\boldsymbol{n}_o$, the two occupancies are compared through an objective function Θ. In particular, we resort to the minimization of a distance function between the occupancies as defined by the χ^2 statistics. Yet, such function is not applied directly on the two vectors under comparison, but rather on their occupancy classes. Let $f(n)$ denote the number of regions chosen by n firms. For instance, $f(0)$ is the number of regions chosen by zero firms, $f(1)$ is the number of regions selected exactly by one firm, and so on. It follows that the sum of all these quantities is equal to the number of available regions, that is $\sum_{n=0}^{N} f(n) = L$. One then identifies a finite partition of the integers, that is a set of increasing integer numbers $\{y_1, y_2, \ldots, y_J\}$, such that each of the resulting classes $C_j = [y_j, y_{j+1})$ counts roughly the same number of regions. Formally, $f(C_j) = f(C_i)$ with $f(C_j) = \sum_{n \in C_j} f(n)$. The histogram plotted with respect to these classes is almost flat, since each bin counts approximately the same number of occurrences (see Figure 2b). Classes are computed on the observed configuration $\boldsymbol{n}_o$, and then they are maintained constant for all other simulated configurations. By doing so, it is ensured that the bins are chosen according to the real data while the cost of defining new classes is limited to the moment in which a new observation $\boldsymbol{n}_o$ is considered. Moreover, this also allows to have an immediate visual hindsight on how the observed and simulated configurations may differ (see Figure 2c).

Given this definition of occupancy classes, the observed and the predicted distributions of firms across regions are compared according to the χ^2 statistics

$$\Theta\left(\boldsymbol{n}'(c), \boldsymbol{n}_o\right) = \chi^2 = \sum_{j=1}^{J} \frac{(h_{o,j} - h_{e,j})^2}{h_{e,j}} \tag{5}$$

where $h_{o,j}$ is the frequency of class j for the observed occupancy vector $\boldsymbol{n}_o$ and $h_{e,j}$ is the corresponding class frequency for the predicted occupancy. With respect to the latter, we do not consider just the "final" occupancy obtained after a certain number of time periods; rather, we regard the average occupancy over a set of time periods $t_{min} + T$ in which the stochastic process has already fully reached its long-run stability. Hence, the class frequency for the predicted distribution is

$$h_{e,j} = \frac{1}{T} \sum_{t=t_{min}}^{t_{min}+T} h_{t,j} \ , \tag{6}$$

where t_{min} is the minimum number of time steps needed to reach convergence, T is the number of time periods of which the average is computed, and $h_{t,j}$ is the frequency of class j for the occupancy vector predicted by the model at time $t \in (t_{min}, t_{min} + T)$.

As usual in inferential analysis, the estimation of the unknown parameter c rests on two steps. First, there is a (numerical) search for the particular value $\hat{c}$ that generates the particular equilibrium occupancy $\boldsymbol{n}'(\hat{c})$ which is closest to the observed occupancy $\boldsymbol{n}_o$ according to the χ^2 distance measure. The search takes place within a predetermined set $[c_{min}, c_{max}]$, which contains 0. Second, the null hypothesis $H_0 : c = 0$ undergoes statistical testing against the alternative $H_1 : c \neq 0$. This means to evaluate whether the null hypothesis $c = 0$ can be rejected, at a given level of statistical significance, based on the observation $\boldsymbol{n}_o$. To conduct the reader toward a deeper understanding of this estimation procedure, it is convenient to go through its illustration in Figure 3.

[Figure 3 about here.]

The first step consists in deriving the point estimate $\hat{c}$ as illustrated in the upper-left panel of Figure 3(a). Starting from the observed configuration $\boldsymbol{n}_o$, the model is simulated with K different values of $c \in [c_{min}, c_{max}]$ and fixed values for the other parameters. For each value of c, the model is run for a transient of t_{min} time steps which are sufficient to reach the equilibrium configuration $\boldsymbol{n}'(t_{min}; c)$ (see Section 4 for further discussion on t_{min}). Then, each configuration is further evolved for a sufficiently large number of steps T. Hence, K independent trajectories $\{\boldsymbol{n}'(t_{min}; c), \boldsymbol{n}'(t_{min} + 1; c), \ldots, \boldsymbol{n}'(T; c)\}$ are obtained for each $c \in \{c_1, \ldots, c_K\}$. The goal is then to identify the particular value $\hat{c}$ whose associated trajectory $\boldsymbol{n}'(\hat{c})$ is the closest to $\boldsymbol{n}_o$. The comparison between each simulated trajectory $\boldsymbol{n}'(c)$ and the observed occupancy $\boldsymbol{n}_o$ is performed according to the objective function Θ in equation (5). Formally, the optimization problem is

$$\hat{c}(\boldsymbol{n}_o, t_{min}, T) = \underset{c \in (c_1, \ldots, c_K)}{\text{argmin}} \ \{\Theta(\boldsymbol{n}'(c), \boldsymbol{n}_o)\} \tag{7}$$

where t_{min} is the minimum number of time steps needed to reach convergence and T is an additional number of time periods after convergence has been reached. The solution to problem (7) yields the point estimate $\hat{c}$.

The second step consists in a statistical test of the null hypothesis $H_0 : c = 0$ as illustrated in the right panels of Figure 3(b-c). Essentially, the inferential analysis relies on a Monte Carlo approach. Initially, the model is run with $c = 0$ for $s \in (1, \ldots, S)$ different realizations, thus obtaining the set of replicas $\{\boldsymbol{n}_1, \ldots, \boldsymbol{n}_S\}$. Then, the same search procedure described above is performed for each replica. Namely, each configuration $\boldsymbol{n}_s$ is further evolved for a time T with different parameter values $c \in (c_1, \ldots, c_K)$, thus obtaining a set of trajectories $\{\boldsymbol{n}'_s(c_1), \ldots, \boldsymbol{n}'_s(c_K)\}$. One of these trajectories will be closest to $\boldsymbol{n}_s$, thus being the solution of problem (7) with $\boldsymbol{n}_o$ replaced by $\boldsymbol{n}_s$. The associated value of c will be identified as the solution for the same minimization problem as (7). Since this procedure is repeated for each replica, one ends up with S independent estimates $\tilde{c}$. From these estimates it is possible to build the empirical distribution function $\hat{F}(c)$ under the null hypothesis $c = 0$. Then the two-side p-value of $\hat{c}(\boldsymbol{n}_o)$ is given by $\hat{F}(-|\hat{c}|) + 1 - \hat{F}(|\hat{c}|)$.

If focusing on the estimate of a single parameter might have eased the exposition thus far, the present estimation method can indeed rely on a multistage procedure to estimate multiple parameters. Basically, such a procedure consists in estimating each unknown parameter conditional on an initial value of the other parameter(s), thus producing an iterative cycle that stops as soon as convergence is reached. As an example, imagine that the vector of intrinsic advantages $\boldsymbol{a}$ were known, while the two externality parameters (b, c) were unknown. In order to obtain an estimate of c as described above, the numerical search method that solves problem 7 needs first to accomplish a set or realizations of the stochastic process entailed by the model. To do so, it needs to assume some value of b. Hence, at an initial stage, the estimate $\hat{c}_0$ is derived conditional on having assumed an initialization value $b = \hat{b}_0$. If $\hat{c}_0$ is such that the null hypothesis $H_0 : c = 0$ is rejected, then the estimate of the parameter b will be revised starting from the initialization value $c = \hat{c}_0$. In turn, this delivers a revised estimate $\hat{b}_1$ conditional on $c = \hat{c}_0$. Now, if $\hat{b}_1$ is statistically different from the previous initialization value $\hat{b}_0$, then also the estimate of c will be revised further. This iterative cycle stops as soon as the last estimate of each parameter is not statistically different from the initialization value adopted at the previous stage.

4 Application

This section shows how the model described in Section 2 and the related estimation method detailed in Section 3 can be applied to detect the presence of a quadratic component in the localized externalities that drive the spatial distribution of firms. More precisely, the aim is to test whether the null hypothesis $H_0 : c = 0$ is to be rejected or not.

This inferential exercise is carried out at the sectoral level using Italian census data for year 2001 (see ISTAT, 2006). To make our results directly comparable with Bottazzi and Gragnolati (2015), we use the number of plants by commuting zone to measure n_l and we adopt their same 2-3 digit NACE sectoral disaggregation. If the null hypothesis $H_0 : c = 0$ is not rejected, then the model presented in Section 2 reduces to the Polya model with linear externalities estimated by Bottazzi and Gragnolati (2015). In this case, their estimates will hold as being sufficiently accurate and quadratic externalities would turn out to be redundant to explain the observed spatial distribution of plants. Otherwise, if the null hypothesis $H_0 : c = 0$ is rejected, the results obtained under linearity would need to be revised.

According to the model in Section 2, the observed distribution of plants across commuting zones is interpreted as the equilibrium outcome of discrete localization choices on the side of firms. Consequently, the effect of non-linear externalities in shaping the observed spatial distribution of plants is detected by estimating the unknown parameter c. The analysis operates at the sectoral level, so that each economic sector taken into account is characterized by a varying number N of plants (reported in the second column of Table 1) located across the $L = 686$ commuting zones that compose Italy. In this sense, two of the parameters of the model in Section 2, namely N and L, are given directly by the data. Other parameters, instead, need to be tuned.

On the one hand, S is a precision parameter. To recall, S defines the number of stochastic realizations that determine the distribution of $\tilde{c}$. Hence, higher values of T and S will ensure greater precision, but with an increasing computational cost. For the purpose of the present application, we set $S = 1000$, which ensures a level of precision in the estimate of p-values in the order of 10^{-2}.

On the other hand, t_{min} and T are tuned in relation to the size of N. To recall, t_{min} defines the minimum number of time steps needed to guarantee the convergence of numerical simulations to the invariant equilibrium distribution $\pi = (\boldsymbol{a}, b, c)$. Whereas, T defines the number of time periods over which equilibrium time averages are computed in problem (7). Therefore, t_{min} can be interpreted as indicating how many times, on average, the N firms have to revise their individual choices in order to reach the equilibrium. As shown by Garibaldi and Scalas (2010), the rate of approach to equilibrium as a function of N is

$$r = \frac{A/b}{N(A + N - 1)} \,. \tag{8}$$

Since the time to reach the equilibrium is of the order $1/r$, we can accordingly set $t_{min} = 1/r$ time steps. In the same way, in order to use the stationarity to compute time averages we can set the necessary steps to $T = 2/r$, that guarantees the use of the ergodic property of the process.

For estimation to be feasible, it is also necessary to know the vector of intrinsic features $\boldsymbol{a} = (a_1, \ldots, a_L)$. Otherwise, according to equation (2), the model would have $L + 2$ unknown parameters, thus making estimation unfeasible. To escape this problem, a_l can be characterized as a function of $H \ll L$ variables that are meant to describe the regions at stake along some relevant dimensions. As a consequence, the number of unknown parameters reduces to $H + 2 \ll L$. To make our results fully comparable with Bottazzi and Gragnolati (2015), we assume their same Cobb-Douglas form

$$a(\boldsymbol{\beta}, \boldsymbol{x_l}) = \exp\left(\sum_{h=1}^{H} \beta_h x_{h,l} + \beta_0\right) \tag{9}$$

$$= \prod_{h=1}^{H} x_{h,l}^{\beta_h} e^{\beta_0} \, . \tag{10}$$

Substituting the specification for a in equation (9) into the common utility component g in equation (2), one can then insert the latter into the probability p_l as defined in equation (1). This yields the probability for commuting zone l to be chosen at given time step, which serves to evolve the system as described in Section 2 and obtain the simulated trajectories that are necessary for the numerical optimization described in Section 3. In the present application, rather than to fully estimate each parameter of the model, our primary objective is to test whether the addition of a quadratic externality term is statistically meaningful as compared to the linear case. Hence, we look particularly at c. To this purpose we introduce a useful normalization: both the numerator and denominator of equation (11) are divided by b. Hence, the probability p_l for commuting zone l to be chosen at given time step becomes

$$p_l = \frac{\frac{a(\boldsymbol{\beta}, \boldsymbol{x}_l)}{b} + n_l + \frac{c}{b} n_l^2}{\sum_{l=1}^{L} \frac{a(\boldsymbol{\beta}, \boldsymbol{x}_l)}{b} + n_l + \frac{c}{b} n_l^2} \, , \tag{11}$$

where $(\boldsymbol{\beta}, c)$ are the $H+1$ unknown parameters to be estimated and it is set $b = 1$ without loss of generality.

Given our focus on non-linearity, we can initialize the stochastic process using the estimates $(\hat{a_1/b}, \ldots, \hat{a_L/b})$ obtained under the linear Polya model to then check whether the numerical search converges elsewhere relative from where it started. Hence, the first step of the multi-stage estimation procedure estimates c conditional on the initialization values $(\hat{a_1/b}, \ldots, \hat{a_L/b})$ obtained under the linear Polya and $b = 1$. Only if the null hypothesis $H_0 : c = 0$ is rejected will the multistage estimation procedure be triggered.

[Figure 4 about here.]

In general, there are multiple intrinsic features of a commuting zone that may be relevant to firm localization, thus entering among the H regressors that shape $a(\boldsymbol{\beta}, \boldsymbol{x}_l)$. Factors such as the local size of final goods and labor markets, infrastructural endowments, the extent of local productive variety, as well as the presence of especially conducive local institutions may all play a role in the localization choices of firms. Indeed, all these different factors were taken into account in the original analysis of Bottazzi and Gragnolati (2015), so that the present one will use exactly the same controls in order to guarantee comparability. This means in particular that the initialization values $(\hat{a_1/b}, \ldots, \hat{a_L/b})$ are derived inserting such variable among the H regressors that shape $a(\boldsymbol{\beta}, \boldsymbol{x}_l)$. Nonetheless, it is worth recalling that population plays by far the most important role, serving as a proxy for the local size of both final goods and labor markets. This fact is discussed at length in Bottazzi and Gragnolati (2015, especially pp. 9–11 and Table

2), where the marginal effect of population in shaping the attractiveness of commuting zones is shown to be orders of magnitude stronger than any other intrinsic feature. To facilitate visual inspection of this key variable, Figure 4a illustrates the spread of population among Italian commuting zones in 2001.

5 Results

Considering that the focus here is on the detection of non-linearity, we look at the estimates of c. To provide a visual summary of the data on which estimation is based, Figure 4 illustrates the spatial distributions for two different sectors together with the spatial distribution of a key control variable such as population. The results of the estimation procedure as applied on a comprehensive set of manufacturing and service sectors are reported in Table 1. As a premise, it should be noted that the estimates can be derived for high values of both N and L, thus allowing to apply the present methodology at virtually any spatial scale or sectoral disaggregation.

The basic result we obtain is a negative one. The estimate of c turns out to be statistically different from zero only in a very small minority of the forty-three sectors under analysis. More precisely, only in four sectors is the null hypothesis $H_0 : c = 0$ rejected at a 90% confidence level, and only in two sectors it is rejected at a 95% confidence level. Hence, at the scale of commuting zones, the geography of production across the various sectors of the Italian economy in year 2001 is *not* generally characterized by quadratic localization externalities. In fact, a linear specification of the Polya model is already accurate enough to capture most of the effect of externalities on firm localization. Therefore, adding an extra firm to a location l that hosts $n_l = 10$ or to a location m that hosts $n_m = 100$ firms has approximately the same *incremental* effect on the probability for each location to further attract other firms at the next localization round. Nonetheless, all other things being equal, positive localization externalities make m more attractive than l in *absolute* terms, precisely because m hosts more firms to begin with. In fact, being systematically unable to reject $H_0 : c = 0$ implies also that congestion effects generally do not show up at this spatial scale. Hence, localization externalities are overall positive. That is why having an extra firm generally increases the attractiveness of a commuting zone.

Relatedly, the multistage procedure that would serve to estimate the other unknown parameters $(\boldsymbol{\beta}, b)$ is almost never triggered. Since c is normally not statistically different from zero, the multistage estimates of the other unknown parameters turn out never being statistically different from the initialization values obtained by estimating the linear Polya model. In this sense, the original marginal effects put froward by Bottazzi and Gragnolati (2015) are not falsified. It follows that also their conclusion about the relative strength of localization and urbanization economies result to be robust to a changing functional specification of localization externalities. Similarly, also other results in the literature that were based on a linear specification of externalities may possibly represent a sufficiently accurate estimate of the determinants of firm localization (see Black and Henderson, 1999, Desmet and Fafchamps, 2006, Devereux et al., 2004, Dumais et al., 2002, Duranton and Overman, 2005, Ellison and Glaeser, 1999, Henderson, 2003, Maurel and Sédillot, 1999, Rosenthal and Strange, 2001).

6 Conclusion

This work has presented an empirical analysis on the functional shape of localization externalities. So far, most of the econometric studies regarding the determinants of the localization of economic activities have adopted a linear specification of externalities. Such an approach may produce an inaccurate measurement of the actual strength of localization externalities, particularly if congestion costs are present. Hence, before jumping to strong conclusions about the weight of externalities relative to other determinants of firm localization, one may want to investigate the effect of an alternative function specification of externalities. In particular, the present work has adopted a quadratic specification. Our main result is of a negative kind. The quadratic externality coefficient is almost never statistically different from zero. Therefore, localization externalities can be approximated sufficiently accurately also with a linear specification.

Notably, our entire analysis is structured so as to allow for a direct comparison with Bottazzi and Gragnolati (2015). On the one hand, this approach allows to further test their results, which however cannot be generally falsified. On the other hand, our focus on comparison has led to adopt their same data. This meant estimating the quadratic Polya model at the scale of commuting zones. Nonetheless, non-linearities in firm localization could possibly be occurring at other spatial scales. In particular, even in the very few sectors in which the quadratic externality coefficient is not statistically equal to zero, its size is extremely small. This may signal that commuting zones are too coarse of a spatial scale to make non-linearities detectable at a sufficient level of statistical precision. For this reason, it could be interesting to apply the analytical framework presented here on finer spatial scales, which may allows to better capture, for instance, the effect of spatial congestion.

[Table 1 about here.]

References

S.P. Anderson, A. De Palma, and J-F. Thisse. Demand for differentiated products, discrete choice models, and the characteristics approach. *Review of Economic Studies*, 56(1):21–35, 1989.

W.B. Arthur, Yu.M. Ermoliev, and Yu.M. Kaniovski. Path-dependent processes and the emergence of macro-structure. *European Journal of Operational Research*, 30(3):294–303, 1987.

D. Black and V. Henderson. Spatial evolution of population and industry in the United States. *American Economic Review*, 89(2):321–327, 1999.

L.E. Blume, W.A. Brock, S.N. Durlauf, and Y.M. Ioannides. Identification of social interactions. *Handbook of Social Economics*, 1:855–966, 2011.

G. Bottazzi and U.M. Gragnolati. Cities and clusters: economy-wide and sector specific effects in corporate location. *Regional Studies*, 49(1):113–129, 2015.

G. Bottazzi and A. Secchi. Repeated choices under dynamic externalities. LEM Working Paper Series, September 2007. URL http://www.lem.sssup.it/WPLem/files/2007-08.pdf.

G. Bottazzi, G. Dosi, G. Fagiolo, and A. Secchi. Modeling industrial evolution in geographical space. *Journal of Economic Geography*, 7(5):651–672, 2007.

G. Bottazzi, G. Dosi, G. Fagiolo, and A. Secchi. Sectoral and geographical specificities in the spatial structure of economic activities. *Structural Change and Economic Dynamics*, 19(3): 189–202, 2008.

W.A. Brock and S.N. Durlauf. Interactions-based models. *Handbook of Econometrics*, 5:3297–3380, 2001a.

W.A. Brock and S.N. Durlauf. Discrete choice with social interactions. *Review of Economic Studies*, 68(2):235–260, 2001b.

K. Desmet and M. Fafchamps. Employment concentration across US counties. *Regional Science and Urban Economics*, 36(4):482–509, 2006.

M.P. Devereux, R. Griffith, and H. Simpson. The geographic distribution of production activity in the UK. *Regional Science and Urban Economics*, 34(5):533–564, 2004.

G. Dosi, Yu. Ermoliev, and Yu. Kaniovski. Generalized urn schemes and technological dynamics. *Journal of Mathematical Economics*, 23(1):1–19, 1994.

G. Dumais, G. Ellison, and E.L. Glaeser. Geographic concentration as a dynamic process. *Review of Economics and Statistics*, 84(2):193–204, 2002.

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

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

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

U. Garibaldi and E. Scalas. *Finitary probabilistic methods in econophysics*. Cambridge University Press Cambridge, 2010.

J.V. Henderson. Marshall's scale economies. *Journal of Urban Economics*, 53(1):1–28, 2003.

ISTAT. Atlante statistico dei comuni. Roma, 2006.

P.R. Krugman. Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499, 1991a.

P.R. Krugman. *Geography and trade*. MIT Press, Cambridge, Mass., 1991b.

H. Luce. *Individual choice behavior*. Wiley, New York, 1959.

A. Marshall. *Principles of economics*. McMillan, London, 1890.

F. Maurel and B. Sédillot. A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, 29(5):575–604, 1999.

D. McFadden. Econometric analysis of qualitative response models. volume 2, pages 1395–1457. Amsterdam,, North-Holland, 1984.

G. Myrdal. *Economic Theory and Underdeveloped Regions*. Duckworth, London, 1957.

M. Raouf Jaibi and Thijs ten Raa. An asymptotic foundation for logit models. *Regional Science and Urban Economics*, 28(1):75–90, 1997.

S.S. Rosenthal and W.C. Strange. The determinants of agglomeration. *Journal of Urban Economics*, 50(2):191–229, 2001.

E. Scalas, U. Garibaldi, and S. Donadio. Statistical equilibrium in simple exchange games. *The European Physical Journal B-Condensed Matter and Complex Systems*, 53(2):267–272, 2006.

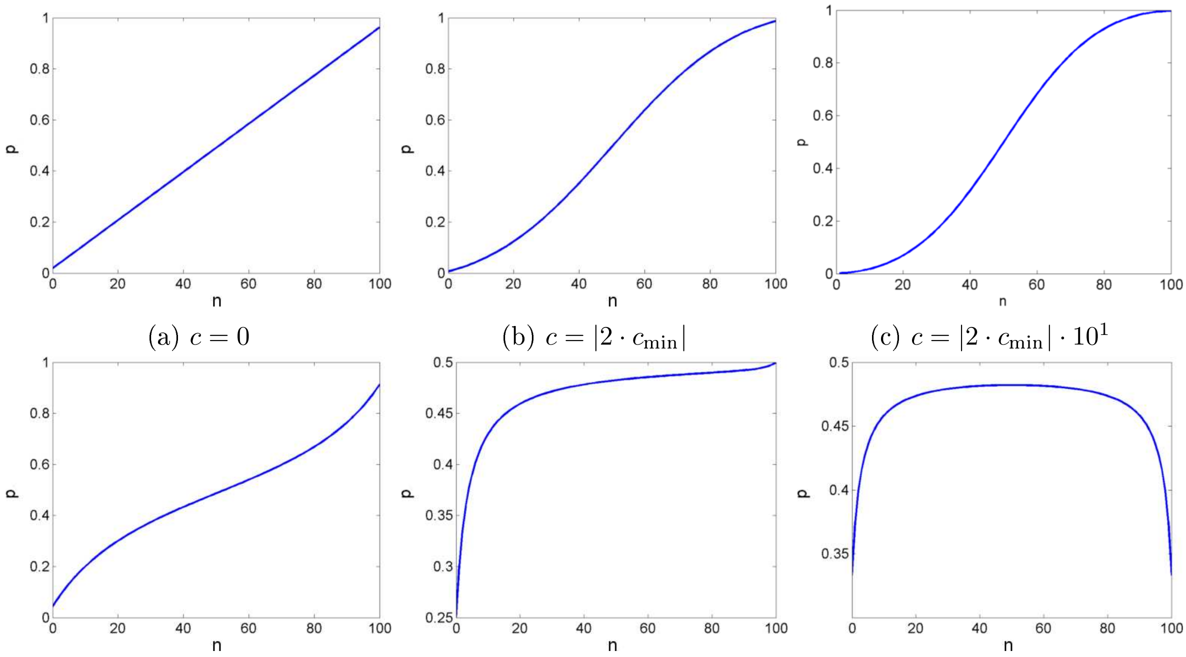
T. Scitovsky. Two concepts of external economies. *Journal of Political Economy*, pages 143–151, 1954.

L.L. Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273–286, 1927.

A.J. Venables. Equilibrium locations of vertically linked industries. *International Economic Review*, 37(2):341–359, 1996.

J.I. Yellott. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15 (2):109–144, 1977.

List of Figures

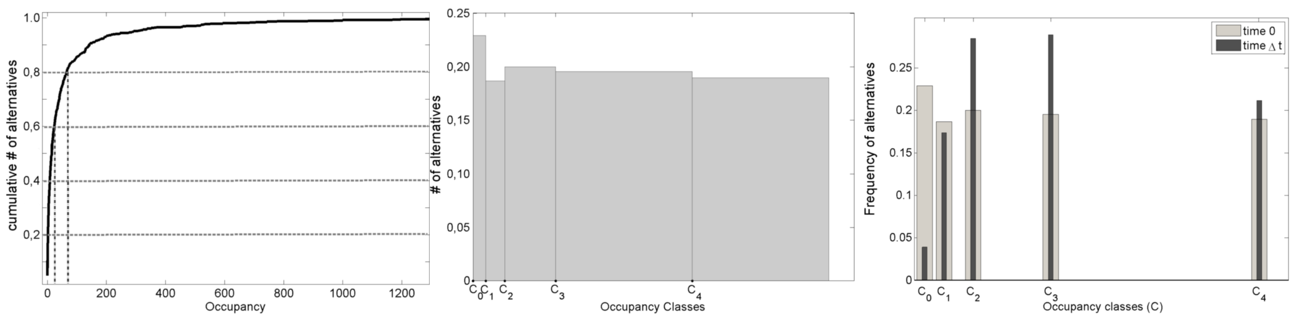


(a) $c = 0$ (b) $c = |2 \cdot c_{\min}|$ (c) $c = |2 \cdot c_{\min}| \cdot 10^1$

(d) $c = -|2 \cdot c_{\min}| \cdot 10^2$ (e) $c = -|0.98 \cdot c_{\min}|$ (f) $c = c_{\min} = -5 \cdot 10^{-3}$

Figure 1: The probability p_l as a function of n_l for varying values of c.

Note: The common parameters of these examples are $L = 2$, $N = 100$, $a_1 = 1$, $a_2 = 2$, $b = 0.5$.



(a) (b) (c)

Figure 2: Binning procedure

Note: The figure illustrates the binning procedure as operated on simulated data obtained via the Polya random generator with the following parameters: $N = 10000$, $L = 700$, $b = 1$, $c = 0$, $a_1 = a_2 = \ldots = a_L = 1$. The total number of bins is $J = 5$. The cumulative distribution in (a) serves to determine a set of bins such that the corresponding occupancy classes are of equal size. The resulting histogram shown in (b) has therefore bins of variable width. By comparing the different histograms as in (c) one gathers a first hindsight on the compatibility of two distributions. The sub-figure (c) compares in particular the Polya distribution obtained with the parameters values listed above against the distribution that is obtained by changing $c = 0$ to $c = 1 \cdot 10^{-3}$ and letting the system evolve for an additional $\Delta t = 1 \cdot 10^5$ time periods.

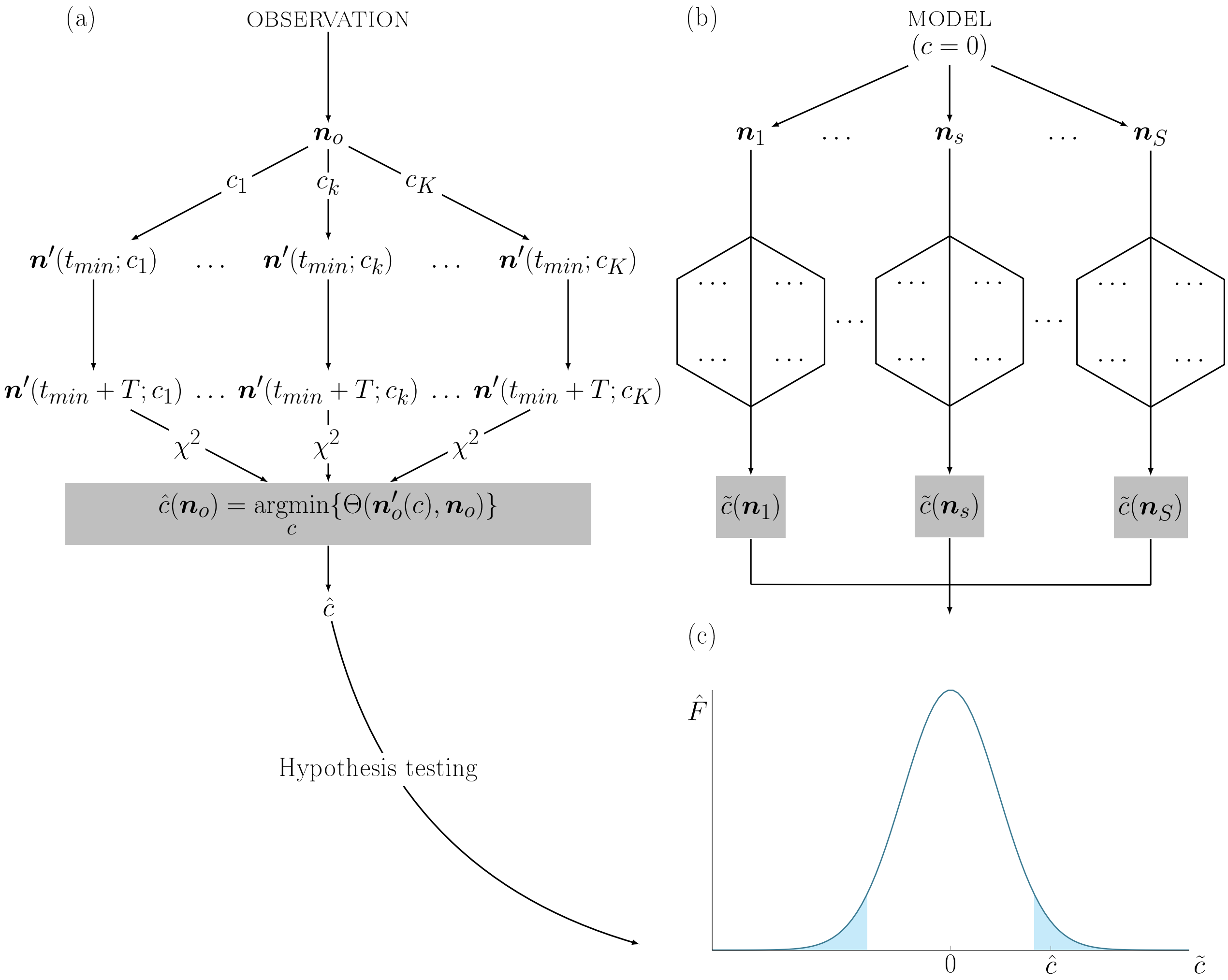
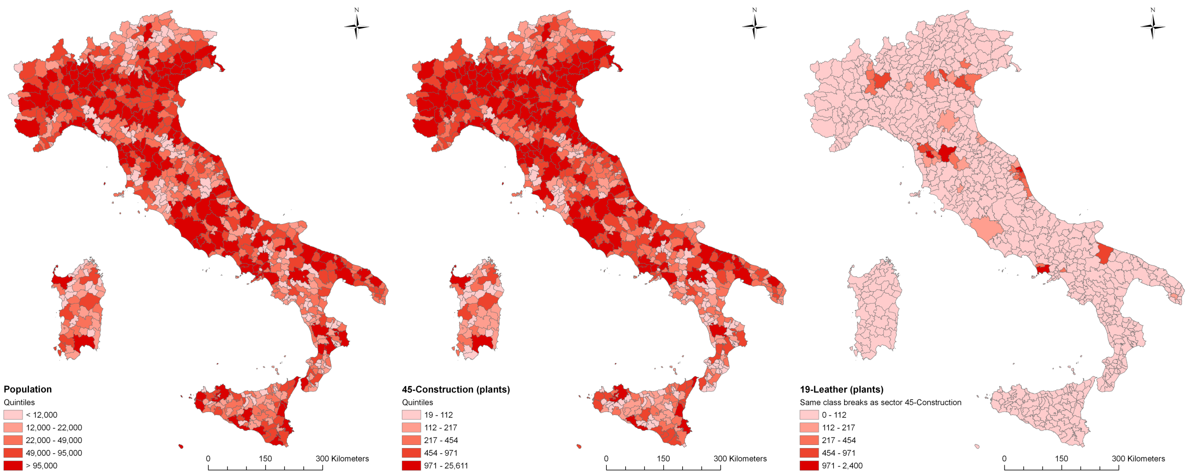


Figure 3: Estimation approach (grid method).



(a) Population. (b) Sector `45-Construction`. (c) Sector `19-Leather`.

Figure 4: Maps of the data (year 2001).

List of Tables

Table 1: Number of plants, estimate of c, and corresponding p-value.

Sectors	N	$\hat{c}$	p-value
15-Food products	73680	6.77E-006	0.460
17-Textiles	31984	-4.63E-006	0.300
18-Apparel	46377	-7.94E-006	0.360
19-Leather products	24195	1.91E-004	0.230
20-Wood processing	50250	-1.73E-005	0.360
21-Pulp and paper	5175	6.82E-004	0.190
22-Publishing and printing	29166	1.91E-005	0.500
23-Coke,petroleum and nuclear fuel	913	1.45E-003	0.110
24-Organic and inorganic chemicals	7721	-3.99E-004	0.210
25-Rubber and plastic products	15115	7.87E-005	0.025
26-Non metallic mineral products	31177	1.10E-004	0.160
27-Basic metals	3984	5.22E-004	0.410
28-Fabricated metal products	102295	-7.67E-006	0.420
29-Industrial machinery	46481	-5.68E-005	0.570
30-Office machinery	1715	-7.84E-004	0.540
31-Electrical machinery	20282	-7.12E-005	0.590
32-Radio TV and TLC devices	9677	9.27E-005	0.330
33-Precision instruments	26244	-6.38E-005	0.320
34-Motor vehicles and trailers	2229	5.41E-004	0.690
35-Other transport equipment	4951	-7.25E-005	0.730
361-Furniture	35784	5.47E-005	0.680
362-Jewelry	10906	2.02E-005	0.790
363-Musical instruments	695	-4.96E-003	0.081
36R-Residual of sector 36	6728	5.97E-008	0.710
40-Electricity and gas	4159	-6.97E-004	0.730
41-Water	1408	-5.66E-003	0.220
45-Construction	529757	1.50E-002	0.015
50-Sale and services of motor vehicles	164079	2.94E-006	0.810
51-Wholesale and commission trade	404278	3.75E-007	0.840
52-Retail trade	772730	1.80E-005	0.011
55-Hotels and restaurants	261304	-2.61E-005	0.320
60-Land transport	135135	-1.56E-005	0.690
61-Water transport	1319	-3.88E-004	0.840
62-Air transport	457	6.22E-004	0.910
63-Auxiliary transport activities	33765	-1.14E-004	0.088
64-Post and telecommunications	18056	-1.98E-004	0.540
65-Financial intermediation	30587	-1.04E-004	0.350
66-Private insurance and pensions	1771	-4.46E-004	0.860
67-Auxiliary financial activities	84677	4.73E-006	0.920
70-Real estate activities	149990	-1.57E-005	0.790
71-Renting of machinery and equipment	13291	4.41E-004	0.110
72-Computer and related activities	84100	-3.39E-005	0.210
74-Business services	216883	-2.32E-005	0.120

Note: For each sector, N is the number of plants and $\hat{c}$ is the estimate of c, which is obtained with the corresponding p-value. The number of regions is fixed to $L = 686$ commuting zones.