

## **USO COMBINATO DEGLI INDICI DI MORAN E DI THEIL: UN'APPLICAZIONE SULLA VULNERABILITA' SOCIALE E MATERIALE A LIVELLO TERRITORIALE<sup>1</sup>**

Simona Ballabio, Alberto Vitalini

### **1. Introduzione**

Molte discipline scientifiche studiano la distribuzione delle caratteristiche socio-economiche all'interno delle unità amministrative o di altre unità territoriali e le differenze tra di esse. Ma spesso si concentrano sulle differenze tra unità amministrative - solitamente macro-aree o regioni - oppure sulla distribuzione spaziale senza tener conto dei vari confini amministrativi.

La variabilità delle caratteristiche della popolazione e la loro distribuzione a livello geografico sono argomenti chiave per lo studio dei fenomeni socio-economici a diversi livelli territoriali. Sono molte le tecniche presenti nel panorama scientifico che indagano questo aspetto, tuttavia non esiste un quadro universale per considerare congiuntamente le suddivisioni amministrative e spaziali nell'analisi dei dati areali.

Il modo più semplice per misurare la variabilità geografica sono le misure statistiche standard di variabilità. Tuttavia, queste misure sono insensibili ai pattern geografici (Fotheringham *et al.*, 2000) e trascurano l'autocorrelazione spaziale portando agli stessi risultati sia quando le unità con alti valori di concentrazione sono adiacenti (alta autocorrelazione) sia quando le unità sono situate in parti anche molto lontane dell'area di studio (bassa autocorrelazione).

Questa situazione però può essere in parte superata utilizzando un approccio che analizza congiuntamente la variabilità e l'autocorrelazione spaziale globale (Netrdová e Nosek, 2017). Tale approccio permette di considerare gli effetti di prossimità nella misurazione della variabilità e di identificare se esiste un modello spaziale significativo. Gli indicatori dell'autocorrelazione spaziale globale misurano l'estensione del raggruppamento spaziale di valori "simili". Ci sono molti modi per misurare l'autocorrelazione spaziale globale a seconda della natura e delle proprietà dei dati (Anselin, 1988), tra questi è frequentemente utilizzato l'indice di Moran, che

---

<sup>1</sup> L'articolo è frutto del lavoro comune degli autori. Il paragrafo 2 è attribuito ad Alberto Vitalini, il paragrafo 3 è attribuito a Simona Ballabio.

si basa sulla covarianza e presenta molte analogie con il coefficiente di correlazione di Pearson.

In linea con quanto proposto da Netrdová e Nosek (2017) nel presente lavoro utilizziamo congiuntamente due tecniche che combinano tra loro sia l'approccio che tiene conto dei confini amministrativi sia l'approccio spaziale nell'analisi della variabilità. In particolare, per il primo approccio utilizziamo l'indice di Theil, (T di Theil), un indice di scomposizione della varianza che quantifica l'effetto dei confini amministrativi sul fenomeno indagato. Mentre per il secondo approccio utilizziamo l'indice di Moran (I di Moran), misura di autocorrelazione spaziale in grado di quantificare il livello di clustering spaziale nell'area di studio.

Il fenomeno indagato è la vulnerabilità sociale e materiale a livello familiare. In particolare, con l'uso combinato dei due approcci, l'obiettivo è quello di delineare dei pattern territoriali di vulnerabilità e di capire quanto questi pattern siano confinati nei limiti amministrativi provinciali all'interno delle singole regioni italiane. Il vantaggio dell'analisi è anche quello di considerare congiuntamente più confini amministrativi, quello provinciale e quello comunale. La comprensione della distribuzione della vulnerabilità a livello territoriale ha evidentemente valenza non solo in termini di mappatura ma anche in termini di implementazione di efficaci politiche locali.

## 2. Dati e metodi

La base dati utilizzata è di tipo censuario e si riferisce al Censimento della Popolazione e delle Abitazioni del 2011 condotto dall'Istituto Nazionale di Statistica Italiano (Istat). Nello specifico, i dati sono stati estratti da 8milaCensus (Istat, 2015), un sistema di diffusione dei dati censuari sintetizzati attraverso una selezione di 99 indicatori che consentono di delineare un profilo del territorio italiano a diversi livelli territoriali.

La variabile considerata per le analisi è l'indice di vulnerabilità sociale e materiale, testato e validato dall'Istat. Si tratta di un indice che cerca di rilevare l'esposizione di alcune fasce di popolazione a situazioni di rischio, di incertezza della propria situazione sociale ed economica che però non necessariamente si traducono in effettive situazioni di disagio materiale e sociale. L'indice è costruito attraverso la sintesi di sette indicatori<sup>2</sup> e tiene conto della multidimensionalità del fenomeno

---

<sup>2</sup> Gli indicatori considerati sono: 1. incidenza percentuale della popolazione di 25-64 anni analfabeta e alfabeto senza titolo di studio; 2. incidenza percentuale delle famiglie con potenziale disagio economico, ossia la quota di famiglie giovani o adulte con figli nei quali nessuno è occupato o pensionato; 3. incidenza percentuale delle famiglie con potenziale disagio assistenziale, ossia la quota di famiglie composte solo da anziani ultra 65-enni con almeno un componente ultraottantenne; 4.

indagato cercando di sintetizzare in un unico valore sia la dimensione materiale sia quella sociale della vulnerabilità,

La metodologia utilizzata per la costruzione dell'indice è basata sull'ipotesi di non "sostituibilità" delle diverse componenti (Mazziotta e Pareto, 2014 e 2015) e consente di produrre un indice sintetico non compensativo confrontabile in termini "assoluti" (Adjusted Mazziotta-Pareto Index – AMPI+/-), in grado di facilitare l'individuazione di potenziali aree di criticità attraverso la comparazione tanto spaziale quanto temporale.

Le tecniche di misurazione dell'autocorrelazione spaziale (approccio spaziale) e della variabilità tra unità amministrative (approccio che tiene conto dei confini amministrativi provinciali) utilizzate sono, nel primo caso, l'indice di Moran (I) e, nel secondo caso, l'indice di Theil (T) e la sua scomposizione. Le formule per l'I di Moran (Equazione 1) e l'indice di Theil T (Equazione 2) possono essere scritte come:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

dove: n = numero di comuni, i e j = indici per due diversi comuni, y = media della variabile in analisi,  $w_{ij}$  = valore del peso spaziale tra due comuni i, j con  $w_{ii} = 0$  (Anselin, 1995);

$$T = \left( \sum_{j=1}^k \frac{n_j y_j}{n y} \ln \frac{y_j}{y} \right) + \left( \sum_{j=1}^k \frac{n_j y_j}{n y} \sum_{i=1}^{n_j} \frac{y_{ij}}{y_j} \ln \frac{y_{ij}}{y_j} \right) = T_B + T_W \quad (2)$$

dove: n = numero di comuni, i = indice per i singoli comuni, j = indice per le province, k = numero di province, y = media della variabile in analisi, per T (indice di Theil complessivo),  $T_B$  = componente interprovinciale dell'indice Theil,  $T_W$  = componente intra-provinciale dell'indice Theil (Elbers *et al.*, 2008).

Mentre la  $T_B$  è una misura della variabilità provinciale, la quota della variabilità provinciale rispetto alla variabilità complessiva ( $T_B/T$ ) misura la variabilità provinciale relativa. Tutti i calcoli riguardanti l'indice di Theil e la sua scomposizione sono stati eseguiti in Stata e tutti i calcoli riguardanti l'I di Moran sono stati eseguiti in GeoDa 1.4.0 (Anselin, 2003; Anselin *et al.*, 2004). Come mostra

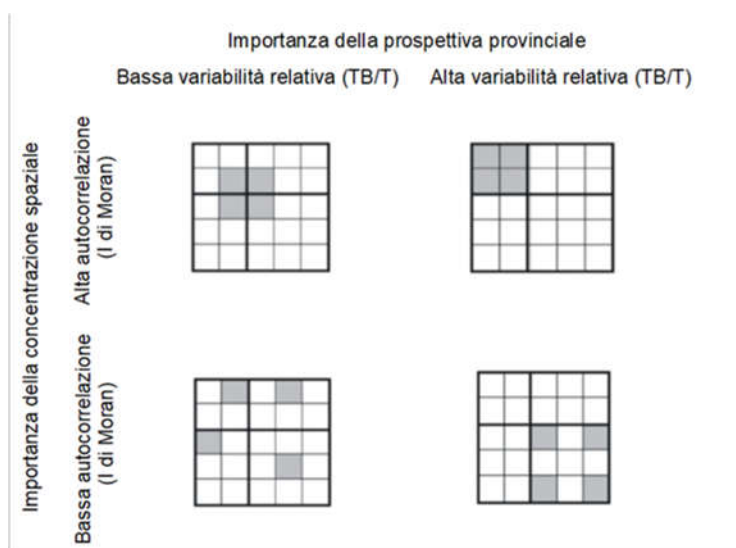
---

incidenza percentuale della popolazione in affollamento grave, ossia il rapporto percentuale tra la popolazione residente in abitazioni con superficie inferiore a 40 mq e più di 4 occupanti o in 40-59 mq e più di 5 occupanti o in 60-79 mq e più di 6 occupanti, e il totale della popolazione residente in abitazioni occupate; 5. incidenza percentuale delle famiglie con 6 e più componenti; 6. incidenza percentuale delle famiglie monogenitoriali giovani (età del genitore inferiore ai 35 anni) o adulte (età del genitore compresa fra 35 e 64 anni) sul totale delle famiglie; 7. incidenza percentuale di giovani di 15-29 anni fuori dal mercato del lavoro e dalla formazione scolastica.

la formula (Equazione 1), per misurare l'autocorrelazione spaziale è necessaria una matrice di peso spaziale ( $w_{ij}$ ) che renda operativa la posizione e la vicinanza delle unità geografiche. In questo lavoro coincide con una matrice di adiacenza, formata da tutte le possibili combinazioni di coppie di comuni, che assume valore "1" (se i due comuni condividono un tratto di confine) o valore "0" (se due comuni non condividono alcun tratto di confine). Per convenzione, ciascun comune non è contiguo a se stesso, pertanto la diagonale principale della matrice è formata da tutti valori pari a zero.

A livello teorico, combinando tra loro le due misure dicotomizzate, I di Moran e misura di variabilità relativa (TB/T) si ricava una tipologia in cui si delineano quattro tipi di risultati (Figura 1).

**Figura 1** – Tipologia dei dati areali basata sulle prospettive spaziali e provinciali.



Note: Ognuno dei quattro quadrati corrisponde ad un tipo di regione. All'interno di ogni tipo di regione i venticinque quadratini piccoli corrispondono ai comuni mentre i quattro quadrati più grandi alle province.

Fonte: rielaborazione personale diagramma da Netrdová e Nosek (2017)

Se sia la variabilità provinciale relativa sia l'autocorrelazione spaziale sono elevate (riquadro in alto a destra), il fenomeno in analisi può essere considerato dipendente dallo spazio e delimitato nelle province. Il fenomeno è concentrato e lo è all'interno dei confini amministrativi predefiniti, le province.

Se il fenomeno si concentra in comuni spazialmente contigui che attraversano i confini provinciali (riquadro in alto a sinistra) la dipendenza spaziale risulta scarsamente correlata con i confini provinciali.

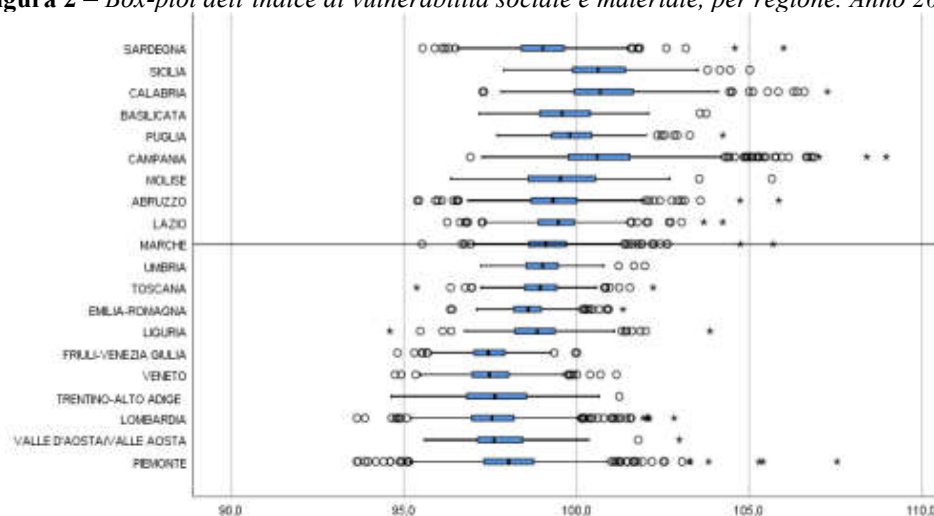
Se il fenomeno non forma concentrazioni spaziali né si concentra nelle province, è indipendente sia dal punto di vista spaziale che provinciale (riquadro in basso a sinistra).

È infine teoricamente possibile che non si osservi alcuna concentrazione spaziale, sebbene il fenomeno si concentri a livello provinciale (riquadro in basso a destra). In quest'ultimo caso il fenomeno sarebbe indipendente dal punto di vista spaziale ma delimitato nelle province.

### 3. Risultati e discussione

I rischi di vulnerabilità sociale e materiale, intesi in termini di incertezza della situazione sociale ed economica familiare che però non necessariamente si traduce in effettive situazioni di disagio, non si distribuiscono in modo omogeneo sul territorio italiano.

**Figura 2** – Box-plot dell'indice di vulnerabilità sociale e materiale, per regione. Anno 2011



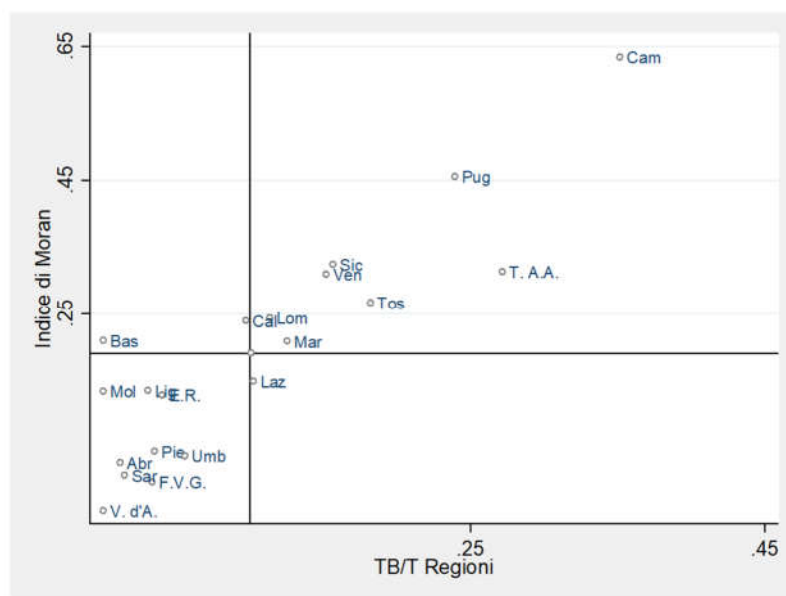
Fonte: nostre elaborazioni su dati 8milacensus.

Come mostra la Figura 2, le condizioni di maggior rischio di vulnerabilità sociale e materiale riguardano le regioni del Sud Italia, in particolare Campania, Calabria e Sicilia. In questi casi, oltre ad un valore mediano di incidenza del fenomeno più

elevato rispetto alle altre regioni, si osserva anche una maggiore variabilità, quindi una maggiore dispersione dei valori osservati per i singoli comuni. Di converso, confermando anche in questo caso il noto divario Nord-Sud, i minori rischi di vulnerabilità si rilevano nelle regioni settentrionali, in particolare in Friuli-Venezia Giulia, Veneto, Lombardia, Valle d'Aosta, Trentino-Alto Adige e Piemonte. Per il Trentino Alto-Adige si osserva però una variabilità del fenomeno, entro i confini regionali, più marcata rispetto alle altre regioni del Nord.

Per analizzare la diffusione dei fenomeni sia sociali sia economici, compresa la vulnerabilità, l'attenzione è spesso posta sulle differenze tra macro-aree o tra regioni. Tuttavia, è anche interessante capire cosa accade all'interno delle unità territoriali regionali. In particolare, come visto nel paragrafo precedente, l'utilizzo congiunto dell'I di Moran e della misura di variabilità relativa (TB/T) consente, per ciascuna regione italiana, di sintetizzare la distribuzione spaziale della vulnerabilità, rilevata a livello comunale, evidenziando sia il peso dei pattern territoriali sia il loro confinamento entro i limiti amministrativi provinciali.

**Figura 3** – Diagramma a dispersione dell'I di Moran e dell'indice TB/T (T di Theil) per regione. Anno 2011



Fonte: nostre elaborazioni su dati 8milacensus.

La figura 3 mostra i risultati. Sull'asse verticale si trova l'I di Moran (I), mentre sull'asse orizzontale è presente la misura di variabilità relativa (TB/T) ricavata

dall'indice di Theil (T). Come abbiamo visto, la prima fornisce un'indicazione sintetica dell'autocorrelazione spaziale – quindi dell'esistenza di pattern geografici – entro le singole regioni, mentre la seconda misura fornisce un'indicazione sintetica della variabilità provinciale o, detto in altro modo, di quanta parte della variabilità regionale è “spiegata” dalle differenze tra le province.

Dal grafico emerge chiaramente una forte relazione positiva tra le due misure. In particolare, richiamando lo schema teorico presentato nel secondo paragrafo e utilizzando i valori medi a livello nazionale delle due misure per la ripartizione nei quattro quadranti del grafico (linea orizzontale per I e linea verticale per TB/T), si delineano due gruppi di regioni. Nel quadrante in basso a sinistra le regioni in cui il fenomeno della vulnerabilità non forma concentrazioni spaziali né si concentra nelle province, pertanto sembra essere indipendente sia dal punto di vista spaziale sia dal punto di vista provinciale. Nel quadrante in alto a destra invece troviamo le regioni in cui il fenomeno è concentrato spazialmente e lo è all'interno dei confini amministrativi provinciali. Le tre regioni con i valori più bassi di entrambe le misure, tralasciando la Valle d'Aosta che ha un'unica provincia, sono Abruzzo, Sardegna e Friuli-Venezia Giulia. Di converso, le tre regioni con i valori più alti sono Campania, Puglia e Trentino-Alto Adige.

Per una corretta interpretazione e comprensione dei risultati empirici, è fondamentale considerare anche il livello comunale. Sebbene entrambi i metodi considerati (spaziale e provinciale) forniscano importanti informazioni sulla variabilità geografica della vulnerabilità rimangono ancora aspetti da approfondire, ad esempio: come si configurano i cluster spaziali?

Per rispondere a questa domanda ricorriamo alle *Cluster Maps* LISA soffermandoci unicamente sulle tre regioni con i valori più alti delle due misure considerate. Le *Cluster Maps* LISA sono cartografie tematiche che mostrano solo i comuni con valori statisticamente significativi della misura LISA (Anselin, 1995), cioè di una misura di autocorrelazione spaziale locale che rileva la somiglianza tra il valore di una variabile misurata in un'unità di analisi areale (p.es. comune) e i valori della stessa variabile nelle unità vicine. Nel caso specifico, il valore di LISA, calcolato per ogni singolo comune, evidenzia se il comune presenta un valore dell'indice di vulnerabilità tendenzialmente simile o differente rispetto a quello dei comuni vicini, cioè quelli che condividono con il comune un tratto di confine. Valori negativi di LISA indicano diversità: il comune presenta un indice di vulnerabilità differente dai comuni vicini. Valori positivi indicano similarità: il comune presenta un indice di vulnerabilità simile a quello dei comuni vicini.

Nelle mappe sotto riportate (Figure 4, 5 e 6) abbiamo evidenziato i comuni con valori significativamente simili della misura LISA. In particolare, nelle cartografie successive, i comuni con valori LISA significativamente simili ai vicini e valori elevati dell'indice di vulnerabilità sono evidenziati in grigio, quelli con valori LISA

significativamente simili ai vicini e valori bassi dell'indice di vulnerabilità sono evidenziati in nero.

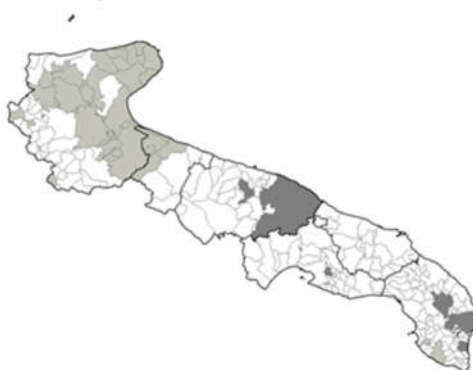
Nella regione Campania i cluster si concentrano nelle province di Caserta e di Napoli e nelle province di Benevento e Avellino. Nel primo caso si tratta di gruppi di comuni con incidenza elevata di vulnerabilità, mentre nel secondo con incidenza bassa. In Puglia i cluster di alta vulnerabilità si concentrano soprattutto nella provincia di Foggia, mentre quelli di bassa vulnerabilità nella zona sud delle province di Bari e di Lecce.

**Figura 4** – LISA Cluster Map, per comune. Campania. Anno 2011



Fonte: nostre elaborazioni su dati Smilacensus.

**Figura 5** – LISA Cluster Map, per comune. Puglia. Anno 2011

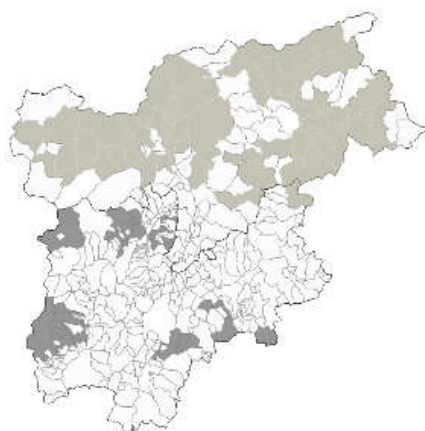


Fonte: nostre elaborazioni su dati Smilacensus.



Il Trentino-Alto Adige evidenzia forti differenze tra le uniche due province che compongono la regione. La provincia di Trento presenta molteplici cluster composti da un numero limitato di comuni caratterizzati da bassa vulnerabilità sociale e materiale. Mentre la provincia di Bolzano presenta estesi cluster caratterizzati da alta vulnerabilità.

**Figura 6** – LISA Cluster Map, per comune. Trentino-Alto Adige. Anno 2011



*Fonte: nostre elaborazioni su dati Smilacensus.*

#### 4. Conclusione

In questo lavoro si è trattato il tema della vulnerabilità sociale dei comuni italiani e si è cercato di mostrare la necessità, per una sua lettura più completa in termini di distribuzione territoriale, di un superamento delle tradizionali analisi che considerano uno alla volta diversi livelli di aggregazione territoriale: regionale, provinciale e comunale, riconducendole ad un approccio globale basato su due fasi.

Nella prima fase si è descritto come i pattern territoriali di vulnerabilità siano confinati nei limiti amministrativi provinciali all'interno delle singole regioni italiane. Sono stati utilizzati congiuntamente, da un lato, l'indice di Theil e la sua scomposizione in variabilità spiegata dalla provincia per quantificare l'effetto dei confini amministrativi provinciali sul fenomeno indagato e, dall'altro, l'indice di Moran per quantificare il livello di autocorrelazione del fenomeno indagato nella regione.

Questo tipo di analisi ha permesso di individuare alcune regioni, ad esempio Campania e Puglia, in cui il fenomeno è concentrato, sia in termini positivi (basso

valore indice di vulnerabilità) sia negativi (alto valore indice di vulnerabilità) solo all'interno di alcune province della regione.

Nella seconda fase, si è focalizzata l'attenzione sul livello comunale ed utilizzando i LISA si è approfondita la configurazione dei cluster spaziali nelle regioni individuate nella fase precedente, identificando cluster ben differenziati di comuni contigui con valori simili dell'indice di vulnerabilità.

La sequenza di tecniche proposte, oltre ad avere il vantaggio di studiare la distribuzione spaziale della vulnerabilità sociale e materiale considerando congiuntamente più confini amministrativi, ha anche una valenza in termini di implementazione di efficaci politiche locali, soprattutto se integrata con ulteriori analisi che pongano in relazione i risultati esposti in questo lavoro con le caratteristiche socio-demografiche del territorio.

Sarebbe, infine, utile un'analisi longitudinale per capire come il fenomeno evolve nel tempo, essendoci limitati a mostrare le potenzialità del nostro approccio solo per l'anno 2011.

### Riferimenti bibliografici

ANSELIN L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

ANSELIN L. 1995. Local Indicators of Spatial Association – LISA, *Geographical Analysis*, Vol. 27, No. 2, pp. 93–115.

ANSELIN L. 2003. *An introduction to spatial autocorrelation analysis with Geoda*. Urbana-Champaign, USA: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, [https://personal.utdallas.edu/~briggs/poec6382/geoda\\_spauto.pdf](https://personal.utdallas.edu/~briggs/poec6382/geoda_spauto.pdf), accesso 16/07/2020.

ANSELIN L., SYABRI I., KHO Y. 2004. *GeoDa: An introduction to spatial data analysis*. Urbana-Champaign, USA: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, <http://www.csiss.org/events/workshops/geodaGA.pdf>, accesso 16/07/2020.

ELBERS C., LAJOUW P. F., MISTIAEN J. A., ÖZLER, B. 2008. Reinterpreting between-group inequality, *The Journal of Economic Inequality*, Vol. 6, No. 3, pp. 231–245.

FOTHERINGHAM A.S., BRUNSDON C., CHARLTON, M. 2000. *Quantitative geography – Perspectives on spatial data analysis*. London: SAGE.

ISTAT. 2015. *Indice di vulnerabilità sociale e materiale*, Roma: Istituto nazionale di statistica, <http://ottomilacensus.istat.it>, accesso 16/07/2020.

NETRDOVÁ P., NOSEK V. 2017. Exploring the Variability and Geographical Patterns of Population Characteristics: Regional and Spatial Perspectives, *Moravian*

*Geographical Reports*, Vol. 25, No. 2, pp. 85-94.

MAZZIOTTA M., PARETO A. 2014. A Composite Index for measuring Italian regions' development over time, *Rivista italiana di economia e demografia*, Vol. LXVIII, No. 3/4.

MAZZIOTTA M., PARETO A. 2015. On a Generalized Non-compensatory composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, pp. 983-1003.

## SUMMARY

### **Combined use of Moran and Theil indices: an application on social and material vulnerability at territorial level**

The variability of the characteristics of the population and families at different territorial levels is an important key to interpreting social phenomena. This work deals with the social vulnerability of Italian municipalities.

In the analyses we use the empirical indicators of vulnerability based on the 2011 census of the Italian population.

This work goes beyond the traditional analyses that consider, one at a time, different levels of territorial aggregation: regional, provincial and municipal, leading them back to a global approach based on two phases.

In the first phase we described how the patterns of vulnerability are confined to the provinces of the individual Italian regions, using jointly: the Theil index and its breakdown into variability explained by the province to quantify the effect of the provincial administrative boundaries on the investigated phenomenon and the Moran index to quantify the level of self-correlation of the investigated phenomenon in the region. This type of analysis made it possible to identify some regions, for example Campania and Puglia, where the phenomenon is concentrated only within some provinces of the region.

In the second phase, attention was focused on the municipal level and using LISAs, we improved the understanding of spatial clusters of municipalities in the regions identified in the previous phase