

\*Isidora Barbaccia  
 \* Erika Ghirardo  
 \*\*Maurizio Festa

## Quotazioni immobiliari: effetti di vicinanza in mercati poco dinamici

**Parole chiave:** sottomercati residenziali, quotazioni immobiliari, *cluster analysis*, autocorrelazione spaziale.

**Abstract** L'identificazione dei sottomercati residenziali risulta di notevole interesse per la corretta programmazione di interventi nelle politiche abitative. Inoltre, negli ultimi anni si è assistito ad una crescente richiesta di informazioni statistiche a livello territoriale sempre più dettagliata che rende necessarie partizioni del territorio sempre più specifiche.

In questo studio si propone un'analisi per giungere ad un'aggregazione di comuni caratterizzati da omogeneità per livelli e andamento delle quotazioni delle abitazioni.

L'analisi empirica viene condotta in una prima fase impiegando strumenti di statistica descrittiva, indici econometrici spaziali, globali e locali, e tecniche regressive per costruire un *set* di indicatori utili. In una seconda fase è applicato un metodo di *clustering* per giungere al raggruppamento finale.

### INTRODUZIONE

Il mercato immobiliare risulta caratterizzato da un elevato grado di complessità che rende i prezzi delle abitazioni molto diversi sia nel tempo sia nello spazio. Il mercato non può quindi essere considerato come un sistema unico ma va necessariamente visto come un insieme di micro mercati tra loro interconnessi.

In letteratura si ravvisa un generale accordo sull'esistenza nel mercato residenziale di sottomercati (Goodman e Thibodeau, 2003 e McCluskey *et al.*, 2002) e la loro identificazione risulta di notevole importanza in molti ambiti.

La letteratura individua principalmente due aree in cui la conoscenza dei sottomercati può apportare un contributo significativo:

- il campo delle politiche residenziali, per comprendere l'effetto localizzazione di una zona rispetto ad un'altra (Pryce e Evans, 2007) o ancora per fare in modo che i piani di intervento tengano debitamente in conto l'andamento del mercato (Bates, 2006);
- la sfera della tassazione della proprietà dove il processo di stima del valore della proprietà richiede l'identificazione di sottomercati (Eckert, 1990).

Nonostante il pieno riconoscimento dell'esistenza e dell'utilità, non esiste una definizione unica e coerente di sottomercato (Watkins, 2001). In ogni caso, l'origine degli studi sui sottomercati va ricondotto al lavoro pionieristico di Grigsby (1963)<sup>1</sup>. In particolare, Grigsby propone il concetto teorico

\* Funzionario Ufficio Statistiche e studi mercato immobiliare - Agenzia del Territorio (isidora.barbaccia@agenziaterritorio.it - erika.ghirardo@agenziaterritorio.it)

\*\* Responsabile Ufficio Statistiche e studi mercato immobiliare - Agenzia del Territorio (maurizio.festa@agenziaterritorio.it)

<sup>1</sup> Come evidenziato nei lavori di Galster (1996) e Megbolugbe *et al.* (1996).

di sostituibilità<sup>2</sup> come base per la definizione di sottomercato. La sostituibilità è rappresentata dalla volontà del mercato di sostituire un'abitazione, con il complesso dei suoi attributi, con un'altra al fine di ottimizzare esigenze abitative, fissando un vincolo determinato di prezzo. Laddove questa capacità di sostituire diventa così debole da non produrre alcun effetto riconoscibile sul prezzo pagato, allora le abitazioni possono essere considerate appartenenti a diversi sottomercati. Quindi, un'area del mercato immobiliare all'interno della quale tutte le unità abitative sono collegate insieme in una catena di sostituzione, nel senso che ogni unità abitativa ubicata in quest'area può essere considerata sostituita di ogni altra unità, costituisce un sottomercato.

Un aspetto critico è se il sottomercato vada definito a livello spaziale o meno, ovvero verso quali fattori va applicato il concetto di sostituibilità. Secondo Watkins (2001) i sottomercati residenziali sono determinati sia da fattori spaziali sia da fattori strutturali.

Dal punto di vista empirico tra i modelli utilizzati per identificare i sottomercati il più diffuso è indubbiamente il modello edonico (Allen *et al.*, 1996; Goodman, 1981) mentre altri studi hanno utilizzato tecniche di statistica multivariata. L'analisi fattoriale è stata utilizzata nel lavoro di Dole-Johnson (1982) e ancora la *cluster analysis* è impiegata ad esempio nei lavori di Abraham *et al.* (1994) e di Hoesli *et al.* (1997). Negli ultimi anni si è fatto ampiamente uso delle tecniche di statistica spaziale come nel lavoro di Basu e Thibodeau (1998) o in quello di Pace e Gilley (1997).

Partendo dalla considerazione già citata di Watkins (2001), ma emersa già in precedenza (Adair *et al.*, 1996) sull'importanza delle caratteristiche sia spaziali sia strutturali, il contributo che si propone sulle analisi territoriali del mercato immobiliare consiste nell'individuare gruppi di mercato omogenei analizzando variabili spaziali, temporali e variabili legate alla struttura del mercato stesso attraverso l'utilizzo di tecniche di *clustering*.

Il lavoro è strutturato come segue. Dapprima si specificano gli obiettivi dell'analisi, poi si illustrano le categorie di fattori ritenuti determinanti. Nel paragrafo seguente si esaminano le fasi che hanno condotto alla creazione del *set* di variabili e si descrive il processo di *clustering*, infine si discutono le conclusioni e gli eventuali sviluppi futuri.

## OBIETTIVO DELL'ANALISI

L'identificazione dei sottomercati residenziali, come visto nell'introduzione, risulta di notevole interesse sotto diversi punti di vista. D'altro lato, negli ultimi anni è notevolmente maturata, la consapevolezza dell'importanza che il mercato immobiliare riveste all'interno dell'economia in generale, con la conseguenza di una crescente richiesta di informazioni statistiche a livello territoriale sempre più dettagliato che rende necessarie partizioni del territorio sempre più specifiche.

Dal 2001 l'Osservatorio del Mercato Immobiliare (OMI) dell'Agenzia del Territorio cura la rilevazione, l'elaborazione e la pubblicazione dei valori immobiliari per tutto il territorio nazionale. In circa 1.300 comuni, che rappresentano più del 70% delle compravendite del mercato immobiliare residenziale, la rilevazione avviene in modo diretto attraverso l'acquisizione di un congruo numero di informazioni sulle singole abitazioni attraverso la compilazione di schede standardizzate uniche per tutto il territorio nazionale.

Nei restanti comuni, soprattutto a causa di una scarsa dinamica di mercato, è effettuata una rilevazione di tipo indiretto attraverso metodi di comparazione e sulla base dell'*expertise* degli uffici che operano in campo tecnico estimale.

Al fine di migliorare la qualità dell'inferenza, si ritiene necessario creare delle aggregazioni di questi comuni scarsamente dinamici, in modo tale da ottenere gruppi il più possibile omogenei per livelli e andamenti delle quotazioni. Si tratta in altri termini di cercare conferme empiriche all'ipotesi di esi-

<sup>2</sup> Va precisato che il concetto di sostituibilità è stato derivato da Grigsby (1963) da uno dei primissimi lavori sui sottomercati riconducibile a Rapkin (1953).

stenza di sottomercati immobiliari locali che territorialmente coinvolgono più comuni con l'obiettivo di rendere più robuste le procedure di elaborazione delle quotazioni.

Per quanto detto nell'introduzione alla luce della letteratura esaminata, l'identificazione dei sottomercati dovrebbe tralasciare qualsiasi riferimento ai confini amministrativi fondando lo studio sulla base delle informazioni sulle singole abitazioni. Tuttavia, per il modo in cui è strutturata l'Agenzia del Territorio che è presente in ciascuna provincia con un Ufficio e per l'obiettivo dell'analisi di aggregare più realtà territoriali identificate nei comuni, nell'elaborato non si potrà prescindere dai confini amministrativi dei comuni e della provincia.

Viene quindi presentato un modo di giungere ad una definizione di sottomercati attraverso una riflessione su una serie di fattori. Il metodo non ha l'ambizione di avere validità generale quanto piuttosto, si propone di presentare uno dei possibili percorsi di identificazione di aree omogenee di quotazioni, partendo da un ragionamento su fattori generali. Lo scopo è quindi quello raggiungere un obiettivo specifico che risponde tuttavia ad esigenze generali di disponibilità di informazioni di dettaglio.

In particolare, in questo articolo si propone uno studio sperimentale su un territorio di media estensione che presenta una complessità contenuta per differenze territoriali. Si è scelto di prendere in esame i 47 comuni che costituiscono la provincia di Modena<sup>3</sup>. Sebbene, come evidenziato, la realtà immobiliare in questa provincia non risulti particolarmente disomogenea, sussistono comunque particolari caratteristiche che non possono essere trascurate quando si esamina un territorio ai fini di una sua partizione in aree omogenee. La provincia di Modena, presenta infatti, comuni pianeggianti, pedemontani e montani nonché comuni attraversati da due affluenti del Po, il Secchia e il Panaro. Non trascurabile è la sua rilevanza industriale con la presenza di particolari distretti. Infatti, qui hanno sede importanti industrie alimentari (grandi salumifici, centri di produzione del parmigiano e della lavorazione del maiale), metalmeccaniche (è considerata il centro più importante per l'automobilismo sportivo), delle ceramiche, del tessile e del biomedicale. E' fuori dubbio che queste caratteristiche influenzano la localizzazione delle residenze e quindi formazione dei prezzi delle abitazioni.

E' necessario, pertanto, costruire un sistema di variabili in grado di cogliere in livelli e in variazioni la diversità di apprezzamenti delle residenze che si possono osservare tra comuni. Si vuole in altri termini cercare di capire, da un lato, quali variabili discriminano l'evoluzione dei prezzi e, dall'altro, utilizzare queste variabili per creare raggruppamenti di comuni quanto più omogenei rispetto alle variabili scelte. A tal fine saranno utilizzate tecniche di *clustering*.

<sup>3</sup> La scelta di analizzare la provincia di Modena è, ovviamente, precedente al verificarsi dei tragici eventi sismici del maggio 2012 che hanno colpito diversi comuni di questa provincia, provocando danni e vittime.

Figura 1 Comuni della provincia di Modena<sup>4</sup>

## FATTORI INFLUENTI E DISCRIMINANTI

La prima fase dell'analisi consiste nell'identificazione delle variabili secondo le quali articolare la partizione dei comuni del modenese.

La base dati di partenza è rappresentata dalle quotazioni medie<sup>5</sup> pubblicate dall'OMI dell'Agenzia del Territorio. Le quotazioni sono disponibili sul sito istituzionale dell'Agenzia del Territorio con il dettaglio territoriale della zona OMI e con cadenza temporale semestrale dal 2004. Per ciascuna zona e per ogni tipologia residenziale (abitazioni signorili, civili, economiche, ville e villini) viene elaborato un intervallo di valori di mercato. Calcolando la media dei valori centrali degli intervalli di tutte le tipologie residenziali presenti in ogni zona OMI appartenente al comune si ottiene la quotazione media comunale. Quest'ultima, così calcolata rappresenta l'elemento iniziale per riflessioni successive. Il valore che la quotazione media assume, in assoluto ma soprattutto in comparazione con le altre quotazioni medie, rappresenta una differenza di apprezzamento delle residenze ubicate nei comuni considerati.

I fattori che incidono sui diversi livelli di quotazione che si riscontrano sono evidentemente molteplici, essendo legati a variabili localizzative territoriali che, a loro volta, comprendono caratteristiche dei fabbricati, presenza di collegamenti ma anche caratteristiche sociali ed economiche. Nei primi modelli apparsi in letteratura, la variabile localizzativa era trascurata a causa delle difficoltà di reperimento dei dati. Si trattava di modelli cosiddetti "location insensitive". Il miglioramento delle

<sup>4</sup> In tratteggiato si è evidenziato Modena in qualità di comune capoluogo della provincia.

<sup>5</sup> Per maggiori dettagli sul processo di formazione delle quotazioni e sulla definizione di zona omogenea si può consultare il Manuale della banca dati dell'Osservatorio del Mercato immobiliare (Agenzia del Territorio, 2008).

tecnologie informatiche ha permesso negli ultimi anni di utilizzare una vasta gamma di indicatori che tengono conto contemporaneamente di caratteristiche geografiche e socio-economiche (ne sono esempi i lavori di Jackson *et al.*, 2007 o di Kestens *et al.*, 2006).

Un numero così elevato di caratteristiche richiede un processo di selezione accurato, in grado di giungere ad un numero ristretto di variabili rilevanti. In questo lavoro, si cerca di risolvere l'ovvia complessità del sistema non attraverso l'applicazione di un metodo fattoriale di riduzione della dimensionalità, come più volte fatto in letteratura (MacLennan e Tu, 1996), ma piuttosto di sintetizzare in un unico indicatore informazioni che a priori si stabilisce appartenere logicamente ad una categoria di fattori. Non avendo la possibilità di considerare tutte le variabili influenti, si sono fissate delle categorie generali piuttosto ampie all'interno delle quali fare delle considerazioni al fine di identificare un'unica variabile a rappresentanza della categoria. E' bene dire che le categorie scelte pur non essendo esaustive, rappresentano comunque un buon punto di partenza per approfondimenti successivi.

## DEFINIZIONE DELLE VARIABILI

Nel seguito del paragrafo saranno descritti brevemente i passi che hanno condotto alla formazione delle variabili che costituiranno il *set* di dati utilizzato, come matrice di *input*, per l'applicazione dell'analisi dei gruppi. Partendo, come detto, da un'analisi dei possibili fattori influenti si sono scelte sei classi di categorie da analizzare:

- fattori di composizione dello *stock* immobiliare sul territorio;
- fattori di intensità e dinamicità del mercato residenziale;
- fattori attrattivi delle residenze;
- fattori di dinamica temporale dei prezzi delle abitazioni;
- livelli dei prezzi delle abitazioni;
- fattori di correlazione spaziale dei livelli di prezzo.

### Fattori di composizione dello *stock* immobiliare sul territorio

La qualificazione di una zona come residenziale piuttosto che industriale o commerciale si rivela, anche intuitivamente, una caratteristica importante nella formazione del prezzo medio di un'area territoriale. Infatti, è evidente che la maggiore presenza di una certa tipologia edilizia modifica l'ambiente circostante con conseguenze sull'apprezzamento delle abitazioni in esso ubicate. Si ritiene quindi opportuno tenere in considerazione, attraverso la scelta di una variabile idonea, la tipologia di unità immobiliari presenti in ciascun comune.

Per apprezzare le differenze si è ritenuto utile valutare lo *stock*<sup>6</sup> immobiliare edificato per ciascun gruppo catastale in termini di unità censite e di rendita. Ai fini della costruzione della variabile di interesse nel presente studio per *stock* immobiliare si intende il numero di unità immobiliari censite negli archivi catastali, distinto per tipologia e per rendita che rappresenta la base imponibile fiscale determinata dal Catasto.

Solitamente nel mondo immobiliare si utilizza la distinzione tra due importanti macro categorie di immobili: gli immobili residenziali e gli immobili non residenziali. L'analisi della composizione delle unità immobiliari offre informazioni sulla tipologia di unità immobiliari presenti che possono qualificare un'area come residenziale, piuttosto che produttiva o commerciale. Tuttavia, il numero inteso in termini di somma di singole unità immobiliari per ciascuna tipologia non è sufficiente a fornire l'informazione ritenuta utile. Riscontrare in un comune una quota maggiore di abitazioni residenziali non indica di per sé che il comune possa essere considerato avere più vocazione residenziale rispetto ad un altro.

<sup>6</sup> Per una migliore definizione di *stock* e di rendita si può fare riferimento al volume STATISTICHE CATASTALI 2010 (CATASTO EDILIZIO URBANO) di cui l'ultima pubblicazione è del 27 ottobre 2011 disponibile sul sito [www.agenziaterritorio.gov.it](http://www.agenziaterritorio.gov.it)

Per tale ragione si è ritenuto più utile considerare l'informazione sulla rendita<sup>7</sup> e in particolare calcolare la quota di rendita attribuibile al settore residenziale che si indica per brevità con  $QR_{res}$ .

Per la costruzione di questa variabile si è considerata da un lato la somma di tutte le rendite degli immobili ritenuti residenziali ( $R_{res}$ ) e dall'altro la somma delle rendite delle unità non residenziali ( $R_{nres}$ ). In  $R_{res}$  sono sommate tutte le rendite delle unità immobiliari censite nel gruppo A ad eccezione delle A10. In  $R_{nres}$  sono state sommate le rendite di diversi gruppi di immobili concettualmente distinti per destinazione d'uso. Per quantificare il peso del settore terziario si sono considerate le unità immobiliari appartenenti alla categoria A10 (uffici) e D5 (istituti di credito). Il settore commerciale è stato identificato con le categorie C1 e C3 (negozi), D8 (edifici commerciali) e D2 (alberghi). Infine per il settore produttivo si sono considerate le rendite delle categorie D1 e D7 (capannoni).

La variabile  $QR_{res}$  è quindi definita come nella relazione seguente:

$$QR_{res} = R_{res} / (R_{res} + R_{nres}),$$

e sarà inserita nella matrice delle variabili utili per la successiva analisi dei gruppi.

### Fattori di intensità e dinamicità del mercato residenziale

Il livello delle quotazioni e l'andamento nel tempo delle stesse dipende anche dal numero di compravendite che sono effettuate nel territorio. A loro volta il numero di scambi è, in generale, funzione del numero di unità immobiliari presenti in una certa ripartizione territoriale. Per fare una riflessione sulla vivacità di un mercato è quindi necessario esaminare congiuntamente il numero di compravendite avvenute nel territorio e il numero di abitazioni presenti nello stesso. Per quanto riguarda il numero di scambi per ciascun comune viene preso in considerazione il numero di compravendite (transazioni) normalizzate per quota compravenduta (NTN)<sup>8</sup>, riferito agli immobili residenziali ovvero quelli appartenenti alle categorie catastali da A1 ad A11 esclusa la categoria A10<sup>9</sup>. Per il numero di unità residenziali presenti nel comune si considera lo *stock* di unità censite per le stesse categorie prese a riferimento per il NTN. Per capire la dinamicità del settore residenziale in ciascun comune è necessario, quindi, depurare il numero assoluto di compravendite osservate dalla grandezza dello *stock* presente. Per tale ragione si effettua un'analisi del rapporto tra NTN e *stock* che si indica con IMI. Per ogni comune dal 2004 al 2010 è calcolato il valore dell'IMI, in formule:

$$IMI_{it} = NTN_{it} / Stock_{it}$$

con *i* indice del comune *i*-esimo e *t* indice dell'anno di riferimento.

L'IMI esprime quindi la quota di abitazioni compravendute in un certo anno e può essere interpretato come una misura della vivacità del mercato in considerazione del fatto che ad un valore dell'IMI più elevato corrisponde una quota maggiore di abitazioni scambiate al netto dell'effetto della dimensione dello *stock*.

Calcolato il valore dell'IMI in ogni comune per ciascun anno, l'analisi della dinamica evidenzia un andamento generalmente decrescente imputabile del tutto al calo delle compravendite ravvisato in tutti comuni.

<sup>7</sup> A rigore la rendita catastale non è l'indicatore più corretto per la misura dell'utilizzo degli immobili sul territorio. Sarebbe più idoneo un parametro che esprima la quota di suolo occupato da immobili residenziali rispetto agli altri utilizzi. Tuttavia non si dispone di tale informazione e quindi aver utilizzato la rendita catastale consente quantomeno di tener conto della superficie delle unità immobiliari secondo il loro utilizzo.

<sup>8</sup> Il valore del NTN in un determinato periodo e per una determinata categoria di immobili viene desunto dalle Banche Dati della Pubblicità immobiliare (ex Conservatorie).

<sup>9</sup> Per un dettaglio degli immobili contenuti in ciascuna categoria si può vedere l'appendice della pubblicazione STATISTICHE CATASTALI 2010 (CATASTO EDILIZIO URBANO).

Per cogliere la dinamica nel tempo della vivacità del mercato si prende in considerazione la differenza dell'IMI tra il periodo iniziale di osservazione (2004) e il periodo finale (2010) ovvero

$$\Delta IMI_{10-04} = IMI_{10} - IMI_{04}$$

L'indicatore così calcolato costituisce la variabile da inserire nella matrice per l'analisi di *cluster*.

### Fattori attrattivi delle residenze

In un certo territorio, un fattore rilevante nella formazione dei prezzi e nella loro dinamica nel tempo è rappresentato dalla capacità dello stesso di suscitare interesse all'acquisto di residenze. L'idea alla base della considerazione dei fattori attrattivi tra le variabili dell'analisi è quella che, in un particolare territorio, la presenza di compravendite immobiliari effettuate da soggetti non residenti nel territorio stesso sia un indizio di appetibilità dall'esterno del mercato immobiliare locale.

Tale interesse può essere imputabile a fattori economici legati alla presenza di attività industriali o del terziario, che offrendo opportunità di lavoro stimolano flussi migratori di persone aumentando la richiesta di abitazioni, o può dipendere da altri fattori quali la facilità di raggiungere un centro più grande oppure da motivazioni di villeggiatura.

Partendo da queste ipotesi, si distinguono fattori attrattivi prevalentemente di due tipi: di natura economica e di natura sociale.

Per quanto concerne i fattori attrattivi di natura economica, si può fare riferimento all'informazione sull'appartenenza o meno di un comune ad un distretto industriale. Questa variabile Attrazione economica ( $A_e$ ) è, dunque, basata sulla specializzazione produttiva del territorio. Sulla base di queste informazioni si è costruita una variabile categorica assegnando a ciascun comune i seguenti valori:

- 0 se il comune non appartiene ad alcun distretto;
- 1 se il comune appartiene al distretto industriale del sistema locale del lavoro SLL Modena (sistemi urbani);
- 2 se il comune appartiene al distretto industriale del SLL Carpi (sistemi del tessile, delle pelli e dell'abbigliamento);
- 3 se il comune appartiene al distretto industriale del SLL Mirandola (altri sistemi del made in Italy).

La variabile legata a fattori di natura sociale che si indica con Attrazione sociale ( $A_s$ ) è costruita come un indice composito che tiene conto congiuntamente della combinazione della percentuale di abitazioni principali e dell'indicazione turistica fornita dall'ISTAT. Per ciascun comune la percentuale di abitazioni principali  $Q\_Abitaz_p$  è calcolata come rapporto tra il numero di abitazioni possedute da persone fisiche che dichiarano di utilizzare l'immobile come abitazione principale<sup>10</sup> nel 2009 e lo stock di abitazioni nello stesso anno.

Da un punto di vista operativo sulla base dei valori assunti dalla variabile  $Q\_Abitaz_p$  viene fissata una soglia empirica pari al 40%; valore poco inferiore al primo quartile della distribuzione della  $Q\_Abitaz_p$  di tutti i comuni pari a circa il 43%.

Dal sottoinsieme dei comuni che presentano un valore inferiore alla soglia prefissata, si considerano solo i comuni definiti turistici in base alla definizione ISTAT.

In tal modo giungiamo alla definizione di una variabile *dummy*  $A_s$  che vale 0 o 1. La variabile assume valore pari a zero se la condizione definita non è verificata e il comune non è attrattivo, mentre assume valore pari ad uno se presenta attrattività (sotto la soglia e comune turistico secondo ISTAT).

Va detto che una percentuale bassa di abitazioni principali potrebbe essere imputata anche al fenomeno di spopolamento del comune. Nel caso dei comuni in esame si ritiene però che il valore vada

<sup>10</sup> Per un'analisi approfondita dell'utilizzo dello immobiliare in Italia si può vedere il volume Gli immobili in Italia. Distribuzione del patrimonio e dei redditi dei proprietari disponibile sul sito [www.agenziaterritorio.it](http://www.agenziaterritorio.it).

letto nel senso positivo di presenza di alloggi del tipo seconde abitazioni. Infatti, molti dei comuni selezionati appartengono al Monte Cimone che è il maggior rilievo dell'Appennino settentrionale e della regione Emilia-Romagna e ospita un'importante stazione sciistica invernale. Il Monte Cimone interessa in particolare i comuni di Fiumalbo, Fanano, Riolunato e Sestola che è forse la più celebre località turistica invernale dell'Emilia-Romagna. Inoltre, i comuni di Pievelego e Fiumalbo confinano con i comuni dell'alta montagna del Pistoiese che comprende alcune note località di villeggiatura di montagna sia invernale sia estiva tra cui la più nota è probabilmente il comune di Abetone.

Quale ipotesi semplificatrice, in questo studio si ipotizza che osservare in un certo comune una quota di abitazioni principali esigua rispetto ai restanti valori della provincia si possa interpretare come indice di capacità del comune di attrarre residenze del tipo seconde case ad uso villeggiatura (per affittarle o ad uso diretto).

Le variabili  $A_e$  e  $A_s$  saranno quelle considerate nella base dati di partenza per l'analisi dei gruppi.

### Fattori di dinamica temporale dei prezzi delle abitazioni

Come noto una successione ordinata di osservazioni nel tempo costituisce una serie storica la cui analisi si propone tipicamente di raggiungere obiettivi di previsione attraverso lo studio delle regolarità insite nell'andamento del fenomeno.

In questo caso, la disponibilità di dati con una cadenza temporale regolare per più di dieci semestri rende possibile qualche riflessione sulla dinamica nel tempo delle quotazioni rilevate. D'altro lato, la relativa brevità della lunghezza della serie storica dei prezzi non permette l'applicazione di modelli statistici avanzati, ma solo un primo esame esplorativo che tuttavia si rivela utile per la costruzione di una variabile che offra informazioni sull'andamento delle quotazioni in ciascun comune. In questo senso, l'analisi delle quotazioni nel tempo viene qui effettuata con l'obiettivo di definire una variabile che dia un'indicazione della dinamica di fondo delle quotazioni nel periodo considerato.

Per giungere alla definizione di questa variabile si è proceduto in due fasi:

- in una prima fase si sono analizzate le serie delle quotazioni separatamente per ciascun comune applicando una regressione temporale;
- in una seconda fase i coefficienti ottenuti sono stati raggruppati in categorie.

L'analisi grafica congiunta delle serie di tutti i comuni ha messo in evidenza per la maggior parte di essi la presenza di un trend lineare e in alcuni casi di un trend quadratico. Si è proceduto, quindi, a stimare per ogni comune una regressione delle quotazioni ( $Q$ ) rispetto ad una variabile temporale  $t$  e ad una variabile temporale al quadrato  $t^2$ . Per il singolo comune il modello di regressione stimato è del tipo:

$$Q_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon,$$

con  $\beta_0$  intercetta del modello,  $\beta_1$  e  $\beta_2$  coefficienti rispettivamente delle variabili  $t$  e  $t^2$  e infine  $\varepsilon$  componente erratica che si ipotizza avere distribuzione normale con media nulla e varianza costante  $\sigma^2$ . La stima dei modelli di regressione ha confermato quanto già evidente dall'analisi grafica. Essenzialmente vi sono comuni caratterizzati da livelli dei prezzi crescenti in tutto il periodo e comuni che mostrano una flessione ravvisabile generalmente nell'ultimo periodo. Ottenute, per ciascun comune, le stime dei coefficienti con il metodo dei minimi quadrati lineari, di cui si riporta il dettaglio nell'appendice A, si è costruita una variabile con 4 modalità assegnando:

- valore nullo nel caso in cui nel comune sia il coefficiente  $\beta_1$  legato al trend sia il coefficiente  $\beta_2$  relativo alla variabile trend al quadrato sono entrambi non significativi;
- valore 1 nel caso in cui i coefficienti  $\beta_1$  e  $\beta_2$  sono entrambi significativi;
- 2 e 3 si riferiscono entrambi al risultato in cui solo la variabile  $t$  è significativa ma si distinguono i comuni in cui il coefficiente  $\beta_1$  assume un valore contenuto, arbitrariamente fissato inferiore a 10. A tali comuni si assegna valore 2 mentre laddove  $\beta_1$  è superiore a 10 si assegna valore 3.

La variabile così definita, che si può indicare genericamente con  $T$ , ad indicazione del *trend* temporale, sarà utilizzata nel processo di *clustering*.

### Livelli dei prezzi delle abitazioni

L'analisi della divergenza dei prezzi che si riscontra tra diverse partizioni territoriali risulta sempre di notevole interesse, tanto che l'argomento costituisce un tema a sé stante per la quantità di studi ad esso dedicati. In generale si può ipotizzare che la differenza tra i prezzi dei comuni sia rappresentativa dei diversi vantaggi o svantaggi o più sinteticamente delle amenità che si hanno nell'abitare in un luogo piuttosto che in un altro in considerazione di un'ampia serie di fattori. In tal senso, si ritiene utile un'analisi di tali differenze al fine di giungere ad una variabile che possa esprimere una misura dei divari tra i livelli dei prezzi.

Un modo semplice per evidenziare le differenze tra i livelli delle quotazioni è quello di considerare il rapporto tra il prezzo di ciascun comune e un prezzo ritenuto particolarmente importante in termini di confronto. Il prezzo al denominatore può essere rappresentato dal valore massimo, dal minimo, dalla media o da un valore assunto da un comune che ha un ruolo di interesse nell'analisi. Il rapporto citato, che si può denominare sinteticamente differenziale, offre un'informazione sulla distanza tra la quotazione media di un comune e quella del comune scelto come base di confronto.

In questa analisi il differenziale ( $Diff$ ) è calcolato per ciascun semestre come rapporto tra la quotazione di ciascun comune ( $Q_i$ ) e la quotazione rilevata a Modena ( $Q_M$ ), quotazione che si ritiene essere rappresentativa essendo, oltre che il capoluogo della provincia, anche il massimo delle quotazioni per quasi tutti i semestri. Quindi, in simboli la variabile è data da:

$$Diff_{it} = Q_{it} / Q_{Mt}$$

con  $i$  indicatore del comune e  $t$  indicatore del semestre. Si ottengono così 15 valori dei differenziali per ogni comune (dal I semestre 2004 al I semestre 2011). Per analizzare l'evoluzione nel tempo dei differenziali si è costruita la distribuzione di frequenza dei differenziali osservati in tutto il periodo fissando un'unica ampiezza delle classi<sup>11</sup> per tutti i comuni. In tal modo si è notato che i differenziali di ciascun comune sono risultati piuttosto stabili nel tempo. La maggior parte dei comuni mostrano una sostanziale permanenza del valore dei differenziali nella stessa classe per tutto il periodo. Si è scelto, quindi di utilizzare la media dei differenziali del periodo,  $Diff_{i,}$ , come variabile che esprima le differenze di apprezzamento tra i comuni.

### Fattori di correlazione spaziale dei prezzi

Molti autori hanno dimostrato che la posizione geografica, e più in generale la dimensione spaziale, hanno un ruolo centrale nell'analisi delle determinanti dei prezzi delle abitazioni (Smith *et al.*, 1988). I valori immobiliari presentano, infatti, una connotazione geografica precisa.

I fattori di dinamica spaziale si esprimono attraverso la costruzione di una variabile basata su di un indicatore di correlazione spaziale.

Il concetto di correlazione spaziale su un principio della geografia quantitativa è che ciò che succede vicino è più importante di ciò che succede lontano. Più precisamente, nell'analisi di un fenomeno spaziale, ci si attende che la correlazione tra due variabili osservate in due locazioni ad una distanza

<sup>11</sup> Sono state fissate cinque classi:

0 -| 0,5;  
0,5 -| 0,7;  
0,7 -| 0,9;  
0,9 -| 1;  
1 -| 1,1.

x tra loro decresca all'aumentare di tale distanza x. Ciò ovviamente pone la necessità di dare una forma analitica a questo fenomeno di “decadimento spaziale” attraverso una matrice di pesi che esprima la forza relativa del legame di correlazione tra variabili osservate a varie distanze. Usualmente nelle analisi empiriche, i legami spaziali sono presi in considerazione attraverso la costruzione della matrice dei pesi spaziali  $W^{12}$ . La scelta della matrice di distanze è in alcuni casi tutt'altro che neutrale. In effetti, come mostrato in Anselin (1988), matrici diverse possono portare a conclusioni diametralmente opposte circa la forza e l'estensione geografica della correlazione spaziale. Non esiste però una regola generalmente valida e la scelta della matrice dipende dalla natura del fenomeno oggetto di studio. Le matrici di contiguità e a bande di distanza rappresentano invece delle scelte più neutrali che vale la pena di implementare quando non si ha un'idea precisa della natura del processo spaziale sottostante. Una particolarità di queste matrici è che, a differenza di quelle funzionali, se ne può usare più d'una nella medesima analisi. Per esempio, si possono costruire matrici per cogliere la correlazione spaziale tra zero e cinquanta chilometri, cinquanta-cento, cento-centocinquanta, ecc., per poi usarle tutte insieme al fine di vedere fino a che distanza si rileva evidenza di correlazione. Rosenthal e Strange (2003) fanno uso di tale tecnica per vedere per esempio fino a che distanza le esternalità marshalliane influiscono sulla nascita di nuove imprese.

Tra le tante scelte possibili, quella più coerente con la natura dei comuni di una provincia è quella delle matrici di distanza, in particolare si è scelto il criterio *K nearest neighbour*. L'estensione geografica dei comuni è, infatti, nella maggior parte dei casi, molto ristretta, per cui sembra logico considerare delle matrici di correlazione in cui non solo ciò che è direttamente contiguo ha un effetto diretto, ma l'effetto contagio si estende anche ai “vicini dei vicini”.

Secondo questo criterio due comuni  $S_i$  e  $S_j$  si possono definire vicini se la distanza tra essi è minore di una soglia prefissata. Data la particolare forma di poligoni irregolari propria dei confini amministrativi comunali, si è scelto un numero di comuni vicini pari a cinque, ovvero considerando contigui i 5 comuni più vicini (nel senso di distanza minore tra i centroidi dei poligoni dei comuni).

In tal modo, partendo dall'ipotesi che le quotazioni immobiliari non sono influenzate solo da ciò che caratterizza l'area di residenza dell'immobile, ma vengono influenzate anche dalle caratteristiche economico-sociali e territoriali delle aree limitrofe a quelle in cui si trovano gli immobili di riferimento, si suppone che le determinanti dei prezzi degli immobili hanno un raggio di azione più ampio rispetto ai confini del comune in cui risiede l'immobile. Si propone, dunque, una variabile che rappresenta l'insieme dei fattori di dinamica spaziale, che tenta di cogliere quella parte del prezzo degli immobili che sfugge alle tradizionali misure statistiche che non contemplano la dimensione spaziale. Va, tuttavia, tenuto conto che l'analisi è limitata, per ovvie esigenze di semplificazione, ai confini amministrativi della provincia e dunque non tiene conto delle eventuali influenze esercitate dai comuni limitrofi che appartengono ad altre provincie.

Al fine della determinazione della variabile Indicatore Spaziale ( $I_s$ ) bisogna esplicitare il concetto di

<sup>12</sup> In merito alla costruzione di tale matrice in letteratura, esistono due approcci, un primo approccio *standard*, fondato sulla contiguità e sulle misure basate sulla distanza, ed un altro recente approccio costituito da tutta una serie di tentativi di misura, tra i quali si ricordano il metodo Bayesiano che ricerca il *best fit*, le tecniche non parametriche, la correlazione spaziale, le tecniche iterative, ecc.. A tal proposito Bhattacharjee e Jensen-Butler (2006) sottolineano che la scelta dei pesi è frequentemente arbitraria, c'è una sostanziale incertezza riguardo la scelta della matrice ( $W$ ) ed i risultati degli studi variano considerevolmente al variare della scelta dei pesi spaziali. La scelta di  $W$  dovrebbe essere quella che porta ad un risultato buono e parsimonioso ma il rischio di un'errata specificazione esiste sempre. Nell'approccio *standard*, la scelta di questa matrice avviene a priori, e si assume sempre che gli elementi  $w_{ij}$  della diagonale della matrice  $W$  siano nulli per evitare di cadere direttamente in problemi di endogeneità. Tra le matrici più comuni vi sono:

- Matrici di contiguità, in cui l'elemento generico  $w_{ij}$  di  $W$  vale uno se le aree  $i$  e  $j$  condividono un confine, o zero altrimenti;
- Matrici a bande di distanza in cui il generico elemento  $w_{ij}$  vale uno se le localizzazioni  $i$  e  $j$  si trovano ad una distanza  $d_{ij}$  inferiore ad una soglia prefissata  $d$ , o zero altrimenti;
- Matrici funzionali in cui il generico elemento  $w_{ij}$  assume un valore pari ad una funzione della loro distanza  $d_{ij}$ .

Tuttavia la distanza stessa non è un concetto semplice, essa può anche essere riferita ad una distanza economica tra le regioni, ad esempio la prossimità tecnologica.

correlazione spaziale. Infatti, una volta scelta la matrice più congrua in funzione del fenomeno da analizzare, si passa alla misurazione della correlazione spaziale.

Per ogni semestre viene calcolato l'indice Moran<sup>13</sup>, per verificare se esiste un'autocorrelazione significativa tra le quotazioni delle abitazioni. Dall'analisi della autocorrelazione spaziale globale sulle quotazioni delle abitazioni si rileva l'esistenza di una significativa dipendenza spaziale tra i prezzi delle case in tutti semestri. Al fine di classificare i sottomercati in base al prezzo delle case, vista la significativa autocorrelazione globale, oltre all'indice di Moran globale viene calcolato anche il corrispondente indicatore locale di associazione spaziale, LISA, il quale misura il grado di autocorrelazione per ciascuna unità territoriale. C'è autocorrelazione se si può individuare qualche *pattern* caratteristico di localizzazione di una variabile. In particolare, se aree vicine sono più simili di quelle lontane, allora si tratta di autocorrelazione positiva. Se aree vicine sono più diverse di quelle lontane, invece l'autocorrelazione è negativa. Se i *pattern* sono casuali abbiamo assenza di correlazione.

Il Moran locale soddisfa due criteri: il LISA per ogni osservazione fornisce informazioni sulla significatività del raggruppamento spaziale di valori simili intorno ad una osservazione; in secondo luogo la somma dei LISA per tutte le osservazioni è proporzionale all'indicatore globale di associazione spaziale.

La versione locale della statistica I di Moran per ogni osservazione i può essere scritta così:

$$I_i = \frac{z_i}{\sum_{i=1}^N z_i} \sum_{j \in J_i} w_{ij} z_j$$

dove  $z_i$  è il valore corrispondente al comune  $i$  della variabile già normalizzata e  $J_i$  è l'insieme delle comuni vicine a  $i$ . L'ipotesi nulla è l'assenza di autocorrelazione spaziale per cui se il test, che si distribuisce come una normale standardizzata, ha valori significativamente positivi, avremo un *cluster* di regioni con caratteristiche simili. Viceversa, valori significativamente negativi indicheranno un *cluster* di regioni diversificate. In altre parole, per ogni unità territoriale sarà possibile indicare il tipo di correlazione (negativa o positiva) e il suo livello di significatività.

Il LISA sulle quotazioni immobiliari ci permette di verificare l'appartenenza di ciascun comune ad un *cluster* che può essere sempre e solo di quattro tipi: 0, non significativo, 1 alto alto (H - H), 2 basso basso (L - L), 3 basso alto (L - H), 4 alto basso (H - L). I risultati di dettaglio sono riportati nell'appendice B.

<sup>13</sup> L'indice Moran, è dato dalla seguente espressione:

$$I = \frac{(N/S_0) \sum_{i,j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Dove  $N$  indica in numero totale delle aree,  $x_i$  rappresenta la variabile quotazioni immobiliari nel comune  $i$ ,  $x_j$  rappresenta la variabile quotazioni immobiliari nel comune  $j$  e  $w_{ij}$  sono i pesi della matrice  $W$  e

$$S_0 = \sum_{i,j=1}^n w_{ij}$$

è la somma di tutti gli elementi della matrice nella quale in base alla distanza tra punti è definita contiguità tra le aree. Tale indicatore misura la correlazione di una variabile con sé stessa nello spazio, ha un campo di variazione compreso tra -1 e 1.

L'analisi ottenuta attraverso il LISA per ciascun semestre evidenzia l'esistenza di due aree con autocorrelazione positiva di tipo L - L (il primo nella parte a nord della provincia di Modena e nei comuni limitrofi; comprendendo i comuni di Concordia sul Secchia, San Possidonio, Cavezzo, Mirandola, San Felice sul Panaro, Medolla e Camposanto; il secondo, nella parte sud-ovest della provincia che include i comuni di Frassinoro, Montefiorino, Lama Mocogno e Polinago). Da segnalare, soprattutto, la presenza di un'area intorno al capoluogo, con autocorrelazione positiva di tipo H - H, comprendente i comuni di Modena, Campogalliano, Formigine, Fiorano Modenese, Maranello, Castelvetro di Modena, Vignola, Spilamberto e Castelfranco Emilia. Alle precedenti si affianca un comune "atipico" San Cesareo sul Panaro, con bassi livelli del valore delle quotazioni immobiliari, ma circondato da centri con valori superiori alla media provinciale (L - H). Al fine di creare un'unica variabile informativa dei risultati ottenuti dal LISA, viene confrontata l'appartenenza di ciascun comune ad un *cluster* per tutti i semestri in esame. Se il comune permane nel *cluster* del tipo specificato per un numero di volte (anche se non consecutive) superiore ad una soglia fissata, si stabilisce che il comune appartiene sicuramente al *cluster*. La soglia è fissata pari a 6 semestri su un totale di 22 semestri analizzati, ovvero per 3 anni.

A questo punto si costruisce una variabile ordinale, che definiamo Indicatore spaziale,  $I_s$ , che assume valori 0, 1, 2, 3 e 4 con i seguenti significati:

- 0 se il comune non ha dimostrato di appartenere a nessun *cluster* per almeno un numero di semestri pari a 6;
- 1 se il comune si trova in un *cluster* di comuni in cui vi sono alte quotazioni e queste si influenzano vicendevolmente (H - H) per un numero di semestri almeno pari a 6;
- 2 se il comune si trova all'interno di un *cluster* di comuni in cui sono presenti basse quotazioni con una correlazione positiva tra loro (L - L) per almeno 6 semestri;
- 3 e 4 se il comune si trova in un'area in cui le quotazioni comunali hanno una correlazione negativa tra loro del tipo 3 basso alto o 4 alto basso per almeno 6 semestri.

Questa variabile rappresenta l'indicatore di correlazione spaziale  $I_s$ .

## RAGGRUPPAMENTO DEI COMUNI

I metodi che consentono di aggregare unità spaziali in modo da partizionare un'area di più ampie dimensioni si rifanno principalmente a procedure di *cluster analysis*. Essa risulta una tecnica utile anche nel caso, quale quello in esame, in cui la formazione di nuove aree in termini di raggruppamento di più comuni ha l'obiettivo di individuare aree omogenee rispetto ad un insieme di variabili per le quali può essere interessante la diffusione periodica di informazioni provenienti da fonti amministrative.

La *cluster analysis* è una tecnica ampiamente utilizzata nelle scienze applicate ed è supportata da una letteratura molto estesa che in questi ultimi anni ha reso disponibili una quantità elevata di metodi nuovi nonché approfondimenti e miglioramenti di strumenti già esistenti. Come sottolineato in una delle ultime revisioni sull'argomento fornita da Jain *et al.* (1999) una disponibilità così vasta di algoritmi di *clustering* può facilmente confondere un ricercatore che tenta di selezionare un algoritmo adatto alla spiegazione di un problema applicato. Nonostante l'ampia produzione sul tema molti problemi restano ancora aperti e la soluzione è il più delle volte lasciata all'esperienza dell'utilizzatore. Nel seguito, quindi, si riportano brevemente i passi seguiti per giungere al raggruppamento finale.

Prima di procedere con la *cluster* si è effettuata un'analisi esplorativa iniziale della matrice dei dati. Per i 47 comuni della provincia di Modena si sono costruite 7 variabili ognuna delle quali offre un'informazione sintetica di un numero più ampio di fattori. Delle 7 variabili 3 sono continue ( $QR_{res}$ ,  $\Delta IMI_{10-04}$  e  $Diff_i$ ) sebbene assumano valori nell'intervallo [0, 1] e 4 variabili sono categoriche ( $A_s$ ,  $A_e$ , T e  $I_s$ ).

In Tabella 1 si riportano le statistiche descrittive delle 3 variabili quantitative mentre in Figura 2 e in Figura 3 sono visualizzate le mappe per ciascuna delle variabili.

Si evince che la quota di rendita attribuibile al settore residenziale è maggiore nei comuni del sud della

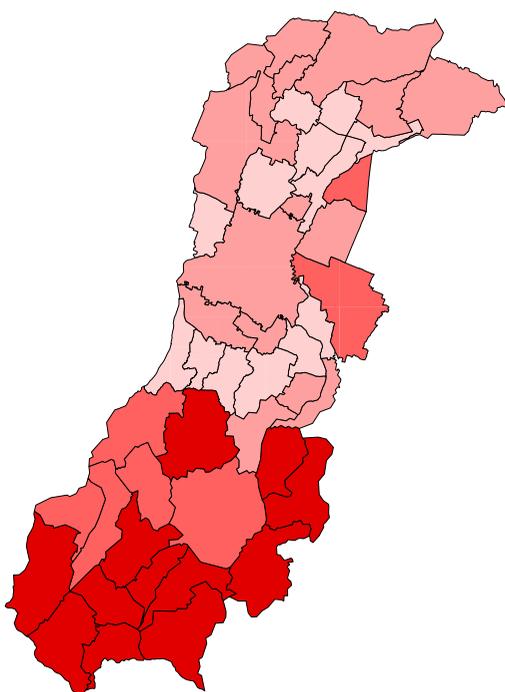
provincia rispetto alla restante porzione di territorio esaminata. In particolare i comuni in cui il peso del settore residenziale è notevolmente maggiore rispetto ai settori terziario, commerciale e produttivo sono situati a Sud di Modena. Per ciò che concerne la differenza di IMI tra il 2004 e il 2010, la mappa evidenzia una generale diminuzione della dinamicità del mercato dovuta alla riduzione delle compravendite. Tra i comuni, a subire i cali più marcati sono quelli situati a Est di Modena.

Da un'analisi della divergenza dei prezzi delle abitazioni si può notare che, nell'ambito della provincia, ci sono significative differenze di apprezzamento tra i diversi comuni. Si possono notare quotazioni simili per i comuni intorno al capoluogo, che mostrano un valore del differenziale nella classe tra 0,7 e 1. Fanno eccezione i comuni di Sestola e Vignola che pur presentando valori analoghi a Modena sono ubicati a Sud della provincia. Valori distanti dalla quotazione di riferimento di Modena si riscontrano indifferentemente sia a Sud sia a Nord del capoluogo. Circa la metà dei comuni non appartengono ad alcun distretto produttivo e sono collocati quasi esclusivamente a Sud del capoluogo ad eccezione di Camposanto e Medolla i quali non presentano caratteristiche di attrazione economica. Gli altri comuni della provincia fanno parte di tre distretti. In particolare 13 comuni, ubicati prevalentemente nella cintura del capoluogo, appartengono al sistema urbano di Modena. Carpi, Novi Modena e Soliera formano il distretto industriale del tessile delle pelli e dell'abbigliamento di Carpi. Infine i 7 comuni più a Nord Est appartengono al distretto del Made in Italy. Da notare che la variabile  $A_g$  conferma quanto già visto dall'analisi della variabile  $QR_{res}$  che mostra una minore presenza di attività non residenziali nella parte a Sud di Modena. Per quanto riguarda la variabile  $A_s$ , mostrata nella seconda mappa di Figura 3, gran parte dei comuni della provincia presenta caratteristiche residenziali del tipo abitazioni principali. La parte Sud Est mostra una capacità di attrazione da un punto di vista immobiliare di seconde abitazioni. Rispetto all'andamento temporale, a differenza delle altre variabili, i comuni non presentano particolari strutture geografiche, mostrando un'evoluzione nel tempo delle quotazioni piuttosto diversificata all'interno della provincia. Nella mappa riferita alla variabile  $I_s$  si può vedere che ben ventidue comuni della provincia di Modena non appartengono a nessun *cluster*. Nove comuni nell'intorno della provincia di Modena fanno parte del *cluster* in cui alte quotazioni nei diversi comuni si influenzano positivamente e vicendevolmente. Il *cluster* più cospicuo per numero di comuni che ne fanno parte è quello caratterizzato da correlazione positiva tra basse quotazioni immobiliari di cui fanno parte dodici comuni; tali comuni si trovano in due aree ben definite, a sud ovest della provincia di Modena e a Nord della provincia. I comuni di Campogalliano, San Cesareo sul Panaro e Savignano sul Panaro presentano valori anomali con basse quotazioni immobiliari, pur trovandosi in un'area caratterizzata da prezzi alti delle abitazioni. Il comune di Mirandola, d'altro lato, rappresenta anch'esso un *outlier* con alte quotazioni immobiliari in un'area in cui prevalgono basse quotazioni.

**Tabella 1** Statistiche descrittive delle variabili quantitative

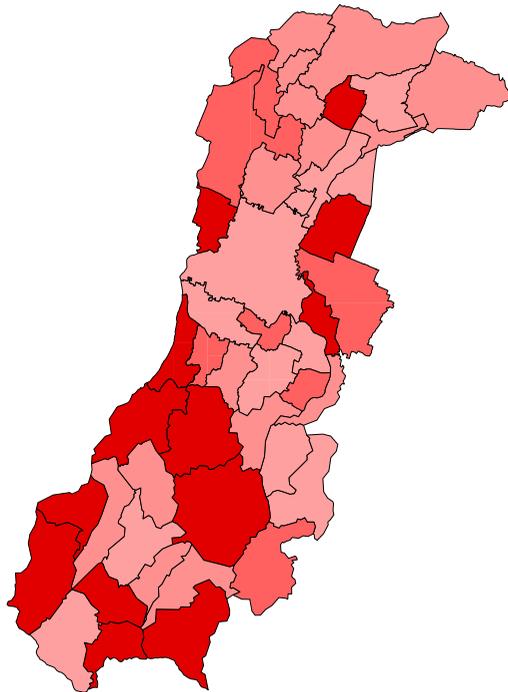
Statistiche	$QR_{res}$	$\Delta IMI_{10-04}$	Diff.
Minimo	0,174	-0,036	0,384
Primo quartile	0,418	-0,018	0,499
Mediana	0,474	-0,013	0,588
Media	0,518	-0,014	0,609
Terzo quartile	0,656	-0,010	0,705
Massimo	0,839	0,002	1,000

Figura 2 Mappe delle variabili quantitative



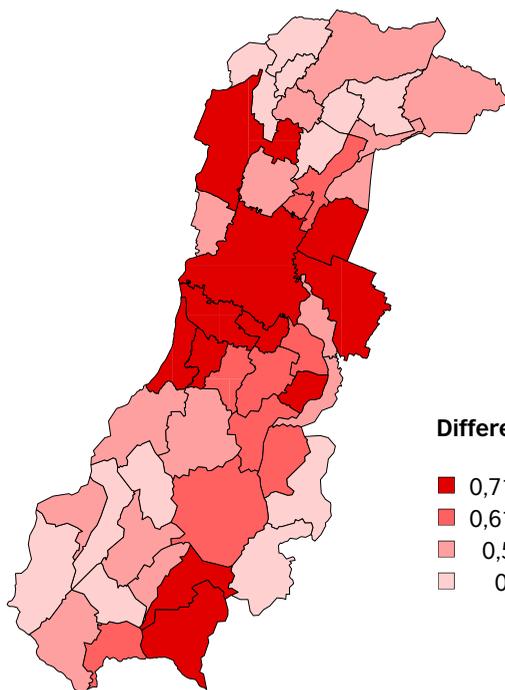
Quota rendita non residenziale,  $QR_{Res}$

- 66% - 84% (12)
- 52% - 66% (7)
- 42% - 52% (15)
- 0 - 42% (13)



Differenza IMI residenziale,  $IMI_{10-04}$

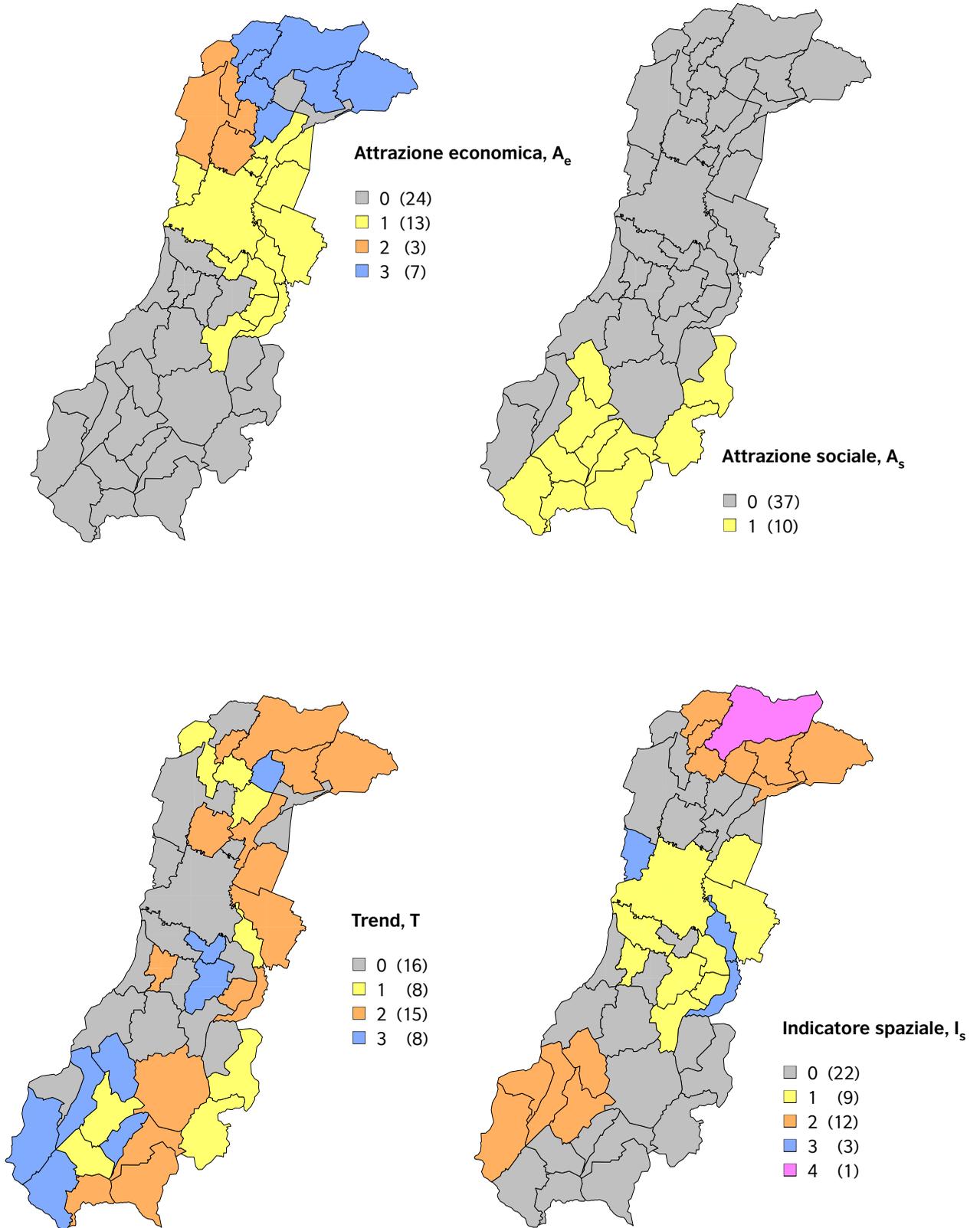
- - 0,01 - 0,003 (12)
- - 0,014 - 0,01 (15)
- - 0,018 - 0,014 (7)
- -0,036 - 0,018 (13)



Differenziale medio,  $Diff_i$

- 0,71 - 1 (11)
- 0,61 - 0,71 (9)
- 0,5 - 0,61 (15)
- 0 - 0,50 (12)

Figura 3 Mappe delle variabili categoriche



Sulla base dei risultati ottenuti con la selezione delle variabili, di cui si sono forniti i dettagli nei paragrafi precedenti, si procede ad applicare una procedura di *clustering*.

Per evitare eventuali problemi di scala e non essendoci alcun particolare motivo per conservarla si è deciso di standardizzare preliminarmente le variabili.

Per la scelta dell'algoritmo si è tenuto conto del fatto che le tecniche non gerarchiche si rivelano più efficaci se si cercano gruppi caratterizzati da alta omogeneità interna e si è optato per un algoritmo del tipo K-medie. Il K-medie, sviluppato per la prima volta da MacQueen nel 1967, appartiene alla classe dei metodi cosiddetti partizionanti, in contrapposizione con i metodi gerarchici. Attraverso una procedura iterativa il K-medie permette di suddividere le  $n$  unità in  $k$  partizioni in modo tale che ogni unità appartenga ad un solo gruppo. I passi<sup>14</sup> dell'algoritmo possono essere così riassunti:

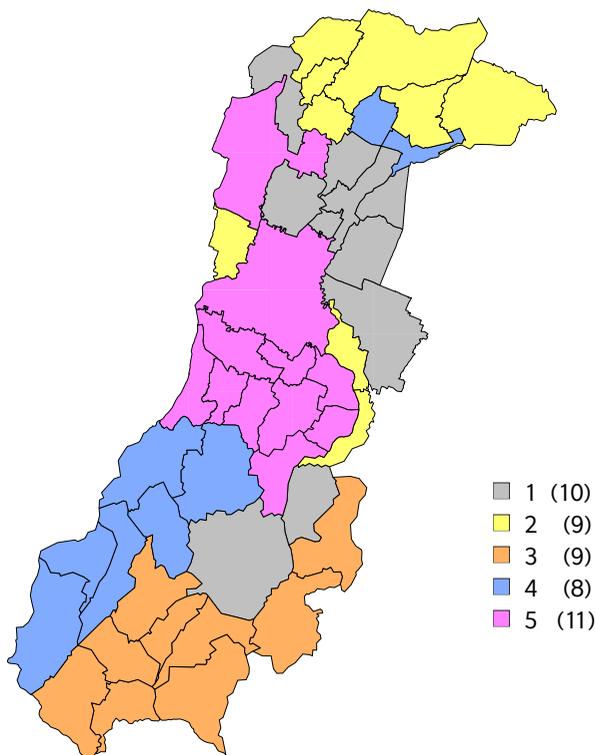
- scegliere il numero  $k$  di *cluster* da determinare;
- scegliere casualmente  $k$  dati dalla matrice di input che costituiscono i centroidi iniziali dei *cluster*;
- assegnare i restanti dati ai centri dei *cluster* più vicini utilizzando una funzione criterio (solitamente una funzione che minimizza la somma dei quadrati delle differenze tra le unità e i centroidi);
- utilizzare i dati in ogni *cluster* per calcolare una nuova media di ogni *cluster*;
- se le nuove medie sono identiche a quelle calcolate in precedenza, il processo è giunto a convergenza e termina, altrimenti, utilizzare le nuove medie come centri dei *cluster* e ripetere i passi da 3 a 5.

Come detto per questo tipo di tecnica il primo passo è la scelta del numero fisso di gruppi. Per effettuare tale scelta si è applicata una tecnica *model based*<sup>15</sup> che restituisce 5 come numero ottimale di gruppi presenti nella base dati. Tra i diversi algoritmi disponibili, per l'elaborazione della matrice dei dati, è stato utilizzato l'algoritmo proposto da Hartigan e Wong (1979), implementato nel *software* R con il quale è stata ottenuta la partizione. Quindi, impostando 5 come numero di gruppi si ottiene la partizione dei comuni visualizzata nella mappa di Figura 4 che mostra i *cluster* ottenuti con il K-medie.

<sup>14</sup> Si può fare riferimento, tra gli altri, a quanto riportato in Jain *et al.* (1999).

<sup>15</sup> In particolare è stata utilizzata la funzione *Mclust* del pacchetto *mclust* disponibile per il software R. Il pacchetto *mclust* versione 3 per R è descritto nell'articolo *MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering*, Technical Report no. 504, Department of Statistics, University of Washington, September 2006.

Figura 4: Mappa dei cluster ottenuti con il K-medie



Per validare la solidità della partizione si procede con l'applicazione alla stessa matrice di dati di una tecnica *fuzzy* impostando 5 gruppi. Nell'ambito della teoria degli insiemi *fuzzy* (la cui origine si fa risalire al lavoro di Zadeh del 1965) una unità può appartenere ad un *cluster* con un certo grado di appartenenza che assume un valore nell'intervallo  $[0, 1]$ . Quindi ciascuna unità, ciascun comune nel caso in analisi, può appartenere a più gruppi contemporaneamente con il vincolo che la somma delle appartenenze è pari ad 1. Ottenute per ciascun comune le probabilità di appartenenza<sup>16</sup> a ciascun gruppo si stabilisce di assegnare le unità utilizzando delle soglie soggettive. I comuni sono assegnati nei gruppi per cui risulta una probabilità elevata di appartenenza,  $> 0,6$  o una buona probabilità di appartenenza, compresa nell'intervallo  $[0,4 \text{ e } 0,6]$ . La Tabella 2 mostra per ciascun comune, attraverso i grafici a barre, la distribuzione delle probabilità di appartenenza a ciascun gruppo mentre la mappa di Figura 5 evidenzia il raggruppamento ottenuto assegnando i comuni che verificano la regola descritta.

Essenzialmente si possono distinguere due tipologie di comuni: quelli che presentano buone probabilità di appartenere ad un gruppo e quelli per i quali l'appartenenza ad un unico gruppo non è netta. Per la prima categoria di comuni il *fuzzy clustering* restituisce risultati del tutto uguali a quelli ottenuti con il K-medie evidenziando quindi anche una robustezza della partizione ottenuta in precedenza. Per i comuni che presentano caratteristiche di similarità con più di un gruppo emerge invece un problema di *overlapping*. Tuttavia, questi comuni generalmente presentano probabilità maggiori di appartenenza a due gruppi e rarefatte per il resto dei gruppi, si può quindi pensare di accettare le partizioni ottenute con il K medie.

<sup>16</sup> Per ottenere la ripartizione *fuzzy* si è utilizzato il pacchetto `e1071` disponibile per il software R fissando 5 come numero di gruppi e utilizzando il metodo di partizione *fuzzy C-Means* introdotto da Bezdek (1981).

Per validare ulteriormente i risultati del *clustering K* medie, si è proceduto applicando un test ANOVA sulle quotazioni. Per non prendere in esame le quotazioni di un singolo periodo si è calcolata la media delle quotazioni di ciascun comune dal I semestre 2004 al I semestre 2011 che si indica genericamente con  $Q_i$  e su tale valore si è applicato il test con l'obiettivo di verificare il sistema di ipotesi che vede sotto la nulla  $H_0$  l'uguaglianza delle quotazioni medie in ciascun gruppo  $C_j$  contro l'ipotesi alternativa  $H_1$  che almeno un gruppo presenti quotazioni medie significativamente diverse. Si ipotizza un modello del tipo:

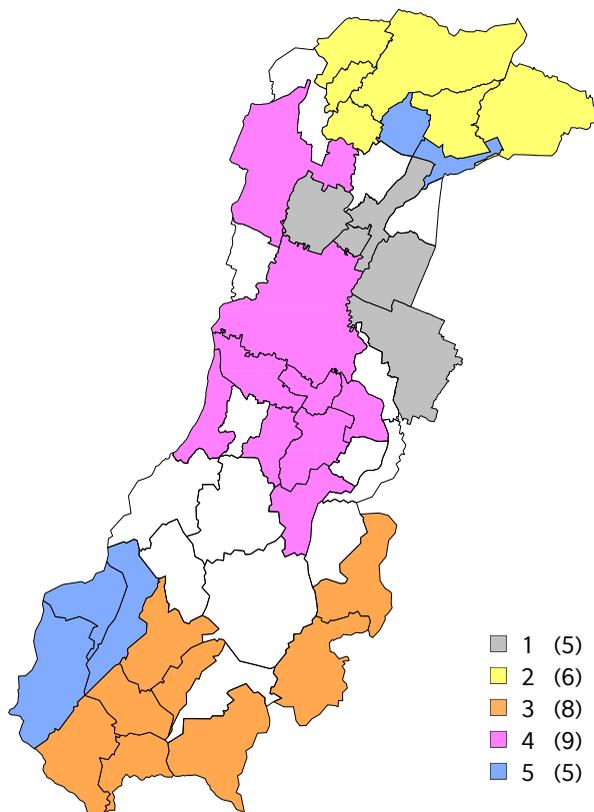
$$Q_{ij} = C_j + \mu_{ij}$$

dove  $Q_{ij}$  indica la quotazione media dell' $i$ -esimo comune, che è risultato appartenere al  $j$ -esimo *cluster*,  $C_j$  è la quotazione media del *cluster*  $j$ -esimo e infine  $\mu_{ij}$  sono gli errori per ipotesi normalmente distribuiti a media nulla e varianza costante  $\sigma^2$ . Si ottiene un valore della statistica  $F$  pari a 12,54 che conduce ad un netto rifiuto di  $H_0$  e permette di sostenere che le quotazioni medie di ciascun *cluster* non sono uguali.

**Tabella 2** Probabilità di appartenere a ciascun gruppo procedura *fuzzy clustering*

Comune	Probabilità di appartenenza $P > 0,6$	Comune	Probabilità di appartenenza $0,4 \leq P < 0,6$	Comune	Probabilità di appartenenza $P < 0,4$
Castelfranco Emilia		Modena		Campogalliano	
Cavezzo		Bastiglia		Fiorano Modenese	
Finale Emilia		Bomporto		Guiglia	
Fiumalbo		Camposanto		Serramazzoni	
Formigine		Carpi		Novi di Modena	
Montefiorino		Castelnuovo Rangone		Pavullo nel Frignano	
Montese		Castelvetro di Modena		Polinago	
Nonantola		Concordia sulla Secchia		Prignano sulla Secchia	
Riolunato		Fanano		Ravarino	
San Felice sul Panaro		Frassinoro		San Cesareo sul Panaro	
San Possidonio		Lama Mocogno		San Prospero	
Sassuolo		Maranello		Savignano sul Panaro	
Spilamberto		Marano sul Panaro		Sestola	
Zocca		Medolla		Vignola	
		Mirandola			
		Montecreto			
		Palagano			
		Pievepelago			
		Soliera			

Figura 5: Mappa dei cluster ottenuti con il fuzzy clustering



## CONCLUSIONI E SVILUPPI

L'analisi empirica è stata condotta analizzando ampie categorie di fattori attraverso l'impiego di semplici strumenti di statistica descrittiva, di indicatori econometrici spaziali globali e locali e di adeguate tecniche regressive. Tale fase ha portato alla costruzione di un set di variabili che utilizzate congiuntamente per effettuare il raggruppamento di comuni con la tecnica multidimensionale della *cluster* si è rivelata utile per la determinazione di gruppi di comuni che presentino al loro interno un certo livello di omogeneità di quotazioni delle abitazioni. Sembra essere quindi confermata l'idea empirica dell'esistenza dei sottomercati e della loro omogeneità. È possibile, quindi, rilevare le quotazioni di una parte dei comuni del *cluster* identificato e tale operazione è giustificata dall'omogeneità riscontrata attraverso l'analisi delle variabili spaziali, temporali, sociali, economiche e immobiliari. La procedura così come impostata sembra essere in grado di cogliere l'effetto combinato di fattori spaziali legati agli aspetti geografici territoriali e strutturali del mercato immobiliare residenziale. Tuttavia, come visto, l'applicazione della tecnica *fuzzy* ha evidenziato l'esistenza di un problema di *overlapping* ovvero della probabilità di alcuni comuni di appartenere a più gruppi. Permangono, quindi, degli elementi che rendono l'analisi non del tutto oggettiva ma influenzata dalle scelte dell'analista. Nonostante questa constatazione, si ritiene che il risultato ottenuto sia un buon punto di partenza per riflessioni successive. Si deve, infatti, tenere in considerazione che i prezzi delle abitazioni oltre a mostrare un'alta variabilità territoriale, nel tempo possono risultare molto diversi a causa, ad esempio, di mutamenti del contesto localizzativo di riferimento. In tal senso anche per i sottomercati, che subiscono influenze reciproche e influenze dagli ambiti territoriali limitrofi, si renderanno

indispensabili valutazioni di verifica della validità. Inoltre, seppur con i suoi limiti quest'esercizio ha analizzato variabili immediatamente disponibili e facilmente riproducibili in altri studi.

In termini di nuovi sviluppi si possono individuare principalmente due ulteriori approfondimenti da perseguire. Una naturale estensione è rappresentata dall'applicazione dello studio ad altri ambiti territoriali per verificare se la metodologia adottata in questa indagine pilota resta valida anche in altre porzioni territoriali. In particolare si potrà provare a superare il limite, evidenziato, dei confini amministrativi delle provincie ed effettuare un'applicazione su porzioni territoriali più estese e quindi più complesse. Di notevole interesse può rivelarsi anche lo studio delle differenze nella determinazione dei prezzi tra i diversi sottomercati individuati.

A questo livello di analisi si può comunque concludere che i risultati ottenuti, utilizzati in combinazione con conoscenze specifiche degli esperti possono essere di utilità per migliorare la conoscenza delle connessioni tra i sottomercati immobiliari di una data ripartizione territoriale e quindi per contribuire a migliorare la qualità dell'inferenza nei comuni non direttamente rilevati.

#### Bibliografia

- Abraham, J., Goetzmann, W. e Watcher, S. (1994). Homogeneous Groupings of Metropolitan Housing Markets. *Journal of Housing Economics* 3, 186-206
- Adair A. S., Berry J. N. & McGreal W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research* Volume 13, Issue 1
- Agenzia del Territorio (2008). *Manuale della banca dati dell'osservatorio del mercato immobiliare*. (pag. 59). Disponibile sul sito [www.agenziaterritorio.it](http://www.agenziaterritorio.it)
- Agenzia del Territorio (2010). *Statistiche Catastali. Catasto Edilizio Urbano*. Disponibile sul sito [www.agenziaterritorio.it](http://www.agenziaterritorio.it)
- Allen, M. T., Springer, T. M., Waller, N. G. (1995). Implicit pricing across residential rental submarkets. *The Journal of Real Estate Finance and Economics* Springer Netherlands Volume 11, Issue 2
- Anselin, L., 1988 *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht
- Basu, A. e T.G. Thibodeau. Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics*, 1998, 17:1, 61-85
- Bates, L. K. 2006. Does Neighborhood Really Matter? Comparing Historically Defined Neighborhood Boundaries with Housing Submarkets. *Journal of Planning Education and Research*. 26(1): 5-17
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum.
- Bhattacharjee A. e Jensen-Butler C., (2006), "Estimation of Spatial Weights Matrix in a Spatial Error Model, with an Application to Diffusion in Housing Demand," CRIEFF Discussion Papers 0519, Centre for Research into Industry, Enterprise, Finance and the Firm.
- Dipartimento delle Finanze e Agenzia del Territorio (2011). *Gli immobili in Italia. Distribuzione del patrimonio e dei redditi dei proprietari*. Disponibile sul sito [www.agenziaterritorio.it](http://www.agenziaterritorio.it)
- Dole Johnson, 1982. An alternative approach housing market segmentation using hedonic price data. *Journal of urban economics* Volume 11
- C. Fraley and A. E. Raftery (2006). MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report no. 504, Department of Statistics, University of Washington
- A.K. Jain, Murty M.N. and Flynn P.J., Data clustering: a review. *ACM computing surveys (CSUR)*, 1999. 31(3): p. 264-323
- McCluskey, W. J., W. G. Deddis, et al. (2002). *The Application of Spatially Derived Location Factors within a GIS environment*. PRRES conference, Christchurch, New Zealand, PRRES Eighth Annual Conference Christchurch New Zealand
- Megbolugbe Isaac F., Hoek-Smit Marja C., Linneman Peter D. Understanding Neighbourhood Dynamics: A Review of the Contributions of William G. Grigsby. *Urban Studies* December 1996 33: 1779-1795
- MacLennan D., Tu Y., 1996. *Economic perspectives on the structure of local housing systems*. *Housing Studies* 11 387-406
- MacQueen J.B., (1967). *Some Methods for Classification Analysis of Multivariate Observations*. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297.
- Eckert, J., Ed. (1990). *Property Appraisal and Assessment Administration*. International Association of Assessing Officers
- Galster, G. (1996). William Grigsby and the Analysis of Housing Sub-markets and Filtering. *Urban Studies* December 1996 vol. 33 no. 10 1797-1805
- Grigsby, W.G. (1963). *Housing Markets and Public Policy*. Philadelphia: University of Pennsylvania Press
- Goodman A. C. (1981). Housing submarkets within urban areas: definitions and evidence. *Journal of Regional Science* Volume 21, Issue 2, pages 175-185
- Goodman, A.C. and T.G. Thibodeau. (2003). Housing Market Segmentation and Hedonic Prediction Accuracy. *Journal of Housing Economics* 12(3):181-201
- Hartigan, J. A. e Wong, M. A. (1979). A *K-means clustering algorithm*. *Applied Statistics* 28, 100-108
- Hoesli, M., e MacGregor, B. (1995). The Classification of Local Property Markets in the UK Using Cluster Analysis. The Cutting Edge: Proceedings of the RICS Property Research Conference 1995, Volume 1, pp 185-204, RICS, London
- Jackson, E., V. Kupke and P. Rossini (2007). *The relationship between socio-economic indicators and residential property values in Darwin*. Thirteenth Annual Pacific-Rim Real Estate Society Conference, Fremantle, Western Australia
- Kestens, Y., M. Thériault, and F. Des Rosiers. 2006. Heterogeneity in hedonic modeling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, 8(1):61-96
- Pace, R. K., e Gilley, O.W. (1997). Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics* 14: 333-340
- Pryce, G. e Evans, G. (2007) *Identifying Submarkets at the Sub-Regional Level in England*. London: Department of Communities and Local Government
- Rapkin, C., Winnick, L. e Blank, D. (1953). *Housing Market Analysis*. Washington, DC: U.S. Housing and Home Finance Agency
- Rosenthal S.S. e Strange W.C., (2003), "Geography, Industrial Organization, and Agglomeration". *Review of Economics and Statistics*, 85 (2): 377-393
- Smith, L., Rosen, K., Fallis, G. (1988). Recent Developments in Economic Models of Housing Markets", *Journal of Economic Literature*, 26(1), 29-64.
- Watkins, C.A. (2001). *The definition and identification of housing submarkets*. *Environment and Planning A*, 33, 2235-2253
- L. A. Zadeh. Fuzzy sets. *Inf.Control* 8, 1965

**APPENDICE A**

**Dettagli sulle stime del trend**

Nella tabella si riportano i risultati ottenuti dall'applicazione del modello di regressione. In particolare per ogni comune si forniscono le stime dei coefficienti, i valori delle statistiche t e il livello di significatività (dove \*\*\* indica  $\alpha=0,01$ , \*\* indica  $\alpha=0,05$  e \* indica  $\alpha =0,1$ ).

**Tabella 3** Stime dei coefficienti, statistiche t e significatività delle regressioni temporali

Comune	$\beta_0$	$\beta_1$	$\beta_2$	Comune	$\beta_0$	$\beta_1$	$\beta_2$	Comune	$\beta_0$	$\beta_1$	$\beta_2$
Modena	1.430,14 -47,23 ***	62,15 7,14 ***	-2,90 -5,47 ***	Fanano	1.128,26 102,79 ***	14,60 4,83 ***	-0,32 -1,73	Bomporto	1.000,43 35,27 ***	16,41 2,48 **	-0,16 -0,41
Guiglia	1.046,36 44,89 ***	13,06 1,95 *	-0,96 -2,37 **	Fiumalbo	773,97 8,05 ***	60,03 2,42 **	-1,24 -0,83	Castelfranco Emilia	1.122,83 49,35 ***	15,74 2,51 **	0,29 0,75
Marano sul Panaro	1.062,40 54,85 ***	18,98 3,41 ***	-0,82 -2,43 **	Lama Mocogno	753,15 40,18 ***	9,65 1,87 *	0,18 0,59	Nonantola	988,18 32,30 ***	33,19 3,90 ***	-0,65 -1,26
Montese	765,42 53,51 ***	8,29 2,02 *	0,03 0,12	Serramazzoni	762,54 57,26 ***	22,92 7,32 ***	-0,79 -4,66 ***	Ravarino	850,52 76,19 ***	15,19 5,21 ***	-0,55 -3,13 ***
Zocca	740,27 49,70 ***	9,81 2,29 **	-0,06 -0,21	Montecreto	739,58 36,39 ***	10,71 1,83 *	0,30 0,83	San Cesario sul Panaro	799,86 52,40 ***	9,59 2,33 **	0,16 0,62
Frassinoro	613,92 33,53 ***	7,90 1,50	0,17 0,52	Pavullo nel Frignano	939,55 49,69 ***	16,87 3,15 ***	-0,04 -0,12	Campogalliano	914,44 100,49 ***	20,98 9,58 ***	-0,78 -5,96 ***
Montefiorino	848,23 78,05 ***	11,69 3,90 ***	-0,55 -3,02 **	Pievepelago	849,65 27,29 ***	14,09 1,60	0,31 0,59	Carpi	1.241,16 54,31 ***	40,59 6,90 ***	-1,93 -5,43 ***
Palagano	530,97 14,73 ***	8,10 0,78	0,64 1,01	Riolunato	718,90 47,40 ***	9,56 2,29 **	0,08 0,32	Novi di Modena	692,45 75,21 ***	7,27 2,93 **	0,09 0,58
Polinago	625,33 27,33 ***	8,19 1,30	0,35 0,92	Sestola	1.256,07 32,12 ***	39,09 3,52 ***	-0,62 -0,92	Soliera	913,66 65,57 ***	14,46 3,70 ***	-0,26 -1,08
Prignano sulla Secchia	697,78 31,97 ***	36,69 6,43 ***	-1,77 -5,17 ***	Castelnuovo Rangone	1.486,62 50,04 ***	4,45 0,53	-0,40 -0,77	Finale Emilia	682,18 17,31 ***	36,86 3,58 ***	-1,09 -1,76
Camposanto	845,02 56,33 ***	17,19 4,16 ***	-0,60 -2,41 **	Castelvetro di Modena	1.126,01 47,12 ***	3,59 0,58	0,37 0,98	Medolla	652,82 13,41 ***	8,11 0,60	1,06 1,30
Cavezzo	767,19 61,55 ***	7,39 2,11 *	0,27 1,28	Fiorano Modenese	1.123,80 74,81 ***	17,62 4,33 ***	-0,45 -1,78	Mirandola	712,42 17,68 ***	42,13 3,92 ***	-1,06 -1,62
Concordia sulla Secchia	697,78 31,97 ***	36,69 6,43 ***	-1,77 -5,17 ***	Formigine	1.153,51 44,98 ***	21,91 3,19 ***	-0,89 -2,12 *	San Felice sul Panaro	561,54 16,28 ***	27,51 2,99 **	-0,34 -0,60
Spilamberto	1.057,78 100,97 ***	14,10 4,89 ***	-0,53 -3,04 **	Maranello	1.024,76 36,08 ***	32,85 4,09 ***	-1,36 -2,78 **	San Possidonio	533,35 15,23 ***	22,83 2,36 **	-0,12 -0,20
Vignola	1.257,38 109,05 ***	14,19 4,18 ***	-0,13 -0,61	Sassuolo	1.183,18 102,40 ***	16,63 5,09 ***	-0,41 -2,05 *	San Prospero	742,55 57,36 ***	9,20 2,53	-0,12 -0,20
Bastiglia	1.028,05 103,75 ***	24,86 8,98 ***	-0,72 -4,26 ***	Savignano sul Panaro	908,07 84,64 ***	13,21 4,71 ***	-0,33 -1,96 *				

## APPENDICE B

### Dettagli sull'autocorrelazione spaziale

Nella tabella presentata di seguito si riportano, per ogni semestre, i valori della statistica di I di Moran, il valore atteso, la deviazione *standard* e il p-value. In particolare l'inferenza sulla statistica I di Moran è ottenuta utilizzando tecniche non parametriche basate sulle permutazioni delle osservazioni.

**Tabella 4** Risultati dell'indice I-Moran

Semestre	I - Moran	E (I)	Sd (I)	p-value	
I 2001	0,42	-0,02	0,07	0,01	**
II 2001	0,45	-0,02	0,08	0,01	***
I 2002	0,39	-0,02	0,09	0,01	**
II 2002	0,39	-0,02	0,09	0,01	**
I 2003	0,39	-0,02	0,08	0,01	**
II 2003	0,40	-0,02	0,08	0,01	**
I 2004	0,42	-0,02	0,08	0,01	**
II 2004	0,41	-0,02	0,08	0,01	**
I 2005	0,40	-0,02	0,08	0,01	**
II 2005	0,39	-0,02	0,08	0,01	**
I 2006	0,39	-0,02	0,09	0,01	**
II 2006	0,38	-0,02	0,08	0,01	**
I 2007	0,36	-0,02	0,08	0,01	**
II 2007	0,31	-0,02	0,08	0,01	**
I 2008	0,31	-0,02	0,09	0,01	**
II 2008	0,25	-0,02	0,08	0,01	**
I 2009	0,25	-0,02	0,08	0,01	**
II 2009	0,27	-0,02	0,08	0,01	**
I 2010	0,27	-0,02	0,09	0,02	**
II 2010	0,27	-0,02	0,08	0,01	**
I 2011	0,27	-0,02	0,08	0,01	**
II 2011	0,30	-0,02	0,09	0,01	**

Nelle mappe che seguono si riportano i risultati dell'indicatore spaziale LISA applicati alla variabile quotazione Q di ogni semestre. Il colore rosso identifica i comuni per i quali il LISA restituisce una significativa correlazione positiva (H - H), il blu una significativa correlazione negativa (L - L), infine ai colori viola e rosa sono associati valori anomali che possono indicare una significativa correlazione negativa accanto a valori positivi (L - H) e una significativa correlazione positiva accanto a valori negativi (H - L).

Figura 6 Mappe del Local Indicator Spatial Association (LISA)

