

UN METODO PER L'ANALISI EMPIRICA DEI FENOMENI DI CONCENTRAZIONE
SPAZIALE DELLE ATTIVITÀ ECONOMICHE BASATO SU DATI MICRO-
GEOGRAFICI

Diego Giuliani¹, Giuseppe Espa²

SOMMARIO

La concentrazione spaziale delle attività economiche, a causa delle implicazioni che può avere per la crescita economica locale e le disparità territoriali, è un fenomeno di grande interesse per l'economia e le scienze regionali. Un recente approccio alla sua misurazione, basato sulle metodologie della *point pattern analysis*, utilizza dati micro-geografici e considera le imprese come punti privi di dimensione distribuiti nello spazio economico. Tale approccio trascura l'aspetto della diversa dimensione delle imprese e quindi ignora il fatto che un grado elevato di concentrazione spaziale possa essere dovuto a un numero elevato di imprese di piccole dimensioni localizzate in un'area geografica ristretta o a poche imprese di grandi dimensioni localizzate vicine. In questo articolo, facendo riferimento alla teoria dei processi di punto marcati (Penttinen, 2006), il problema viene affrontato adattando la funzione K di Ripley in modo tale che tenga conto della dimensione dei punti. Per illustrare il metodo proposto, viene presentata un'applicazione empirica all'analisi della distribuzione spaziale delle imprese manifatturiere ad *alta* e *medio-alta* intensità tecnologica nei comuni di Milano e Torino.

¹ Università di Trento, Via Inama 5, 38100, Trento, e-mail: diego_giuliani@yahoo.it.

² Dipartimento di Economia, Università di Trento, Via Inama 5, 38100, Trento, e-mail: giuseppe.espa@economia.unitn.it.

1 Introduzione

L'intensificarsi dei fenomeni di globalizzazione, accompagnato allo sviluppo crescente delle tecnologie dei trasporti e delle telecomunicazioni, ha reso i processi decisionali delle imprese meno dipendenti dallo spazio geografico. Negli ultimi dieci anni, infatti, la distanza spaziale ha rappresentato sempre meno un fattore limitante della circolazione dei beni, dei capitali e della conoscenza. Ciò avrebbe dovuto spingere verso una distribuzione territoriale diffusa delle attività economiche. Paradossalmente la tendenza delle imprese a concentrarsi nello spazio sembra, al contrario, essersi intensificata (si veda, per esempio, Storper e Venables, 2003; Enright, 2000).

Le scelte localizzative degli agenti economici che portano all'agglomerazione industriale non sono casuali o fortuite. Concentrandosi in alcune aree specifiche le imprese possono ottenere numerosi vantaggi, che hanno poi ricadute significative per la crescita economica locale e le disparità regionali e, di conseguenza, sono oggetto di studio per l'economia, l'organizzazione industriale, la geografia e oggetto di interesse per i decisori pubblici.

È stato tuttavia riconosciuto che un ostacolo rilevante allo studio dei fenomeni di agglomerazione delle imprese è rappresentato dalla mancanza di metodi appropriati per la misurazione della loro concentrazione spaziale (Combes e Overman, 2004; Combes *et al.*, 2008). Le misure comunemente più utilizzate dagli economisti (quali gli indici di Gini, di Herfindhal o di Ellison-Glaeser), infatti, non sono completamente affidabili. Il loro limite metodologico principale riguarda l'utilizzo di aggregati regionali costruiti sulla base di definizioni arbitrarie delle unità spaziali, coincidenti generalmente con i confini amministrativi (quali i comuni, le province o le regioni), che introducono distorsioni statistiche dovute alla definizione scelta dello spazio. Questo problema può essere superato ricorrendo ad un approccio continuo allo spazio, in cui i dati sono rilevati al livello massimo di disaggregazione spaziale, ossia ciascuna impresa è identificata dalle sue coordinate geografiche, e la concentrazione spaziale è misurata in riferimento alla distribuzione delle distanze tra le attività economiche. Le tecniche di riferimento per l'approccio in parola sono quelle della *point pattern analysis*.

In una serie di recenti articoli, Duranton e Overman (2005), Marcon e Puech (2010) e Arbia *et al.* (2011) hanno adottato questo approccio e costruito test formali per la presenza di fenomeni di concentrazione spaziale considerando la distribuzione geografica delle attività economiche come una mappa di punti e applicando misure di sintesi basate sulla funzione K di Ripley (Ripley, 1976). Nella costruzione di test basati sull'impiego di dati micro-geografici sulle imprese, un elemento che spesso è necessario tenere in considerazione è la dimensione di ciascuna impresa (misurata, per esempio, in termini di numero di occupati, di fatturato o di valore aggiunto). Un livello elevato di concentrazione spaziale, infatti, può essere identificato,

come evidenziato in Devereux *et al.* (1999), da almeno due fatti stilizzati molto diversi (si veda Figura 1). Segnatamente, un'industria può presentare elevata concentrazione spaziale se:

- un numero elevato di imprese di piccole dimensioni è localizzato nella medesima area geografica ristretta (*Caso 1*), oppure
- l'attività economica è concentrata in poche imprese di grandi dimensioni localizzate vicine (*Caso 2*).

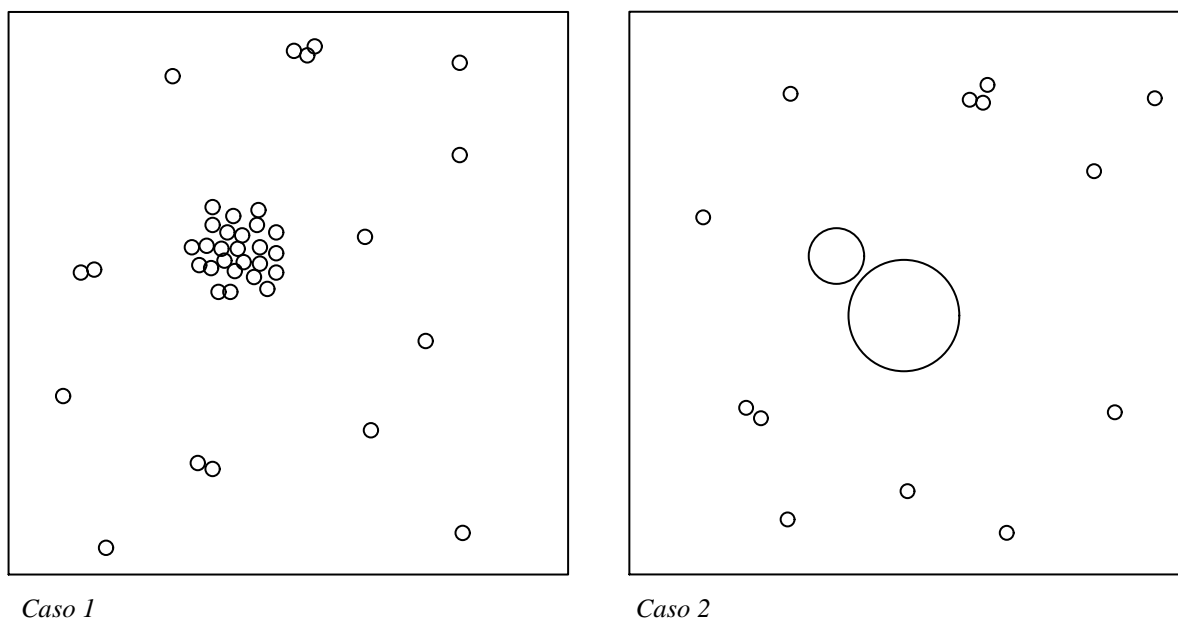


Figura 1 - Due situazioni paradigmatiche estreme di concentrazione spaziale dell'attività economica

Una misura adeguata di concentrazione spaziale deve quindi essere in grado di riconoscere come *cluster* spaziali di imprese entrambi questi due casi paradigmatici estremi, così come tutte le possibili situazioni intermedie. Una tale misura deve pertanto tenere in considerazione l'impatto della dimensione delle imprese sull'agglomerazione industriale.

Nel rispetto di questa proprietà, Marcon e Puech (2010) e Duranton e Overman (2005) hanno sviluppato delle misure di concentrazione spaziale basate su estensioni della funzione K di Ripley. In particolare, tali estensioni consentono di tenere conto della dimensione delle attività economiche in una struttura in cui la dimensione dell'impresa è trattata come un peso associato a ciascun punto della mappa geografica delle imprese. In entrambi i lavori citati, le deviazioni dalla casualità spaziale sono rappresentate da concentrazioni più elevate delle imprese appartenenti a una specifica industria rispetto alla distribuzione spaziale delle imprese dell'intera economia. In altre parole, questi autori hanno sviluppato misure relative in cui la concentrazione (o dispersione) spaziale delle attività economiche emerge quando il *pattern* di uno specifico settore è più aggregato (o diffuso) del *pattern* dell'intera economia. Sebbene garantiscano il controllo delle caratteristiche specifiche del territorio in esame, le misure

relative non consentono il confronto diretto tra economie e territori diversi (si veda Haaland *et al.*, 1999 e Mori *et al.*, 2005 per una trattazione dettagliata del problema).

Nel presente lavoro proponiamo un'estensione simile della funzione K che porti a una misura *assoluta* (anziché *relativa*) di concentrazione spaziale delle attività economiche e che pertanto garantisca la comparabilità tra realtà geografiche e industriali differenti. Più precisamente, in riferimento alla teoria statistica dei processi di punto marcati, sviluppiamo un meccanismo stocastico per la generazione di mappe di punti pesati di imprese che possano adeguatamente rappresentare i fatti stilizzati dei diversi fenomeni di localizzazione che si verificano nei casi reali di studio (essenzialmente: casualità spaziale o concentrazione spaziale nella prospettiva indicata nel “Caso 1” e nel “Caso 2” definiti sopra). I valori che la misura proposta assume per questi *pattern* paradigmatici costituiscono i valori appropriati dell'ipotesi nulla che consente di testare formalmente le deviazioni dalla casualità spaziale.

L'articolo è organizzato nel modo seguente. Il paragrafo 2 introduce brevemente la funzione K di Ripley, che rappresenta il nostro punto di partenza per lo sviluppo di misure di concentrazione spaziale più adatte per lo studio dei fenomeni di localizzazione industriale. Il paragrafo 3 presenta il meccanismo stocastico, basato sulla teoria dei processi di punto marcati, che consente di sviluppare un test per la presenza di concentrazione spaziale assoluta delle attività economiche. Il paragrafo 4 è dedicato all'applicazione del metodo a un caso di studio riguardante la distribuzione spaziale dell'industria manifatturiera ad *alta* e *medio-alta* intensità tecnologica nelle aree metropolitane di Milano e Torino. Il paragrafo 5, infine, conclude con alcune considerazioni sui possibili sviluppi futuri del metodo proposto.

2 Misurare la concentrazione senza tenere conto della dimensione delle imprese: la funzione K di Ripley

Non è un'esagerazione affermare che la funzione K di Ripley (Ripley, 1976 e 1977) è attualmente la misura basata sulle distanze più utilizzata per la sintesi della distribuzione spaziale degli eventi nell'ambito di dati micro-geografici. Tale misura si è dimostrata uno strumento molto utile per testare statisticamente la presenza di concentrazione spaziale significativa per *pattern* di punto in cui ciascun evento è trattato come un punto privo di dimensione. La funzione K è stata infatti diffusamente utilizzata in contesti applicativi diversi quali, per esempio, la geografia, l'ecologia, l'epidemiologia e, più recentemente, l'economia (si veda Arbia ed Espa, 1996; Marcon e Puech, 2003). Per un elenco esaustivo delle applicazioni della funzione K che si possono trovare in letteratura si rimanda a Illian *et al.* (2008).

In termini euristici, la funzione K può essere definita nel modo seguente:

$$K(d) = \lambda^{-1} E \{ \text{numero di punti localizzati entro una distanza } \leq d \text{ da un punto arbitrario} \} \quad (1)$$

dove $E\{\cdot\}$ indica l'operatore aspettativa e λ è il parametro rappresentativo dell'*intensità* spaziale, che corrisponde al numero medio di punti per area unitaria. Per definizione, quindi, $\lambda K(d)$ può essere interpretato come il numero atteso di punti aggiuntivi localizzati entro una distanza d da un punto arbitrario del processo generatore dei dati (Ripley, 1977). Nel contesto di un processo di punto omogeneo, la funzione K quantifica il livello di dipendenza spaziale tra i punti a ciascuna distanza d .

Lo sviluppo di test, basati sulla funzione K , per la presenza di concentrazione spaziale assoluta è facilitato dal fatto che per molti processi stocastici è possibile calcolare l'aspettativa del lato destro dell'Equazione (1), in modo tale che $K(d)$ possa essere scritta in forma chiusa (Dixon, 2002). Il processo di punto che genera distribuzioni spaziali di eventi completamente casuali (ossia i punti sono uniformemente e indipendentemente distribuiti nello spazio) è il cosiddetto *processo omogeneo di Poisson*. Si può dimostrare che, per un *pattern* di punto generato da un processo omogeneo di Poisson, $K(d)$ tende ad avere forma pari a πd^2 (si veda Diggle, 2003). Ovvero,

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

rappresenta adeguatamente l'ipotesi nulla di localizzazione casuale degli eventi. Le deviazioni significative da questo valore di riferimento rappresentano, al contrario, le ipotesi alternative di dipendenza spaziale. Nello specifico, per $K(d) > \pi d^2$ abbiamo dipendenza positiva e quindi *concentrazione* (circostanza in cui i punti tendono ad *attirare* la localizzazione di altri punti), per $K(d) < \pi d^2$ abbiamo invece dipendenza negativa e quindi *inibizione* (circostanza in cui i punti tendono a *respingere* la localizzazione di altri punti). Di conseguenza, allo scopo di verificare formalmente se i punti osservati tendono a concentrarsi nello spazio, possiamo verificare se, per qualche distanza d , il valore assunto da $K(d)$ è significativamente più elevato di πd^2 . I valori critici, necessari per la formazione della decisione di rifiuto dell'ipotesi nulla, possono essere calcolati attraverso procedure Monte Carlo di simulazione di processi omogenei di Poisson (si veda Besag e Diggle, 1977).

Un test statistico per la presenza di concentrazione assoluta basato sulla funzione K può essere adeguatamente utilizzato per identificare il grado di concentrazione delle attività economiche solamente nel caso in cui si possa considerare che le imprese abbiano la medesima dimensione. Di fatto, in un contesto in cui le attività economiche differiscono rispetto alla dimensione e, per esempio, si rileva la presenza di piccole, medie e grandi imprese, un *pattern* di punto potrebbe non essere più una rappresentazione appropriata della struttura localizzativa delle attività economiche e, di conseguenza, la funzione K potrebbe non essere più uno strumento di sintesi adeguato della distribuzione spaziale. Per esempio, la funzione K non

sarebbe in grado di identificare come *cluster* spaziale di imprese una situazione come quella rappresentata dal “Caso 2”, riportata nella Figura 1.

3 Misurare la concentrazione tenendo conto della dimensione delle imprese: la funzione K pesata

3.1 La funzione K pesata

La funzione K pesata, che qui indichiamo con $K_{mm}(d)$, è una misura proposta da Penttinen (2006) che consente di sintetizzare le caratteristiche di un *pattern* di punto marcato quantitativo (ossia un *pattern* di punto in cui a ciascun punto è associato un peso che rappresenta il valore di una variabile quantitativa). Tale funzione è stata originariamente proposta come una generalizzazione naturale della funzione K di Ripley. Nel presentarla può risultare conveniente partire dalla funzione K non pesata, riformulata nel modo seguente:

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

dove il termine d_{ij} indica la distanza Euclidea tra l’ i -esimo e il j -esimo punto arbitrario, n è il numero totale di punti e $I(d_{ij} \leq d)$ rappresenta la funzione indicatore che assume valore pari a 1 se $d_{ij} \leq d$ e valore pari a 0 altrimenti. In riferimento a questa notazione, la funzione K pesata ha una forma funzionale simile in cui, però, i pesi associati ai punti sono tenuti in considerazione:

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

Nell’Equazione (2) m_i e m_j sono i pesi associati, rispettivamente, all’ i -esimo e al j -esimo punto arbitrario e μ è la media dei pesi. Il termine $\lambda \mu^2 K_{mm}(d)$ può quindi essere interpretato come la media della somma dei prodotti del peso dell’ i -esimo punto arbitrario e dei pesi di tutti gli altri punti localizzati entro il cerchio di raggio d centrato in esso. Quindi la funzione K pesata misura, a ciascuna distanza d , la distribuzione congiunta cumulata dei pesi e dei punti. Per quanto concerne la stima, in Penttinen (2006) è dimostrato che uno stimatore approssimativamente corretto di $K_{mm}(d)$ è

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

dove $\hat{\lambda} = n/|A|$ è l'intensità spaziale stimata, $|A|$ è la superficie dell'area di studio e $\hat{\mu}$ è la media dei pesi osservati. Per la presenza degli effetti di bordo, che sono dovuti all'arbitrarietà nella definizione dei confini dell'area di studio, è necessario introdurre il fattore di aggiustamento w_{ij} al fine di evitare potenziali distorsioni nelle stime in prossimità dei confini dell'area di studio (per una trattazione esaustiva del problema si rimanda a Diggle, 2003). Nello specifico, il termine w_{ij} indica il reciproco della proporzione dell'area di un cerchio centrato sul punto i -esimo e passante per il punto j -esimo collocato all'interno dell'area di studio A (Boots e Getis, 1988).

Nell'ambito della localizzazione delle attività economiche, in cui i pesi sono i valori di una variabile quantitativa che rappresenta la dimensione delle imprese, la funzione K pesata potrebbe essere utilizzata per sviluppare un test per la presenza di concentrazione spaziale assoluta. È necessario, però, derivare il valore di riferimento assunto dalla funzione che possa considerarsi rappresentativo dell'ipotesi nulla di casualità spaziale. A tal fine, il prossimo paragrafo è dedicato alla costruzione di un modello stocastico per la generazione di *pattern* di punto marcati di imprese che possa rappresentare adeguatamente i fatti stilizzati di casualità spaziale e di concentrazione spaziale nella prospettiva indicata dal “*Caso 1*” (ossia un numero elevato di imprese piccole che si concentra nello spazio) e dal “*Caso 2*” (ossia poche imprese di grandi dimensioni che si concentrano nello spazio).

3.2 Un modello per l'ipotesi nulla di casualità spaziale

L'idea di base sui cui si fonda la costruzione del modello è che la concentrazione spaziale delle attività economiche (nella prospettiva indicata dal “*Caso 1*” e dal “*Caso 2*”) sia dovuta a una qualche forma di correlazione tra l'intensità spaziale dei punti e i pesi. L'esistenza di una tale correlazione implicherebbe, per esempio, che in regioni caratterizzate da un'elevata intensità dei punti i pesi tendano ad essere sistematicamente grandi se la correlazione è positiva o, al contrario, piccoli se la correlazione è negativa.

Il modello che presentiamo, che considera una tale struttura di correlazione, è una versione semplificata del modello introdotto in Arbia *et al.* (2010) che si basa sull'idea, già elaborata in Ho e Stoyan (2008), di un processo di Cox a intensità marcata, in cui l'intensità dei punti è regolata da un processo di Cox e i pesi sono realizzazioni di un processo con parametri condizionati dai valori dell'intensità spaziale dei punti.

3.2.1 Il processo di Cox log-normale dell'intensità spaziale dei punti

Assumiamo innanzitutto che l'intensità spaziale dei punti possa essere formalizzata come un processo di Cox log-normale (un tipo specifico di processo di Cox proposto da Møller *et al.*, 1998). In riferimento a questo modello, ciascun *pattern* di punto rappresenta una realizzazione parziale di un processo non omogeneo di Poisson caratterizzato da una funzione di intensità

spaziale $\lambda(x)$, dove x rappresenta le coordinate spaziali di un punto arbitrario (si veda Diggle, 2003). I valori di $\lambda(x)$ rappresentano, a loro volta, la realizzazione di un *random field* $\{\Lambda(x)\}$ tale che $\Lambda(x) = \exp\{S(x)\}$, in cui $\{S(x)\}$ è un *random field* normale con media μ_s , varianza σ_s^2 e funzione di correlazione $\rho_s(d)$. $\{\Lambda(x)\}$ è noto come processo di Cox log-normale.

Assumere che i punti siano generati da un processo di Cox log-normale è particolarmente utile in quanto rende possibile derivare le espressioni matematiche esplicite dell'intensità e della struttura della covarianza. In base alla funzione generatrice dei momenti della distribuzione log-normale, infatti, l'intensità λ di un processo di Cox log-normale $\{\Lambda(x)\}$ può essere esplicitata nel modo seguente:

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

Per quanto riguarda la struttura della covarianza, per ogni coppia di punti arbitrari (x e x'), $\Lambda(x)\Lambda(x') = \exp\{S(x) + S(x')\}$, in cui $S(x) + S(x')$ segue ancora una distribuzione normale con media $m = 2\mu_s$ e varianza $v = 2\sigma_s^2[1 + \rho_s(d)]$ (d è la distanza Euclidea tra x e x'). Di conseguenza, $E[\Lambda(x)\Lambda(x')] = \exp(m + v/2)$, e pertanto:

$$E[\Lambda(x)\Lambda(x')] = \lambda^2 \exp\{\sigma_s^2 \rho_s(d)\}.$$

3.2.2 Il processo generatore dei pesi

Nel definire il processo generatore dei pesi assumiamo che la realizzazione di un peso generico $m(x_n)$, associato al punto arbitrario x_n generato dal processo di Cox log-normale, sia condizionata dall'intensità del processo di Cox stesso. Formalmente, quindi, assumiamo che

$$m(x_n) = a\Lambda(x_n) + E(x_n) \quad (3)$$

in cui $\Lambda(x_n)$ è il valore dell'intensità spaziale in corrispondenza del punto x_n e il valore $E(x_n)$ è dovuto a un processo residuo tale che $E(x) = \exp\{R(x)\}$, dove $R(x)$ è un *random field* normale con media μ_R , varianza σ_R^2 e funzione di correlazione $\rho_R(d)$. Quindi il valore atteso del processo $E(x)$, che indichiamo con ε , è pari a $\varepsilon = E[\exp\{R(x)\}] = \exp\left\{\mu_R + \frac{1}{2}\sigma_R^2\right\}$.

La costante a , che appare nell'Equazione (3), è il parametro di interesse del modello. Questo parametro regola la correlazione tra il processo dell'intensità spaziale dei punti e il processo generatore dei pesi. In particolare, quando $a = 0$ i pesi sono indipendenti dall'intensità spaziale dei punti. Altrimenti, quando $a > 0$ il processo dei pesi genera pesi che tendono ad

essere più grandi (ossia imprese più grandi) in aree caratterizzate da un'elevata intensità spaziale dei punti. Al contrario, in quei casi in cui $a < 0$ i pesi tendono ad essere piccoli (e quindi le imprese di piccole dimensioni) in aree caratterizzate da un'elevata intensità spaziale dei punti. Il processo $E(x_n)$ sintetizza la presenza di quegli elementi non rilevabili che “disturbano” il fenomeno di correlazione regolato da a .

Assumere una distribuzione di probabilità log-normale rende il calcolo del valore atteso del processo generatore dei pesi matematicamente trattabile. Infatti, abbiamo che

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

Sarebbe naturale identificare il valore atteso del processo dei pesi nell'espressione $a\lambda + \varepsilon$. Tuttavia, in riferimento a quanto sostenuto da Ho e Stoyan (2008), il vero valore atteso corretto è $\mu = a\lambda \exp\{\sigma_s^2\} + \varepsilon$, che è maggiore di $a\lambda + \varepsilon$ quando $a > 0$, e minore quando $a < 0$. Per una spiegazione dettagliata riguardante l'introduzione di questo fattore di correzione si veda Ho e Stoyan (2008).

Nell'applicazione di questa struttura metodologica al problema dell'identificazione empirica della concentrazione spaziale delle attività economiche, i *pattern* di punto marcati generati quando $a = 0$ rappresentano l'ipotesi nulla di localizzazione casuale delle imprese. Allo stesso modo, $a < 0$ e $a > 0$ identificano le ipotesi alternative di concentrazione spaziale delle attività economiche secondo le prospettive indicate, rispettivamente, dal “Caso 1” e dal “Caso 2” illustrati nel Paragrafo 1.

3.2.3 L'ipotesi nulla di casualità spaziale

Per il modello stocastico generatore di *pattern* di punto marcati di imprese appena introdotto, è possibile scrivere in forma chiusa la forma teorica della funzione K pesata. La funzione K pesata teorica di questo modello, infatti, è (per $d > 0$) la seguente:

$$K_{mm}(d) = 2\pi \int_0^d u \frac{a^2 \lambda^2 \exp\{2\sigma_s^2 + 3\sigma_s^2 \rho_s(u)\} + 2a\lambda \exp\left\{\sigma_s^2 + \frac{3}{2}\sigma_s^2 \rho_s(u)\right\} \varepsilon + \varepsilon^2 \exp\{\sigma_R^2 \rho_R(u)\}}{[a\lambda \exp\{\sigma_s^2\} + \varepsilon]^2} du \quad (4)$$

La derivazione matematica dell'Equazione (4) è riportata in Appendice. L'Equazione (4) permette di sviluppare un test statistico per la presenza di concentrazione assoluta delle attività economiche basato sulla funzione K pesata, in cui l'ipotesi nulla di localizzazione casuale delle imprese è rappresentata dai valori assunti da $K_{mm}(d)$ quando $a = 0$. In particolare, quando $a = 0$ abbiamo che

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

L'Equazione (5) può pertanto rappresentare adeguatamente il *benchmark* appropriato per la verifica della presenza di concentrazione spaziale delle attività economiche.

3.2.4 La significatività statistica dell'ipotesi di casualità spaziale: un test Monte Carlo

Per identificare la presenza di concentrazione spaziale assoluta in un pattern osservato di localizzazione delle imprese, possiamo dunque valutare la significatività statistica delle deviazioni della funzione K pesata stimata, $\hat{K}_{mm}(d)$, dai valori della funzione teorica rappresentata dall'Equazione (5). La distribuzione di probabilità esatta o asintotica dello stimatore $\hat{K}_{mm}(d)$ non è nota ed è pertanto necessario ricorrere a tecniche di simulazione Monte Carlo. Per costruire le bande di confidenza per i valori di $K_{mm}(d)$ sotto l'ipotesi di casualità spaziale, simuliamo quindi un numero n di *pattern* di punto marcati dal modello descritto nell'Equazione (3) con $a = 0$. I *pattern* simulati devono avere un numero di punti condizionato al medesimo numero di punti del *pattern* osservato e i valori dei parametri del modello (μ_s , σ_s^2 , $\rho_s(d)$, μ_R , σ_R^2 e $\rho_R(d)$) devono essere adeguatamente stimati dai dati. Per ciascun pattern simulato, consistente con l'ipotesi nulla in quanto generato dal modello con $a = 0$, possiamo calcolare $\hat{K}_{mm}(d)$. In questo modo, siamo in grado di ottenere bande di confidenza approssimate a un livello del $n/(n+1) \times 100\%$ prendendo in considerazione, per ogni distanza d , i valori più elevati e meno elevati delle diverse funzioni $\hat{K}_{mm}(d)$ calcolate sulle n simulazioni sotto l'ipotesi nulla di casualità spaziale. Se la curva della funzione $\hat{K}_{mm}(d)$ del *pattern* osservato giace, per qualche distanza d , al di fuori delle bande di confidenza, vi è deviazione significativa dall'ipotesi nulla.

Per quanto concerne la stima dei parametri del modello sui dati osservati, i valori di μ_s , σ_s^2 e $\rho_s(d)$ del processo dell'intensità spaziale dei punti, $\Lambda(x) = \exp\{S(x)\}$, possono essere stimati correttamente ricorrendo al metodo del *Minimum Contrast* basato sulla funzione K (Diggle e Gratton, 1984; Møller e Waagepetersen, 2003). Tale metodo è una tecnica generale di stima che consente di determinare i parametri del modello generatore di punti che meglio si adatta a un *pattern* di punto osservato. La procedura di stima consiste nell'individuare i parametri del modello che realizzano lo scostamento minimo tra la funzione K teorica del modello e la funzione K osservata. Nello specifico, seguendo Waagepetersen (2007), i valori dei parametri di un processo di Cox log-normale $\{\Lambda(x)\}$ che garantiscono tale scostamento minimo sono quei valori che minimizzano la seguente funzione criterio:

$$D(\mu_s, \sigma_s^2, \rho_s(d)) = \int_0^{d_{max}} \left[\left\{ \hat{K}(u) \right\}^{1/4} - \left\{ K(\mu_s, \sigma_s^2, \rho_s(u); u) \right\}^{1/4} \right]^2 du$$

dove $K(\mu_s, \sigma_s^2, \rho_s(d); d) = 2\pi \int_0^d u \exp\{\sigma_s^2 \rho_s(u)\} du$ indica la funzione K teorica di un processo di Cox log-normale $\{\Lambda(x)\}$ e $\hat{K}(d)$ indica lo stimatore della funzione K calcolato sui dati. La funzione criterio $D(\mu_s, \sigma_s^2, \rho_s(d))$, così formalizzata, è l'integrale della differenza al quadrato tra le radici quarte delle due funzioni (Waagepetersen, 2007).

I valori dei parametri μ_R , σ_R^2 e $\rho_R(d)$ del *random field* normale $\{R(x)\}$ del processo residuo $E(x) = \exp\{R(x)\}$ possono essere stimati, condizionatamente alla scelta di una forma funzionale parametrica per la funzione di correlazione spaziale $\rho_R(d)$, con il metodo di stima della massima verosimiglianza. Per una trattazione esaustiva dei metodi di massima verosimiglianza applicati ai *random field* normali si rimanda a Diggle e Ribeiro (2007). In questo contesto, il *random field* $\{R(x)\}$ di interesse può essere adeguatamente stimato in riferimento al modello di regressione seguente:

$$\ln\{m(x_i)\} = R(x_i) + Z(x_i): i = 1, \dots, n$$

dove l'indice $i = 1, \dots, n$ si riferisce alle n localizzazioni osservate, $m(x_i)$ sono i pesi rilevati alle localizzazioni x_i e le $Z(x_i)$ sono variabili casuali $N(0, \tau^2)$ mutuamente indipendenti (Diggle e Ribeiro, 2007).

4 Un caso di studio: la localizzazione dell'industria manifatturiera ad *alta* e *medio-alta* intensità tecnologica nelle aree metropolitane di Milano e Torino

La parte empirica di questo articolo riguarda l'analisi della distribuzione spaziale delle imprese manifatturiere ad *alta* e *medio-alta* intensità tecnologica³, operanti tra il 1996 ed il 2004, localizzate, rispettivamente, nei comuni di Milano e Torino. L'insieme di dati è un sotto-insieme dell'archivio AIDA del Bureau Van Dijk's e fornisce la localizzazione geografica e la dimensione (in termini di numero di addetti) degli stabilimenti produttivi di 162 e 24 società di capitali operanti, rispettivamente, a Milano e a Torino. Tali imprese, prese qui in considerazione, sono caratterizzate dall'avere un'unica localizzazione produttiva. La Figura 2 mostra la distribuzione spaziale delle imprese nei comuni di Milano (Figura 2a) e Torino (Figura 2b) mediante *pattern* di punto marcati in cui la localizzazione di ciascuna impresa è identificata da un cerchio di raggio proporzionale al numero di addetti dell'impresa. Per entrambi i *pattern* di localizzazione delle imprese manifatturiere ad *alta* e *medio-alta* intensità tecnologica, è stata testata la presenza di fenomeni di concentrazione spaziale, controllando per la dimensione delle imprese, attraverso l'utilizzo della funzione K pesata descritta nel paragrafo precedente.

³ L'identificazione dei settori di appartenenza delle imprese è stata compiuta in base allo schema di classificazione OCSE/EUROSTAT, il quale considera "industrie manifatturiere ad alta o medio-alta tecnologia" le divisioni ATECO 24, 29, 30, 31, 32, 33, 34 e 35.

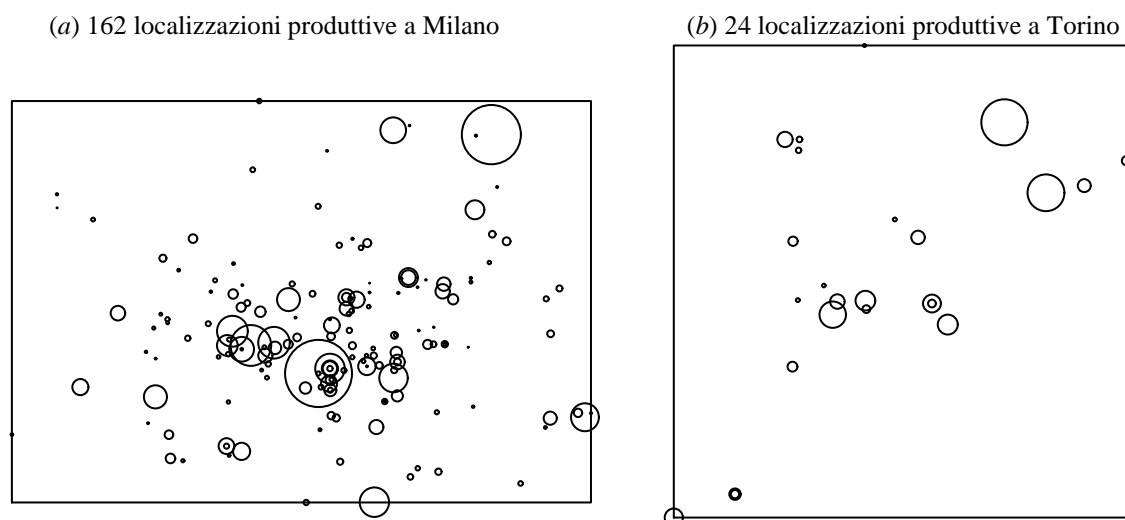


Figura 2 - Distribuzione spaziale delle imprese manifatturiere ad alta e medio-alta intensità tecnologica nei comuni di Milano e Torino nel periodo 1996-2004

Per valutare la significatività statistica delle deviazioni della funzione K pesata osservata dall'ipotesi di localizzazione casuale delle imprese, identificata dall'Equazione (5), abbiamo calcolato inviluppi di confidenza approssimati al 99.9% da 999 realizzazioni simulate del modello generatore di *pattern* di punto marcato. Nello specifico, ad ogni passo della procedura di simulazione, è stato generato un *pattern* di punto marcato dal modello dell'Equazione (3) con $a = 0$ e parametri stimati dai dati con i metodi, descritti nel paragrafo precedente, del *Minimum Contrast* e della massima verosimiglianza. Sul *pattern* così generato è stata calcolata la funzione K pesata, $\hat{K}_{mm}(d)$. Ripetendo questo passo 999 volte e prendendo, per ogni distanza d , il valore minimo e il valore massimo della sequenza di valori delle 999 $\hat{K}_{mm}(d)$, è stato possibile determinare le bande di confidenza per l'ipotesi di localizzazione casuale delle imprese e quindi di assenza di fenomeni di concentrazione spaziale significativi. I due grafici riportati in Figura 3 mostrano il comportamento della funzione $\hat{K}_{mm}(d)$ alle varie distanze d per le imprese manifatturiere ad *alta* e *medio-alta* intensità tecnologica localizzate, rispettivamente, a Milano e Torino. Questi grafici riportano anche le bande di confidenza per l'ipotesi nulla di assenza di concentrazione spaziale ad un livello di significatività pari a $\alpha = 0.001$.

I valori di $\hat{K}_{mm}(d)$ sono essenzialmente interpretabili in termini grafici. I valori d in corrispondenza di picchi delle funzioni $\hat{K}_{mm}(d)$ al di fuori degli inviluppi di confidenza identificano le distanze rispetto alle quali vi è concentrazione spaziale significativa. Osservando i grafici di Figura 3, si può quindi notare che le imprese manifatturiere ad *alta* e *medio-alta* intensità tecnologica localizzate a Milano tendono a concentrarsi nello spazio per distanze superiori a 0.5 chilometri. Il *pattern* di localizzazione di questo tipo di attività economiche a Torino, invece, non è caratterizzato da fenomeni di concentrazione spaziale significativi.

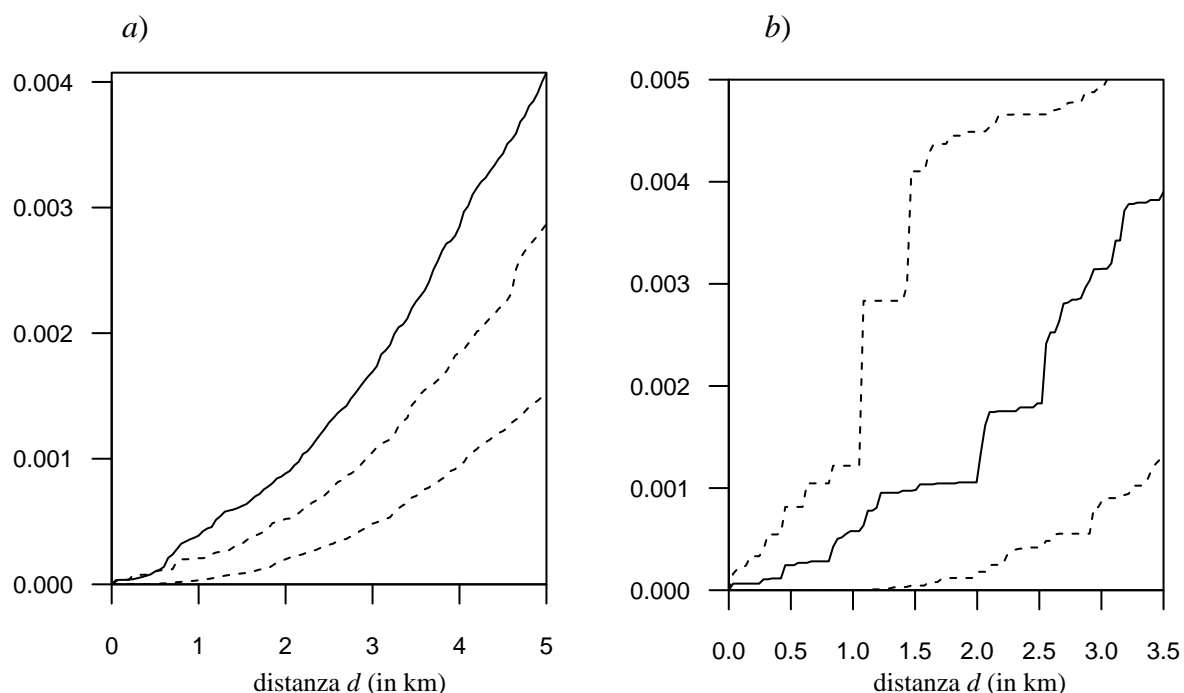


Figura 2 - Comportamento della funzione K pesata stimata (curva continua) e delle corrispondenti bande di confidenza al 99.9% (curve tratteggiate) per le imprese manifatturiere ad alta o medio-alta tecnologia localizzate a Milano (grafico a) e Torino (grafico b)

Nella letteratura di economia spaziale, geografia economica e scienze regionali si possono identificare almeno due fenomeni di localizzazione generali e contrapposti riguardanti l'industria manifatturiera ad utilizzo intensivo della tecnologia e delle innovazioni: un fenomeno porta alla concentrazione spaziale nelle grandi aree metropolitane (quali quelle di Milano e Torino); l'altro determina *pattern* di localizzazione in cui le interazioni spaziali tra gli agenti economici sono irrilevanti (Arbia *et al.*, 2010).

Per effetto della presenza di *spillover* di conoscenza e della circolazione di conoscenza tacita (Storper e Venables, 2003; Yeung *et al.* 2007) o per l'esistenza di *milieux* di innovazione (Camagni, 1991; Capello, 1999), gli agenti economici possono tendere a localizzarsi in prossimità di altre imprese allo scopo di sfruttare vantaggi di agglomerazione. Il *pattern* di localizzazione di Milano sembra essere consistente con questo fatto stilizzato.

In contrasto, con l'evoluzione delle tecnologie di comunicazione, lo spazio geografico potrebbe avere invece un ruolo limitato nella formazione delle scelte localizzative degli agenti economici (Sassen, 1994; Castells, 1996; Cairncross, 2001). La diminuita necessità dell'interazione fisica per la trasmissione di conoscenza e informazioni potrebbe ridurre i vantaggi localizzativi dovuti alla prossimità spaziale. Il *pattern* di localizzazione di Torino sembra essere consistente con questo secondo fatto stilizzato. Un'indagine più approfondita

sulle differenze territoriali tra Milano e Torino in termini di vantaggi agglomerativi richiederebbe l'analisi di altre variabili oltre alle sole informazioni riguardanti le localizzazioni produttive delle imprese. Rimandiamo pertanto questo obiettivo a lavori futuri.

5 Conclusioni

La concentrazione spaziale delle attività economiche, a causa delle implicazioni che può avere per la crescita economica locale e le disparità territoriali, è un fenomeno di grande rilievo per l'economia e le scienze regionali. In questo articolo è stato introdotto un approccio alla misurazione della concentrazione spaziale basato sullo spazio continuo e sull'utilizzo di dati micro-geografici e in grado di tenere in considerazione la dimensione delle imprese (misurata in termini di numero di addetti, fatturato, capitale, ecc.). In particolare, è stata sviluppata una misura basata su un'estensione della funzione K di Ripley che consente di testare statisticamente la presenza di concentrazione spaziale assoluta. Per derivare l'ipotesi nulla di localizzazione casuale delle imprese, è stato costruito un modello stocastico per la generazione di *pattern* di punto marcati di imprese rappresentativi dei diversi fatti stilizzati che potrebbero emergere nei casi empirici. Il metodo inferenziale introdotto consente solamente di valutare la significatività statistica delle deviazioni dall'ipotesi nulla di localizzazione casuale e non dà la possibilità di identificare le ipotesi alternative rilevanti, ossia non permette di individuare le caratteristiche del fenomeno di concentrazione spaziale rilevato. La concentrazione spaziale può essere dovuta, infatti, a un numero elevato di imprese di piccole dimensioni localizzate in un'area geografica ristretta o a poche imprese di grandi dimensioni localizzate vicine. Un'interessante sviluppo ulteriore del metodo proposto potrebbe andare in questa direzione.

6 Riconoscimenti

G. Espa gratefully acknowledges the partial financial contribution of PRIN # 2007JRTXFL (Analysis and modelling of efficiency, productivity and public policies at the micro-area level).

D. Giuliani ringrazia sentitamente il Prof. Marco Corsino per la disponibilità dei dati.

7 Bibliografia

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

Arbia G., Espa G., Giuliani D. (2010) *Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration*, Discussion Paper n.12/2010, Università degli Studi di Trento, Dipartimento di Economia.

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

Arbia G., Espa G., Giuliani D., Mazzitelli A. (2011) Clusters of firms in an inhomogeneous space: The high-tech industries in Milan, *Economic Modelling*, doi:10.1016/j.econmod.2011.01.012.

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

Cairncross F. (2001) *The death of distance 2.0; how the communications revolution will change our lives*, Harvard Business School Press, Cambridge.

Camagni R. (1991) *Local milieu, uncertainty and innovation networks: towards a new dynamics theory of economic space*, in: Camagni, R. (Eds.), *Innovation Networks: Spatial Perspective*, Belhaven Press, London.

Capello R. (1999) Spatial transfer of knowledge in high technology milieux: learning versus collective processes, *Regional Studies*, 33(4), 353–365.

Castells M. (1996) *The Rise of the Network Society*, Blackwell Publishing Ltd, Malden, MA.

Combes P.-P., Mayer T., Thisse J. F. (2008) *Economic geography: the integration of regions and nations*, Princeton University Press, Princeton.

Combes P.-P., Overman H. G. (2004) *The spatial distribution of economic activities in the European Union*, in: Henderson J. V., Thisse J.-F. (Eds.), *Handbook of Urban and Regional Economics*, Elsevier, North Holland, Amsterdam.

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

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

Diggle P.J., Gratton R.J. (1984) Monte Carlo methods of inference for implicit statistical models, *Journal of the Royal Statistical Society, series B*, 46, 193–212.

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.

Enright M. J. (2000) *The globalization of competition and the localization of competitive advantage: policies towards regional clustering*, in: Hood N., Young, S. (Eds.), *The globalization of multinational enterprise activity and economic development*, Macmillan, Glasgow.

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.

Marcon E., Puech F. (2010) Measures of the geographic concentration of industries: improving distance-based methods, *Journal of Economic Geography*, 10(5), 745–762.

Moller J., Waagepetersen R. (2003) *Statistical inference and simulation for spatial point processes*, Chapman and Hall/CRC, Boca Raton.

Mori T., Nishikimi K., Smith T. E. (2005) A divergence statistic for industrial localization, *Review of Economics and Statistics*, 87, 635–651.

Penttinen A. (2006) Statistics for Marked Point Patterns, *The Yearbook of the Finnish Statistical Society*, 70–91.

Porter M. E. (2000) Location, Competition, and Economic Development: Local Clusters in a Global Economy, *Economic Development Quarterly*, 14(1), 15–34.

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.

Sassen S. (1994) Cities in a World Economy, *Pine Forge Press*, London.

Storper M., Venables A.J. (2003) Buzz: face-to-face contact and the urban economy, *Journal of Economic Geography*, 4(4), 351–370.

Waagepetersen R. (2007) An estimating function approach to inference for inhomogeneous Neyman-Scott processes, *Biometrics*, 63, 252–258.

Yeung H.W., Coe N., Kelly, P. (2007) *Economic Geography. Introduction to Contemporary*, Blackwell Publishing.

Appendice: derivazione analitica della funzione K pesata teorica

La funzione K pesata teorica $K_{mm}(d)$ equivale all'integrale della *mark correlation function* $k_{mm}(d)$ (Illian *et al.*, 2008), ossia

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

La *mark correlation function* può essere definita come

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

dove $E_{or}[m(o)m(t)]$ indica la media condizionata sotto la condizione che vi siano punti nelle due localizzazioni arbitrarie separate da una distanza d , le quali sono considerate come origine o e destinazione t . $m(o)$ e $m(t)$ sono i pesi associati ai punti localizzati, rispettivamente, in o e t . Al denominatore, μ rappresenta la media dei pesi. Quindi $k_{mm}(d)$ può essere interpretato

come la media normalizzata del prodotto dei pesi di una coppia di punti separati da una distanza d .

Seguendo Ho e Stoyan (2008), il numeratore di $k_{mm}(d)$ soddisfa la condizione per cui

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

Se $\Lambda(x)$ è definito come nel sottoparagrafo 3.2.1 e $m(x)$ è dato dall'Equazione (3) allora

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

e

$$E[\Lambda(o)\Lambda(t)] = \lambda^2 \exp\{\sigma_s^2 \rho_s(d)\}$$

L'Equazione (8) può dunque essere riscritta nel modo seguente:

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

Di conseguenza, poiché $\mu = a\lambda \exp\{\sigma_s^2\} + \varepsilon$, la *mark correlation function* assume la seguente forma funzionale:

$$k_{mm}(d) = \frac{a^2 \lambda^2 \exp\{2\sigma_s^2 + 3\sigma_s^2 \rho_s(u)\} + 2a\lambda \exp\left\{\sigma_s^2 + \frac{3}{2}\sigma_s^2 \rho_s(u)\right\} \varepsilon + \varepsilon^2 \exp\{\sigma_R^2 \rho_R(u)\}}{[a\lambda \exp\{\sigma_s^2\} + \varepsilon]^2}, \quad d > 0. \quad (9)$$

Infine, sostituendo l'Equazione (9) nell'Equazione (6) otteniamo, per $d > 0$, la forma esplicita della funzione K pesata:

$$K_{mm}(d) = 2\pi \int_0^d u \frac{a^2 \lambda^2 \exp\{2\sigma_s^2 + 3\sigma_s^2 \rho_s(u)\} + 2a\lambda \exp\left\{\sigma_s^2 + \frac{3}{2}\sigma_s^2 \rho_s(u)\right\} \varepsilon + \varepsilon^2 \exp\{\sigma_R^2 \rho_R(u)\}}{[a\lambda \exp\{\sigma_s^2\} + \varepsilon]^2} du$$

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 recent approach to its measurement based on the point pattern analysis methodologies, which does not suffer from the problem of the arbitrariness of the regional boundaries, makes use of micro-geographic 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. The point pattern analysis methodologies based approach disregards the aspect of the different firm dimension and ignores the fact that a high degree of spatial concentration may result from both the case of many small firms clustering in definite portions of space and from only few large firms clustering together. The present paper aims at tackling this problem by adapting the K -function (Ripley, 1977) to account for the point dimension using the framework of marked point process theory (Penttinen, 2006). We present an empirical application of the proposed method to the spatial distribution of high-tech industries in Milan and Turin (Italy).