



# A Framework for Modelling Economic Regional Location Processes Under Uncertainty

Roy Cerqueti<sup>1,2</sup>  · Eleonora Cutrini<sup>3</sup>

Accepted: 20 July 2021  
© The Author(s) 2021

## Abstract

This paper deals with the theoretical analysis of the spatial concentration and localization of firms and employees over a set of regions. In particular, it provides a simple site-selection theoretical model to describe the probabilistic framework of the location patterns. The adopted quantitative tool is the stochastic theory of urns. The model moves from the empirical evidence of the deviation of the spatial location of companies from the uniform distribution and of employees from the distribution of firms. Factors leading to such deviations are taken into consideration. Specifically, we formalize a decision problem grounded on the economic attributes of the regions and also on the distribution of the existing firms and employees in the territory. To our purpose, the site-selection model is presented as a stepwise process.

**Keywords** Spatial concentration of firms · Localization process · Urn model

**JEL Classification** R12 · C02

## Introduction

In the framework of spatial economics, a theme of paramount relevance is represented by analyzing how firms and employees decide to locate in a specific region.

Understanding the settlement process of individuals is also essential for planners that should define appropriate policies able to attract new employees in their

---

✉ Roy Cerqueti  
roy.cerqueti@uniroma1.it  
Eleonora Cutrini  
eleonora.cutrini@unimc.it

<sup>1</sup> Department of Economic and Social Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5., 00185 Rome, Italy

<sup>2</sup> School of Business, London South Bank University, 103 Borough Rd, London SE1 0AA, UK

<sup>3</sup> Department of Economics and Law, University of Macerata, Via Crescimbeni, 20., 62100 Macerata, Italy

jurisdictions. Obviously, where the new location will occur and how much employment will be generated are strictly intermingled. Still, in the literature, the crucial link between location choices and the firm's size has not been investigated in depth. Few exceptions can be found. The most influential for the subsequent literature on location choices is the work of Carlton (1983), who simultaneously models (via duality theory) both the decisions of where to locate and how many employees to hire. Recently, in the context of transport and land-use modelling, Hensher et al. (2019) present a formal model in which the number of firms influences the number of jobs at a specific location. De Palma et al. (2019) model the origin of cities and urban hierarchy within a framework with two types of agents characterized by heterogeneous preferences.

On the theoretical side, different strands of the literature have clearly recognized that location processes are related to the interplay between, on one side, various sources of external economies that contribute to agglomeration, and on the other side, congestion costs, competition and the spatial distribution demand that tend to facilitate dispersion. Besides, within the evolutionary perspective, it is asserted that economic systems are pervaded by endemic uncertainty and that regional entrepreneurship tends to be a path-dependent process, being also the outcome of heterogeneous behaviours of agents characterized by bounded rationality. There is overwhelming evidence that new firms do not opt for optimal locations in terms of cost minimization. Still, they are affected by local structures that originated in the past, social interactions, and the regional knowledge base (new firms typically exploit local knowledge and skills) (Boschma and Frenken 2018).

In the empirical literature, given the discrete nature of location choices, they are typically modelled through the McFadden–Carlton conditional logit approach (McFadden 1974; Carlton 1979, 1983). This approach provides sound theoretical underpinnings of empirical models, even if they refer to a theoretical objective function of a representative profit-maximizing agent. Moreover, as Carlton (1983) suggested, the idiosyncratic reasons behind each choice of the preferred location are in the error term, and the reasons behind the labour demanded and firm size.

Starting from these premises, this paper aims to provide a more general framework that allows deriving spatial patterns of firms and employees from individual site selections based on bounded rationality. We consider that decision-makers are different in their attitude regarding centripetal and centrifugal forces, and this heterogeneity within the same group of economic agents can lead to different spatial structures. In particular, we here adopt an evolutionary location approach by considering that the selection of a region by firms and employees can be modelled, in a probabilistic context, as a ball extraction from urns.

Urn models are particularly versatile for the modelling of preferential attachment problems, as in our case (for an overview, see Johnson and Kotz 1977). Hence, in the proposed setting, we adopt such a perspective; specifically, we follow a Polya urn approach (see Mahmoud 2008). Polya urns are constructed as standard boxes with coloured balls, with the relevant characteristics that additional balls of some colours are added or removed before the phase of extraction. In doing so, such probabilistic models provide a time-dependent probability of taking a given color from the urn—hence, describing the paradigmatic cases of the changing of the conditions in the attachment

process. The versatility of the Polya urn models explains why they are so popular in economic theory, and why they are suitable for applications in the field of economic geography with special reference to path-dependence that is considered a typical property of industrial dynamics (see e.g. Arthur 1994; Bottazzi et al. 2007; Martin 2012). Interestingly, Polya urns intervene also in the theoretical explanation of the rank-size laws when the position of the sizes at given ranks is viewed as a discrete attachment process (see e.g. Ausloos and Cerqueti 2016; Cerqueti and Ausloos 2015).

As we will see, extraction procedure and urns composition will be suitably conceptualized to describe the location decision criteria of the actors. In so doing, urn theory allows modelling agents' preferences in a flexible way, accounting for differences in how heterogeneous agents put a different relative to positive externalities due to spatial interaction instead of local congestion costs, competition and the spatial distribution of demand.

This general framework generates either:

Equally spaced firms, so that positive externalities and natural advantages do not influence location patterns and firms are distributed over the regions accordingly to a uniform distribution; or, differently, uneven distribution of firms across regions, hence pointing to the presence of aggregation effects.

Equally sized firms, so that the distribution of firms and employees are overlapping; or, differently, disproportionality between employees' distribution and firms' distributions for each sector. In this second case, we may infer that firms' size heterogeneity within the same sector affects this scenario.

Thus, our setting allows us to account for a rich set of location patterns despite its great simplicity.

Equally important, it provides a background for two main features usually observed in the empirical distributions of firms and employees. First, the divergence of firms from the regional uniform distribution in a specific industry. Second, the divergence between the spatial patterns of firms and employees in a specific industry.

The rest of the paper is organized as follows. Section [Related Literature. How to Place our Contribution in the Field of Economic Geography?](#) reviews the related literature, Section [The Model](#) is devoted to the conceptualization and development of the model. The cases of firms and employees are conveniently distinguished. Section [A Discussion of Advantages and Limitations of Urn Theory with Reference to the Literature](#) discusses the benefits of the model concerning the research gap we are addressing, its limitations and possible extensions. Section [Conclusions and future research](#) offers some concluding remarks, along with the identification of future lines of research.

## **Related Literature. How to place Our Contribution in the Field of Economic Geography?**

The present contribution can be placed at the interface between the traditional New Economic Geography and the institutional and evolutionary approaches on the theoretical side. The urn model provides microeconomic foundations to the spatial

distribution of firms and employees based on the interplay between, on one side, different sources of external economies that contribute to agglomeration, and on the other side, to congestion costs and competition that facilitate dispersion. In the well-known synthesis of the New Economic Geography, location patterns are modelled as the results of factors such as demand distribution in space, transport costs and agglomerative forces such as increasing returns to scale (internal to the firm). Among these factors, the complex role of agglomeration economies has received much interest, even in the early contribution of the traditional location theory (Weber 1929; Hoover 1948; Marshall 1964), in theories of central place systems (Christaller 1933, Lösch 1954) and in the less-known von Thünen's work (von Thünen 1826) on the mechanism of industrial agglomeration and city formation (see Fujita 2012). For recent surveys of the various agglomeration and dispersion forces considered in the literature, see Duranton and Puga (2004), Fujita and Thisse (2013), and Behrens and Robert-Nicoud (2015).

Besides, our modelling approach borrows several insights from the study of the location of new firms from a micro-evolutionary perspective based on the hypothesis of bounded rationality (Pred 1967) and the Behavioural Location Theory (Townroe 1991). The evolutionary economic perspective provides a rich view for understanding the spatial distribution of firms and employees and especially its dynamics (Boschma and Frenken 2006, 2007, 2018). Entrepreneurship not only tends to be a geographically localized phenomenon, but it is also a spatially uneven process that tends to persist over time. This implies regional entrepreneurship tends to be a path-dependent process (see Boschma and Martin 2012; Martin and Sunley 2012). The intention is not to deny that location-specific characteristics may matter (see, e.g. Boschma and Lambooy 1999; Martin and Sunley 2012; Boschma and Martin 2012). However, instead of saying that locations matter in a deterministic manner, an evolutionary approach to spatial clustering accounts for dynamic processes and pre-existing structures in regions are expected to condition but not determine their spatial outcome. Our approach also falls under evolutionary economic geography since it accounts for path dependence in which each event changes the probability of a subsequent event to occur (David 2000; Arthur 1989). In this view, path dependence does not simply arise from the assumption of increasing returns, as is the case in NEG models.

On the empirical ground, alongside advancements in theory, empirical studies addressing the role of several location determinants continue to increase. In the empirical literature on regional and urban economics, the determinants of industrial location decisions have been investigated following two main methods: discrete choice modelling and count data modelling. The former method includes both the characteristics of the decider (firms, employees, etc.) and those of the available regions and evaluates them as distinct factors; the latter deals with the territories' features. Specifically, discrete choice models are used when the unit of analysis is the economic agent, while counting data models are adopted when micro-geographical data are not available and, therefore, the units of study are regions (see Arauzo-Carod et al. 2010 and Bhat et al. 2014 for a critical assessment of the methods and results). Although mainly developed in the context of standard theories, empirical modelling of location decisions has gone beyond fully deterministic approaches (e.

g. conditional logit). It now allows for the incorporation of taste variation in decision-makers (“mixed logit” or “random parameter multinomial logit” models). Still, a systematic approach to industrial location modelling has not been found. Discrete microeconomic data were typically analyzed following the random utility maximization (RUM) approach put forward by McFadden’s (1974) seminal article. Carlton (1979, 1983) applied the conditional logit model (CLM) and first demonstrated that location choices of branch plants can be modelled in a random utility maximization setting, as suggested by McFadden. Therefore, the main advantage of the McFadden–Carlton CLM approach is that the empirical setting is explicitly embedded in the theoretical random utility maximization framework.

The McFadden–Carlton CLM approach has become commonly adopted in the literature to investigate location choices of firms and households (for reviews, see Schmidheiny and Brülhart 2011 and Arauzo-Carod et al. 2010).

Although, as suggested by Carlton (1983) and more recently by Guimaraes et al. (2004) and Arauzo-Carod et al. 2010, the Independence of Irrelevant Alternatives (IIA) assumption of the error term constitute a limitation. In other words, once controlling for the observable characteristics of regions and individuals included in the model, the error term captures all the rest, that is the unobservable characteristics. It may lead to infer that all locations are similar to the decision-makers. If this assumption is violated, estimated coefficients can be biased. The proposed urn theory approach overcomes this severe limitation.

Recent studies have also resorted to the Poisson (count) models to tackle the problem of modelling complex choice scenarios with many small spatial units. Its theoretical justification based on the random utility maximization approach was provided by Guimaraes et al. (2004).

Yet, there are still open questions. For example, Schmidheiny and Brülhart (2011) show that the CLM and the Poisson models differ starkly in terms of their implied predictions and that all intermediate cases can be represented as a nested logit model. Moreover, differently from the Poisson model, the conditional logit does not appear appropriate if one wants to model the location process as a dynamic one since it implies.

“that the aggregate number of agents is fixed and that differences across locations affect only the distribution of those agents across those locations” (Schmidheiny and Brülhart 2011). Nevertheless, as pointed out by Bartik (1985, p. 16).

“the conditional logit approach remains attractive because of its computational feasibility compared with other alternative approaches to the discrete choice problem.”.

## The Model

### Motivation: In Search of a Theoretical Foundation for Well-known Empirical Regularities

The present location model unifies the settlement process of firms and employees in a stochastic setting. In the real world, entrepreneurs may have limited information about the actual conditions of regional attributes driving the choice of where to open a new branch. Moreover, the firm generally is active for many years during which all the relevant location factors change over time. Accordingly, the actual location choice should be thought at least as an optimization process subject to uncertainty, or even a decision-making process following satisfying principles.

The implications of uncertainty on individual location choices have been well recognized in regional science and urban studies. Nevertheless, while stochastic models of location are not completely new, they were adopted to investigate the decisions of various parts of the population, separately—e.g. entrepreneurs, migrants, criminals—like in the study of the spatial variations of crime (e.g. Deutsch and Epstein 1998), or in the migration decisions to relocate or not (e.g. O’Connell 1997; Schmidt 2014). In the context of the logistics literature, stochastic models of location were first applied to facilities location problems (see, among many others, Jucker and Carlson 1976, Hodder and Dincer 1986; see also Krarup and Pruzan 1983, Snyder 2006 for surveys). These modeling approaches of location choices, while considering the plant location decisions, do not consider the settlement of employees.

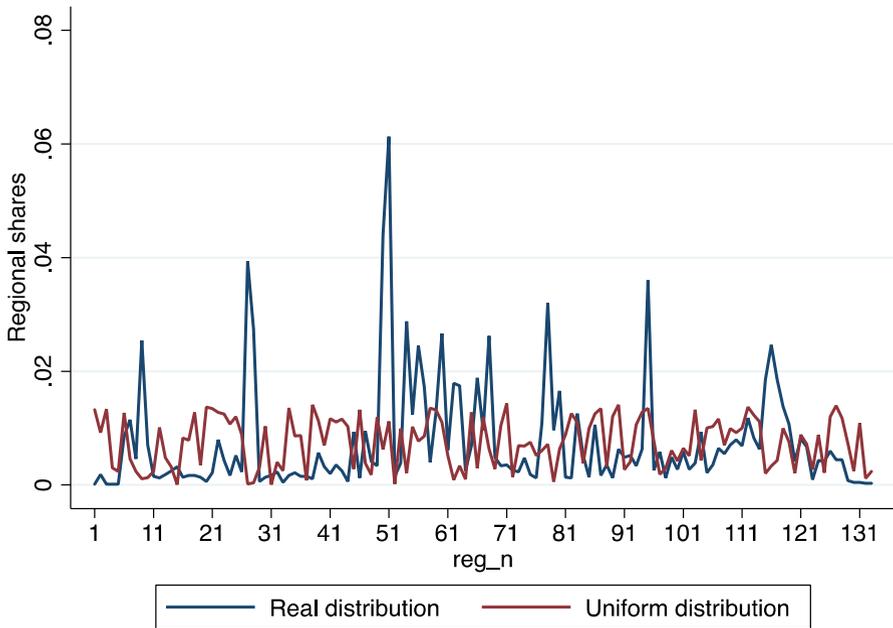
This paper represents an attempt to provide a theoretical ground of the spatial location of firms and employees under uncertainty, considering their interdependency.

Particularly, we start from two empirical regularities: the deviation between the actual distribution of firms and the uniform distribution and the discrepancy between the actual distribution of employees and the one of the existing firms. As for the first empirical regularity, a “no localization” state—wherein even “first nature” location advantages are irrelevant—can be associated to the uniform distribution of firms across a homogeneous space.

To give an example of the first empirical regularity, Fig. 1 compares the interregional settlement of firms in the office machinery and computers industry across 133 EU regions (blu line) with a randomly generated distribution based on a uniform law over [1,400], which is the variation range of the considered phenomenon (red line).

We perform a Kolmogorov–Smirnov test of the equality of distributions with a null distribution of this statistic calculated under the hypothesis that the sample of data is drawn from the reference uniform distribution. Results confirm that the factual distribution is far from following a uniform law (Table 1), as the graphical evidence clearly suggest (Fig. 1).

Several factors beyond “first nature advantages”, that concur to such a deviation have been identified in the literature. Papageorgiou and Smith (1983) claims that the uniform distribution is unstable if subjective attitudes regarding human



**Fig. 1** Firms in the office machinery and computers industry in Europe—Regional shares. The real distribution (blue line) is compared with a randomly generated distribution on the basis of the uniform law on [1,400] (color figure online)

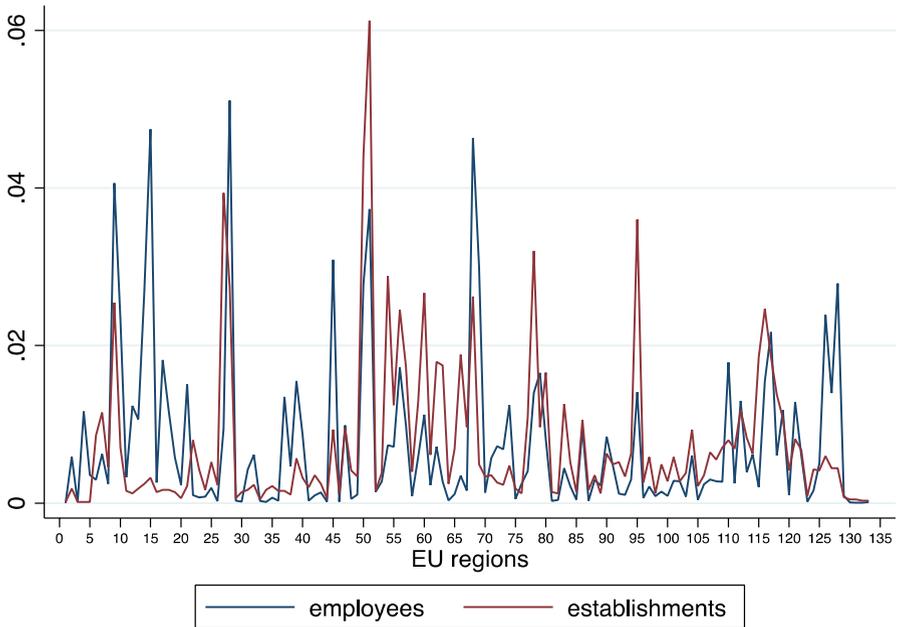
**Table 1** One-sample Kolmogorov–Smirnov test against theoretical uniform distribution on [1,400]

Smaller group	D	P value
N. of firms:	– 0.9699	0.000
Cumulative:	– 4.0e+02	0.000
Combined K-S:	397.5489	0.000

interactions are accounted for. Moreover, within the evolutionary economic literature, further reasons can be identified, including the imitative behaviour of a newly located firm in respect of the existing firms (Arthur 1994; David 2000). We, in fact, have assumed that clustering and establishments’ location is path-dependent (for a similar view, see Boschma and Lambooy 1999).

The second empirical regularities that the urn model refers to is the divergence between the spatial patterns of firms and employees in a specific industry. Figure 2 serves as an example, and compares the interregional settlement of employees in the office machinery and computers industry across 133 EU regions (blu line) with the spatial distribution of plants (red line).

The visual representation indicates that the distribution of employees does not mimic the one of firms, and the results of the Kolmogorov–Smirnov test of the equality of distributions clearly suggest that they cannot be considered as drawn from the same distribution.



**Fig. 2** Employees and establishments in the office machinery and computers industry in Europe—Regional shares

**Table 2** One-sample Kolmogorov–Smirnov test of employment against the distribution of firms

Smaller group	D	P value
Employment:	– 0.2180	0.000
Cumulative:	– 4.0e+02	0.000
Combined K-S:	399.0376	0.000

The null distribution of this statistic is calculated under the hypothesis that the sample of data is drawn from the reference, i.e. the distribution of establishments across space. Results clearly imply that the null hypothesis has to be rejected (see Table 2).

We can summarize the motivations behind our study. In the context of well-established theories, such as the New Economic Geography and even the Evolutional Economic Geography, the issue of uncertainty in location choices by firms and employees have been only marginally considered. Stochastic location models are not entirely new in the literature, but they do not consider the interdependency of the location choices of firms and employees.

More importantly, we aimed to provide a theoretical framework able to reconcile the empirical regularities with a more realistic hypothesis of bounded rationality of entrepreneurs and workers.

In this perspective, location choices are viewed as an optimization process subject to uncertainty.

The offered theoretical perspective is related to the preferential attachment problem, where deciders select one of the available alternatives based on a prefixed criterion. Specifically, as we will see in detail below, we describe the process of the placement of employees and firms over a set of regions by modelling employees and firms as coloured balls and regions as urns. The probability that a given employee (or firm) places in a specific region is the same as extracting a particularly coloured ball from the urn related to the region. The complexity of the environment lets the available opportunities be not identical so that the randomness of the extractions is biased by special events and external factors. We admit balls additions and removals at each step of the process to capture this property of the location dynamics, hence letting our approach be particularly appropriate in our context.

## **A General View and Basic Assumptions of the Model**

We suggest that the evidence highlighted in Sect. [Motivation: In Search of a Theoretical Foundation for Well-known Empirical Regularities](#) can be explained by the interplay between agglomeration economies and countervailing factors acting toward dispersion.

In particular, accordingly to the arguments developed above, we assume that natural advantages do not play any role in driving the regional selection procedure or firms location process. This assumption has a great impact mainly at the beginning of the location process—i.e., when regions are empty and the selection of a region obeys a uniform distribution. However, as we will see below, deviations from the uniform distribution in successive steps occur, reasonably due to the presence of scale economies, intra-regional competition and other sources of economic factors.

In the context of the distribution of the employees, following Carlton (1983), we consider that it should be tailored on the one related to the existing firms. In fact, the plant location choice is usually taken in a long-run perspective, as a part of a more complex investment decision making process, involving, among other key strategic choices-like technology, financing, marketing and distribution etc., the plant firm size (see Hayter 1997 for a similar viewpoint). Hence, the spatial distribution of workers in one industry is certainly linked to the decisions taken by entrepreneurs about the number of employees to hire.

We basically assume that employees choose their location as soon as firms have settled in space. This is reasonable in a world where mobility of labour across sectors and space is allowed and firms are the basic source of economic growth and thus of jobs.<sup>1</sup>

---

<sup>1</sup> Since we postulate, at least in the first step of the model, that firms create jobs and employees follow, we depart from the classical assumptions in the NEG models where it is generally assumed that firms are created and destroyed according to a process driven by the locational choices made by workers who vote with their feet. In NEG models migration is central to define the equilibrium scenarios and it is driven by interregional real wage differences, in a world dominated by firms producing under increasing returns and transportation costs. In the Muthian chicken-and-egg controversy (Muth 1971), NEG seems to answer that firms should follow the employees (Fujita and Thisse 2013). Nevertheless, even in standard Core-Periphery models firms can freely substitute one location for another (see Fujita and Thisse 2013), and some models also recognize the possibility that interregional distribution of economic activi-

In particular, under the assumption that firms belonging to the same sector are equivalent in terms of their sizes, an employee will select the locating area on the basis of the relative number of firms belonging to the available regions. The size of the firm is predetermined at the beginning of the process of site selection by firms and employees but it can change over time. Thus, for what concerns the location process related to the employees, we here take for us the evidence that it is driven by the size of the firms within the same sector. In so doing, we move from the reasonable intuition that employees spread over the available regions proportionally to the typical dimension of firms of the industrial sector—which is measured as the average firm size of the sector. This stated, we consider as grounding assumption that the counterfactual regional distribution of the employees is the one of the firms—hence, implicitly assuming that all the firms belonging to a given sector share the same size. In saying this, we are pointing out that the divergence between firms and employees depends only on the heterogeneity in firm size.

### The Location Choice/Urn Extraction Procedure

We assume that the economic system is composed by a finite number of regions,  $Z$ , and  $R$  is the set of possible alternatives that firms consider as feasible for establishing a plant<sup>2</sup> ( $R \subset Z$ ).

We consider an urn containing a finite number of coloured balls. We denote such colours as  $C_1, C_2, \dots, C_R$ . The  $R$  regions can be viewed as the colors of the balls in the urn.

For both cases of employees and firms, the process of region selection is modelled by a sequential extraction of a ball from a suitably defined urn, and the colour of the extracted ball represents the selected region.

We consider  $S$  industrial sectors forming the entire economic system.<sup>3</sup> We aim at modelling and discussing different types of relationships between the distribution— at a regional level—of the employees and of the firms, both for each sector and for the overall economic activity. The distribution of employees and firms is described as a dynamic settlement process through an urn model. The final outcome of the preference optimization of each economic agent is a probability law over the available regions, and it is strongly dependent on the agent taking the decision. In the following section we enter the details and explain how the urn model can be applied for modelling the rational location choices of firms and employees, respectively.

---

Footnote 1 (continued)

ties is driven by the locational decisions made by profit-maximizing firms, and workers adjust instantaneously (e.g. Picard et al. 2004).

<sup>2</sup> A firm may consider as feasible locations only those regions where there is a pool of relevant services available.

<sup>3</sup> This model should be intended as a general location model, it can be applied to the entire economic system or a sub-set of it (e.g. the manufacturing industry).

## Firms

We consider a set of  $K$  firms.

The probability that the  $k$ -th firm locates in the  $r$ -th region coincides with the probability that the  $k$ -th individual extracts from the urn  $U$  a ball with color  $C_r$ .

In general, we assume that the rational location choice of the  $k$ -th firm is affected by the following characteristics of each region  $r$ :

- (1) presence in the region  $r$  of firms of the same sector to take advantage of industry-specific knowledge spillovers, skilled workers in the local labour market, and within-industry input–output linkages;
- (2) presence in the region  $r$  of firms operating in ancillary activities to exploit complementarities with upstream and downstream firms along the value chain and knowledge spillovers across sectors;
- (3) presence in region  $r$  of infrastructures and services leading to a more profitable business environment, included public supports for firm formation or relocation (urbanization economies).
- (4) Intra-regional competition
- (5) Congestion costs
- (6) the spatial distribution of demand for the product of the firm to save on transport costs

The first three factors are related to the advantages of agglomeration economies, and they favour the spatial concentration of economic activities, while factors (4) and (5) and (6) work in the opposite direction, and they are intended to capture the dispersion forces usually considered in the literature. Accordingly, to each region  $r$  it is associated a possible score, which is a state defined by the combined effect of the above-mentioned regional attributes. As we will see, the scores of the regions lead to a probability distribution which gives information on how likely a firm will select a given region when locating. In particular, the single firm is endowed with a cardinal utility function  $u_k$ , representing the “value” or “profit” placed by the entrepreneur to the envisaged score associated to each region—hence, to the probability distribution induced by the scores. The higher the score, the higher the worthiness of the region for the individual firm.

The probability that the  $k$ -th firm locates in the  $r$ -th region increases as the ratio between the number of balls with colour  $C_r$  and the total balls contained in  $U$  does. Moreover, the heterogeneity among the firms and their preferences based on the interplay between agglomerative effects and dispersion forces is modelled through the modification of the urn configurations at each firm’s location decision. In particular, we consider a multistage procedure, in which the  $k$ -th stage is associated to the selection of the region by the  $k$ -th firm.

At the initial stage, the colours are assumed as being identically distributed in the urn. This means that the first extraction from the urn (i.e.: the location choice of the first firm) is implemented according to a uniform distribution over the available regions. The uniform hypothesis of the initial configuration of the urn stands for the

irrelevance of natural advantages, which are the only mean to let an empty region be preferable than another empty one and are not considered here.

In the dynamic of this site selection model, the economic agents choose to be situated in a particular location because of its specific combination of the relevant characteristics highlighted above. Whatever the specific preference structure of each firm, what is particularly relevant for the firms is the distribution of the firms itself, in the sense that the  $k$ -th firm's choice are affected by the location of the previous  $k - 1$  firms.

So, in general, while the role of "historical accidents" is accounted for in the first step of the selection process, the uniform distribution of the colors in the urn is an invalid condition in each step after the first one.

At the first stage, the firm labeled with 1 implements the selection of the region. In the context of colored balls in the urn, the individual labeled with 1 extracts a ball from  $U$ .

The drawn color corresponds to the region where the 1-st firm locates. So, the number of the firms in the  $R$  regions after the 1-st drawn from the urn is a vector  $(n_1^{(1)}, n_2^{(1)}, \dots, n_R^{(1)})$ , where the subscript indicates the color/region. Of course, if the extracted ball has color  $C_{\bar{r}}$ , then it results  $n_{\bar{r}}^{(1)} = 1$  and  $n_r^{(1)} = 0$  for each  $r \neq \bar{r}$ .

After the extraction, the drawn ball is reinserted in the urn. At the same time, some other colored balls are inserted in or removed from the urn. At the end of this procedure, the number of the balls with colours  $C_1, C_2, \dots, C_R$  is  $C_1^{(2)}, C_2^{(2)}, \dots, C_R^{(2)}$ , respectively. The resulting new configuration of the balls in the urn changes—in general—the probability of extracting a ball with a specific color in the next step, according to the preference structure of the decision maker (see below the details on this for the general case of the  $k$ -th step). This new urn is the one used by the firm labeled with 2 for the ball drawn procedure.

Now, the firm labelled with 2 takes a ball from the urn. The number of the firms in the regions after the 2-st drawn from the urn is  $(n_1^{(2)}, n_2^{(2)}, \dots, n_R^{(2)})$ , where the subscript indicates the color/region. In accord to the previous step, the non-empty regions are all the ones selected in the 1-st and 2-nd step, and we have  $n_1^{(2)} + n_2^{(2)} + \dots + n_R^{(2)} = 2$ .

The procedure goes on according to this rule: at the  $k$ -th stage, the firm extracts from a urn such that the number of balls with colors  $C_1, C_2, \dots, C_R$  is  $C_1^{(k)}, C_2^{(k)}, \dots, C_R^{(k)}$ , respectively.

As already mentioned above, the probability  $p_{\bar{r}}^{(k)}$  of taking a specific color/region  $C_{\bar{r}}$  from the urn is the relative number of the balls with color  $C_{\bar{r}}^{(k)}$  in the urn, so that

$$p_{\bar{r}}^{(k)} = \frac{C_{\bar{r}}^{(k)}}{\sum_{r=1}^R C_r^{(k)}}, \forall k = 1, 2, \dots, K \text{ and } \bar{r} = 1, 2, \dots, R. \tag{1}$$

In general,  $p_{\bar{r}}^{(k)} \neq p_{\bar{r}}^{(k-1)}$ , for each  $k = 1, 2, \dots, K$  and  $\bar{r} = 1, 2, \dots, R$ .

After the  $k$ -th drawn, the number of firms in the regions is  $(n_1^{(k)}, n_2^{(k)}, \dots, n_R^{(k)})$ .

The balls addition/removal procedure implemented at each stage—and the associated probability distribution  $p^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_R^{(k)})$ —has economic reasoning.

In fact, the  $k$ -th firm takes its rational choice on the basis of its utility  $u_k : \mathbf{D} \rightarrow \mathbb{R}$ , where  $\mathbf{D}$  is the set collecting all the discrete probability distributions of the type  $p = (p_1, p_2, \dots, p_R)$  over the set  $\{1, 2, \dots, R\}$ , where

$$p_{\bar{r}} \in [0, 1]; \forall \bar{r} = 1, 2, \dots, R \text{ and } \sum_{r=1}^R p_r = 1.$$

The meaning of such probabilities in our settlement process is

$$p_{\bar{r}} = \text{Prob}(\text{to draw a ball of color } C_{\bar{r}}).$$

At each stage, the probability distribution reveals the preference structure of the  $k$ -th firm, according to the specific cardinal utility function  $u_k$  which places a value to each region taking into consideration the interplay between the perceived regional attributes (1)–(6) described above. The value of the cardinal utility would be different for different firms. For example, a firm that displays a strong attitude towards interaction and put a relatively lower weight to congestion costs will make a different location choice compared to a firm that displays a strong aversion to competition relative to the advantages it perceives in terms of positive externalities related to the presence of other firms in the same region. Example 1 below is suitably designed to show how the location choices may differ for agents with heterogenous preferences with respect to the interplay between agglomeration and dispersion forces.

Hence, the probability to extract a region  $r$  can be conceived as the result of a procedure of rational choice by the firm (See Simon 1955). The key assumption here is that the  $k$ -th firm has incomplete information about the final state  $s$  of the region in which it will end up operating, in terms of the presence of other firms.

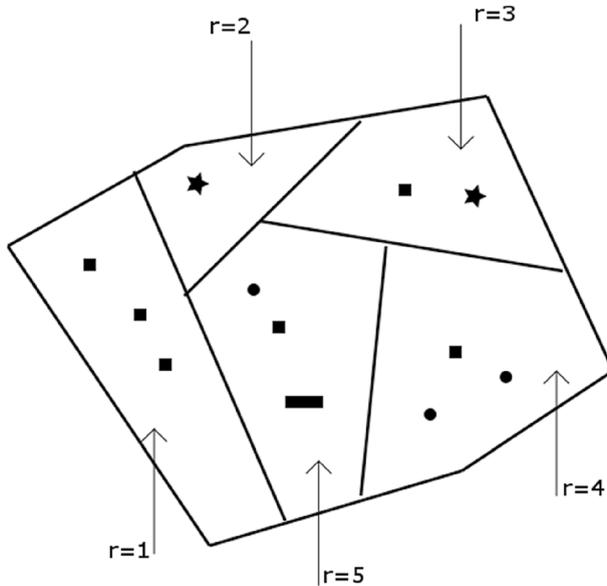
This is one main source of uncertainty that we consider in the present model: the  $k$ -th firm can observe the spatial distribution of existing firms but it cannot know how many other firms will locate in the same region or elsewhere.

After the rational choice procedure, the colored balls are added (removed) in (from) the urn, so that the distribution of the balls in the urn is in accord to the rational choice probability distribution, which is exactly  $p^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_R^{(k)})$  as in formula (1).

The specific shape of the utility function depends also on the sectors of activities of the previously located firms and on the locating one. We assume that firms are shared among  $S$  sectors of activity and denote the sectors by  $s = 1, 2, \dots, S$ .

To better explain how the settlement process works, an illustrative example is needed. We refer here to the manufacturing industry and a rational location choice by firms only driven by the possibility to exploit upstream and downstream complementarities across industries and industry-specific Marshallian economies.

**Example 1** Consider five regions ( $R = 5$ ) and suppose that eleven firms have already chosen their regional location. So, the 12-th firm must now extract a ball from the urn and locate (i.e.:  $k = 12$ ).



**Fig. 3** Situation at the beginning of the 12-th step, describing the localization of the eleven firms among the five regions. Different symbols stand for different sectors of activities. Circles represent the Textiles, stars are associated to the Motor vehicles, square is Machinery while the rectangle indicates the Basic metals

Let us also suppose that the located firms belong to four sectors (i.e.:  $S = 4$ ). Specifically, three of them are in the Textiles, two of them are in the Motor vehicles, six of them are in the Machinery and one in the Basic metals.

The regional distribution of the firms is as in Fig. 3.

Suppose now that the 12-th firm belongs to the Textiles and it is a firm with a strong attitude towards interaction within the same sector, in other words, for the entrepreneur the most relevant positive feature for choosing a location is given by the possibility to take advantage of localization economies. He also values the possibility to activate backward and forward linkages across sectors, but to a lesser extent. Moreover, local congestion costs and competition are of negligible relevance in his preference structure. The firm is aware about the location configuration presented in Fig. 3. Then, it ranks regions in accord to its preferences. An example of plausible reasoning could be the following: regions  $r = 1, 2, 3$  are free of firms coming from the Textiles sectors. Then, placing in one of such regions could not be a good choice. However,  $r = 1$  is better than the other two, since the presence of a higher number of firms is for sure associated to a higher level of infrastructures and urbanization economies, hence facilitating the business activity. Besides, the existing firms are operating in a sector, which may be important for the firm, because of the possible complementarities with an upstream sector producing specialized machinery for textile products. Nevertheless, region  $r = 4$  should be a better choice for the possibility to exploit Marshallian economies. Region  $r = 5$  has also some attractive features due to the presence of one firm from the Textiles. In fact, the new established firm can also take advantage of the already established

business environment in the Textiles sector fostered by the existing firm (e.g. presence of skilled labor, and/or specialized services for the textile industry). Said this,  $r = 5$  is perceived to be better than  $r = 1, 2, 3$ .

To conclude, the 12-th firm decides to locate in a region different from  $r = 1, 2, 3$ . In a preference scale, it decides to assign mark 10 to the preferred region  $r = 4$  and, by comparison, mark 7 to  $r = 5$  and mark 5 to  $r = 1$ . Regions  $r = 2$  and  $r = 3$  have mark 3. Such scores lead to a probability distribution  $p^{(12)} = (p_1^{(12)}, p_2^{(12)}, p_3^{(12)}, p_4^{(12)}, p_5^{(12)}) = (\frac{5}{28}, \frac{3}{28}, \frac{3}{28}, \frac{10}{28}, \frac{7}{28})$ .

The probability distribution  $p^{(12)}$  may be viewed as the solution of a utility-based constrained optimization problem. The utility function is suitably defined as  $u_{12} : D \rightarrow \mathbb{R}$ , so that

$$u_{12}(p^{(12)}) = u_{12}\left(\frac{5}{28}, \frac{3}{28}, \frac{3}{28}, \frac{10}{28}, \frac{7}{28}\right) = \max_{p \in D} u_{12}(p),$$

and  $D$  is the admissible region.

The removal/addition procedure is then in accord to the optimization probability  $p^{(12)}$ . Specifically, if we suppose that, for example, the number of the colored balls in the urn at the 11-th step was configured as  $(C_1^{(11)}, C_2^{(11)}, C_3^{(11)}, C_4^{(11)}, C_5^{(11)}) = (8, 3, 2, 14, 25)$ , then the (not unique) urn for the 12-th extraction could be created by: adding two balls of color  $C_1$ , adding three balls of color  $C_2$ , adding four balls of color  $C_3$ , adding six balls of color  $C_4$  and removing eleven balls of color  $C_5$ . The resulting configuration of the colors is  $(C_1^{(12)}, C_2^{(12)}, C_3^{(12)}, C_4^{(12)}, C_5^{(12)}) = (10, 6, 6, 20, 14)$ , and the probability distribution associated to the extraction of a colored ball from the urn is exactly  $p^{(12)}$ .

At this point, the 12-th extraction takes place. Consistently with the above example, consider that the drawn ball is of color  $C_4$ . Then, we have  $(n_1^{(12)}, n_2^{(12)}, n_3^{(12)}, n_4^{(12)}, n_5^{(12)}) = (3, 1, 2, 4, 3)$ , where the bold indicates the change with respect to the location of Fig. 3.

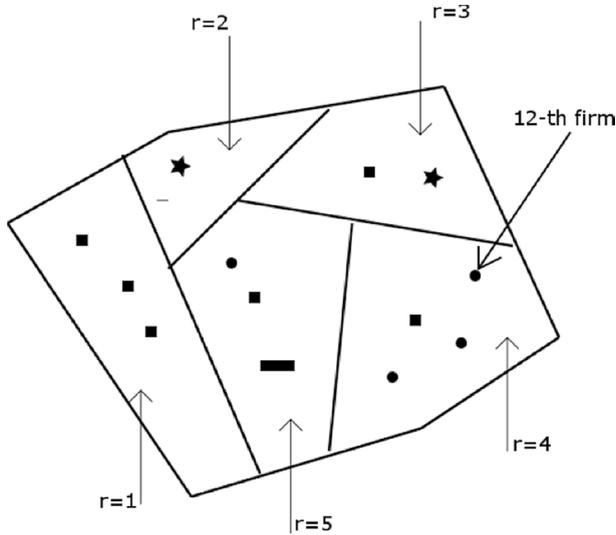
The regional distribution of the firms at the end of the 12-th step is illustrated in Fig. 4.

Notice that the location outcome would have been different if the 12-th firm were a textile firm placing a relatively higher value to the variety of the local environment, in terms of the presence of firms operating in different sectors, compared to the possibility to exploit sector-specific external economies. In this case, the entrepreneur would have chosen region 5 as the preferred site to open a new branch. Higher scores would have been attributed to regions with a greater variety of sectors, and an associated probability distribution  $p^{(12)}$  could have been as follow:

$$u_{12}(p^{(12)}) = u_{12}\left(\frac{3}{28}, \frac{2}{28}, \frac{6}{28}, \frac{7}{28}, \frac{10}{28}\right) = \max_{p \in A} u_{12}(p),$$

The process of location of firms in the regions stops at the  $K$ -th stage.

The final distribution of the  $K$  firms in the  $R$  regions will be denoted by  $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_R)$ , and it is the outcome of how firms located during the  $K$  steps. Hence, it can be derived in a natural way from the number of firms in the regions at the end of the  $K$ -th drawn from the urn, namely  $(n_1^{(K)}, n_2^{(K)}, \dots, n_R^{(K)})$ , as follows:



**Fig. 4** Situation at the end of the 12-th step. The location of the 12-th firm is highlighted. Symbols are those of Fig. 3

$$\tilde{p}_r = \frac{n_r^{(K)}}{K}, \forall r = 1, 2, \dots, R.$$

It is possible to derive the final regional distribution of the firms belonging to a specific sector by the settlement procedure implemented above.

At each step of the location procedure, we identify the sector of the located firm so that  $(n_1^{(s,k)}, n_2^{(s,k)}, \dots, n_R^{(s,k)})$  is the vector collecting the number of the firms of sector  $s$  in the  $R$  regions. The final distribution of the firms belonging to sector  $s$  is then a vector  $\tilde{p}^{(s)} = (\tilde{p}_1^{(s)}, \tilde{p}_2^{(s)}, \dots, \tilde{p}_R^{(s)})$ ,

where

$$\tilde{p}_r^{(s)} = \frac{n_r^{(K,s)}}{\sum_{r=1}^R n_r^{(K,s)}}, \forall r = 1, 2, \dots, R \text{ and } s = 1, 2, \dots, S.$$

The final distribution of firms normally does not follow a uniform law because of path dependence and the location factors described above (i.e. industry-specific Marshallian economies, complementarities between upstream and downstream firms across sectors along the value chain, urbanization economies, the distribution of demand).

### Employees

We now deal with the employees' distribution, and assume an economic system populated by  $L$  employees.

Also in this case, we assume that an employee locates in the  $r$ -th region with the same probability of extracting a ball with color  $C_r$  from a urn.

As explained in Sect. [Motivation: In Search of A Theoretical Foundation for Well-known Empirical Regularities](#) the settlement procedure of employees is conditioned to the spatial distribution of firms' establishments. As in the Ellison and Glaeser (1997)'s approach, we consider that, under the initial random location scenario, firms locate independently from each other but employees do not. Particularly, the lumpiness of establishments is incorporated in the model by assuming that, in the theoretical case (homogeneous regions), firms are identical in terms of firm size, an expression of the strength of internal scale economies typical of the sector.

Hence, our grounding assumption is that firms belonging to the same sector  $s$  hire the same number of employees, say  $L_s$ . So, the number of employees of sector  $s$  in region  $r$  is proportional to the number of firms of sector  $s$  in region  $r$ , and the proportionality factor is exactly  $L_s$ . This means that, under the initial random location scenario, the spatial distribution of the firms and the spatial distribution of employees of sector  $s$  should coincide.

Deviations from this configuration is the result of two main forces: (a) the heterogeneity of firms' size, which could emerge dynamically, leading to several firms departing from the typical industrial structure of the sector, and (b) a different site selection process of economic agents (firms vs employees) since the structure of preferences of workers is usually different from those of entrepreneurs. For examples workers may assign higher attractiveness to some regions (e.g. metropolitan regions for the easier access to urban amenities, regions with higher wages).

However, the settlement process is of sequential type: the first employee is assumed to select a region in accord to the initial assumption, since the industrial system is not affected at the beginning by the effects (a)–(b). After the first step, the economical-geographical activity proceeds, and effects (a)–(b) need to be taken in full consideration. We want to stress, that the identification of the sector is of paramount relevance at the beginning of the location process.

There are  $S$  urns, one for each sector. We denote them as  $U_1, U_2, \dots, U_S$ .

At the beginning of the location process, each urn contains balls of colors  $C_1, C_2, \dots, C_R$ , so that the distribution of the colored balls in  $U_s$  coincides with the final distribution of the firms of sectors, i.e.  $\tilde{p}^{(s)} = (\tilde{p}_1^{(s)}, \tilde{p}_2^{(s)}, \dots, \tilde{p}_R^{(s)})$ , for each  $s = 1, 2, \dots, S$ .

Suppose that the first employee (first extraction from the urn) belongs to the sector  $s_1 \in \{1, 2, \dots, S\}$ . Then, she/he extracts her/his region by drawing a ball from the sector-specific urn  $U_{s_1}$ . Then, the employee locates, according to the color of the extracted ball. The number of the employees belonging to sector  $s$  in the  $R$  regions after the 1-st draw is a vector  $(m_1^{(1,s)}, m_2^{(1,s)}, \dots, m_R^{(1,s)})$ , for  $s = 1, 2, \dots, S$ .

After the extraction, coloured balls are added/removed from each urn (and not only from  $U_{s_1}$ ). This hypothesis captures the dependence structure among different sectors, which actually may interact.

By adopting the same notation of the previous section, at the end of the addition/removal procedure the number of the balls with colours  $C_1, C_2, \dots, C_R$  in urn  $U_s$  is  $C_1^{(2,s)}, C_2^{(2,s)}, \dots, C_R^{(2,s)}$ , respectively, for each  $s = 1, 2, \dots, S$ . These urns are ready for the second extraction (i.e., the second employee), even if only one of the urns will play an active role (the one associated to the sector of activity of the employee labeled with 2).

Indeed, the second employee is assumed to belong to the sector  $s_2 \in \{1, 2, \dots, S\}$ . Then, she/he extracts her/his region by drawing a ball from  $U_{s_2}$ , while the remaining urns are not considered.

Recursively: at the  $l$ -th stage, the number of the balls with colors  $C_1, C_2, \dots, C_R$  in urn  $U_s$  is  $C_1^{(l,s)}, C_2^{(l,s)}, \dots, C_R^{(l,s)}$ , respectively, for each  $s = 1, 2, \dots, S$ . The number of the employees belonging to sector  $s$  in the  $R$  regions after the  $l$ -th drawn is  $(m_1^{(l,s)}, m_2^{(l,s)}, \dots, m_R^{(l,s)})$ , for  $s = 1, 2, \dots, S$  and  $l = 1, 2, \dots, L$ .

The procedure stops at the  $L$ -th stage, when the last employee extracts and locates in a region.

Also in this case, the addition and removal of coloured balls from the urn is due to a rational choice over the set  $D$  of the discrete probability distributions on  $\{1, 2, \dots, L\}$ .

Specifically, the  $l$ -th employee, belonging to sector  $s_l$ , is assumed to have utility function  $u_{|s_l} : D \rightarrow \mathbb{R}$ . The utility maximization over  $D$  leads to a removal/addition of coloured balls from each urn, and not only from  $U_{s_l}$ , to obtain a distribution of colored balls identical to the preferred probability distribution  $\bar{p}^{(l,s)} = (\bar{p}_1^{(l,s)}, \bar{p}_2^{(l,s)}, \dots, \bar{p}_R^{(l,s)})$ , i.e.

$$\bar{p}_r^{(l,s)} = \frac{C_r^{(l,s)}}{\sum_{r=1}^R C_r^{(l,s)}}, \forall l = 1, 2, \dots, L; s = 1, 2, \dots, S \text{ and } \bar{r} = 1, 2, \dots, R.$$

As already stated above, all the urns change their configurations because the action of a sector might change the equilibria also in different sectors, and let a region be more or less attractive than another one also for employees belonging to sectors different from  $s_l$ .

At each stage, the probability distribution reveals the preference structure of the  $l$ -th worker, similarly as in the case of the site selection process by firms. As for their specific selection criteria, we may suggest that, given the spatial distribution of firms that mainly delineate job opportunities, workers are assumed to be interested in locating in urban areas or more populated areas. Thus, the second main source of uncertainty that we consider in the present model relates to the incomplete information by the  $l$ -th worker about the final population density of the region, while he can observe the existing distribution of firms across space and the regions already chosen by workers at earlier stages.

The final distribution of the employees belonging to sector  $s$  over the  $R$  regions is a vector  $\tilde{q}^{(s)} = (\tilde{q}_1^{(s)}, \tilde{q}_2^{(s)}, \dots, \tilde{q}_R^{(s)})$ , where

$$\tilde{q}_{\bar{r}}^{(s)} = \frac{m_{\bar{r}}^{(L,s)}}{L_s}, \forall \bar{r} = 1, 2, \dots, R \text{ and } s = 1, 2, \dots, S.$$

By the sectorial analysis we can also infer the distribution of the employees in the overall manufacturing activity. Such a distribution is denoted as  $\tilde{q} = (\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_R)$ , where

$$\tilde{q}_{\bar{r}} = \frac{\sum_{s=1}^S m_{\bar{r}}^{(L,s)}}{L}, \forall \bar{r} = 1, 2, \dots, R$$

The deviation between  $\tilde{q}^{(s)} = (\tilde{q}_1^{(s)}, \tilde{q}_2^{(s)}, \dots, \tilde{q}_R^{(s)})$  and  $\tilde{q} = (\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_R)$  can be viewed as the final result of a cumulative process in the location of firms and employees.

The dissimilarity between the spatial distribution of firms and employees is essentially due to the heterogeneity of firms in term of plant size, according to the arguments proposed above on the correspondence between equal size and uniform distribution.

## A Discussion of Advantages and Limitations of Urn Theory with Reference to the Literature

The objective of this section is to show the usefulness of urn theory for the study of location processes of firms and employees under uncertainty. The discussion will be carried out by taking also a close look at the reference literature.

First, the modelling approach allows the researcher to portray the complexity of spatial clustering without radically simplifying it. Particularly, the treatment of space is midway between the formalized schematic and the more realistic narrative. Hence, it can be adopted as a background theoretical framework to rationalize the empirical evidence of the geographic concentration of economic activities in the context of spatial economics. In the dartboard approach, the measurement of localization explicitly involves a comparison between the distributions of employment and firms and it is anchored to a theoretical model of location (Ellison and Glaeser 1997). While other approaches over the discrete space, such as those based on relative entropy measures (e.g. Brühlhart and Träeger 2005; Cutrini 2009, 2010, among others), make use only of lattice employment data, with the underlying location choice of firms completely on the shadow. Given these premises, we believe that the present framework can be adopted for the development of a testing procedure able to identify a significant departure from a random location scenario.

Second, location choices of agents are considered under more general assumptions than optimization criteria and they are modelled in a flexible way, so that a wide range of spatial patterns of firms and employees can be rationalized. In this respect, it is worth stressing that we do not explain the location choices of businesses and employees on the basis of deterministic factors that guide the choices as it is the standard approach in the literature. We have moved from a rather different

idea. In fact, we have assumed that individual preferences can be viewed as perturbations of the random configuration, and the urn composition at each stage is a proxy of the preference structure of the single firm that is taking the location choice. In this respect, we build on Papageorgiu and Smith (1983) who maintained that even the uniform distribution of identical individuals across space is unstable if one considers that individuals may vary in their attitude regarding spatial interaction.

Third, we have started from the view that the chosen location of a firm and the chosen number of people employed are two source of information that are not independent from each other. Our approach is able to consider both location choices and firm size, as suggested by Carlton (1983). We have assumed that, at the beginning of the site selection process, all firms are equally sized within each sector, and this hypothesis is justified by the need to account for the heterogeneity in minimal optimal scale and industrial structure across industries. Moreover, in the successive steps, the model allows for heterogeneity of firm sizes even within each industry.

Fourth, the modelling approach features a role for history and 'historical accidents' that determine path-dependence and persistence, and it is basically consistent with the evolutionary perspective in economic geography (David 2000; Arthur 1989; Boschma and Lambooy 1999; Boschma and Martin 2012; Martin and Sunley 2012). This is one of the main advantages of the model, that is, it considers the dynamics of the localization process. In fact, agglomeration is modelled not only as the result of increasing returns at the plant level, as is the case in NEG models, but it is also a product of history. The current choices of firms depend on the past choices of previously settled firms. In this respect, it is worth mentioning the Independence of Irrelevant Alternatives (IIA) assumption of the error term and how it constitutes a limitation to consider cumulative processes in empirical studies. In other words, once controlling for the observable characteristics of regions and those of individuals, the error term captures all the rest, that is the unobservable characteristics not included in the model and it may lead to infer that all locations are similar to the decision makers. If this assumption is violated, estimated coefficients can be biased. The proposed urn theory approach overcomes this severe limitation.

Finally, the interaction between positive externalities and congestion costs (competition) is explicitly considered in the model through heterogenous preferences of the decision makers and the resulting rational choices (see the example provided in Sect. [A General View and Basic Assumptions of the Model](#)).

For all these reasons, we claim that the urn model can be a useful device in the context of location choice modelling. Nevertheless, we do not contend that the approach presented here is free from limitations and, in this respect, we can suggest that its outstanding flexibility allows for possible variations and several refinements that may overcome some of them.

To give an example, we have selected firms and employees as the relevant economic actors in our model. Our choice was driven by the goal of providing a theoretical background for the two empirical regularities highlighted in Sect. 3.1.1. It goes without saying that, thanks to the high flexibility of the model, it is possible to consider other actors, depending on the aim of the study. To give an example, to conceptualize the role of demand in the location choice of firms, it would be more appropriate to consider firms and consumers as the relevant economic agents.

## Conclusions and Future Research

This paper develops a stochastic model for describing the location process of firms and employees among a set of regions. We postulate spatial structures as emerging from the micro-behaviours of heterogeneous economic agents. As a first step, we consider two different key economic agents: entrepreneurs and workers, each individual within each subset of the population is endowed with a specific preference structure about the desirability to locate in a particular region.

The model is grounded on urn theory and it is able to capture the wide number of aspects driving the location choice. Particularly, it allows conceptualizing the process of localization under more general assumptions than the traditional non-random patterns driven by optimization of economic agents. Uncertainty is implicitly assumed because the decision maker is not certain that a locational choice would be profitable or satisfying until regional attributes completely unfold. A key insight from our model is that the location divergence between firms and employees is due to the divergences in terms of sizes of the firms belonging to the same sector.

The framework presented here considers the problem of translating the theory of location choice behaviour into concrete models suitable for the empirical analysis of the spatial distribution of firms and workers. It can be seen as a first step to the development of a systematic theorization of the location processes under uncertainty. In this respect, our setting can be further explored in several directions, and we give here only two noticeable suggestions: first, an explicit reference to the decision time and its effects on the location process could be included. Specifically, it seems to be reasonable to assume that firms can take the opportunity of selecting the appropriate time for locating. As an example, think at a firm producing a patented product and at how the patent expiration date influences the managerial decisions. In this context, the extraction from the urn occurs at (random) times whose identification is a complex task (for instance, it might come out from an optimal stopping problem); second, the presence of an exogenous factor driving the location process could be an interesting aspect to be discussed. Suppose that there are some rumours on a new regulation which incentives investments in a specific region. The presence of uncertainty leads to a further addition/removal of coloured balls in the urn, whose number and distribution is unknown. In this case, a Bayesian-type analysis or scenarios simulations could be of some usefulness.

**Acknowledgements** The authors are grateful to Masahisa Fujita, Tomoya Mori, Jacques-François Thisse, Kristian Behrens, Philipp Ushchev for very helpful comments.

**Funding** Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is

not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arauzo-Carod, J.M., D. Liviano-Solis, and M. Manjón-Antón. 2010. Empirical studies in industrial location: an assessment of their methods and results. *Journal of Regional Science* 50 (3): 685–711.
- Arthur, W.B. 1989. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal* 99: 116–131.
- Arthur, W.B. 1994. *Increasing returns and path dependence in the economy*. University of Michigan Press.
- Ausloos, M., and R. Cerqueti. 2016. A universal rank-size law. *PLoS One* 11 (11): e0166011.
- Bartik, T.J. 1985. Business location decisions in the US: estimates of the effects of unionization, Taxes, and Other characteristics of states. *Journal of Business and Economic Statistics* 3: 14–22.
- Behrens, K., and F. Robert-Nicoud. 2015. Agglomeration theory with heterogeneous agents. In *Handbook of regional and urban economics*, vol. 5, ed. G. Duranton, J.V. Henderson, and W.C. Strange, 171–245. Amsterdam: Elsevier.
- Bhat, C.R., R. Paleti, and P. Singh. 2014. A spatial multivariate count model for firm location decisions. *Journal of Regional Science* 54 (3): 462–502.
- Boschma, R.A., and K. Frenken. 2006. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography* 6 (3): 273–302.
- Boschma, R.A., and K. Frenken. 2007. Application of evolutionary economic geography. In *Applied evolutionary economics and economic geography*, ed. K. Frenken. Cheltenham: Edward Elgar.
- Boschma, R., and K. Frenken. 2018. *Evolutionary economic geography*. The New Oxford Handbook of Economic Geography.
- Boschma, R.A., and J.G. Lambo. 1999. Evolutionary economics and economic geography. *Journal of Evolutionary Economics* 9 (4): 411–429.
- Boschma, R., and R. Martin. 2012. *The handbook on evolutionary economic geography*. Edward Elgar.
- Bottazzi, G., G. Dosi, G. Fagiolo, and A. Secchi. 2007. Modeling industrial evolution in geographical space. *Journal of Economic Geography* 7 (5): 651–672.
- Brühlhart, M., and R. Träger. 2005. An account of geographic concentration patterns in Europe. *Regional Science and Urban Economics* 35 (6): 597–624.
- Carlton, D.W. 1979. Why new firms locate where they do: an econometric model. In *Interregional movements and regional growth*, ed. William Wheaton, 13–50. Washington, DC: Urban Institute.
- Carlton, D.W. 1983. The location and employment choices of new firms: an econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics* 65: 440–449.
- Cerqueti, R., and M. Ausloos. 2015. Cross ranking of cities and regions: population versus income. *Journal of Statistical Mechanics: Theory and Experiment* 2015 (7): P07002.
- Christaller, W. 1933. *The central places of Southern Germany*. Englewood Cliffs: Prentice-Hall.
- Cutrini, E. 2009. Using entropy measures to disentangle regional from national localization patterns. *Regional Science and Urban Economics* 39 (2): 243–250.
- Cutrini, E. 2010. Specialization and concentration from a twofold geographical perspective: evidence from Europe. *Regional Studies* 44 (3): 315–336.
- David, P. 2000. *Path dependence, its critics and the quest for 'historical economics.'* Oxford and Stanford University.
- de Palma, A., Y.Y. Papageorgiou, J.-F. Thisse, and P. Ushchev. 2019. About the origin of cities. *Journal of Urban Economics* 111 (C): 1–13.
- Deutsch, Joseph, and G.S. Epstein. 1998. Changing a decision taken under uncertainty: the case of criminal's location choice. *Urban Studies* 35: 1335–1343.
- Duranton, G., and D. Puga. 2004. Micro-foundations of urban increasing returns: theory. In *Handbook of regional and urban economics*, vol. 4, ed. J.V. Henderson and J.-F. Thisse, 2063–2117. Amsterdam: Elsevier.

- Ellison, G., and E.L. Glaeser. 1997. Geographic concentration in US manufacturing industries: a dashboard approach. *Journal of Political Economy* 105 (5): 889–927.
- Fujita, M. 2012. Thünen and the new economic geography. *Regional Science and Urban Economics* 42: 907–912.
- Fujita, M., and J.-F. Thisse. 2013. *Economics of agglomeration industrial location and globalization cities*. Cambridge: Cambridge University Press.
- Guimaraes, P., O. Figueiredo, and W. Douglas. 2004. Industrial location modelling: extending the random utility framework. *Journal of Regional Science* 44 (1): 1–20.
- Hayter, R. 1997. *The dynamics of industrial location; the factory, the firm and the production system*. Chichester: Wiley.
- Hensher, D.A., C.Q. Ho, and R.B. Ellison. 2019. Simultaneous location of firms and jobs in a transport and land use model. *Journal of Transport Geography* 75: 110–121.
- Hodder, J.E., and M.C. Dincer. 1986. A multifactor model for international plant location and financing under uncertainty. *Computers & Operations Research* 13 (5): 601–609.
- Hoover, E.M. 1948. *The location of economic activity*. New York: McGraw-Hill.
- Johnson, N.L., and S. Kotz. 1977. *Urn models and their application*. New York: John Wiley.
- Jucker, J.V., and R.C. Carlson. 1976. The simple plant-location problem under uncertainty. *Operations Research* 24 (6): 1045–1055.
- Krarrup, J., and P.M. Pruzan. 1983. The simple plant location problem: Survey and synthesis. *European Journal of Operational Research* 12 (1): 36–57.
- Lösch, A. 1954. *The economics of location*. New Haven: Yale University Press.
- Mahmoud, H. 2008. *Pólya urn models*. CRC Press.
- Marshall, A., 1964. Principles of economics, an introductory volume. Macmillan, London, 9th edition (First edition published 1890).
- Martin, R. 2012. (Re) placing path dependence: a response to the debate. *International Journal of Urban and Regional Research* 36 (1): 179–192.
- Martin, R., and P. Sunley. 2012. Path dependence and regional economic evolution. *Journal of Economic Geography* 6 (4): 395–437.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.
- Muth, R.F. 1971. Migration: chicken or egg? *Southern Economic Journal* 37: 295–306.
- O’Connell, P.G.J. 1997. Migration under uncertainty: “Try Your Luck” or “Wait and See.” *Journal of Regional Science* 37 (2): 331–347.
- Papageorgiou, Y.Y., and T.R. Smith. 1983. Agglomeration as local instability of spatially uniform steady-states. *Econometrica* 51: 1109–1119.
- Picard P.M., Thisse J.-F., and Toulemonde E. 2004 Economic geography and the role of profits. *Journal of Urban Economics*, 56(1): 144–167
- Pred, A. 1967. Behavior and location: foundations for a geographic and dynamic location theory. Part I. *Lund Studies in Geography*, 28.
- Schmidheiny, K., and M. Brühlhart. 2011. On the equivalence of location choice models: conditional logit, nested logit and Poisson. *Journal of Urban Economics* 69: 214–222.
- Schmidt, C. 2014. Optimal commuting and migration decisions under commuting cost uncertainty. *Urban Studies* 51 (3): 477–492.
- Simon, H.A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 69 (1): 99–118.
- Snyder, L.V. 2006. Facility location under uncertainty: a review. *IIE Transactions* 38 (7): 547–564.
- von Thünen, J.H., 1826. *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg, Perthes. English translation by C.M. Wartenberg: von Thünen’s Isolated State, Oxford, Pergamon Press (1966).
- Townroe, P.M. 1991. Rationality of the industrial location decision. *Urban Studies* 28: 383–392.
- Weber, A. 1929. *Theory of the location of industries*. University of Chicago Press.