UNIVERSITY
OF TRENTO - Italy
Department of Economics

# Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration

Giuseppe Espa, Diego Giuliani, Giuseppe Arbia

n. 12/2010

UNIVERSITY
OF TRENTO - Italy
Department of Economics

# Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration

Giuseppe Espa, Diego Giuliani, Giuseppe Arbia

n.
12/2010

# Weighting Ripley's *K*-function to account for the firm dimension in the analysis of spatial concentration[1]

**Giuseppe Espa**[†][2], **Diego Giuliani**[*], **Giuseppe Arbia**[§]
[†]*Department of Economics, University of Trento, Via Inama, 5, 38100 Trento, Italy*
[*]*University of Rome "La Sapienza", Piazzale Aldo Moro, 5, 00185 Rome, Italy*
[§]*Department of the Business, Statistical, Technological and Environmental Sciences, University "G. d'Annunzio" of Chieti–Pescara, Viale Pindaro, 42, 65127 Pescara, Italy*

**Abstract:** The spatial concentration of firms has long been a central issue in economics both under the theoretical and the applied point of view due mainly to the important policy implications. A popular approach to its measurement, which does not suffer from the problem of the arbitrariness of the regional boundaries, makes use of micro data and looks at the firms as if they were dimensionless points distributed in the economic space. However in practical circumstances the points (firms) observed in the economic space are far from being dimensionless and are conversely characterized by different dimension in terms of the number of employees, the product, the capital and so on. In the literature, the works that originally introduce such an approach (e.g. Arbia and Espa, 1996; Marcon and Puech, 2003) disregard the aspect of the different firm dimension and ignore the fact that a high degree of spatial concentration may result from both the case of many small points clustering in definite portions of space and from only few large points clustering together (e.g. few large firms). We refer to this phenomena as to *clustering of firms* and *clustering of economic activities*. The present paper aims at tackling this problem by adapting the popular *K*-function (Ripley, 1977) to account for the point dimension using the framework of marked point process theory (Penttinen, 2006).

**Keywords**: Agglomeration, Marked point processes, Spatial clusters, Spatial econometrics.

**JEL classification codes:** ***C21 · D92 · L60 · O18 · R12***

[1] This paper was presented at the *IVth World Conference of the Spatial Econometrics Association*, Purdue University, Chicago, June 9–12, 2010. We wish to thank the participant to the discussion for the useful comments received.
[2] Corresponding author: Giuseppe Espa (giuseppe.espa@economia.unitn.it).

## 1. Introduction

Spatial economics theories show that economic integration may boost spatial concentration of economic activities and industrial specialization both at a regional and at an international level (Bickenbach and Bode, 2008). Furthermore, due to the external increasing returns driven by the spatial concentration, the core regions (where spatial clusters of firms are more likely to occur) may reach higher levels of economic growth than the peripheral regions (see Krugman, 1991 and Fujita *et al*., 1999 among others). As a consequence, the phenomenon of spatial concentration is of paramount importance to explain the determinants of growth and development on one hand and regional disparities on the other.

Fostered by the centrality of these issues under the theoretical and the practical point of view, a variety of empirical studies have tried to develop proper indices and statistical tests to measure the degree of spatial clustering in real industrial situations. Under this respect, a series of recent papers (Arbia *et al*, 2008, 2010; Marcon and Puech, 2003, 2009; Duranton and Overman, 2005, 2008) have introduced the use of distance-based methods. These methods are more robust than the traditional measures of spatial concentration (such as Gini index (Gini, 1912, 1921) or Ellison-Glaeser index (Ellison and Glaeser, 1997)), which make use of regional aggregates and thus depend on the arbitrariness of the definitions of the spatial units. The distance-based methods, conversely, make use of micro economic data, treating each firm as a point on a map and studying their spatial distribution with the methods borrowed from the so called *point pattern analysis* (Diggle, 2003).

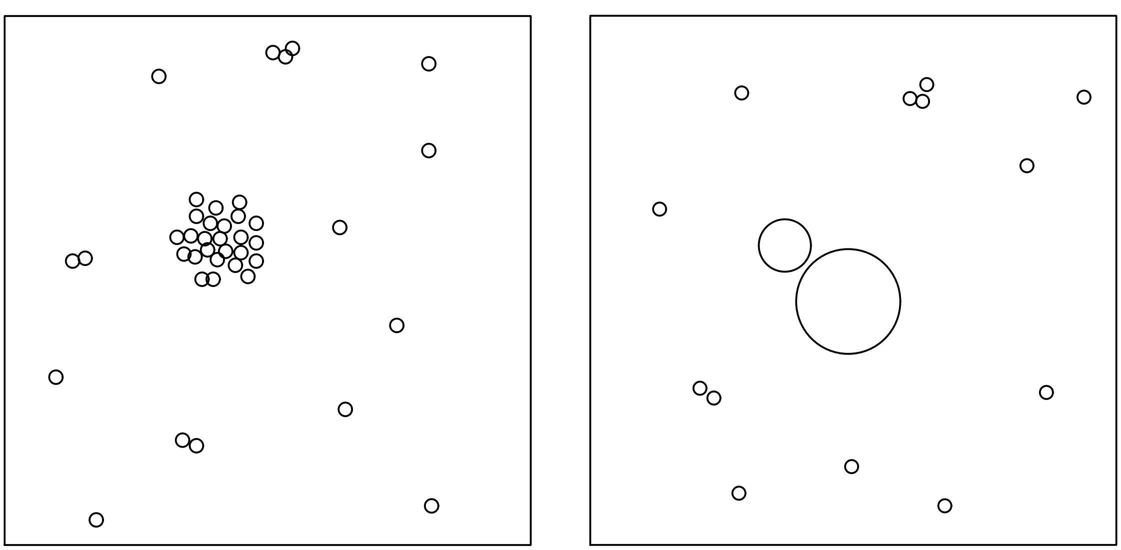
In many empirical circumstances where the presence of spatial clusters of firms is tested by using micro-geographical data, an important element to be taken into account is represented by the firm dimension.

Indeed a high level of spatial concentration can be due to two very different phenomena (see Figure 1). Namely,

- *Case 1*: many small firms clustering in space, and
- *Case 2*: few large firms (in the limit just one firm) clustering in space.

We can refer to the first case as to the case of *clustering of firms* and to the second as to the case of *clustering of economic activities*.

**Figure 1:** Two extreme paradigmatic situations of spatial concentration.



*Case 1: clustering of firms* *Case 2: clustering of economic activities*

A proper test for the presence of spatial clusters should thus consider the impact of the firm dimension on industrial agglomeration by clearly distinguishing these two cases.

Under this respect, Marcon and Puech (2009) and Duranton and Overman (2005) have extended the use of Ripley's *K*-function (Ripley, 1977) considering firm size treating it as a weight attached to each of the points constituting the pattern. Both quoted papers developed *relative* measures of the spatial concentration, detecting the extra-concentrations of firms belonging to a specific industry with respect to the distribution of firms of the whole economy. Following this procedure a positive (or negative) spatial dependence between firms is detected when the pattern of a specific sector is more aggregated (or more dispersed) than the pattern of the whole economy. Although measures of relative spatial concentration are very useful in controlling for the idiosyncratic characteristics of the territories under study, on the other hand they do not allow comparisons across different economies (see Haaland *et al.*, 1999 and Mori *et al.*, 2005 for a more detailed discussion).

In this paper we propose a similar extension of Ripley's *K*-function which leads to an *absolute* (rather than a *relative*) measure of the industrial agglomeration and which allows comparability amongst different empirical situations. More specifically, referring to the theory of *marked point processes*, we develop a stochastic mechanism which generates weighted point patterns of firms representing stylized facts of the different phenomena occurring in real cases (essentially: spatial randomness or spatial concentration in the sense indicated in "*Case 1*" or "*Case 2*" above). The values assumed by the proposed measure in the various cases constitute the benchmark that allows us to formally test the departure from spatial randomness.

We will present our new approach along the following lines. In Section 2 we will briefly discuss the classical Ripley's *K*-function which represents the starting point to develop more sophisticated measures of spatial concentration. Section 3 will be devoted to introduce the stochastic mechanism based on the marked point processes theory which allows us to develop a test for the presence of absolute spatial concentration of firms and economic activities. In this section we will introduce the new model, we will discuss the meaning of the model's parameters in the context of spatial concentration of firms and economic activities and we will present some simulation results to better illustrate how the model works in practice. Finally, Section 4 contains a discussion of the results, some conclusions and directions for further studies in the field.

## 2 Measuring the spatial concentration of firms disregarding size: the basic *K*-function

It is probably fair to say that Ripley's *K*-function (Ripley, 1976 and 1977) is currently the most popular distance-based measure to summarize the cumulative characteristics of a spatial distribution of events in the context of micro-geographic data. It has indeed proved a very versatile tool to test for the presence of spatial concentration within a stationary point pattern where each event is considered as a dimensionless point. As a consequence, the *K*-function has been largely applied in various fields such as geography, ecology, epidemiology and, more recently, economics (see Arbia and Espa, 1996; Marcon and Puech, 2003).

The *K*-function is defined as follows:

$$K(d) = \lambda^{-1} E\{\text{number of points falling at a distance} \leq d \text{ from an arbitrary point}\} \qquad (1)$$

with $E\{.\}$ indicating the expectation operator and $\lambda$ representing the mean number of events per unitary area, a parameter called *intensity*. Therefore, $\lambda K(d)$ can be interpreted as the expected number of further points within a distance *d* of an arbitrary point of the process (Ripley, 1977). In case of a homogeneous field (where the probability of hosting a point is constant across the study area), the *K*-function quantifies the level of spatial dependence between points at each distance *d*.

In order to develop a test for the presence of absolute spatial concentration, we can rely on the fact that for many stochastic processes, it is possible to compute the expectation in the right-hand side of Equation (1), so that $K(d)$ can be written in a closed form (Dixon, 2002). A point process generating a spatial distribution of events completely at random (that is, points are distributed uniformly and independently on space) is the so-called homogeneous Poisson process. It can be shown that if a point pattern is a realisation of a *homogeneous Poisson process* then $K(d)$ tends to be equal to $\pi d^2$ (see Diggle, 2003). Therefore:

$$K(d) = \pi d^2,\ d > 0$$

represents the null hypothesis of random location of events. Significant departures from this benchmarking value represent the alternative hypothesis of spatial dependence. More precisely, for $K(d) > \pi d^2$ we have positive dependence and hence *clustering* (where points tend to attract each other), for $K(d) < \pi d^2$ we have negative dependence and hence *inhibition* (where points tend conversely to repulse each other). Therefore, to formally test whether the observed points tend to cluster in space we can verify if, for some *d*, $K(d)$ is significantly greater than $\pi d^2$. Critical values can be computed by Monte Carlo simulation of homogeneous Poisson processes (see Besag and Diggle, 1977).

The test for the presence of absolute concentration based on Ripley's *K*-function, however, can be used to detect industrial agglomeration only if firms can be considered to have the same dimension. Indeed, in a context where economic activities are different in terms of dimension with the presence of small, medium and large firms, a point pattern is not a good representation of the location pattern of economic activities and, as a result, the *K*-function is no more a proper tool to summarize the spatial distribution. For instance, the simple *K*-function cannot recognize a situation like the one reported in Figure 1 as "*Case 2*" as a cluster. In other words, the test do not "*control for the overall agglomeration of manufacturing*" (Duranton and Overman, 2005).

In such a context, in order to define a proper test, we need to refer to the concepts and methods of the *marked point process* statistics, which is a branch of spatial statistics devoted to analyse sets of events scattered in space, where each event is not only defined by its spatial location, but also by a *mark*, that is a supplementary set of information which might be either quantitative or qualitative (Illian *et al*., 2008).

## 3 Measuring the spatial concentration of firms considering size: the mark-weighted *K*-function

### *3.1 The mark-weighted K-function*

The mark-weighted *K*-function, indicated as $K_{mm}(d)$, is an explorative tool proposed by Penttinen (2006) to summarize the cumulative characteristics of a homogeneous quantitative marked point pattern (that is a pattern where a quantitative mark is attached on each point). It has been proposed as a natural generalization of Ripley's *K*-function. In order to introduce it let us first rewrite the classical *K*-function as:

$$K(d) = E\left[\sum_{i=1}^{n}\sum_{j\neq i} I\left(d_{ij} \leq d\right)\right] \Big/ \lambda$$

where the term $d_{ij}$ is the Euclidean distance between the *i*th and *j*th arbitrary points, *n* is the total number of points and $I(d_{ij} \le d)$ represents the indicator function such that $I$ = 1 if $d_{ij} \le d$ and 0 otherwise. Following this notation, the mark-weighted *K*-function has a similar form but the marks are now taken into account:

$$K_{mm}(d) = E\left[\sum_{i=1}^{n}\sum_{j \ne i} m_i m_j I(d_{ij} \le d)\right] \bigg/ \lambda \mu^2 . \qquad (2)$$

In Equation (2) $m_i$ and $m_j$ are the marks attached to the *i*th and *j*th points, respectively, and $\mu$ is the mean of the marks. Thus the term $\lambda\mu^2 K_{mm}(d)$ can be interpreted as the mean of the sum of the products formed by the mark of the *i*th arbitrary point and the marks of all other points in the circle *d* centred in it (Illian *et al*., 2008). Therefore, the mark-weighted *K*-function measures the joint cumulative distribution of marks and points at each distance *d*.

Turning now to the estimation aspects, following Penttinen (2006), a proper approximately edge-corrected unbiased estimator of $K_{mm}(d)$ is

$$\hat{K}_{mm}(d) = \left(\sum_{i=1}^{n}\sum_{j \ne i} m_i m_j w_{ij} I(d_{ij} \le d)\right) \bigg/ n\hat{\lambda}\hat{\mu}^2$$

where $\hat{\lambda} = n/|A|$ is the estimated spatial intensity, $|A|$ is the area of the study region and $\hat{\mu}$ is the mean of the observed marks. Due to the presence of edge effects arising from the arbitrariness of the boundaries of the study region, the adjustment factor $w_{ij}$ is introduced thus avoiding potential biases in the estimates in proximity to the boundaries of the study region. More precisely, the weight function $w_{ij}$ expresses the reciprocal of the proportion of the area of a circle centred on the *i*th point, passing through the *j*th point, which lies within the study region *A* (Boots and Getis, 1988).

In an economic context, in which the marks are the values of a quantitative variable representing the firms size, the mark-weighted *K*-function might be used to develop a test for the presence of absolute spatial concentration. However, we need to derive the benchmark value of the function representing the null hypothesis of spatial randomness. For this reason the next paragraph is devoted to derive a stochastic model to generate marked point patterns of firms which is able to represent the stylized situations of spatial randomness and concentration in the meaning of "*Case 1*" (i.e., many small firms clustering in space) and "*Case 2*" (i.e., few large firms clustering in space).

### *3.2 A model for the null hypothesis of spatial randomness*

The basic idea we follow is that the spatial concentration of economic activities (in the sense of "*Case 1*" and "*Case 2*") can be originated by some form of correlation between the spatial point intensity and the marks. This would imply, for instance, that in regions characterized by high spatial point intensity the marks tend to be systematically large if such a correlation is positive or, conversely, small if such correlation is negative.

To define a model which incorporate such a correlation structure we refer to the design, already explored by Ho and Stoyan (2008), of an *intensity-marked Cox process*, where the spatial point intensity is driven by a Cox process and the marks are realizations of a process whose parameters are conditioned by the values of the spatial point intensity.

#### *3.2.1 The log Gaussian Cox process for the spatial point intensity*

To start with we assume that the spatial point intensity can be modelled as a log Gaussian Cox process (a specific kind of Cox process proposed by Møller *et al*., 1998). According to this model each point pattern represents a partial realization of an inhomogeneous Poisson process characterized by a spatial intensity function $\lambda(x)$, with *x* representing the spatial coordinates of an arbitrary point (see Diggle, 2003). The values of $\lambda(x)$ constitute a realization of a positive random field $\{\Lambda(x)\}$ such that $\Lambda(x) = exp\{S(x)\}$, where $\{S(x)\}$ is a Gaussian random field with mean $\mu_S$, variance $\sigma_S^2$ and correlation function $\rho_S(d)$. $\{\Lambda(x)\}$ is known as a log Gaussian Cox process.

The log Gaussian assumption is particularly useful because explicit expressions can be derived for the intensity and covariance structure of the point process. Indeed, according to the moment generating function of a log Gaussian distribution, the intensity $\lambda$ of a log Gaussian Cox process $\{\Lambda(x)\}$ can be written as:

$$\lambda = E[\Lambda(x)] = E[exp(S(x))] = exp\left(\mu_S + \frac{1}{2}\sigma_S^2\right).$$

Concerning to the covariance structure, for any arbitrary pairs of points (say $x$ and $x'$), $\Lambda(x)\Lambda(x') = exp\{S(x) + S(x')\}$, and $S(x) + S(x')$ is also Gaussian with mean $m = 2\mu_S$ and variance $v = 2\sigma_S^2[1 + \rho_S(d)]$ where $d$ is the Euclidean distance between $x$ and $x'$. As a result, $E[\Lambda(x)\Lambda(x')] = exp(m + v/2)$, and hence:

$$E[\Lambda(x)\Lambda(x')] = \lambda\, exp\{\sigma_S^2 \rho_S(d)\}.$$

*3.2.2 The marks process*

Our model assumes that the mark $m(x_n)$ attached to the point $x_n$ generated by the log Gaussian Cox process depends on the intensity of the process itself. More formally we have:

$$m(x_n) = a\Lambda(x_n) + b\mathrm{E}(x_n) \qquad (3)$$

where $\Lambda(x_n)$ is the value of the spatial intensity at point $x_n$ and $\mathrm{E}(x_n)$ is due to a residual process such that $\mathrm{E}(x) = exp\{R(x)\}$, where $R(x)$ is a Gaussian random field with mean $\mu_R$, variance $\sigma_R^2$ and correlation function $\rho_R(d)$. Thus, the expected value of process $\mathrm{E}(x)$, indicated with $\varepsilon$, is $\varepsilon = E[exp\{R(x)\}] = exp\left\{\mu_R + \frac{1}{2}\sigma_R^2\right\}$[3].

The two constants *a* and *b* appearing in Equation (3) are the model parameters. It is important to understand the role of these two parameters in the generation of the patterns of firms and the way in which they can model the relationship between the intensity with which firms are distributed in space and their dimension. More specifically, *a* is the parameter driving the correlation between the spatial point intensity process and the marks process. When $a = 0$ the marks are independent of the spatial intensity. Conversely when $a > 0$ the marks process generates marks that tend to be larger (that is larger firms) in regions characterized by a high spatial point intensity. Finally, in those cases where $a < 0$ the marks tend to be smaller (and hence the firms of smaller dimension) in regions

[3] In order to avoid any misunderstanding, note that the greek letter $\mathrm{E}$, used to indicate the residual process, and the expectation operator $E$ are different symbols.

characterized by a high spatial point intensity. On the other hand the parameter *b* represents the perturbation effect of the residual process on the correlation between marks and intensity. The larger is *b* in absolute value, the more the residual process disturbs the phenomenon of correlation controlled by *a*.

The log Gaussian assumption makes the computation of the expected value of the marks process mathematically tractable, indeed we have:

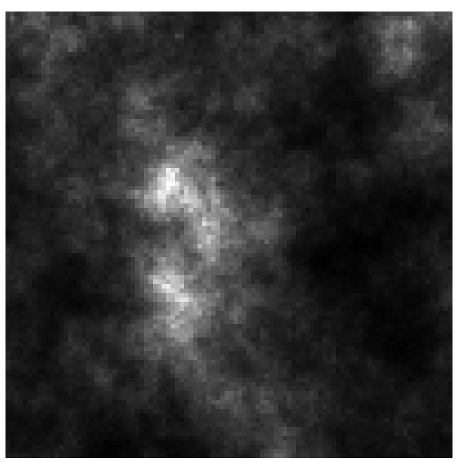
$$\mu = E[m(x)] = a\lambda \, exp\{\sigma_S^2\} + b\varepsilon .$$

It is easy to show that the expected value of the marks process would be $a\lambda + b\varepsilon$. However, following Ho and Stoyan (2008), the true unbiased expected value is $\mu = a\lambda \, exp\{\sigma_S^2\} + b\varepsilon$, which is larger than $a\lambda + b\varepsilon$ when $a > 0$, and smaller when $a < 0$. For a detailed explanation of this bias correction see Ho and Stoyan (2008).

The model proposed here is particularly interesting having in mind economic application and specifically the study of firm location. In fact in the application of the present methodological framework to the problem of assessing industrial agglomeration, the marked point patterns generated when $a = 0$ represent the null hypothesis of spatial randomness of firms. Similarly, $a > 0$ and $a < 0$ refer to the alternative hypothesis of spatial concentration of economic activities in the sense expressed in "*Case 1*" and "*Case 2*", respectively, in Section 1.

To better illustrate how the model works, in the reminder of this section we will show some realizations of a marked point process. In what follows all the generated patterns are obtained using the same random seed so that all realizations are directly comparable and the differences between the patterns can be ascribed only to differences in the model parameters. Figure 2 shows the realization of the underlying spatial point intensity process given as $\Lambda(x) = esp\{S(x)\}$ on the unit square, with mean $\mu_S = 5$, variance $\sigma_S^2 = 0.25$ and correlation function $\rho_S(d) = exp\{-d/0.25\}$[4]. As we can see, in this particular realisation, the spatial point intensity tends to be higher (light grey colours) towards the centre of the unitary area.

**Figure 2:** A realization of the underlying spatial point intensity (grey-scale image).



In order to illustrate the role of parameter *a* in driving the correlation between the spatial point intensity and the marks Figure 3 displays different realizations of the marked point process with different values for *a*. The six simulated marked point patterns appearing in Figure 3 show the net

[4] This specific form of the correlation function is known as the *exponential function*; see Diggle and Ribeiro (2007) for details.

effect of parameter *a* since *b* is always set to zero. In each pattern the marks are rescaled to the unit interval and each point is represented by a circle with radius proportional to its rescaled mark. Figure 3 shows quite clearly that, for positive values of *a*, the marks tend to be larger where the spatial point intensity is higher, that is approximately at the centre of the unitary area (see pattern *i*, *iii* and *v*). On the other hand, for negative values of *a*, the marks tend to be smaller where the spatial point intensity is higher (see pattern *ii*, i*v* and *vi*). The two kind of clustering situation – namely, "*Case 1*" and "*Case 2*" – tend to be more evident when *a* increases in absolute value.

Figure 4 shows six simulated marked point patterns with different values for *b* which illustrate the role of this parameter in disturbing the correlation between the spatial point intensity and the marks. In all six cases the residual process $\mathrm{E}(x)$ is characterised by $\mu_R = 5$, $\sigma_R^2 = 0.25$ and $\rho_R(d) = exp\{-d/0.25\}$ and *a* is set to be equal to 0.25. To understand how the parameter *b* disturbs the effect of the parameter *a*, we can compare the patterns of Figure 4 with the pattern of Figure 3(*i*) where *a* = 0.25. As *b* increases in absolute terms, the residual process becomes relatively more important in generating the marked point patterns. In this situation the correlation between the spatial point intensity and the marks depicted by the pattern reported in Figure 3(*i*) becomes less strong.

*3.2.3 The benchmark value of the mark-weighted K-function*

Because of the mathematical tractability of the model defined above, the corresponding theoretical mark-weighted *K*-function can be derived in a closed form. Indeed, for such a marked log-Gaussian Cox process (for *d* >0), the mark-weighted *K*-function assumes the form:

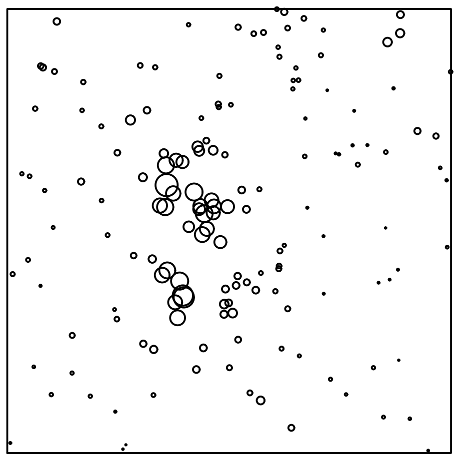
$$K_{mm}(d) = 2\pi \int_0^d u \frac{a^2\lambda^2 exp\{2\sigma_S^2 + 3\sigma_S^2\rho_S(d)\} + 2ab\lambda\, exp\left\{\sigma_S^2 + \frac{3}{2}\sigma_S^2\rho_S(d)\right\}\varepsilon + b^2\varepsilon^2 exp\{\sigma_R^2\rho_R(d)\}}{\left[a\lambda\, exp\{\sigma_S^2\} + b\varepsilon\right]^2} du \tag{4}$$

The formal derivation of Equation (4) is reported in the Appendix. Equation (4) above allows us to develop a test for the presence of absolute concentration of economic activities using the mark-weighted *K*-function, in which the null hypothesis of spatial randomness of firms is represented by the values of $K_{mm}(d)$ when *a* = 0. In fact, when *a* = 0, then we have:

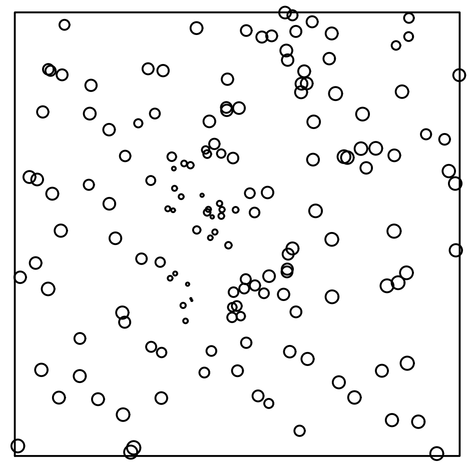
$$K_{mm}(d) = 2\pi \int_0^d u\, exp\{\sigma_R^2 \rho_R(d)\} du\,. \tag{5}$$

To help the visualization, Figure 5 shows the mean of $\hat{K}_{mm}(d)$ for 1000 marked point patterns generated in the unit square from model (3) with parameters $\mu_S = 5$, $\sigma_S^2 = 0.25$, $\rho_S(d) = exp\{-d/0.25\}$, $\mu_R = 0$, $\sigma_R^2 = 0.25$, $\rho_R(d) = exp\{-d/0.25\}$, *a* = 0 and *b* = 1. Since the theoretical function (dashed line), given by Equation (5), lies within the confidence envelopes (resulting from the highest and lowest values of $\hat{K}_{mm}(d)$ calculated from the 1000 simulations) and very close to the mean of $\hat{K}_{mm}(d)$ (solid line), the graph confirms that Equation (5) may well represent the proper benchmark to verify the presence of spatial concentration of economic activities.

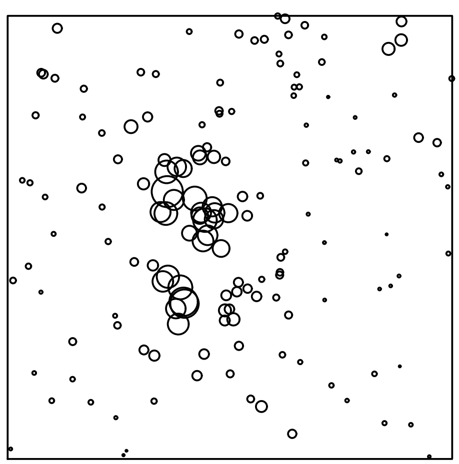
**Figure 3:** Simulated patterns of marks according to model (3). The figure illustrates the role of parameter *a* when *b* = constant = 0.



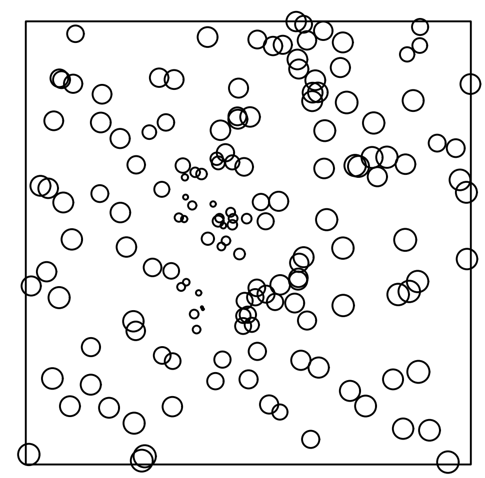
*i*) *a* = 0.25; *b* = 0



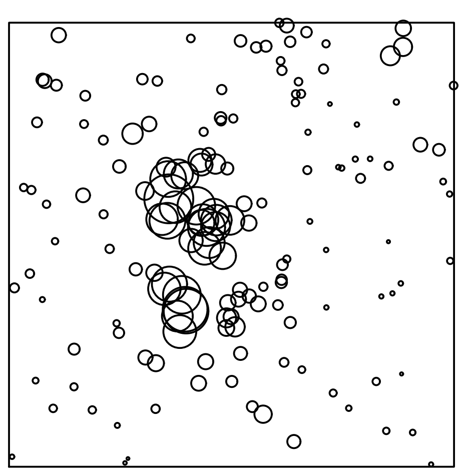
*ii*) *a* = – 0.25; *b* = 0



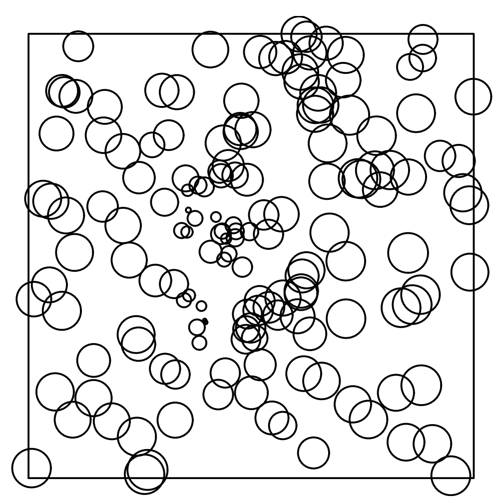
*iii*) *a* = 0.5; *b* = 0



i*v*) *a* = – 0.5; *b* = 0

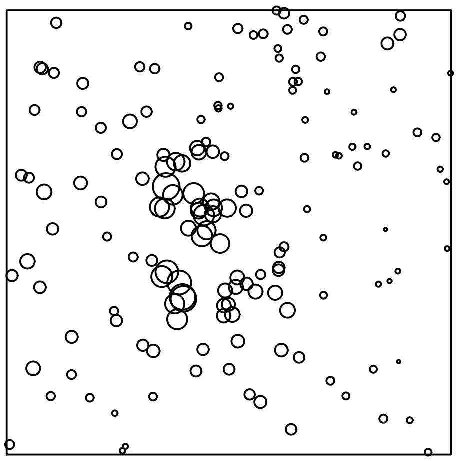


*v*) *a* = 1; *b* = 0

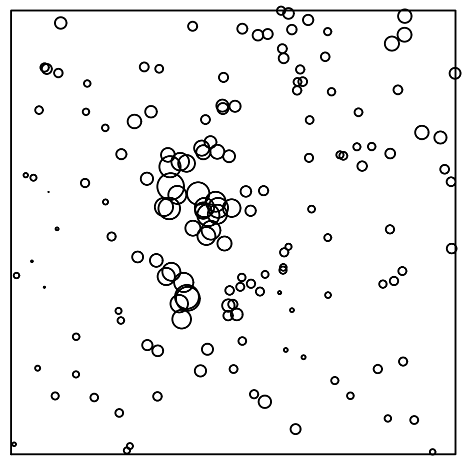


*vi*) *a* = – 1; *b* = 0

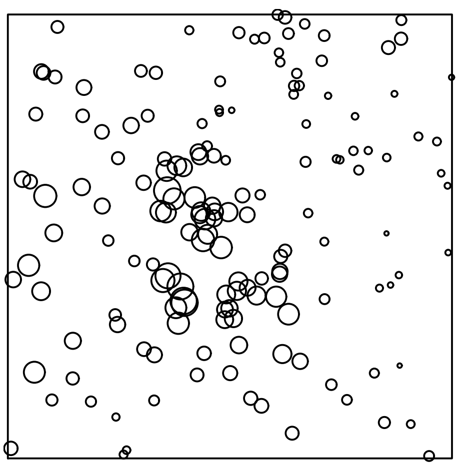
**Figure 4:** Simulated patterns of marks according to model (3). The figure illustrates the role of parameter *b* when *a* = constant = 0.25.



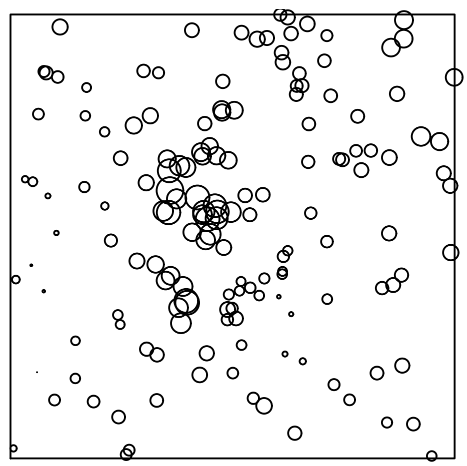
*i*) *a* = 0.25; *b* = 0.25



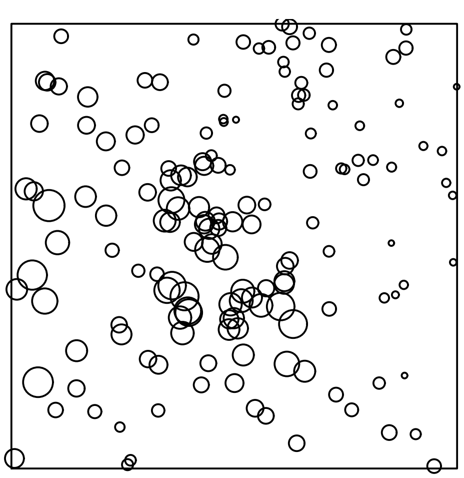
*ii*) *a* = 0.25; *b* = – 0.25



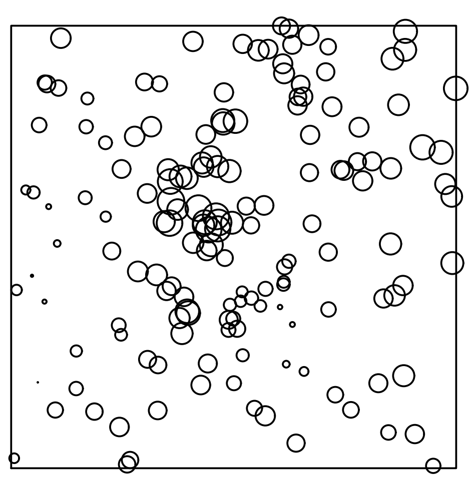
*iii*) *a* = 0.25; *b* = 0.5



*iv*) *a* = 0.25; *b* = – 0.5

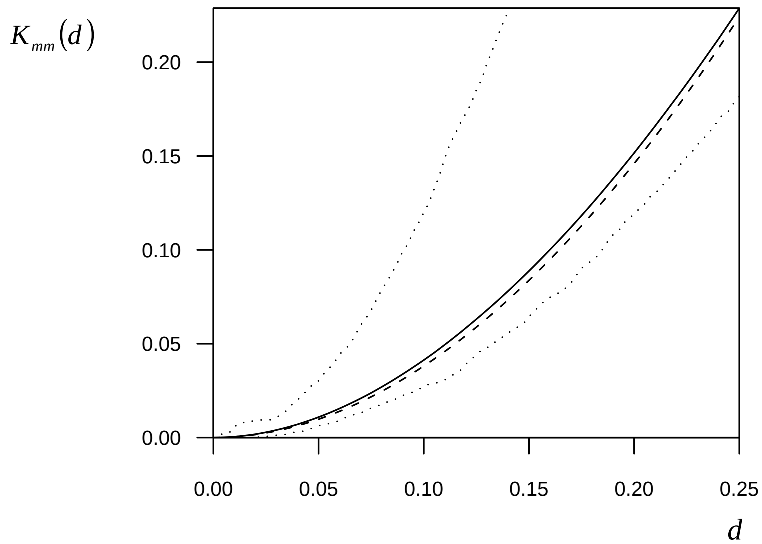


*v*) *a* = 0.25; *b* = 1



*vi*) *a* = 0.25; *b* = – 1

**Figure 5:** Mean of $\hat{K}_{mm}(d)$ estimated from 1000 simulations of the marked point process following model (3) with parameters $a$ = 0 and $b$ = 1. The behaviour of the empirical mean is represented by the solid line. The theoretical function given by (5) is reported in the graph as a dashed line.




## 4. Discussion and conclusions

The spatial concentration of firms has long been a central issue in economics both under the theoretical and the applied point of view due mainly to the important policy implications. An approach to its measurement, that became recently very popular, makes use of micro data and looks at the firms as if they were dimensionless points distributed in the economic space. This approach is very attractive because it does not suffer from the problem of choosing an arbitrary partition of the economic space (such as e.g. regions, counties or countries). However in practical circumstances this is an excessive simplification since the points (firms) observed in the economic space are far from being dimensionless and are conversely characterized by different dimension measured in terms of the number of employees, the product, the capital and so on. In the literature, the papers that introduced such an approach (e.g. Arbia and Espa, 1996; Marcon and Puech, 2003) disregard the aspect of the different firm dimension and ignore the fact that a high degree of spatial concentration may result from the case of many small points clustering in definite portions of space (as it is usually considered in the literature), but also from only few large points clustering together (e.g. few large firms). In other words they are not able to distinguish between two very different issues, namely the *clustering of firms* and the *clustering of economic activities.* The aim of this paper was to introduce absolute measures of spatial concentration of firms based on an extension of Ripley's *K*-function that accounts for the different firm dimension. In order to derive the null hypothesis of spatial randomness in this more complex environment, we developed a new stochastic model that generates marked point patterns of firms and is able to describe the various situations that could arise in empirical cases. In our model the firm dimension is expressed as a function of the spatial intensity of the point process. According to the different values assumed by the model parameters, this could result either in larger points located in areas with high intensity or, conversely, smaller points located in areas characterized by high intensity. The first case is more grounded under the economic point view where we can postulate that the same conditions that lead to a higher clustering of firms in some portions of space may also lead to the growth of the dimension of the existing firms. A good example is constituted by the action of the three Marshallian forces fostering agglomeration (Marshall, 1920). In his seminal work Marshall emphasized that industrial agglomeration can be explained by the fact that firms try to locate near suppliers to save shipping costs, by the theory of labor market pooling and by the theory of knowledge spillovers. If some of the services are internalized in one leading big company than the same forces could produce a growth of the firms' dimension rather than an increase in the number of firms located in the area. We would expect therefore that in most practical cases the parameter *a* in Equation (3) will be positive and large in absolute value. Similar arguments reinforcing this empirical expectation may be found in Krugman (1991).

On the basis of the stochastic model introduced here we derived the corresponding mark-weighted *K*-function and, by making use of some simulated pattern, we presented evidence that this tool represents a proper mean to detect the presence of absolute concentration of firms keeping their dimension into account.

The problem of calibrating the values of the model's parameters in practical cases is complex and it is not undertaken here where we restricted ourselves to only the presentation of the stochastic mechanism. The inferential aspects would involve the estimation of the parameters *a* and *b* in Equation (3) and also of the parameters characterising the two log Gaussian processes $\Lambda(x) = esp\{S(x)\}$ and $\mathrm{E}(x) = exp\{R(x)\}$ introduced in Section 3.2. A closed form for the likelihood of the model is not yet available at current state of the literature and currently the only viable possibility appears to be to exploit (as it is usual practice in such instances) a pseudo-likelihood approach as indicated in Besag (1974). We will undertake such an approach in some future work.

**References**


Arbia G., Espa G. (1996) *Statistica economica territoriale*, Cedam, Padua.

Arbia G., Espa G., Quah D. (2008) A class of spatial econometric methods in the empirical analysis of clusters of firms in the space, *Empirical Economics*, 34, 81–103.

Arbia G., Espa G., Giuliani D., Mazzitelli A. (2010) Detecting the existence of space-time clustering of firms, *Regional Science & Urban Economics*, 40, 311–323.

Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussions), *Journal of the Royal Statistical Society*, B, 36, 192–236.

Besag J., Diggle P.J. (1977) Simple Monte Carlo tests for spatial pattern, *Applied Statistics*, 26, 327–333.

Bickenbach F., Bode E. (2008) Disproportionality Measures of Concentration, Specialization, and Localization, *International Regional Science Review*, 31, 359–388.

Boots B.N., Getis A. (1988) *Point pattern analysis*, Sage Scientific Geography Series, Vol. 8, Sage Publications, London.

Diggle P.J. (2003) *Statistical analysis of spatial point patterns*, 2nd edition, Edward Arnold, London.

Diggle P. J., Ribeiro Jr P. J. (2007) *Model-based Geostatistics*, Springer, New York.

Dixon P. (2002) Ripley's *K*-function, In: El-Shaarawi A.H. and Piergorsch W.W. (eds.) *The encyclopedia of environmetrics*, pp. 1976–1803, John Wiley & Sons Ltd, New York.

Duranton G., Overman H.G. (2005) Testing for localisation using micro-geographic data, *Review of Economic Studies*, 72, 1077–1106.

Duranton G., Overman H.G. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data, *Journal of Regional Science*, 48, 213–243.

Ellison G., Glaeser E.L. (1997) Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105, 889–927.

Fujita M., Krugman P., Venables A. (1999) *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge.

Gini C. (1912) *Variabilità e mutabilità*, Reprinted in *Memorie di metodologica statistica*, Pizetti E. and Salvemini T. (eds.), Rome.

Gini C. (1921) Measurement of Inequality of Incomes, *The Economic Journal*, 31, 121, 124–126.

Haaland J.I., Kind H.J., Midelfart-Knarvik K.H., Torstensson J. (1999) What determines the

economic geography of Europe? *Centre for Economic Policy Research*, Discussion paper, 2072.
Ho L. P., Stoyan D. (2008) Modelling marked point patterns by intensity-marked Cox processes, *Statistics & Probability Letters*, 78, 1194–1199.
Illian J., Penttinen A., Stoyan H., Stoyan D. (2008) *Statistical analysis and modelling of spatial point pattern*, John Wiley & Sons, Chichester.
Krugman P. (1991) *Geography and Trade*, MIT Press, Cambridge.
Marcon E., Puech F. (2003) Evaluating the geographic concentration of industries using distance-based methods, *Journal of Economic Geography*, 3, 409–428.
Marcon E., Puech F. (2009) Measures of the geographic concentration of industries: improving distance-based methods, *Journal of Economic Geography*, doi: 10.1093/jeg/lbp056.
Marshall A. (1920) *Principles of Economics*, revised edition, Macmillan, London.
Mori T., Nishikimi K., Smith T.E. (2005) A divergence statistic for industrial localization, *Review of Economics and Statistics*, 87, 635–651.
Møller J., Syversveen A.R., Waagepetersen R.P. (1998) Log Gaussian Cox processes, *Scandinavian journal of statistics*, 25, 451–482.
Penttinen A. (2006) Statistics for Marked Point Patterns, *The Yearbook of the Finnish Statistical Society*, 70–91.
Ripley B.D. (1976) The second-order analysis of stationary point processes, *Journal of Applied Probability*, 13, 255–266.
Ripley B.D. (1977) *Modelling Spatial Patterns* (with discussion), *Journal of the Royal Statistical Society*, B, 39, 172–212.

**Appendix: Analytical derivation of the theoretical mark-weighted *K*-function**

The mark-weighted *K*-function $K_{mm}(d)$ can be conceived as the integral of the mark correlation function $k_{mm}(d)$ (Illian *et al.*, 2008), i.e.

$$K_{mm}(d) = 2\pi\int_0^d u k_{mm}(u)du \, . \tag{6}$$

The mark correlation function can be given by:

$$k_{mm}(d) = \frac{E_{ot}[m(o)m(t)]}{\mu^2} \tag{7}$$

where $E_{ot}[m(o)m(t)]$ denotes the conditional mean under the condition that there are points in two arbitrary locations separated by a distance *d*, which are considered as the origin *o* and the destination *t*. *m*(*o*) and *m*(*t*) are the marks attached to the points located in *o* and *t* respectively. The term in the denominator $\mu$ represents the mean of the marks. Therefore $k_{mm}(d)$ can be interpreted as the normalized mean of the product of the marks of a pair of points separated by a distance *d.*

According to Ho and Stoyan (2008), the numerator of $k_{mm}(d)$ satisfies the condition that:

$$E_{ot}[m(o)m(t)] = \frac{E[m(o)m(t)\Lambda(o)\Lambda(t)]}{E[\Lambda(o)\Lambda(t)]} \, . \tag{8}$$

If $\Lambda(x)$ is defined as in section 3.2.1 and $m(x)$ is given by equation (3) then

$$
\begin{aligned}
E[m(o)m(t)\Lambda(o)\Lambda(t)] &= E[[a\,exp\{S(o)\}+b\,exp\{R(o)\}][a\,exp\{S(t)\}+b\,exp\{R(t)\}]exp\{S(o)S(t)\}] \\
&= a^2E[exp\{2S(o)+2S(t)\}]+abE[exp\{2S(o)+S(t)+R(t)\}] \\
&\quad +abE[exp\{S(o)+2S(t)+R(o)\}]+b^2E[R(o)+R(t)+S(o)+S(t)] \\
&= a^2\lambda^4\,exp\{2\sigma_S^2+4\sigma_S^2\rho_S(d)\}+2ab\lambda^3\,exp\left\{\sigma_S^2+\frac{5}{2}\sigma_S^2\rho_S(d)\right\}\varepsilon \\
&\quad +b^2\lambda^2\,exp\{\sigma_S^2\rho_S(d)\}\varepsilon^2\,exp\{\sigma_R^2\rho_R(d)\}
\end{aligned}
$$

and

$$E[\Lambda(o)\Lambda(t)]=\lambda^2\,exp\{\sigma_S^2\rho_S(d)\}$$

Therefore Equation (8) can be written as

$$E_{ot}[m(o)m(t)]=a^2\lambda^2\,exp\{2\sigma_S^2+3\sigma_S^2\rho_S(d)\}+2ab\lambda\,exp\left\{\sigma_S^2+\frac{3}{2}\sigma_S^2\rho_S(d)\right\}\varepsilon+b^2\varepsilon^2\,exp\{\sigma_R^2\rho_R(d)\}$$

As a result, since $\mu=a\lambda\,exp\{\sigma_S^2\}+b\varepsilon$, the mark correlation function has the following form:

$$k_{mm}(d)=\frac{a^2\lambda^2\,exp\{2\sigma_S^2+3\sigma_S^2\rho_S(d)\}+2ab\lambda\,exp\left\{\sigma_S^2+\frac{3}{2}\sigma_S^2\rho_S(d)\right\}\varepsilon+b^2\varepsilon^2\,exp\{\sigma_R^2\rho_R(d)\}}{\left[a\lambda\,exp\{\sigma_S^2\}+b\varepsilon\right]^2},\ d>0. \tag{9}$$

Finally, by substituting Equation (9) in Equation (6) we obtain, for $d$ > 0, the explicit form of the mark-weighted $K$-function:

$$K_{mm}(d)=2\pi\int_0^d u\,\frac{a^2\lambda^2\,exp\{2\sigma_S^2+3\sigma_S^2\rho_S(d)\}+2ab\lambda\,exp\left\{\sigma_S^2+\frac{3}{2}\sigma_S^2\rho_S(d)\right\}\varepsilon+b^2\varepsilon^2\,exp\{\sigma_R^2\rho_R(d)\}}{\left[a\lambda\,exp\{\sigma_S^2\}+b\varepsilon\right]^2}du \tag{10}$$

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

2008.1 *A Monte Carlo EM Algorithm for the Estimation of a Logistic Auto-logistic Model with Missing Data,* by Marco Bee and Giuseppe Espa.

2008.2 *Adaptive microfoundations for emergent macroeconomics*, Edoardo Gaffeo, Domenico Delli Gatti, Saul Desiderio, Mauro Gallegati.

2008.3 *A look at the relationship between industrial dynamics and aggregate fluctuations*, Domenico Delli Gatti, Edoardo Gaffeo, Mauro Gallegati.

2008.4 *Demand Distribution Dynamics in Creative Industries: the Market for Books in Italy*, Edoardo Gaffeo, Antonello E. Scorcu, Laura Vici.

2008.5 *On the mean/variance relationship of the firm size distribution: evidence and some theory*, Edoardo Gaffeo, Corrado di Guilmi, Mauro Gallegati, Alberto Russo.

2008.6 *Uncomputability and Undecidability in Economic Theory*, K. Vela Velupillai.

2008.7 *The Mathematization of Macroeconomics*: *A Recursive Revolution*, K. Vela Velupillai.

2008.8 *Natural disturbances and natural hazards in mountain forests: a framework for the economic valuation*, Sandra Notaro, Alessandro Paletto

2008.9 *Does forest damage have an economic impact? A case study from the Italian Alps,* Sandra Notaro, Alessandro Paletto, Roberta Raffaelli.

2008.10 *Compliance by believing: an experimental exploration on social norms and impartial agreements,* Marco Faillo, Stefania Ottone, Lorenzo Sacconi.

2008.11 *You Won the Battle. What about the War? A Model of Competition between Proprietary and Open Source Software*, Riccardo Leoncini, Francesco Rentocchini, Giuseppe Vittucci Marzetti.

2008.12 *Minsky's Upward Instability: the Not-Too-Keynesian Optimism of a Financial Cassandra,* Elisabetta De Antoni.

2008.13 *A theoretical analysis of the relationship between social capital and corporate social responsibility: concepts and definitions,* Lorenzo Sacconi, Giacomo Degli Antoni.

2008.14 *Conformity, Reciprocity and the Sense of Justice. How Social Contract-based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence,* Lorenzo Sacconi, Marco Faillo.

2008.15 *The macroeconomics of imperfect capital markets. Whither saving-investment imbalances?* Roberto Tamborini

2008.16 *Financial Constraints and Firm Export Behavior*, Flora Bellone, Patrick Musso, Lionel Nesta and Stefano Schiavo

2008.17 *Why do frms invest abroad? An analysis of the motives underlying Foreign Direct Investments Financial Constraints and Firm Export Behavior*, Chiara Franco, Francesco Rentocchini and Giuseppe Vittucci Marzetti

2008.18 *CSR as Contractarian Model of Multi-Stakeholder Corporate Governance and the Game-Theory of its Implementation*, Lorenzo Sacconi

2008.19 *Managing Agricultural Price Risk in Developing Countries*, Julie Dana and Christopher L. Gilbert

2008.20 *Commodity Speculation and Commodity Investment*, Christopher L. Gilbert

2008.21 *Inspection games with long-run inspectors*, Luciano Andreozzi

2008.22 *Property Rights and Investments: An Evolutionary Approach*, Luciano Andreozzi

2008.23 *How to Understand High Food Price*, Christopher L. Gilbert

2008.24 *Trade-imbalances networks and exchange rate adjustments: The paradox of a new Plaza*, Andrea Fracasso and Stefano Schiavo

2008.25 *The process of Convergence Towards the Euro for the VISEGRAD – 4 Countries*, Giuliana Passamani

2009.1 *Income Shocks, Coping Strategies, and Consumption Smoothing. An Application to Indonesian Data*, Gabriella Berloffa and Francesca Modena

2009.2 *Clusters of firms in space and time*, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani e Andrea Mazzitelli

2009.3 *A note on maximum likelihood estimation of a Pareto mixture*, Marco Bee, Roberto Benedetti e Giuseppe Espa

2009.4 *Job performance and job satisfaction: an integrated survey*, Maurizio Pugno e Sara Depedri

2009.5 *The evolution of the Sino-American co-dependency: modeling a regime switch in a growth setting*, Luigi Bonatti e Andrea Fracasso

2009.6 *The Two Triangles: What did Wicksell and Keynes know about macroeconomics that modern economists do not (consider*)? Ronny Mazzocchi, Roberto Tamborini e Hans-Michael Trautwein

2009.7 *Mobility Systems and Economic Growth: a Theoretical Analysis of the Long-Term Effects of Alternative Transportation Policies*, Luigi Bonatti e Emanuele Campiglio

2009.8 *Money and finance: The heterodox views of R. Clower, A. Leijonhufvud and H. Minsky*, Elisabetta de Antoni

2009.9 *Development and economic growth: the effectiveness of traditional policies*, Gabriella Berloffa, Giuseppe Folloni e Ilaria Schnyder

2009.10 *Management of hail risk: insurance or anti-hail nets?*, Luciano Pilati, Vasco Boatto

2010.1 *Monetary policy through the "credit-cost channel". Italy and Germany pre- and post-EMU*, Giuliana Passamani, Roberto Tamborini

2010.2 *Has food price volatility risen*? Chrisopher L. Gilbert and C. Wyn Morgan

2010.3 *Simulating copula-based distributions and estimating tail probabilities by means of Adaptive Importance Sampling*, Marco Bee

2010.4 *The China-US co-dependency and the elusive costs of growth rebalancing*, Luigi Bonatti and Andrea Fracasso

2010.5 *The Structure and Growth of International Trade*, Massimo Riccaboni and Stefano Schiavo

2010.6 *The Future of the Sino-American co-Dependency*, Luigi Bonatti and Andrea Fracasso

2010.7 *Anomalies in Economics and Finance*, Christopher L. Gilbert

2010.8 *Trust is bound to emerge (In the repeated Trust Game)*, Luciano Andreozzi

2010.9 *Dynamic VaR models and the Peaks over Threshold method for market risk measurement: an empirical investigation during a financial crisis*, Marco Bee e Fabrizio Miorelli

2010.10 *Creativity in work settings and inclusive governance: Survey based findings from Italy*, Silvia Sacchetti, Roger Sugden e Ermanno Tortia

2010.11 *Food Prices in Six Developing Countries and the Grains Price Spike*, Christopher L. Gilbert

2010.12 *Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration*, Giuseppe Espa, Diego Giuliani, Giuseppe Arbia

The Discussion Paper series provides a means for circulating preliminary research results by staff of or visitors to the Department. Its purpose is to stimulate discussion prior to the publication of papers.

Requests for copies of Discussion Papers and address changes should be sent to:
Dott. Luciano Andreozzi
e-mail: luciano.andreozzi@unitn.it
Department of Economics
University of Trento
via Inama, 5 - 38122 Trento (IT)