Bottazzi, Giulio; Gragnollati, Ugo M.; Vanni, Fabio

Working Paper

Non-linear externalities in firm locations: A computational estimation method

LEM Working Paper Series, No. 2014/01

Provided in Cooperation with:
Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Bottazzi, Giulio; Gragnollati, Ugo M.; Vanni, Fabio (2014) : Non-linear externalities in firm locations: A computational estimation method, LEM Working Paper Series, No. 2014/01, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:
https://hdl.handle.net/10419/119823

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 locations: A computational estimation metho

Giulio Bottazzi *
Ugo M. Gragnollati §
Fabio Vanni *

* Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
§ BETA, University of Strasbourg, France

2014/01 **January 2014**

Non-linear externalities in firm locations: A computational estimation method

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

[1]Scuola Superiore Sant'Anna , Institute of Economics , Pisa, Italy.
[2]BETA, University of Strasbourg , Strasbourg, France.
[3]Scuola Superiore Sant'Anna , Institute of Economics , Pisa, Italy.

January 15, 2014

Abstract

A stochastic discrete choice model and its related estimation method are presented which allow to disentangle non-linear externalities from the intrinsic features of the objects of choice and from the idiosyncratic preferences of agents. Having verified for the ergodicity of the underlying stochastic process, parameter estimates are obtained through numerical methods and so is their statistical significance. In particular, optimization rests on successive parabolic interpolation. Finally, the model and its related estimation method are applied to the case of firm localization using Italian sectoral census data.

JEL codes: C12, C13, C46, C52, R12.
Keywords: Externalities, Heterogeneity, Computational methods, Firm localization.

1 Introduction

A variety of individual choices are determined by the intrinsic features of the object of choice as well as by the choices of other individuals. For instance, a consumer might choose to buy a product partly for its qualities and partly as a function of the number of other consumers possessing the same product. This is clearly the case for network goods like telephones or social medias, as well as in any other market in which consumers have some preference for conformity. Similarly, a firm might choose where to place its plants looking also at the localization of other firms, be it to cluster with them in a Silicon Valley style, or to be as far as possible from them so as to escape local competition. In all these cases, the number of agents operating a certain choice today affects how many agents will make the same choice tomorrow, thus generating dynamic externalities among agents. Notably, as shown by the firm localization example, there is no straightforward prior about the functional shape of such externalities.

The present work introduces a framework to detect non-linear externalities, net of given intrinsic features of the object of choice. In this sense, it generally contributes to tackling the problem of identifying social interactions, as discussed in Blume et al. (2011). More specifically, the endogenous choice model developed by Bottazzi and Secchi (2007) and the associated estimation method in Bottazzi and Gragnolati (2012) are here extended to the case of quadratic externalities. While still allowing to estimate the presence of externalities in high dimensional problems, the extension to the quadratic case impedes to derive a closed-form likelihood function. Therefore, the parameter estimates are here obtained computationally, and so is their variance. In particular, point estimates are the result of χ^2 minimization resting on successive parabolic interpolation (see Brent, 1973), while their variance is estimated through Monte Carlo simulations. It follows that the present framework is able to evaluate the statistical significance of the parameter estimates by evaluating their corresponding p-value. This entire methodology is included in the MATLAB (2011) functions `minoptfun_SQA.m` and `minoptfun_GRID.m` that accompany this paper.

After a general presentation of the underlying endogenous choice model and of its corresponding estimation method, this paper presents also an application to the case of firm localization. Using sectoral census data on Italian commuting zones, the aim is to detect whether the localization of firms across commuting zones displays quadratic externalities. In this case, a positive estimate of the quadratic parameter would signal that the probability for a firm to choose a certain location grows more than linearly in the number of firms already settled in the same location. Vice versa, a negative estimate would signal the presence of spatial congestion.

2 Model

At each time step, one of N agents is chosen at random to die and make room for a new entrant. Given an individual utility function $u_i(\cdot)$ mapping into $\mathbb{R}_{\geq 0}$, the entrant i chooses the option l_i^* that satisfies

$$l_i^* = \underset{l}{\operatorname{argmax}} \left\{ u_i(g_l, \varepsilon_{l,i}) \mid l \in \{1, ..., L\} \right\} \ , \tag{1}$$

where L is the number of available options, g_l is a component common to all agents, and ε_l is a random variable capturing preference heterogeneity across the N agents. Following either Yellott (1977, Th.6 §4) and Luce (1959, Axiom 2.1) or Raouf Jaibi and ten Raa (1997), Bottazzi and Secchi (2007) demonstrate that the probability p_l for the new entrant to select option l is

$$p_l = \frac{g_l}{\sum_{j=1}^{L} g_l} \ , \tag{2}$$

which is independent from the idiosyncratic preferences $\varepsilon_{l,i}$. Therefore, the result of the decision process is entirely characterized, in probability, by the vector $\boldsymbol{g}$.

It follows that the different drivers of individual choices enter in the composition of g_l. In particular, it is postulated that g_l is a function of the intrinsic features of the object of choice

a_l, and of the number of agents that have already chosen it, $n_{l,t-1}$:

$$g_t(a_l, n_{l,t-1}) \ , \tag{3}$$

where t indexes time periods. To give some concrete examples, a_l would be determined by the physical qualities of products, if looking at consumer choices; or by the level of local demand and infrastructural endowments of regions, if looking at the localization choices of firms. On the other hand, conformity to other $n_{l,t-1}$ agents might attract or repulse the choice of the new entrant, thus determining whether externalities are positive or negative.

Generally, a_l is a multivariate function of H variables, each being a relevant intrinsic feature of the object of choice. According to this idea, a_l is here specified in the standard Cobb-Douglas form

$$a_l = \prod_{h=1}^{H} x_{h,l}^{\beta_h} \ \exp(\beta_0) \ . \tag{4}$$

This specification has a probabilistic interpretation: if, on average, the probability to choose l according to variable h is proportional to $x_{h,l}^{\beta_h}$, and if the effects of the different factors can be assumed as roughly independent, then the combined probability for the agent to choose this location is given by expression (4).

Also, the dynamics of the system can be studied by specifying a functional form for $g_l(\cdot)$. Assuming linearity in $g(n_l)$, Bottazzi and Secchi (2007) prove that the dynamical system is a finite Markov chain converging to an ergodic distribution $\pi(\cdot)$, which can be derived in closed form. This allows to estimate the underlying parameters of $\log \pi(\cdot)$ by maximum likelihood. In the present case, however, the assumption of linearity in $g(n_l)$ is abandoned in order to assume a quadratic form and test for non-linearity. The price to pay for this extension is ergodicity cannot be demonstrated analytically, nor can $\log \pi(\cdot)$ be derived in closed form. Therefore, numerical simulations will serve here to verify ergodicity, while estimating the parameters through computational methods.

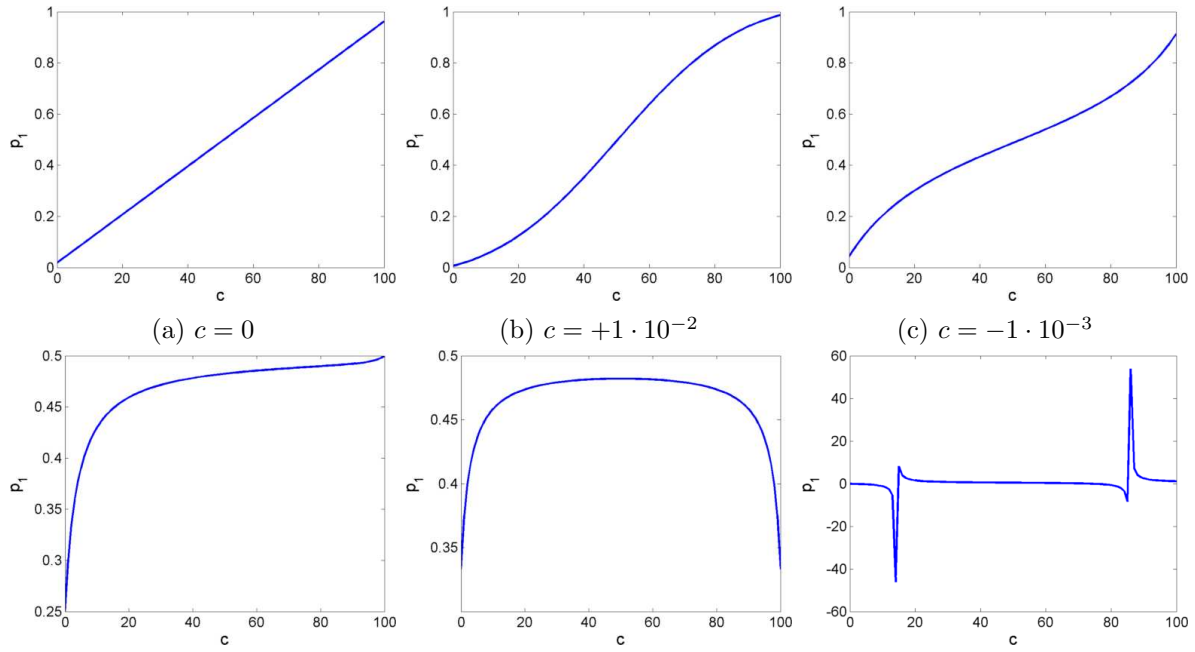
Using a quadratic specification to equation (3), at each time step the probability for l to be chosen is determined by

$$p_l = \frac{\sum_h^H \beta_h x_{l,h} + \gamma n_l + \theta n_l^2}{\sum_l^L \sum_h^H \beta_h x_{l,h} + \gamma n_l + \theta n_l^2} \ , \tag{5}$$

where (β, θ, γ) is a set of unknown parameters. Notably, when the aim is just to detect non-linearity in n rather than to measure and compare its strength relative to other variables, equation (5) can be reduced parametrically by dividing numerator and denominator by a common parameter (for instance β_1). Such a simplification reduces the number of parameters to be estimated to $H+1$. This is relevant here because the application proposed in Section 5 presents $H = 1$, so that only $\hat{b/\beta_1}$ and $\hat{c/\beta_1}$ will be estimated. Accordingly, the relevant expression for p_l becomes

$$p_l \sim x_{l,1} + b n_l + c n_l^2 \ , \tag{6}$$

where $b = \gamma/\beta_1$ and $c = \gamma/\beta_1$. Hence, at each time step, the distribution of agents across the L alternatives will be in agreement with equation (6), thus determining the configuration



(a) $c = 0$ (b) $c = +1 \cdot 10^{-2}$ (c) $c = -1 \cdot 10^{-3}$

(d) $c = -4.9 \cdot 10^{-3}$ (e) $c = -5 \cdot 10^{-3} \equiv c_{\min}$ (f) $c = -7 \cdot 10^{-3} < c_{\min}$

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

Note: The parameters of this example are $L = 2$, $N = 100$, $a_1 = 1$, $a_2 = 2$, $b = 0.5$.

$\boldsymbol{n} = \{n_1, \dots n_L\}$. Such vector constitutes the configuration of the system at each discrete unit of time.

3 Numerical simulations

According to equation (6), the present model describes the evolution of the choices of a population across alternatives in probabilistic terms. Given the vector of intrinsic features $\boldsymbol{x} = (x_{1,1}, \dots, x_{L,1})$, the realization of the stochastic process can be simulated numerically through an algorithm that is governed by the parameters (b, c). In the case of linear externalities (i.e. $c = 0$), such distribution is ergodic, which is a crucial requirement to estimate the model (see Bottazzi and Secchi, 2007). Therefore, the primary concern of this section is to verify whether ergodicity is preserved with non-linear externalities, so that the model can still have empirical applications. In fact, ergodicity turns out to be preserved even with non-linear externalities, as verified through the comparison of time averages and variances with ensemble ones. The following subsections define first the domain over which it is relevant to study the behavior of the stochastic process, then verify its ergodicity, and finally describe the computational cost of the simulations.

3.1 Domain definition

The study of non-linearity in the present model is meaningful only within a given domain of its parameter values. In particular, since all the parameters enter in the definition of the probability p_l, it must be verified that the parameter values maintain $p_l \in [0, 1]$ for all alternatives.

Moreover, since $u_l(\cdot)$ maps into $\mathbb{R}_{\geq 0}$, it must generally hold that $g_l \geq 0$ for any $l \in \{1, \ldots, L\}$. These various constraints suffice to identify the lower and upper bound of the parameter c, which inform the successive numerical simulations.

On the one hand, the lower bound of c follows straightforwardly from the constraint $g_l \geq 0$.

$$c \geq -\frac{a_l + bn_l}{n_l^2}\ , \quad \forall l = 1, \ldots, L\ , \tag{7}$$

$$\geq -\frac{\min\{a_l\} + b\max\{n_l\}}{\max\{n_l\}^2}\ , \tag{8}$$

$$c_{\min} \equiv -\frac{\min\{a_l\} + bN}{N^2}\ . \tag{9}$$

Equation (9) corresponds to "full concentration", that is when all N agents choose the same alternative l. If the condition $g_l \geq 0$ is respected for the alternative l that is chosen by all agents under this extreme scenario, then it will also be respected for all other alternatives under any other scenario. In parallel, conditions (7)–(9) imply a restriction also on the denominator of equation (5), namely

$$A + bN + c\sum_{l=1}^{L} n_l^2 \geq 0\ , \tag{10}$$

$$A - \min\{a_l\} > 0\ , \text{with } c \geq c_{\min}\ , \tag{11}$$

where $A = \sum_{l=1}^{L} a_l$.

On the other hand, the upper bound of the parameter c is obtained considering that the interest of the present analysis is on small deviations from linearity. As a consequence, the quadratic coefficient has to be of the order $|c| \lesssim 1/n_l$, thus fixing the upper bound to

$$c_{\max} \equiv \max_{l=1,\ldots,L}\left\{\frac{1}{\overline{n}_l}\right\}\ , \tag{12}$$

where $\overline{n}_l$ is the time average of n_l.

Given these constraints, Figure 1 shows the behavior of p_l as a function of n_l for varying values of c. Clearly, when $c = 0$ like in Figure 1a, the probability p_l grows linearly in n_l according to equation (5), thus delineating the standard Polya model exposed in Bottazzi and Secchi (2007). By converse, as c moves away from 0, the probability p_l becomes a non-linear function of n_l (see Figures 1b–1e). In particular, while $c < 0$ identifies some degree of congestion, such condition is not sufficient to render p_l strictly non-monotone. For instance, in Figures 1c–1d the probability p_l never decreases despite c being negative. In Figure 1e, instead, p_l decreases beyond a certain level of n_l. Notice also that Figure 1e is drawn using the lower bound $c_{\min}$ expressed in equation (7); not surprisingly, then, the values of p_l escape the probability boundaries $[0, 1]$ as soon as c drops below $c_{\min}$.

Table 1: Comparison between time averages and ensemble averages.

c	$\overline{n}_1(t)$	$\overline{n}_2(t)$	$\mathcal{R}_{11}^{1/2}(\tau)$	$\mathcal{R}_{22}^{1/2}(\tau)$	$\langle n_1 \rangle$	$\langle n_2 \rangle$	$\mathrm{Var}(n_1)^{1/2}$	$\mathrm{Var}(n_2)^{1/2}$
10^{-3}	15.94	84.05	15.16	15.16	15.77	84.23	18.96	18.96
0	29.76	70.23	17.77	17.77	31.68	68.32	18.30	18.30
-10^{-3}	41.40	58.59	12.22	12.22	42.13	57.87	12.62	12.62

Note: The columns report, respectively (1) the values of parameter c; (2-3) the time averages of locations 1 and 2; (4-5) the autocorrelation of the time series for locations 1 and 2; (6-7) the ensemble averages for locations 1 and 2; (8-9) the ensemble standard deviations for locations 1 and 2. The common parameters of these examples are $L = 2$, $N = 100$, $a_1 = 1$, $a_2 = 2$, $b = 0.5$. For each value of c, computations are made over 100 realizations of the process. Time averages are computed over $T = 10^5$, while ensemble averages are computed at $\tilde{t} = 5 \cdot 10^4$.

3.2 Ergodicity

Also in the non-linear case the model may be considered an ergodic with the same approach used for the linear case in Bottazzi and Secchi (2007).

Anyway a double-check has been presented and recover this conditions numerically.

Given the relevant domain of c, it is necessary to verify whether the stochastic process that characterizes the present model is ergodic. For stochastic process to be ergodic, its moments have to consistently estimable through a single realization. In practical applications, this condition is required in particular for the first two moments. That is, the ensemble mean and variance have to be equivalent, respectively, to the average and autocorrelation of the time series.

Labeling the stochastic process with X and being $p(X)$ the corresponding marginal distribution of agents across alternatives for a given set of parameters $(\boldsymbol{a}, b, c)$, the following equations define the time average, the serial autocorrelation, and the first two moments of the stochastic process:

$$\overline{X}_t = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t')dt' \,, \tag{13}$$

$$\mathcal{R}_{XX}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t')X(t'+\tau)dt' \,, \tag{14}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} X\, p(X)dX \,, \tag{15}$$

$$\mathrm{Var(X)} = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Cov}(X, X) = \mathbb{E}[X, X] \,. \tag{16}$$

For X to be ergodic, it must satisfy

$$\overline{X}_t = \mathbb{E}[X] \,, \tag{17}$$

$$\mathcal{R}_{X,X}(\tau) = \mathbb{E}[X, X] \,. \tag{18}$$

The attainment of conditions (17)–(18) must be verified empirically. To do so, the ensemble moments are computed at a time $\tilde{t}$ at which the process has reached its equilibrium. On the

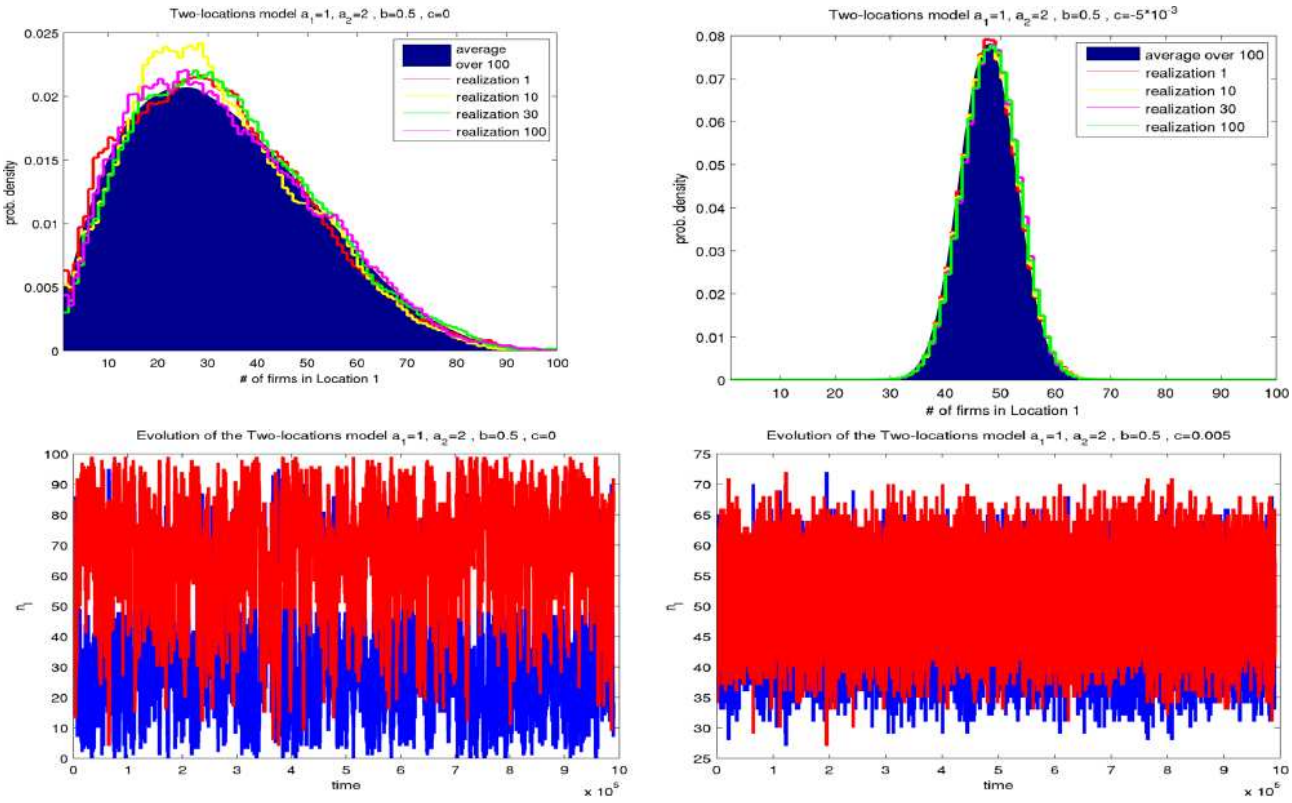


Figure 2: Distributions of agents across alternatives for different values of c.

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

other hand, time averages are computed over the entire span of the simulation $T > \tilde{t}$. As an hindsight, Table 1 shows the results of this exploration for varying values of the non-linear coefficient c. For the sake of consistency, the other parameter values underlying the results of Table 1 are the same as those in Figure 1. The first thing to notice in Table 1 are the results for $c = 0$. In this case, the model has linear externalities, and thus its ergodic distribution follows the Polya shape derived in Bottazzi and Secchi (2007). As a consequence, when $c = 0$, it is possible to have direct theoretical predictions of the ensemble averages, these being respectively $\langle \tilde{n_1} \rangle = 33.\bar{3}$ and $\langle \tilde{n_2} \rangle = 66.\bar{6}$ given the parameter values adopted in Table 1. Through these theoretical predictions one can sense how close averages actually have to be in order to be regarded as non-different for practical purposes. In particular, $\langle \tilde{n_1} \rangle = 33.\bar{3}$ is to be regarded as compatible to $\langle n_1 \rangle = 31.68$ and $\overline{n}_1 = 29.76$, as reported in Table 1 for $c = 0$. Using this benchmark, the comparison of ensemble averages with time averages in Table 1 suggests that the stochastic process underlying the present model should be compatible with condition (17). Similarly, also ensemble variances and serial autocorrelation appear to be close enough to be compatible with condition (18).

More compelling evidence on the ergodicity of the stochastic process underlying the present model comes from comparing the simulated configuration $\boldsymbol{n} = (n_1, \ldots, n_L)$, both across different realizations and among different parameter values. Figure 2 illustrates this exercise for two revealing parameter sets. In the left panels, the model is simulated for the case $c = 0$, in which ergodicity is proved analytically (see Bottazzi and Secchi, 2007). In the right panels, simulations are run tuning the non-linear coefficient to $c = c_{\min}$. Notably, the introduction

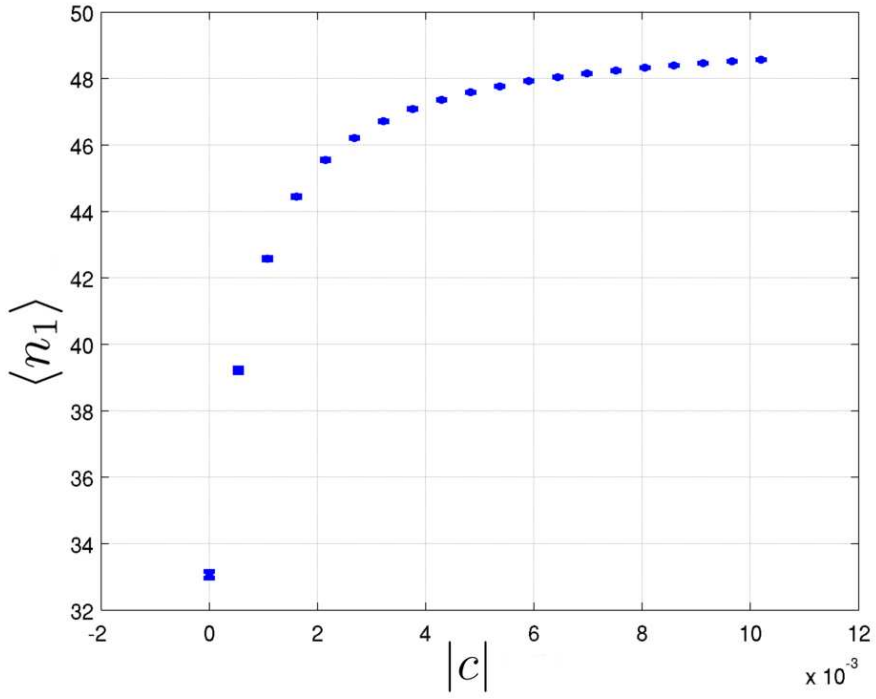


Figure 3: Average number of agents in $l = 1$ for $c < 0$.

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

of the non-linear term reduces fluctuations and stabilizes $\boldsymbol{n}$ as compared to the linear case. Therefore, it seems safe to regard the process as ergodic.

It is also worth to discuss the effect of the non-linear term on the configuration $\boldsymbol{n}$. Intuitively, the non-linear term reinforces concentration as long as $c > 0$. Conversely, the distribution among alternatives becomes gradually less skewed the more c plunges below 0. To illustrate this latter effect, Figure 3 shows the evolution of $\langle n_l \rangle$ for values of $c < 0$. Again, this exercise is carried out using the same parameter values as in Table 1.

4 Estimation

The final target of this study is to investigate whether some observed distribution of agents across alternatives can be accurately "predicted" by the discrete choice model described in Section 2. Practically, this translates in estimating the parameters (b, c) that govern the model through equation (6), while testing also whether these estimates are statistically different from zero. Therefore, the null hypothesis to be tested is $H_0 : b, c = 0$. The parameter estimates and their statistical significance are then evaluated as stylized in Figure 4.

First, the estimate of the relevant parameter, say c, is obtained from the observed configuration of agents across alternatives, $\boldsymbol{n}_o$. Starting from the configuration $\boldsymbol{n}_o$, the model is simulated for a time Δt with varying values of c, searching algorithmically for the one that generates the closest configuration to the observed one. To give an intuition, this would correspond to evolving the model from $\boldsymbol{n}_o$ into a number of new configurations governed by different values of c, so as to identify the value of c that generates the closest configuration to $\boldsymbol{n}_o$. More pre-

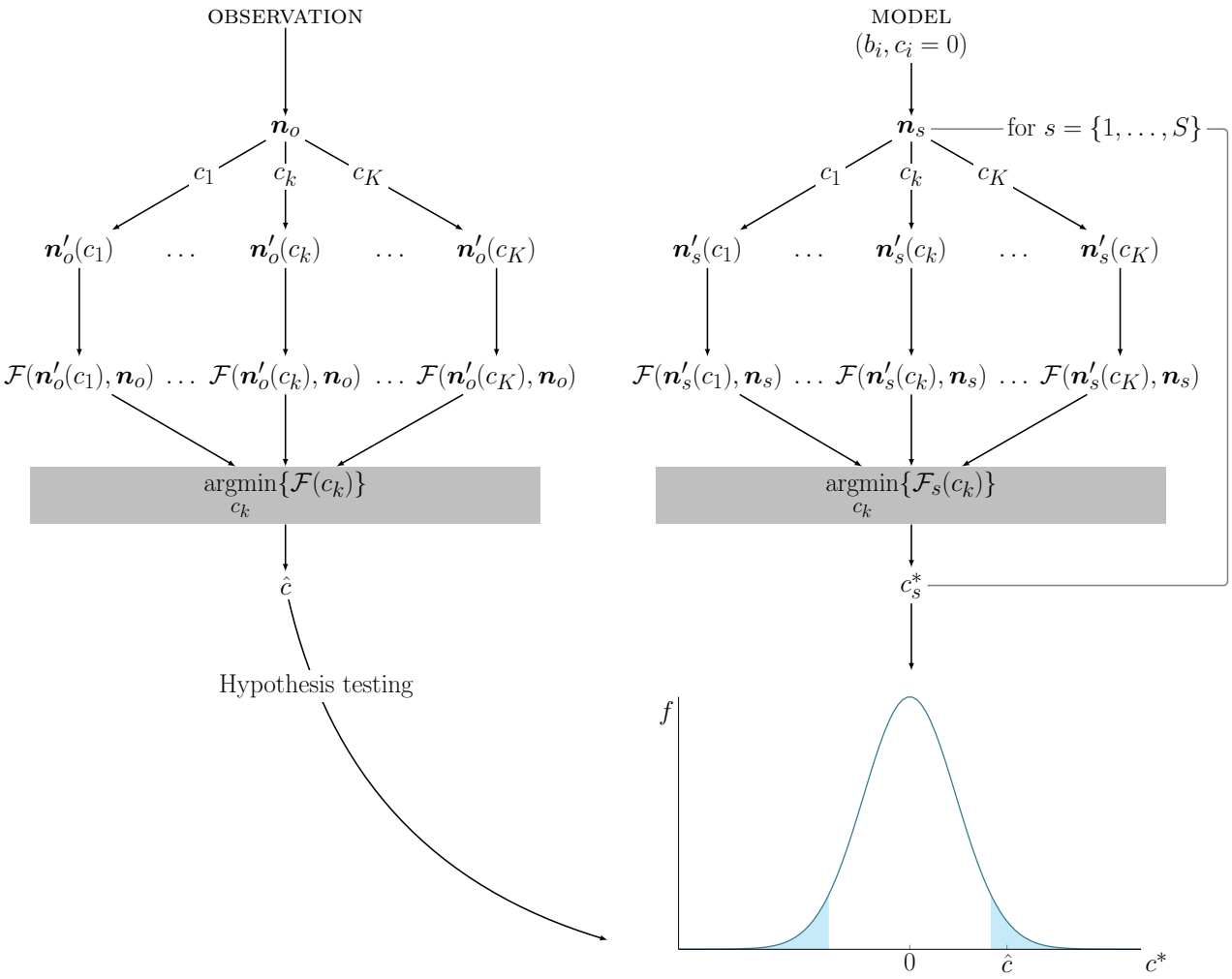


Figure 4: Estimation approach

cisely, defining $\boldsymbol{n}_o \to \boldsymbol{n}'_o = \boldsymbol{n}_o(\Delta t, c_k)$, the objective function $\mathcal{F}$ measures the distance between $\boldsymbol{n}'_o(c_k)$ and $\boldsymbol{n}_o$. The parameter estimate is then identified as

$$\hat{c} = \underset{c_k}{\text{argmin}}\{\mathcal{F}(\boldsymbol{n}'_o(c_k), \boldsymbol{n}_o)\} \quad \forall k = 1, \ldots, K. \tag{19}$$

Second, the statistical significance of $\hat{c}$ is tested through Monte Carlo simulations. Initially the model is simulated setting $c = 0$, which corresponds to the null hypothesis to be tested statistically. Under this hypothesis, the model generates the configuration $\boldsymbol{n}_s$. Starting from there, the model is further simulated for a time Δt with varying values of c, again searching for the value c^*_s that minimizes the distance of the final configuration from the initial $\boldsymbol{n}_s$. This process is repeated for S stochastic realizations, thus having

$$\begin{aligned} \boldsymbol{n}_{s=1} &\overset{\Delta t}{\to} \boldsymbol{n}_{s=1}(\Delta t) \mapsto c^*_1 \\ \boldsymbol{n}_{s=2} &\overset{\Delta t}{\to} \boldsymbol{n}_{s=2}(\Delta t) \mapsto c^*_2 \\ &\vdots \\ \boldsymbol{n}_{s=S} &\overset{\Delta t}{\to} \boldsymbol{n}_{s=S}(\Delta t) \mapsto c^*_S \ , \end{aligned} \tag{20}$$

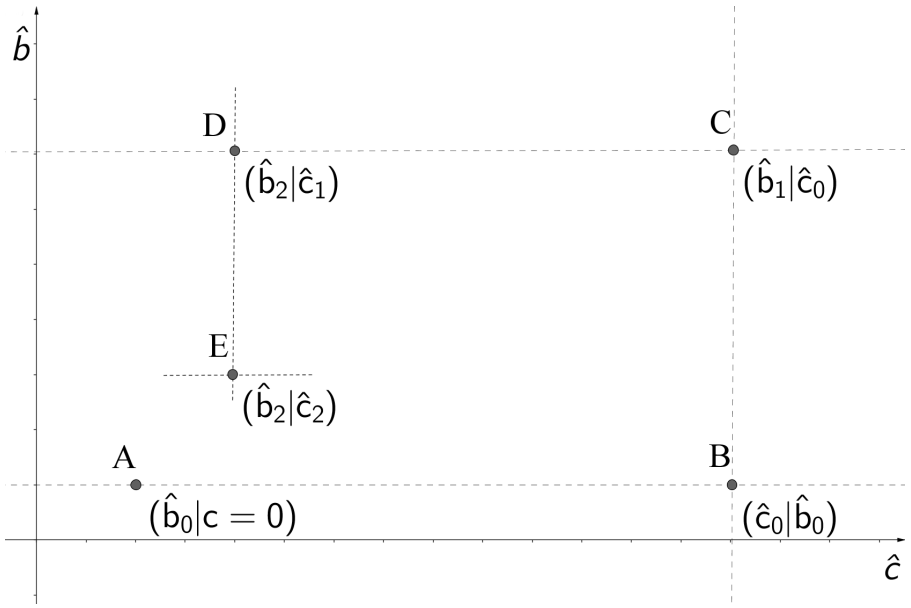


Figure 5: Multistage estimation procedure.

As a result, one ends up with a distribution of c^* revealing the probability to measure $\hat{c}$ under the null hypothesis $c = 0$. Hence, the corresponding p-value is used to test the null hypothesis.

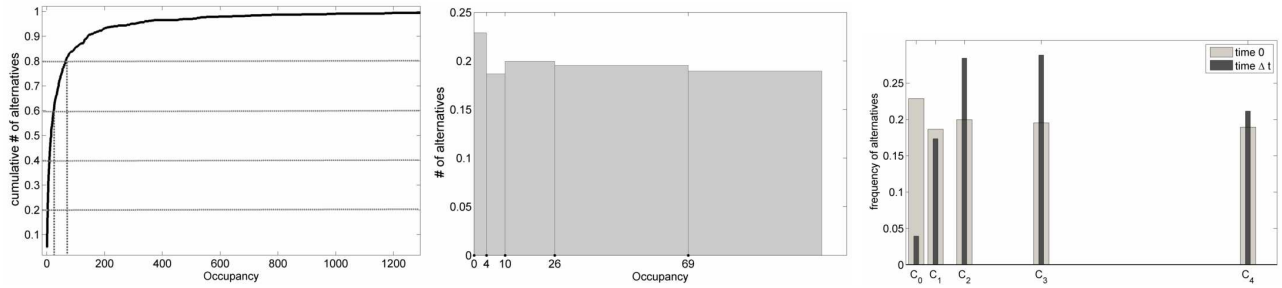
As a remark, the present estimation method relies on a multistage procedure. Generally, the kind of estimation exercises that are relevant for the model presented in Section 2 concern a multidimensional space. Each dimension corresponds to one of the parameters to be estimated, thus determining a complex parameter space. In the present case, however, the parameter space has been intentionally reduced to two dimensions to keep the exposition as simple and effective as possible without losing any substantial generality. In any case, multidimensionality leads to adopt a multistage procedure. That is, each single parameter estimate is obtained conditionally to an initial value of the other parameter(s), in an iterative cycle that stops as soon as convergence is reached. Figure 5 gives an illustration of this procedure. Point **A** identifies the estimate $\hat{b}_0$ obtained under the parametric restriction $c = 0$. From there the algorithm searches for an estimate of c under the restriction $b = \hat{b}_0$, thus reaching $\hat{c}_0$ and identifying point **B**, and so on to points **C**, **D**, and **E**. In particular, the algorithm stops when the new set of estimates is not statistically different from the set encountered at the previous search round.

Having provided a general picture of the present estimation method, it is now time to focus in greater on its two key elements, that is the objective function and the optimization algorithm through which $\hat{c}$ and c_s^* are selected.

4.1 Objective function

As mentioned above, the objective function $\mathcal{F}$ serves here to measure the distance between two configurations of the system. In particular, this happens here through the χ^2 function applied on occupancy classes.

An occupancy $f(n)$ is defined as the number of alternatives chosen by n agents. For instance, $f(0)$ is the number of alternatives not selected by any agent, $f(1)$ is the number of alternatives



(a) (b) (c)

Figure 6: Binning procedure.

selected exactly by one agent, and so on. It follows that the sum of all occupancies is equal to the number of available alternatives, that is $\sum_{n=0}^{N} f(n) = L$. Occupancies are then grouped into classes $C_1, \ldots, C_J$ of variable width and constant size:

$$\begin{cases} C_j = [c_j, c_{j+1}) & j = 0, 1, \ldots, J, \\ f(C_j) = \sum_{n \in C_j} f(n), & \\ f(C_j) = f(C_i) & \forall i, j = 0, 1, \ldots, J, \end{cases} \tag{21}$$

The resulting histograms tends to be flat, since each bin counts the same number of occurrences (see Figure 6). Notice that classes are computed only on the observed configuration $\mathbf{n}_o$, 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 $\mathbf{n}_o$ is considered. Moreover, this also allows to have an immediate visual hindsight of how two configurations may possibly differ (see Figure 6c). Given this definition of occupancy classes, the objective function used to compare configurations reads

$$\mathcal{F} \equiv \chi^2(\Delta t) = \sum_{j=1}^{J} \frac{\left(h_j^{(\Delta t)} - h_j^{(0)}\right)^2}{h_j^{(0)}} \tag{22}$$

where $h_j^{(0)}$ is the frequency of class j at time 0, $h_i^{(\Delta t)}$ is its counterpart at time Δt, and J is the total number of classes.

4.2 Optimization methods

Problem (19) can be solved numerically searching for the argument that minimizes χ^2, either via the classic grid method or via successive parabolic interpolation. It is worth to describe both approaches in some detail to motivate why one is preferable relative to the other in the applications that can be of interest for the model presented here.

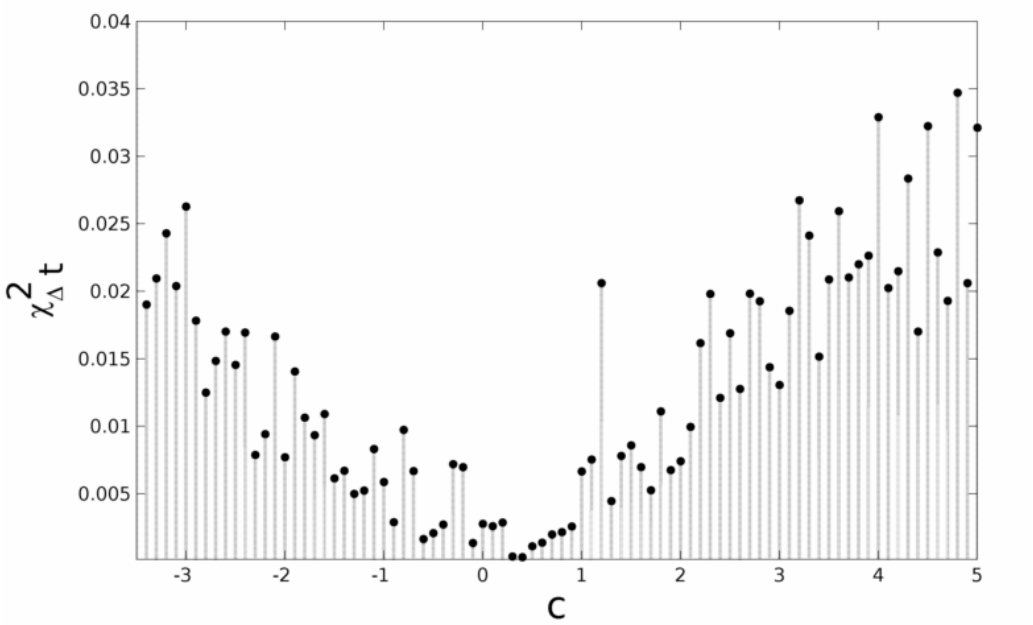


Figure 7: Grid search method.

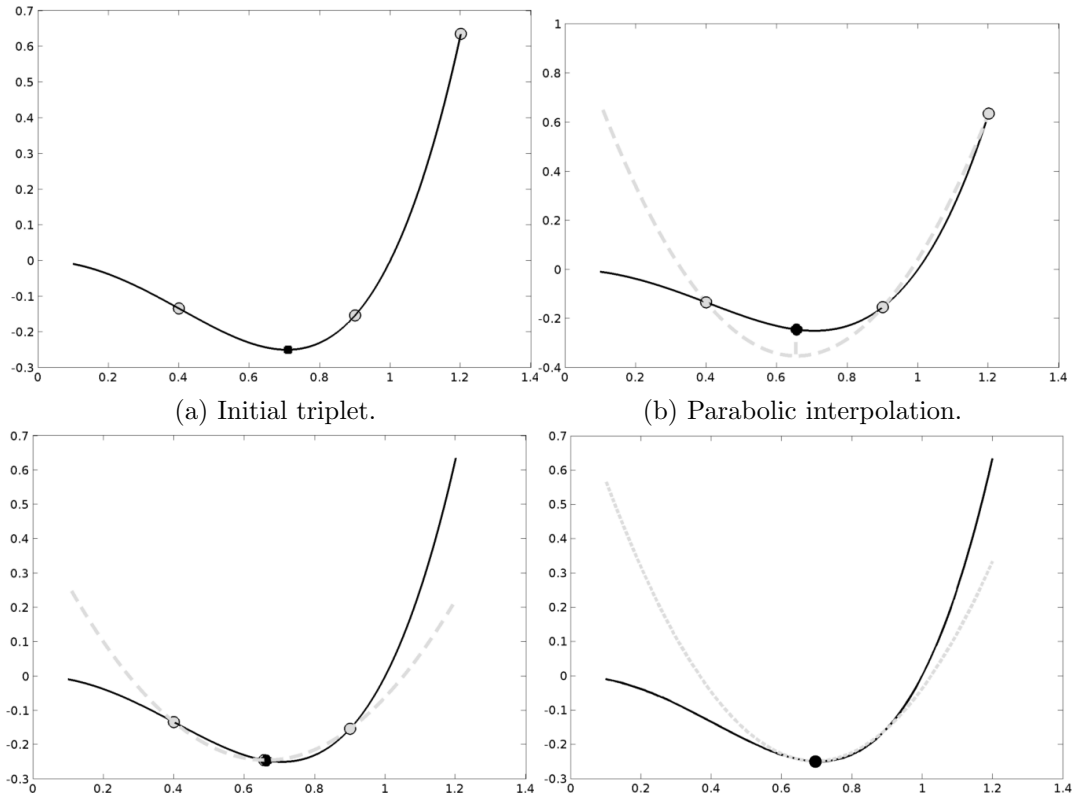
4.2.1 Grid

With the classic grid method, problem (19) is tackled as represented in Figure 4. Starting from $\boldsymbol{n_o}$, multiple evolutions are generated with different values of c. Each of these evolutions is then compared to the initial configuration through the objective function, thus identifying the parameter value that minimizes χ^2. In particular, the minimization problem (19) is faced limiting the search in the interval $I_c = [c_{min}, c_{max}]$, as defined in Section 3.1. Having set these boundaries, the central choice to be made concerns the step size to be adopted in the search.

Intuitively, a finer step size tends to guarantee a higher definition in the measure of the parameter estimates. Nonetheless, the process under scrutiny is stochastic, thus making the value of χ^2 noisy. As a consequence, an excessively fine step size would be useless, since it would not manage to discern values beyond a certain resolution. Figure 7 gives and example in this direction: the relatively fine step size brings highlights the oscillations deriving from the random nature of the process.

In addition, the time of search R is proportional to the number of points in the interval. Therefore, when generating a distribution of the statistics of c^* with S samples, the cost of the computation is proportional to $S \cdot R \sim S \cdot O(T)$, where T is the length of the temporal evolution of the stochastic process in Code 1. Clearly, this implies that the step size cannot be too fine if time-efficiency is to be guaranteed for practical purposes.

On the other hand, a coarser step size entails a loss in definition as well as greater fluctuations around the "true" value of the parameter estimate. Such fluctuations increase the variance of the estimate, and thus the probability not to reject the null hypothesis. Overall, then, the grid method entails a choice in the step size which cannot be easily controlled.



(a) Initial triplet. (b) Parabolic interpolation.

(c) New triplet and interpolation. (d) Convergence.

Figure 8: Successive parabolic interpolation.

4.2.2 Successive parabolic interpolation

An alternative optimization method rests on exploiting the quadratic behavior of the objective function χ^2 around its minimum. Relying on this, the method of successive parabolic interpolation implements a numerical technique based on function evaluation only. In particular, the algorithm does not require the function to be differentiable, nor its derivative to be known or easily computable.

To give a visual intuition, Figure 8 illustrates the method of successive parabolic interpolation in its four salient moments. The algorithm aims at finding the minimum of a function, that is the black dot on the solid line in Figures 8a-8d. The search starts by drawing three points lying on the function to be optimized, as represented by the gray dots in Figure 8a. This initial triplet is then interpolated by a parabola (the dashed line in Figure 8b), whose minimum is determined analytically together with the value of the abscissa at which it occurs (the black dot in Figure 8b). By taking on board this new point while dropping the furthest one from it, a new triplet is formed and a new parabola is interpolated (see Figure 8c), thus identifying a new minimum. This process continues iteratively until it converges to a stable point, which means that the previous and the successive minimum correspond within a certain range of tolerance. The convergence point is the approximate solution to problem (19).

The fundamental property supporting this method is that a unimodal function can approximated by a parabola over an interval that includes the minimum. More precisely, having a triplet of points x_0, x_1, x_2 and their function values $f(x_0), f(x_1), f(x_2)$, the second-order Lagrange interpolation is used to construct a second-degree polynomial which approximates the

parabolic function:

$$q(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2) \ , \quad (23)$$

Clearly, the critical point of equation (23) is given by the first order condition $dq(x)/dx = 0$. For practical purposes, however, equation (23) can also be seen as a quadratic polynomial

$$q(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) \ , \quad (24)$$

where a_0, a_1, a_2 have to be such that equation (24) agrees with $f(x)$ at the three points. That is $f_0 \equiv f(x_0) = q(x_0)$, $f_1 \equiv f(x_1) = q(x_1)$ and $f_2 \equiv f(x_2) = q(x_2)$. In follows that

$$a_0 = f_0 \ , \quad a_1 = \frac{f_1 - f_0}{x_1 - x_0} \ , \quad a_2 = \frac{1}{x_2 - x_1}\left(\frac{f_2 - f_0}{x_2 - x_0} - a_1\right) \ , \quad (25)$$

thus making equation (24) equivalent to equation (23). In this alternative form, the critical point to be sought be the algorithm is

$$\bar{x} = \frac{x_1 + x_0}{2} - \frac{a_1}{2a_2} \ . \quad (26)$$

Indeed, we use formulations (24)-(26) in the implementation of the algorithm. In doing so, one needs to consider at least two things. First, the sign of a_2 determines the concavity of our parabola, thus making the critical point a candidate for a minimum ($a_2 > 0$) or for a maximum ($a_2 < 0$). Second, the three points involved in equation (26) lie on a line if $a_2 = 0$, thus imposing to change the step size between the triplet.

Given this parabolic interpolation, another important detail of the algorithm concerns the stopping criterion. Having to deal with a stochastic process, the objective function χ^2 is noisy, and this introduces the necessity to fix a tolerance criterion based on the variance of the process. To this purpose, the standard deviation of the χ^2 function is calculated computationally. Consequently, the algorithm is set to stop when the fluctuations of our minimum values are of the same order of magnitude of the estimated standard deviation, that is

$$|f(\bar{x}) - f(x_{min})| \leq \sigma(f_{\chi^2}) \quad (27)$$

Finally, convergence is generally fast. As discussed in Brent (1973), the convergence rate for a deterministic function is $r \approx 1.324$, hence defining a superlinear convergence. In the stochastic case, however, it can happen that the interpolation parabola has a local convex behavior for a given triplet, as shown in Figure 9. This is due to the random nature of the function. In this case, the quadratic interpolation is repeated with the same triplet but recalculating the stochastic function value of the central point of the triplet. Such recalculation is repeated until that point is associated to a function value that is under the straight line passing through the two external points of the triplet.

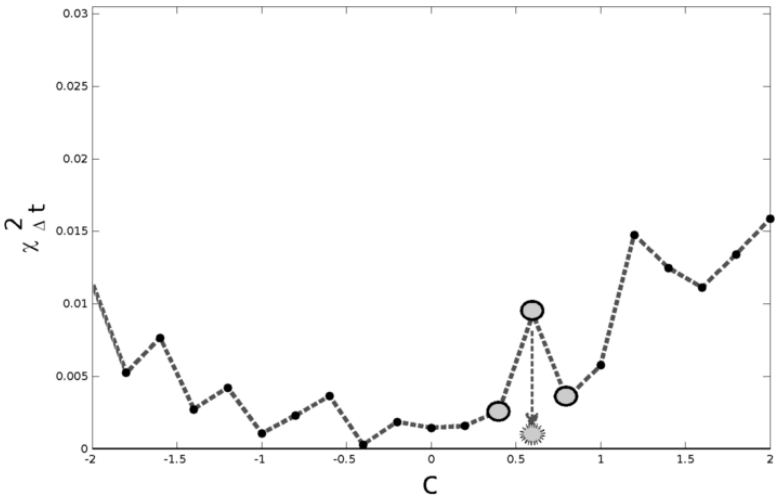
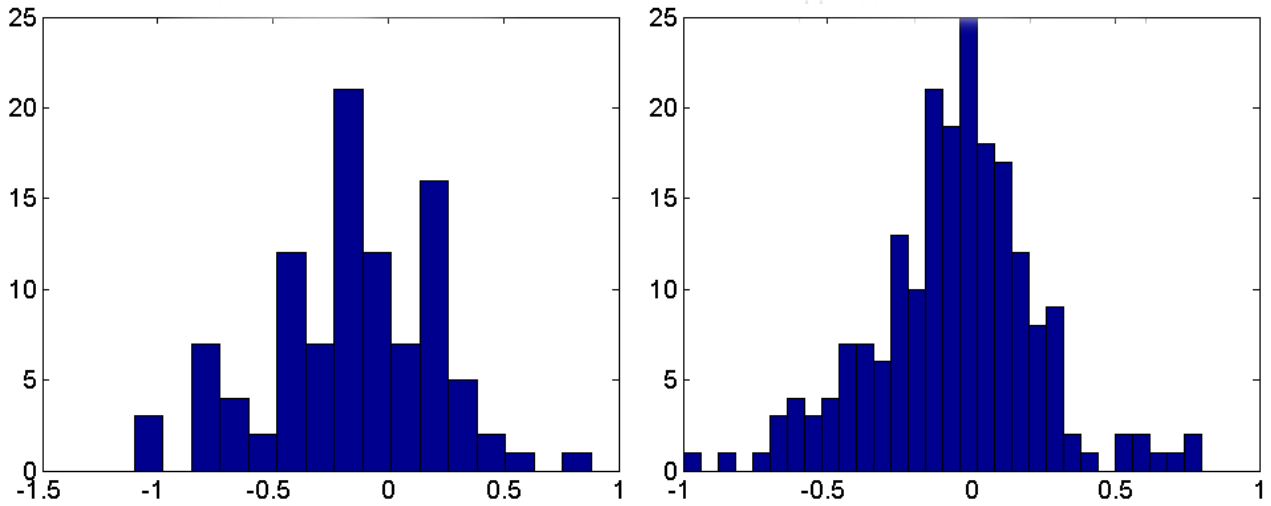


Figure 9: Local convexity problem.

4.2.3 Time performance of the optimization methods

The performance of the two optimization method can be evaluated according to their speed of convergence. As mentioned above, successive parabolic method can have convergence limitations related to far solutions, point alignment, and convexity. Each of these issues needs *ad hoc* control conditions, which also entail some cost in terms of efficiency. Nonetheless, successive parabolic estimation is much faster than the grid method. In particular, we performed a test for a given vector of $\mathbf{x} = \{x_1, \ldots, x_L\}$, an elevated number of alternatives $L = 1 \cdot 10^3$, and conspicuous number of agents $N = 1 \cdot 10^4$. With these parameter values, the grid method managed to construct the the histogram of c^* in Figure 10a in 120 minutes. On the other hand, successive parabolic interpolation allowed to construct the corresponding histograms in Figure 10b in 20 minutes, reaching convergence in $8--10$ steps on average. Due to this remarkable difference in performance, successive parabolic interpolation is adopted as the default optimization method to carry out estimation.

More generally, the algorithm of the stochastic process has asymptotic computational complexity $O(T)$, where T is the length of evolution of the process, which thus generates T configurations. Hence, if the first $t = 10^4$ steps take 0.7 seconds to be simulated, the computation of $t = 10^5$ steps is achieved in 7 seconds. The performance and the time cost of the algorithm depends on many factors (programming language, CPU, the type of data) and, besides that, its efficiency depends also on the RAM of the calculator and the way data is stored. As a consequence, the behavior of the algorithm complexity can only be measured qualitatively. In particular, the estimates of time costs are meaningful only if the following conditions are met, namely: *(i)* the use of the same programming language; *(ii)* the use of the same calculator machine; and the *(iii)* the use of the same dataset as input. In the presents work, we have used the Matlab language, on a 64 bit Linux OS within a machine with Quad Core Intel i7 3612QM @ 2.10GHz, and a RAM of 8,00 GB Dual-Channel DDR3 @ 798MHz.



(a) Grid method.

(b) Successive parabolic interpolation.

Figure 10: Histogram of c^*.

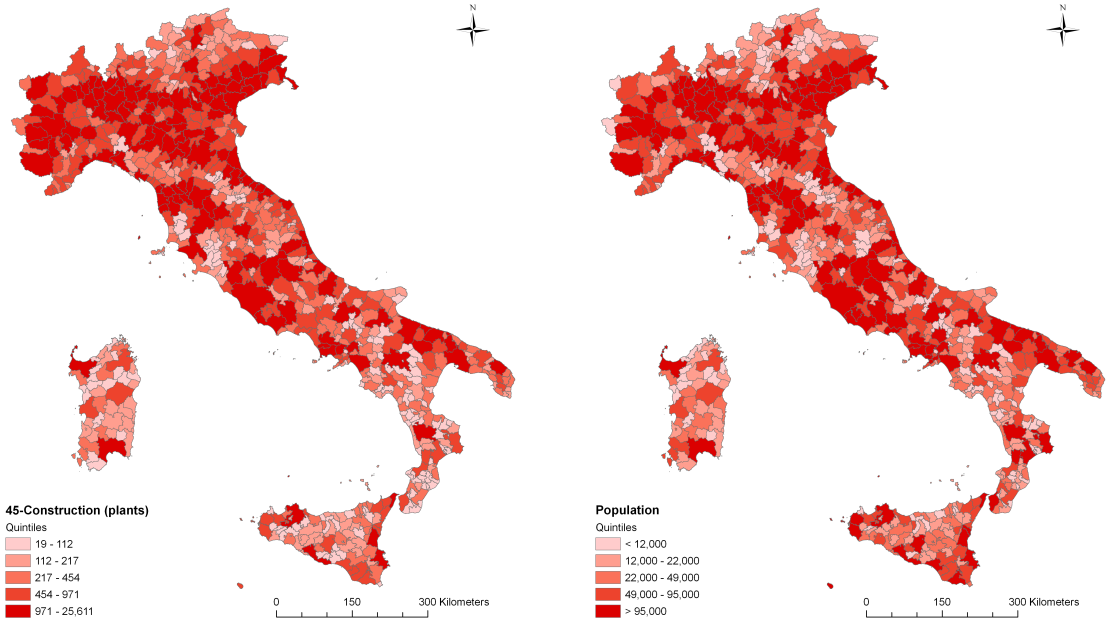
Note: The parameters of this example are $L = 1 \cdot 10^3$ and $N = 1 \cdot 10^4$.

5 Application

The model presented in Section 2 and its associated estimation method discussed in Section 4 are here applied to the case of firm location. The N agents of the model are identified with the plants of various manufacturing and service sectors in the Italian economy, while the alternatives correspond to the $L = 686$ commuting zones across which plants are localized. Is the observed distribution of plants across spatial units affected by non-linear externalities? In the present context, answering this question means to estimate whether the parameter c in equation (6) is statistically different from 0.

This inquiry is relevant under at least two dimensions. On the one hand, and regardless of the sign of $\hat{c}$, it serves to understand whether the inclusion of a non linear term actually adds explanatory power to the model. If the null hypothesis $c = 0$ is never rejected across sectors, then the addition of a non-linear terms would prove to be rather irrelevant for the geography of firm location. On the other hand, it is especially important to verify if c is often negative and statistically significant across sectors. In that case, congestion costs would turn out to play a non-negligible role in the localization choices of firms. This would open a breach in the literature, as most studies investigating the effect of localized externalities ignore entirely the possibility of congestion costs, both at a theoretical and empirical level (see Arthur, 1990, Black and Henderson, 1999, Bottazzi and Gragnolati, 2012, Bottazzi et al., 2007, 2008, Desmet and Fafchamps, 2006, Devereux et al., 2004, Dumais et al., 2002, Duranton and Overman, 2005, Ellison and Glaeser, 1997, 1999, Henderson, 2003, Maurel and Sédillot, 1999, Rosenthal and Strange, 2001).

Previous analysis on Italian data has revealed that the two main drivers of the spatial distributions of firms are population and localized externalities (see Bottazzi and Gragnolati, 2012). All other factors are generally orders of magnitude weaker. Therefore, the present application will simplify the analysis by depicting the intrinsic features of spatial units as being



(a) Sector `45-Construction`. (b) Population.

Figure 11: Maps of the data (year 2001).

shaped only by population (i.e. $a_l = x_l$ so that $H = 1$). It follows that equation (5) can be simplified to equation (6), thus focusing the estimation only on parameters b and c. Considering that the interest here is just on the detection of non-linearity, only the estimates of c will be actually reported and discussed. To provide a visual summary of the data, Figure 11 illustrates the spatial distributions of plants for the construction sector, as well as the spatial distribution of population. Moreover, the second column of Table 2 reports the number of plants N for each of the sectors taken into account. The data come from the Italian census of manufacturers and services, and they refer to year 2001 (see ISTAT, 2006).

The results of the estimation procedure as applied on some manufacturing and service sectors are reported in Table 2. 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. Given this premise, at least two main considerations emerge from Table 2. First, it is not the case that the null hypothesis $c = 0$ is always, or never, rejected. Indeed, the statistical significance of $\hat{c}$ varies across sectors, at least according to the standard confidence levels used for hypothesis testing (i.e. 95% or 99%). Second, when the estimate is statistically significant, the sign of $\hat{c}$ varies across sectors. For instance, using a 95% confidence level, the localization of plants in sector `17-Textiles` results to be affected by a by positive quadratic term, while in sector `20-Wood processing` such term is negative. Overall, these results entail that no general rule can be derived about the way in which non-linear externalities affect firm location, as the effects are markedly sector-specific. In some sectors, firms are co-localizing with other similar firms without reaching a saturation point, while in other sectors spatial congestion is possibly playing a role.

Table 2: Number of plants and parameter estimates by sector.

NACE-Sector	N	$\langle c \rangle$	$\hat{c}$	p-value
15-Food and beverages	73680	-0.24	-0.34	0.66
17-Textiles	31984	-0.07	1.76	0.01
18-Apparel	46377	-0.12	1.90	< 0.01
19-Leather products	24195	0.03	1.70	< 0.01
20-Wood processing	50250	-0.11	-0.62	0.02
22-Publishing and printing	29166	0.00	1.68	0.01
23-Coke, petroleum and nuclear fuel	913	-0.01	0.72	0.22
25-Rubber and plastic products	15115	0.01	1.60	0.06
26-Non-metallic mineral products	31177	0.02	-0.54	0.08
27-Basic metals	3984	0.10	0.56	0.33
30/32-33-Computers and electronics	37636	-0.07	1.70	< 0.01
45-Construction	529757	-0.51	2.00	0.12
55-Hotels and restaurants	261304	0.27	0.24	0.41
64-Post and telecommunications	18056	-0.41	-1.70	0.46

Note: For each sector, N is the number of plants, $\langle c \rangle$ is the average value of c^* across simulations, and $\hat{c}$ is the estimate of c. The number of alternatives is fixed to $L = 686$ commuting zones.

6 Conclusion

This paper has spelled out a discrete choice model accounting for different determinants of the agents' decisions. In particular, the model disentangles the effect of externalities from other factors such as the intrinsic characteristics of the object of choice and the idiosyncratic preferences held by agents. Relative to previous works in the same line of inquiry, the present one has taken a step forward in allowing for non-linear externalities among agents. Among other things, this extension allows to check for congestion effects, which may possibly act as a boundary to positive externalities.

Together with the theoretical model, also an entire estimation framework has been developed. Numerical simulations have first served to unravel the stochastic process underlying the model and verify its ergodicity. Then, they have been used to estimate the unknown parameters through a multistage procedure and test their statistical significance. This involved different numerical optimization techniques, at last finding a preferable approach in successive parabolic interpolation.

Finally, the model and its related estimation method have been applied to the case of firm localization. Using Italian sectoral census data disaggregated by commuting zones, non-linear externalities of localization have been detected in some sectors and not in others. But even where the non-linear term is statistically significant, its direction might change across sectors. This indicates that some economic activities are not hit by spatial congestion, while others

are. Overall, then, the empirical results point to sectoral specificities rather than to general features in the localization behavior of firms. This might possibly have to do with the varying technological regimes underlying the different sectors.

7 Acknowledgement

We thanks prof. Enrico Scalas for his precious feedbacks. The present work has been supported by the Italian Ministry of University and Research as a part of the PRIN 2009 program (grant protocol number 2009H8WPX5).

Code 1: Stochastic evolution of the model.

```
function n=stochastic_evolution(a,b,c,n0,T)
% (a,b,c) are the three parametrs of the non-linear equation for the probability
% the output is the configuration after the time interval T
N = sum(n0);
L=length(a);
b=b0.*ones(1,L);
% Inizialization
a=a(:)';
n0=n0(:);
n=zeros(L,T);
n(:,1)= repmat(n0,[1  1]);
t=1;
%  stochastic evolution:
for t=2:T
        ra = ceil(N * rand(1,1));
        [k,~]=find(ra>[0; cumsum(n(:,t-1))] ); %the exit position
        m=k(end);
        n(m,t-1)=n(m,t-1)-1; % remove the dead agent
        pD=sum(a)+b*n(:,t-1)-b(m) +sum(c.*(n(:,t-1)'- (m==(1:length(b)))).^2);
        pN= a+ b.*n(:,t-1)'- b.*(m==(1:length(b))) + ...
        + c.*(n(:,t-1)'- (m==(1:length(b)))).^2;
        p=pN./pD ;
        r=rand(1,1);
        [~,kk]=find(r>[0, cumsum(p)] );
        in=kk(end);
        n(:,t)=n(:,t-1);
        n(m,t-1)=n(m,t-1)+1;
        n(in,t)=n(in,t)+1;
end
```

Code 2: Re-binning procedure

```
function [e_bin,h,CF,a]=crea_unibin(x)
a=(0:max(x));
i=1;
nB=5;% number of bins
CF=zeros(1,length(a));
while i<=length(a),
    CF(i)=sum(x<=a(i));
    i=i+1;
end
percent=1+nB;
e_vec=ceil(linspace(min(CF),length(x),percent)); % scrolling vector
e=1;
e_bin=zeros(1,length(e_vec));
    while e<=length(e_vec)
        E=sum( CF <= e_vec(e));
        e_bin(e)=E;
        e=e+1;
    end
e_bin(1)=0; % since [0;min(CF)] is empty
h=zeros(length(e_bin)-1,1); % it is the frequency equidistributed inside e_bins
for k=1:length(e_bin)-1
     h(k)=sum(x>=e_bin(k) & x<e_bin(k+1))/length(x) ;
end
```

References

W.B. Arthur. 'Silicon Valley' locational clusters: When do increasing returns imply monopoly? *Mathematical Social Sciences*, 19(3):235–251, 1990.

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*, 2012. forthcoming.

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.

R. Brent. *Algorithms for minimization without derivatives*. Prentice-Hall Inc., 1973.

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. 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.

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

ISTAT. Atlante statistico dei comuni. Roma, 2006.

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

MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2011.

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.

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.

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.