

Non-linear externalities in firm localization

Giulio Bottazzi, Ugo Gragnolati, Fabio Vanni

▶ **To cite this version:**

Giulio Bottazzi, Ugo Gragnolati, Fabio Vanni. Non-linear externalities in firm localization. Regional Studies, 2017, 51 (8), pp.1138 - 1150. 10.1080/00343404.2016.1237770 . hal-01405780

HAL Id: hal-01405780
https://paris1.hal.science/hal-01405780

Submitted on 9 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

January 9, 2017

Abstract

This paper presents a model of firm localization in which the intrinsic advantages of regions are disentangled from localized externalities, while this latter force is allowed to have a quadratic shape. We verify through inferential analysis whether the quadratic component of localized externalities is statistically different from zero. Such term can reflect more-than-linear positive feedbacks as well as congestion effects, so that the sign of the interdependencies stemming from localization is not assumed a priori to be positive. Our main result is that the quadratic term is virtually never statistically different from zero across Italian sectors observed at the scale of commuting zones, so that localized externalities seem to be well approximated by a linear specification.

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 also drive the firms that use such factors to mimic their spatial distribution. For example, extractive industries may be clustered around mines, gushers, or gas fields, while being virtually absent where these natural resources are unavailable. Many other industries, however, are also strongly concentrated in space despite not being bound to immobile factors of production. 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 why 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 as to promote regional development. Yet, the policies that may be thought to govern pecuniary externalities differ from those that could target technological externalities (see Martin and Sunley, 1996, Ottaviano, 2003).

Pecuniary externalities can lead to self-reinforcing dynamics and 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 symmetrical 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 with fiercer local competition. Hence, under suitable transportation costs, imperfectly competitive markets, and economies of scale, 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 by 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, bk 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 and comparing the strength of localized externalities acting *within* sectors against the pull of intrinsic regional characteristics. The idea is that if a particular sector is very concentrated where other sectors are not, this may stem either from particular local advantages or from sector-specific externalities. Once the former are controlled for, the latter can then be properly measured. In particular, the intrinsic attractiveness of locations is defined by a number of controls possibly acting *across* sectors, such as local demand, industrial variety and infrastructural advantages; yet, the controls also include a sector-specific dummy indicating the presence of an industrial district. In this way, the measurement of within-sector externalities is meant to be cleared from those pecuniary effects that may act across sectors – such as consumer demand – and from those that may be at least partly sector-specific – such as specialized labor market pooling. Hence, the type of effect that remains to be captured on the side of localized externality lends itself to be interpreted as a technological externality.

Bottazzi and Gragnolati (2015) base their analysis 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 externalities and population size are by far the most important determinants of firm localization, and these two drivers are comparable with each other in terms of their magnitudes. In fact, externalities are found to have an even stronger effect than population size across most sectors. If one interprets sector-specific externalities as being mostly technological, such a 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 (see Fujita et al., 2001). 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* [pp.61–62, italics added].

By contrast, the results discussed in Bottazzi and Gragnolati (2015) suggest that technological externalities may 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 (see Bottazzi and Dindo (2013) for a model investigating the role of technological externalities in the context of New Economic Geography core-periphery framework).

There is, however, at least one reason to suspect that such a conclusion would 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 (see Fagiolo, 2005, for an interpretation based on endogenous network formation).

The present work extends the localization model originally presented by 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 disentangle localized externalities from the effect of other location-specific variables. Second, the localization choices of firms converge in the long-run toward an ergodic invariant distribution. 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 apply numerical simulations to derive the equilibrium distribution of firms across regions. Such a theoretical distribution is then compared with the observed one through χ^2 minimization, so as to estimate its unknown the parameters. More specifically, the main objective of this procedure is to test whether the quadratic externality coefficient is statistically different from zero. If not, localized externalities would result that are 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 one 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 one extra firm decreases the probability for that same region to attract others 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 by Brock and Durlauf (2001a,b). The equilibrium distribution entailed by the E-B model is in fact equivalent to a finite time Polya process; but, being ergodic, it can be directly estimated from the data.

Notably, ergodicity allows one to overcome the concerns on identification presented by Blume et al. (2011). The present paper introduces an extension of the E-B model to include quadratic externalities while still allowing one to estimate them in high dimensional problems.

A population of N firms has to choose among L regions, and each generic firm i chooses location l with probability

$$p_l \sim a_l + bn_l + cn_l^2 \ , \tag{1}$$

where n_l is the number of firms already present in l, $\boldsymbol{a} = (a_1, \ldots, a_L)$ is an L-dimensional vector assumed with positive components and $b \geq 0$ and c are scalars. The probabilistic outcome of the choice is meant to capture the heterogeneity of firms and the consequently idiosyncratic nature of their preferences. One has to think of these probabilities as an average across all different firms composing the economy. Expression (1) has two distinct components: on the one hand, terms such qs a_l capture the intrinsic attractiveness of the different locations; and, on the other, the term n_l establishes an interdependence between the choice of firm i and those made by other firms, thus capturing externalities.

When $c = 0$, the probability in (1) is linear in n_l and the invariant distribution of the model is the multivariate Polya distribution (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 $c \neq 0$, (1) is not linear and one cannot analytically exploit the detailed balance condition to obtain the explicit expression for the invariant distribution. Therefore, the second approach is used instead by assuming a sequence of choices and then studying numerically their convergence properties.

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, alternative non-random rules could be adopted to select who is called on 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 $\boldsymbol{n}'$ at $t+1$ with $n'_m = n_m - 1$, $n'_l = n_l + 1$ and $n'_k = n_k$ for any $k \neq l, m$ 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 on 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 in (1), conditional to the fact that the firm is no longer among those opting for m, that is without considering self-interaction

$$P\{\boldsymbol{n}' | \boldsymbol{n}\} = \frac{n_m}{N} \frac{a_l + b(n_l - \delta_{l,m}) + c(n_l - \delta_{l,m})^2}{\sum_{i=1}^{L} \{a_i + b(n_i - \delta_{i,m}) + c(n_i - \delta_{i,m})^2\}} \ , \tag{2}$$

where the Kronecker term $\delta_{l,m}$ is 1 if $l = m$ and 0 otherwise. The term $\delta_{l,m}$ is what makes the above expression 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 2. 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 with 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:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \delta_{\boldsymbol{n}_t,\boldsymbol{n}} = \pi(\boldsymbol{n}, \boldsymbol{a}, b, c) \ , \tag{3}$$

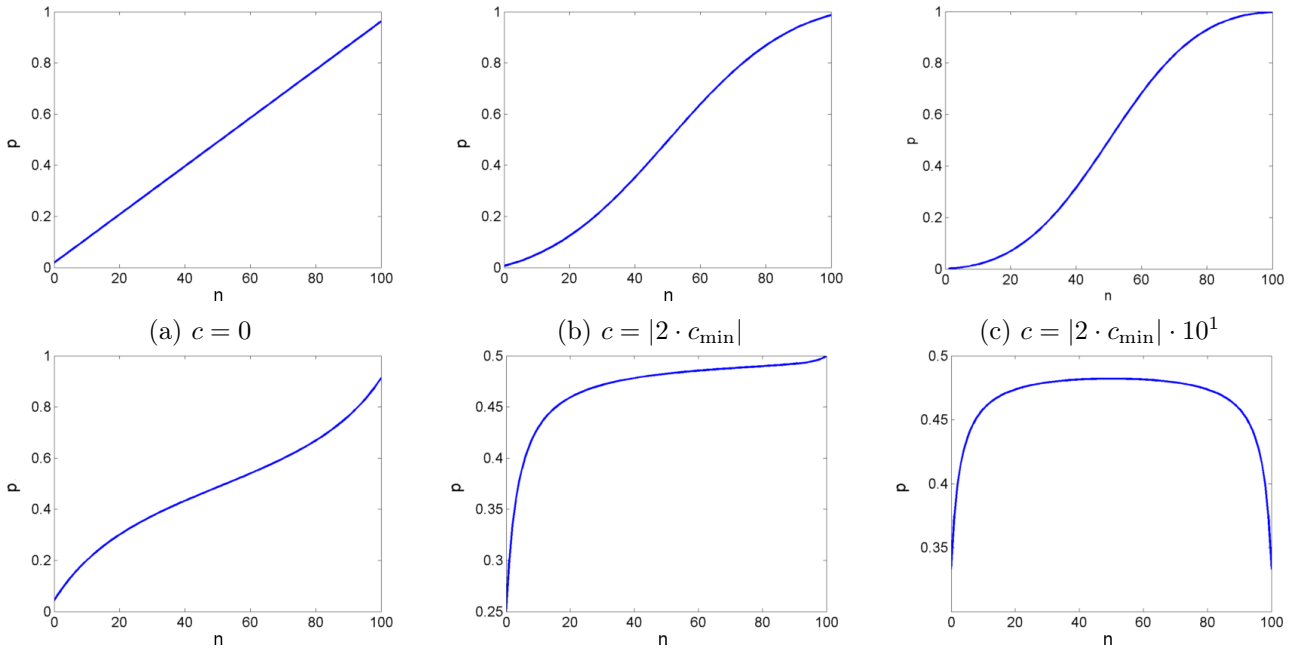
where the Kronecker term $\delta_{\boldsymbol{n}_t,\boldsymbol{n}}$ is 1 if $\boldsymbol{n}_t = \boldsymbol{n}$ and 0 otherwise. Equation 3 is precisely the method adopted here to compute the distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, b, c)$.[1]

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 can be safely assumed to be non-negative in the context of the present analysis. In 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 below in Section 4. 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 outperform systematically the model without externalities, even after penalizing parametric numerosity. For this reason, it is safe to assume that $b \geq 0$ in the specific context of the present work. Conversely, more attention needs to be placed on the behavior of the model for varying values of c.

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. When $c = 0$ as in Figure 1a, the probability p_l grows linearly in n_l according to equation (1), thus delineating the standard EhrenfestâĂŞBrillouin model. Conversely, as c moves away from 0, as 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 trivial 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.

Before moving on to the inferential analysis, it is worth making a remark on the applicability of the present model to empirical data. A time tick in the model corresponds to a localization choice rather than to a unit of real time; hence, the equilibrium distribution $\pi(\boldsymbol{n}; \boldsymbol{a}, b, c)$ describes the state of the system after the occupancy vector has been revised a sufficiently large

[1] 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.



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

(d) $c = -|\frac{c_{\min}}{2}|$ (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 $L = 2$, $N = 100$, $a_1 = 1$, $a_2 = 2$, and $b = 0.5$.

number of times. Practically, this occurs both when incumbents move their establishments and in the case of genuine entry and exit. Considering that the gross yearly turnover rate in any sector is typically between 15% and 20% in OECD countries (see Bartelsman et al., 2005, p.378), the real time needed to reach the invariant distribution may well be relatively short.

3 Inferential analysis

Under the assumption that the model described above represents the true data-generating process, its estimation consists in selecting the set of parameters $(\boldsymbol{a}, b, c)$ that generate the invariant distribution for the occupancy vector $\boldsymbol{n}(\boldsymbol{a}, b, c)$ which is most statistically "compatible" with the observed occupancy vector $\boldsymbol{n}_o$. Specifically, the parameters are estimated by minimizing the chi-square statistic of the implied invariant distribution of occupancy vectors $\pi(\boldsymbol{a}, b, c)$ with respect to the observed occupancy. The "categorization" required by the chi-square statistics is obtained by binning the locations according to their occupancy to obtain a binned occupancy distribution. More precisely, this is how the procedure works.[2]

Let $f(n)$ denote the number of regions chosen by n firms. This can be either the number observed in the data or the expected number implied by the invariant distribution of the process. Thus, $f(0)$ is the number of regions chosen by zero firms, $f(1)$ is the number of regions selected exactly by one firm, and so on. The sum of these quantities is equal to the number of available regions, that is $\sum_{n=0}^{N} f(n) = L$. Given a finite partition of the integers, that is a set of increasing integer numbers $C = \{y_1, y_2, \ldots, y_J\}$, let h_j with $j = 0, 1, \ldots, J$ denote the number of regions

[2]The use of maximum likelihood might appear as more appropriate. However, it poses enormous numerical difficulties as the simulation length necessary to assign reliable probabilities to all system configurations soon becomes soon unfeasible with the increase of the number of locations and firms. This problem is efficiently overcome by considering a binned statistics and reverting to chi-square minimization. Dubious readers are reminded that chi-square minimization in general possesses an efficient asymptotic behavior and, in any case, it is not more severely affected by small sample biases than maximum likelihood.

with occupancy in the class (or bin) $[y_j, y_{j-1})$, that is:

$$h_j = \sum_{n=y_j}^{y_{j+1}-1} f(n)$$

with $y_0 = 0$ and $y_{J+1} = +\infty$. The set $\{h_0, h_1, \ldots, h_J\}$ defines the binned occupancy distribution of the regions with respect to their occupancy. Notice that $\sum_{j=0}^{J} h_j = L$. Keeping fixed the partitions C, one can compute the observed occupancy distribution h^o using the observed distribution of firms across regions and also the theoretical occupancy $h^\pi(\boldsymbol{a}, b, c)$ implied by the invariant distribution of the occupancy vector $\pi(\boldsymbol{a}, b, c)$. Then the parameters can be estimated solving the problem

$$(\hat{\boldsymbol{a}}, \hat{b}, \hat{c}) = \underset{\boldsymbol{a},b,c}{\operatorname{argmin}} \left\{ \chi^2(h^\pi(\boldsymbol{a}, b, c), h^o) = \sum_{j=0}^{J} \frac{(h_j^\pi(\boldsymbol{a}, b, c) - h_j^0)^2}{h_j^\pi(\boldsymbol{a}, b, c)} \right\} . \tag{4}$$

To improve the efficiency of the method, we define the partition C in such a way that the different bins contain a roughly constant number of regions when the observed occupancy is considered. This implies that the histogram of the observed binned occupancy distribution is roughly flat (see Figure 2b).

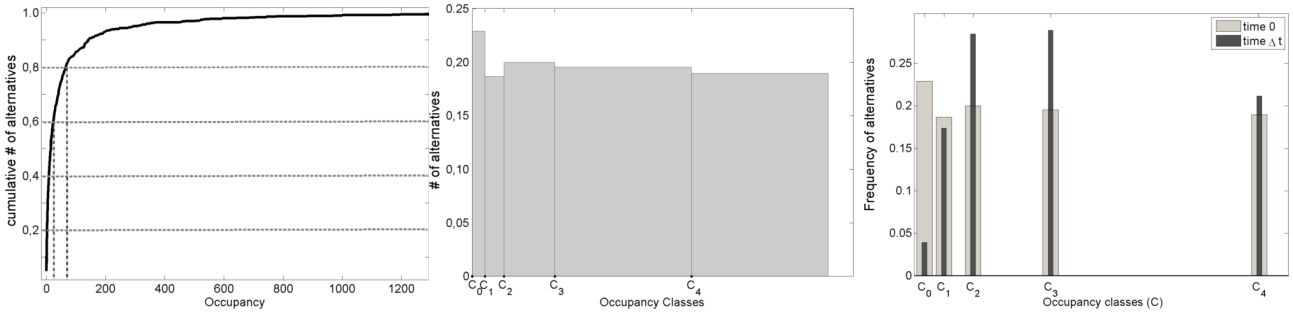
When $c = 0$, the estimation procedure can be carried out more easily because the theoretical occupancy distribution can be directly obtained from the analytic expression of the invariant distribution π (Bottazzi and Gragnolati, 2015). In the general case, however, the invariant distribution is not known and one has to revert to numerical methods, specifically Monte Carlo techniques, to obtain an estimate of the theoretical occupancy distribution.[3] Since the main interest here is to test whether or not the null hypothesis $H_0 : c = 0$ can be statistically rejected, the following exposition we 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 invariant distribution predicted by the model will be labeled as $\pi(c)$ and $h^\pi(c)$ is the implied occupancy distribution. In fact, the values of $\boldsymbol{a}$ and b are initialized according to the estimates obtained under $c = 0$ and are successively re-estimated if the null hypothesis $H_0 : c = 0$ is rejected (see below).

As usual in inferential analysis, the estimation of the unknown parameter c rests on two steps. First, one searches for the particular value $\hat{c}$ that generates the occupancy distribution which is closest, on average, to the observed occupancy h^o according to the chi-square statistics. The search takes place within a predetermined set $[c_{min}, c_{max}]$, which contains 0.[4] Second, the null hypothesis $H_0 : c = 0$ undergoes statistical testing against the alternative $H_1 : c \neq 0$. This means one has 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 understand better the estimation procedure, it is convenient to go through its illustration in Figure 3.

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 $c \in [c_{min}, c_{max}]$. For each value of c, the model is run for an initial number of steps t_{min} to eliminate transient effects (see Section 4 below for further discussion on t_{min}).

[3]The Monte Carlo techniques adopted here are characterized by very poor performances when a maximum likelihood approach is adopted. The problem arises every time the attractiveness of two locations is similar. In this case any redistribution of firms across locations generates a variation in the likelihood function which is, however, spurious. Washing away this spurious effect requires a computation time which is unfeasible for practical applications. For more details, see Bottazzi and Vanni (2014).

[4]To guarantee the convergence of the algorithm and improve its speed, the initial search interval should be tuned on the initial values of the parameters $\boldsymbol{a}$ and b. Essentially, one has to avoid that the search algorithm probes the objective function outside its domain of definition. For more details, see Bottazzi and Vanni (2014).



(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.

Then, each configuration is further evolved for a sufficiently large number of steps T. Hence, K independent trajectories $\{\boldsymbol{n}(t_{min}; c_k), \boldsymbol{n}(t_{min}+1; c_k), \ldots, \boldsymbol{n}(T; c_k)\}$ are obtained for each $c \in \{c_1, \ldots, c_K\}$. Along each trajectory the average occupancy distribution is computed:

$$h_j(c_k) = \frac{1}{T} \sum_{t=t_{min}}^{t_{min}+T} h_{t,j}(c_k) \ , \tag{5}$$

where $h_{t,j}(c_k)$ is the occupancy distribution obtained from the occupancy vector $\boldsymbol{n}(t_{min}+1; c_k)$. If t_{min} and T are large enough, this provide a good numerical approximation of the theoretical distribution $h^{\pi}(c_k)$. Finally, the value of $\hat{c}$ is chosen solving:

$$\hat{c}(\boldsymbol{n}_o, t_{min}, T) = \underset{c \in (c_1, \ldots, c_K)}{\operatorname{argmin}} \left\{ \chi^2(h(c), h^o) \right\} \ . \tag{6}$$

By successive refinements of the grid one can achieve the desired precision.[5]

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). Initially, the model is evolved 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 by considering each replica as the initial observed occupancy. In this way a set of independent estimates $\{\tilde{c}_1, \ldots, \tilde{c}_S\}$ is obtained. From these estimates, we build the empirical distribution function $\hat{F}_0(c)$. This is an approximation of the theoretical distribution of estimated parameters under the null $c = 0$, so that the two-sided p-value of $\hat{c}$ is given by $\hat{F}_0(-|\hat{c}|) + 1 - \hat{F}_0(|\hat{c}|)$.

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 (for the details on the coordinate descend method, see Bertsekas, 1999, sect. 1.2.1). 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

[5]The precision is however limited by the finite computation time. As a robustness check, a minimization algorithm based on successive parabolic interpolations was also employed, obtaining consistent results.

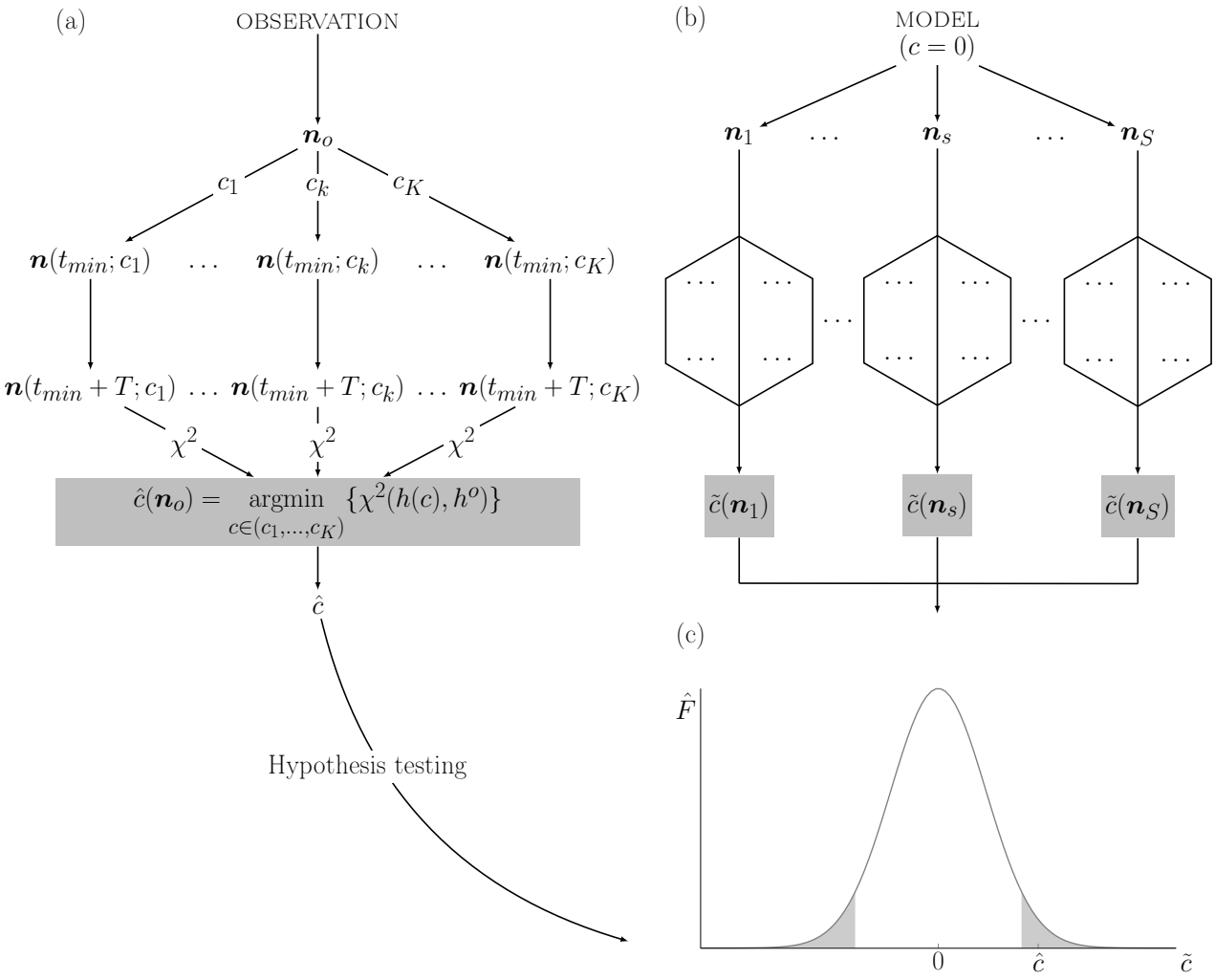


Figure 3: Estimation approach (grid method).

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. Starting with initial values (b_0, c_0), a first estimate $\hat{c}_1$ is obtained keeping the value of b fixed. If $\hat{c}_1$ is significantly different from c_0, one estimates a new value $\hat{b}_1$ keeping $c = \hat{c}_1$ fixed. If $\hat{b}_1$ is significantly different from b_0, one re-estimates c with $b = \hat{b}_1$ and so on until the procedure convergences to a couple of values $(\hat{b}, \hat{c})$.

4 Application

This section shows how the model described and the related estimation method detailed above 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 or not the null hypothesis $H_0 : c = 0$ is to be rejected in the different sectors.

This inferential exercise is carried out at the sectoral level using Italian census data for year 2001 (see ISTAT, 2006). To make the present results directly comparable with those of Bottazzi and Gragnolati (2015), the number of plants by commuting zone is used to measure n_l and their same 2-3 digit NACE sectoral disaggregation is adopted. 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, the authors 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 (6). 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

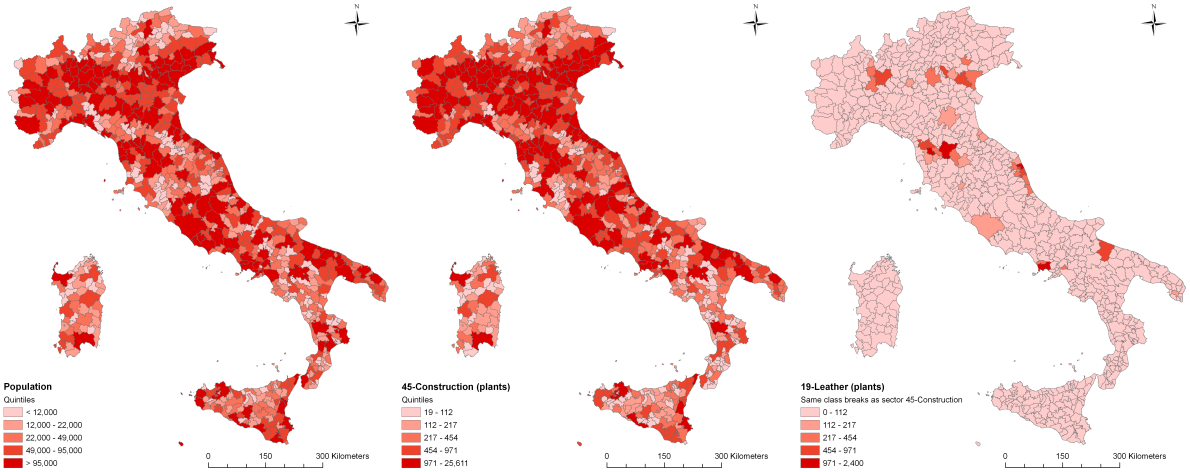
$$r = \frac{A/b}{N(A + N - 1)} \, . \tag{7}$$

Since the time to reach the equilibrium is of the order of $1/r$, one can accordingly set $t_{min} = 1/r$ time steps. In parallel, the number of time periods on which to compute time averages is set to $T = 2/r$, which 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 (1), 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 the results fully comparable with those of Bottazzi and Gragnolati (2015), their same Cobb-Douglas form is assumed:

$$a(\boldsymbol{\beta}, \boldsymbol{x_l}) = \exp\left(\sum_{h=1}^{H} \beta_h \log(x_{h,l}) + \beta_0\right) = \prod_{h=1}^{H} x_{h,l}^{\beta_h} e^{\beta_0} \, . \tag{8}$$

When substituting the specification for a in (8) into the expression of p_l in (1), one obtains 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 estimate fully each parameter of the model, our primary objective is to test whether the addition of a quadratic externality term is statistically meaningful as compared with the linear case. Hence, one looks particularly at c. To this purpose, one introduces a useful normalization: by dividing numerator and denominator by b the expression for the probability p_l of selecting



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

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

commuting zone l becomes

$$p_l = \frac{\frac{a(\boldsymbol{\beta},\boldsymbol{x}_l)}{b} + n_l + \frac{c}{b}n_l^2}{\sum_{l=1}^{L} \frac{a(\boldsymbol{\beta},\boldsymbol{x}_l)}{b} + n_l + \frac{c}{b}n_l^2}, \tag{9}$$

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

Given the focus on non-linearity, one can then initialize the stochastic process using the estimates $(\hat{a}_1/\hat{b}, \ldots, \hat{a}_L/\hat{b})$ obtained under the linear Polya model to check whether the numerical search converges elsewhere relative from where it started. Hence, the first step of the multistage estimation procedure estimates c conditional on the initialization values $(\hat{a}_1/\hat{b}, \ldots, \hat{a}_L/\hat{b})$ 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.

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 study will use exactly the same controls in order to guarantee comparability. Namely, this paper considers population, spatial extension, the presence of transportation infrastructures, aggregate consumption and average labor productivity of the location, an index of production variety plus two dummies for the presence of industrial districts and metropolitan areas (see Bottazzi and Gragnolati (2015) for further details). 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 by Bottazzi and Gragnolati (2015, esp. pp. 9–11, tab. 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, one looks 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 one to apply the present methodology at virtually any spatial scale or sectoral disaggregation.

The basic result we obtained is a negative one. The estimate of c turns out to be statistically different from zero only in a very small minority of the 43 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. These numbers are consistent with the overall validity of the null hypothesis, as they correspond to the 10% and 5% of the independent estimates, respectively. 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.

Moreover, in the rare instances in which it is significantly different from zero, $\hat{c}$ is positive. This result rules out the presence of congestion effects at the scale of commuting zones, especially if one takes into account that the condition $\partial p_l / \partial n_l < 0$ would need a *sufficiently* negative value of c in order to be met (see the related discussion above in Section 2). In fact, albeit rarely, firm localization might be affected by more-than-linear positive feedbacks, thus making the overall effect of localization externalities at the scale of commuting zones 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 to be never 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 also follows that their conclusion about the relative strength of localization and urbanization economies result to be robust to a changing functional specification of localization externalities. Similarly, other results in the literature that were based on a linear specification of externalities may also 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 paper 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, par-

ticularly if congestion effects 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. The main result is of a negative kind. The quadratic externality coefficient is almost never statistically different from zero. Therefore, localization externalities are approximated accurately enough by a linear specification and, as already discussed, by a specification that suggests the presence of positive technological externalities that increase with the number of co-located firms. If one accepts the assumption that pecuniary agglomerating factors act more or less commonly across the entire economy while technological ones are somehow limited in scope inside a specific sector, the analysis confirms a strong presence of the latter in basically all sectors under scrutiny, reinforcing the conclusions put forward by Bottazzi and Gragnolati (2015).

Notably, the entire analysis is structured so as to allow for a direct comparison with Bottazzi and Gragnolati (2015). On the one hand, this approach allows one to test their results further, which however cannot be generally falsified. On the other hand, the focus here on comparison has led to the adoption of the 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 at 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 capture better, for instance, the effect of spatial congestion.

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.

References

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.

E. Bartelsman, S. Scarpetta, and F. Schivardi. Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3):365–391, 2005.

D.P. Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

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 and F. Vanni. Non-linear externalities: A computational estimation method. LEM Working Paper Series, May 2014. URL http://www.lem.sssup.it/WPLem/files/2014-01.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.

Giulio Bottazzi and Pietro Dindo. Gobalizing knowledge: How technological openness affects output, spatial inequality, and welfare levels. *Journal of Regional Science*, 2013.

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.

G. Fagiolo. Endogenous neighborhood formation in a local coordination model with negative network externalities. *Journal of Economic Dynamics and Control*, 29(1):297–319, 2005.

M. Fujita, P.R. Krugman, and A. Venables. *The spatial economy: Cities, regions, and international trade*. MIT press, 2001.

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.

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

Ron Martin and Peter Sunley. Paul Krugman's geographical economics and its implications for regional development theory: a critical assessment. *Economic geography*, pages 259–292, 1996.

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.

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

G Ottaviano. Regional policy in the global economy: Insights from New Economic Geography. *Regional Studies*, 37(6-7):665–673, 2003.

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.

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