

\*Erika Ghirardo  
 \*Raffaele Ciula  
 \*\*Maurizio Festa

## Prezzi delle abitazioni: un'analisi della dinamica evolutiva sul mercato italiano

**Parole chiave:** prezzi delle abitazioni, co-movimenti, convergenza, dati panel, componenti principali.

**Abstract** L'articolo propone uno studio dell'evoluzione dei prezzi delle abitazioni in Italia, sia per il breve che per il lungo periodo, considerando le serie storiche delle quotazioni a livello regionale. Per il breve periodo si indagano "co-movimenti" delle serie regionali e si analizzano le componenti del tasso di apprezzamento dei prezzi tentando di cogliere l'effetto nazionale, i fattori locali nonché l'esistenza di persistenze. D'altro lato, per il lungo periodo si propone una riflessione sull'esistenza di un processo di "convergenza". Per entrambi gli orizzonti temporali, l'analisi empirica è condotta applicando la tecnica delle componenti principali e stimando una relazione per i tassi di apprezzamento dei prezzi delle abitazioni a un *panel* di osservazioni rese disponibili dall'Osservatorio del Mercato Immobiliare dell'Agenzia delle Entrate. A questo livello di analisi, i risultati ottenuti mostrano l'esistenza di un *trend* comune tra le serie dei prezzi seppure attenuato, per molte regioni, da effetti specifici locali. Infine, l'evidenza di convergenza risulta debole e non pienamente confermata dai test effettuati.

### INTRODUZIONE

Negli ultimi tempi è maturata la consapevolezza che l'evoluzione dei prezzi delle abitazioni ha un forte impatto su molti aspetti macroeconomici dell'economia e in tal senso è di estremo interesse per famiglie, responsabili politici e per tutti coloro che sono coinvolti nel mercato immobiliare. La recente crisi economica, che trae origine dal settore immobiliare, ha dimostrato che una quota rilevante del rischio finanziario è legata a complessi strumenti finanziari il cui scopo è quello della tutela del rischio. Il mercato immobiliare è in grado di innescare crisi sistemiche che generano effetti nel settore reale; coglierne le tendenze è quindi importante per tentare di comprendere in anticipo orientamenti generali dell'economia.

La conoscenza della dinamica dei prezzi delle abitazioni è quindi di fondamentale importanza per valutare adeguatamente, evitando di sovra o sottostimare, la direzione della serie. Inoltre, va tenuto in considerazione che l'evoluzione dei prezzi è influenzata da movimenti sotto-territoriali che ne indirizzano il *trend*. I prezzi delle abitazioni mostrano, infatti, più di qualsiasi altro bene, differenze locali legate soprattutto all'estrema varietà di caratteristiche che si riscontrano nel territorio. Lo studio dell'evoluzione delle serie territoriali congiuntamente considerate ovvero l'approfondimento dei loro "co-movimenti" permette di comprendere come le serie rispondono a eventuali *shock* che si determinano nel sistema.

\*Funzionario Ufficio Statistiche e Studi Mercato Immobiliare – Agenzia delle Entrate

\*\*Responsabile Ufficio Statistiche e Studi Mercato Immobiliare – Agenzia delle Entrate

Partendo da queste premesse, in questo articolo si propone un'analisi delle serie storiche delle quotazioni delle abitazioni per il periodo dal 2001 al 2011 nelle regioni italiane.

L'obiettivo principale è la verifica dell'esistenza di un fattore comune che, "guidando" le serie storiche delle quotazioni regionali verso una direzione comune, possa tendere a creare un loro avvicinamento. Nel dettaglio l'analisi è distinta sotto l'aspetto temporale separando il breve e il lungo periodo. L'analisi di breve periodo si focalizza sullo studio dei "co-movimenti" delle serie cercando di determinare le componenti della variazione dei prezzi delle abitazioni e le relazioni tra gli andamenti di tutte le serie. Per il lungo periodo si accenna a un percorso di riflessione tentando di indagare un'ipotesi di convergenza.

Per entrambe le analisi si applicano essenzialmente due tecniche: si stima un modello *panel* e si validano i risultati con un'ulteriore applicazione utilizzando la tecnica delle componenti principali. La regressione, applicata al *panel* di dati, segue una proposta di analisi apparsa nell'articolo Gyourko & Voith del 1992 che introduce una relazione che scinde la variazione dei prezzi delle abitazioni in tre componenti: un effetto globale temporale, un parametro locale e infine un coefficiente di persistenza. Traendo spunto da numerose applicazioni in ambito finanziario, per il breve periodo, e parte del contributo di Holmes & Grimes (2008) per l'analisi del lungo periodo, la struttura della correlazione tra le serie storiche è approfondita utilizzando la tecnica delle componenti principali.

Il resto del contributo è organizzato come segue. Nel secondo paragrafo viene descritta la letteratura inerente al tema della convergenza dei prezzi delle abitazioni, nel terzo paragrafo si presenta il modello di *panel data* utilizzato per stimare la dinamica dei prezzi. Le stime econometriche del *panel* e dell'approccio alle componenti principali sono discusse nel quarto e quinto paragrafo. L'ipotesi di convergenza nel lungo periodo è discussa nel sesto paragrafo; infine, l'ultimo paragrafo è dedicato alla descrizione dei principali risultati, delle problematiche e delle future linee di ricerca inerenti al tema in discussione.

## RICHIAMI DI LETTERATURA

Il concetto di convergenza è molto noto nel filone della letteratura che si occupa dei modelli di crescita economica che prende avvio dalle teorie neoclassiche (Solow, 1956 o Romer, 1986) che mirano ad indagare sotto vari aspetti la crescita del reddito *pro-capite* attraverso l'esame dei fattori che la determinano.

Anche nel mercato immobiliare l'andamento dinamico dei prezzi delle abitazioni nei diversi ambiti territoriali subnazionali e la loro possibile convergenza costituiscono da molto tempo un tema ampiamente dibattuto. In particolare, in letteratura, si possono riscontrare essenzialmente due tipi di convergenza. Il primo concetto si riferisce alla *time convergence* o *ripple effect* (discusso soprattutto in Meen, 1999 e Holmes e Grimes, 2008) in cui la convergenza è definita come la tendenza dei prezzi a rispondere in modo analogo a uno *shock*. In tal senso si ipotizza la presenza di un fattore latente che in un certo lasso di tempo porta i prezzi relativi delle abitazioni a convergere verso una situazione di stabilità. Più in particolare il *ripple effect* o effetto onda rappresenta la trasmissione degli *shock* dei prezzi regionali in tutte le regioni causando, in tal modo, un loro co-movimento nel lungo periodo. Di solito questo effetto viene attribuito a vari fattori tra i quali l'immigrazione delle famiglie, l'arbitraggio spaziale, e ancora la presenza di *pattern* spaziali analoghi nelle determinanti dei prezzi delle abitazioni. Seguendo questa definizione di convergenza non si assume che i prezzi tra le varie aree indagate siano uguali ma si ipotizza che una variazione, o più in generale un cambiamento che si verifica in una o più di queste aree, si diffonda in modo molto simile anche nelle altre aree. Ad esempio nell'articolo di Zhang (2010) viene analizzata la possibilità di un *ripple effect* nei prezzi regionali nel mercato delle abitazioni del Regno Unito utilizzando lo strumento della cointegrazione frazionale. Tale metodo sfrutta il fatto che le variabili in questione (il rapporto dei prezzi di ciascuna regione) siano non stazionarie ed abbiano un

*trend* comune che le tenga unite. I risultati mostrano che con alcune eccezioni le regioni hanno una dipendenza di lungo periodo che tende a farle convergere rispetto al prezzo nazionale.

L'altra nozione di convergenza è quella che si può denominare *cross-sectional convergence*, e ipotizza che le aree in cui si osservano, a un periodo iniziale, bassi valori di prezzo tendono ad avere tassi di crescita più elevati rispetto alle regioni che presentano prezzi iniziali maggiori, così da creare una convergenza nel lungo periodo. In tal senso in letteratura vi sono articoli che indagano i processi di convergenza tra aree metropolitane di uno stesso stato o tra nazioni diverse. A questo concetto appartiene la nozione di convergenza relativa "*t-convergence*" introdotta da Philips e Sul (2007) che ipotizza una diminuzione della dispersione dei prezzi delle abitazioni nel tempo. Questa definizione di convergenza concerne l'andamento dei prezzi delle abitazioni tra ambiti territoriali nel tempo e viene ripresa nell'articolo di Kim e Rous (2012), i quali cercano di testare l'ipotesi di convergenza in un *panel* di 48 stati americani. Avendo riscontrato una mancanza della "*t-convergence*" tra tutti gli stati, gli autori verificano invece una forte convergenza di questo tipo tra sottoinsiemi di stati.

Dal punto di vista empirico le tecniche maggiormente utilizzate per la verifica della convergenza fanno riferimento all'analisi di cointegrazione tra le serie storiche. La *fractional cointegration* è utilizzata nell'articolo di Barros *et al.* (2012) per verificare il rapporto tra i prezzi delle abitazioni tra gli stati negli U.S.A. Il contributo di Gupta e Miller (2012) ha come scopo quello di analizzare l'ipotesi di convergenza fra aree metropolitane nel sud della California attraverso il test di cointegrazione tra gli indici di prezzo delle abitazioni. Inoltre viene applicato il test di causalità di Granger per testare l'ipotesi di causalità statistica temporale tra le varie aree metropolitane.

Per verificare l'ipotesi di convergenza nel mercato delle abitazioni in Gran Bretagna tra i prezzi delle varie regioni, Holmes (2007) utilizza un modello *seemingly unrelated regression estimation* S.U.R.E. Ampiamente utilizzati sono anche i modelli per dati *panel*. Un esempio è costituito dall'articolo di Hiebert e Roma (2010) che si prefiggono di analizzare la convergenza dei prezzi delle abitazioni tra le città di alcune nazioni europee. Un altro articolo da cui questo elaborato trae maggiormente spunto è quello di Gyourko e Voith (1992), in cui è presentato un modello statistico per la scomposizione dei tassi di crescita dei prezzi delle abitazioni in tre componenti: una componente aggregata globale, una variazione locale e una parte legata alle singole serie. Gli autori effettuano un'analisi dei tassi di variazione rilevati nei prezzi delle abitazioni in 56 aree metropolitane degli Stati Uniti nel periodo dal 1971 al 1989 con l'obiettivo di testare l'influenza della componente nazionale quanto di specifici fattori locali legati alle amenità proprie di ciascun territorio. In particolare, Gyourko e Voith evidenziano l'esistenza di una significativa componente nazionale comune nei tassi di apprezzamento dei prezzi di lungo periodo in tutte le aree metropolitane.

L'utilizzo congiunto di metodi di cointegrazione e tecniche di analisi multivariata come l'approccio in componenti principali è impiegato nell'articolo di Holmes & Grimes (2008). Per verificare l'ipotesi di convergenza dei prezzi nel lungo periodo tra le regioni inglesi gli autori utilizzano le componenti principali, applicate al rapporto tra ciascun prezzo regionale ed il valore medio nazionale, e testano la stazionarietà della prima componente. I risultati suggeriscono la tendenza dei prezzi a convergere verso un equilibrio di lungo periodo. Uno *shock* sui prezzi delle abitazioni che si propagasse da una qualsiasi regione darebbe luogo ad un effetto onda (*ripple effect*) che avrebbe lo stesso effetto su tutti i prezzi regionali.

## MODELLO EMPIRICO

Il modello proposto da Gyourko e Voith (1992) ipotizza, in via teorica, un legame tra il tasso di variazione dei prezzi delle abitazioni, una componente nazionale comune a tutte le serie territoriali, un fattore specifico che coglie gli aspetti peculiari di ciascuna area territoriale, e infine alcuni parametri che colgono la persistenza delle singole serie.

Sulla base di questo modello e di uno sviluppo successivo (MacLean, 1994) per le variazioni dei prezzi delle regioni si assume il modello seguente:

$$100 * \ln(P_{it}/P_{it-1}) = B_t T_t + A_i + C_i * 100 * \ln(P_{it-1}) + E_{it} \quad [1]$$

in cui:

- $P_{it}$  è la variabile prezzo delle abitazioni. In particolare si tratta del prezzo deflazionato della regione  $i$  al tempo  $t$ , calcolato come media dei prezzi dei soli capoluoghi di regione come sarà meglio descritto;
- $T_t$  è un indice temporale del semestre di riferimento per ogni tempo  $t$ .

All'interno del modello sono stimati i coefficienti  $B_t$ ,  $A_i$ ,  $C_i$ . I coefficienti  $B_t$  legati alla variabile  $T_t$  forniscono indicazioni sui movimenti comuni a tutte le serie delle variazioni dei prezzi. In termini economici possono essere, quindi, interpretati come variazione nazionale. I coefficienti  $A_i$  sono effetti fissi di ciascuna regione. Infine i  $C_i$  essendo legati al ritardo della variabile prezzo vanno intesi come persistenze e rappresentano un modo per verificare se il livello dei prezzi diverge o converge nel tempo. A completamento del modello e al fine di effettuare i successivi test si ipotizza che il termine di errore  $E_{it}$  sia indipendente ed identicamente distribuito (IID) con media nulla e varianza costante pari a  $\sigma^2$  ovvero  $E_{it} \sim \text{IID}(0, \sigma^2)$ .

Ognuna delle tre componenti della relazione ha quindi la funzione di cogliere uno specifico effetto della variazione dei prezzi. La componente nazionale rappresentata dai coefficienti  $B_t$  è pensata per individuare un effetto comune a tutte le serie che in qualche modo le “guida” tutte insieme nella stessa direzione. Verificare l'esistenza significativa di questo *trend* nazionale significherebbe osservare variazioni nazionali ripercuotersi anche nelle serie regionali facendo in modo che i movimenti delle stesse siano analoghi. Il vettore  $B_t$  cattura quindi le differenze tra le variazioni dei prezzi nel tempo. E' tuttavia ovvio che le serie possono mostrare deviazioni dal *trend* nazionale in considerazione di specificità che caratterizzano ciascuna regione. Nel modello formulato tali effetti sono tenuti in considerazione introducendo la componente  $A_i$  che coglie solo le differenze tra le regioni ignorando la parte temporale.

I coefficienti  $C_i$  intendono, invece, cogliere la parte della variazione imputabile al fatto che per ciascuna regione si dispone di una serie storica e quindi le osservazioni possono esibire una dipendenza dalle osservazioni relative a periodi precedenti.

La verifica della significatività delle tre componenti è effettuata attraverso l'applicazione di una serie di test di significatività congiunta di sottoinsiemi di coefficienti del modello stimato. Per questa verifica si utilizza l'usuale test  $F$  che permette di confrontare la somma dei quadrati dei residui nel modello completo con la stessa quantità nel modello vincolato ovvero quello stimato omettendo i regressori su cui si impone la restrizione.

La presenza di una componente nazionale comune a tutte le serie va testata verificando che tutti i coefficienti  $B_t$  siano congiuntamente non nulli attraverso il sistema di ipotesi che pone sotto l'ipotesi nulla  $H_0: B_t = 0$  per ogni periodo  $t$ . Il rifiuto dell'ipotesi nulla implica che esiste una variazione globale che influenza tutte le serie. Al contrario l'impossibilità di rifiutare  $H_0$  sarebbe un'indicazione dell'assenza di una componente di variazione globale legata al tempo e che le variazioni dei prezzi delle abitazioni possono essere imputate solo a fattori locali.

La significatività della componente locale va testata verificando che tutti i coefficienti  $A_i$  siano congiuntamente non nulli. Se  $A_i = 0$  per tutte le  $i$ , allora le regioni non mostrano deviazioni dall'andamento nazionale o comunque le specificità locali non sono così influenti da riflettersi in una particolare dinamica. Se l'ipotesi nulla è rifiutata, invece, le caratteristiche locali sono significative e possono causare deviazioni dall'andamento nazionale.

La significatività delle differenze tra i tassi di apprezzamento delle regioni va verificata conducendo un test che sotto l'ipotesi nulla prevede l'uguaglianza di tutti i coefficienti  $A_i$  ovvero  $H_0: A_i = A_j$  per tutte le  $i \neq j$ . Nel caso di non rifiuto della nulla le differenze tra gli effetti locali non possono essere ritenute significative, in altri termini i tassi di apprezzamento dei prezzi delle abitazioni tra le regioni sono uguali.

Il test  $H_0: C_i = 0$  per tutte le regioni fornisce indicazioni circa la significatività degli *shock*. Se i coefficienti  $C_i$  sono congiuntamente non significativi allora le serie non mostrano persistenza e deviazioni dal *trend* generale dovute a *shock* casuali hanno effetto in un solo periodo senza ripercussioni nei periodi successivi. Se al contrario i coefficienti sono significativamente diversi da zero allora le serie tendono ad "avere memoria" del proprio passato in quanto gli scostamenti dall'andamento globale non esauriscono i loro effetti in un solo periodo.

Infine, testando l'ipotesi di uguaglianza dei coefficienti  $C_i$ , ovvero  $H_0: C_i = C_j$  per tutte le  $i \neq j$ , si verifica l'uguaglianza delle velocità di aggiustamento tra i tassi di apprezzamento rilevati nelle regioni.

Riassumendo, quindi, si conducono i seguenti test:

- $H_0: B_t = 0$  per ogni  $t$ ,  $H_1$ : altrimenti;
- $H_0: A_i = 0$  per ogni  $i$ ,  $H_1$ : altrimenti;
- $H_0: A_i = A_j$  per ogni  $i \neq j$ ,  $H_1$ : altrimenti;
- $H_0: C_i = 0$  per ogni  $i$ ,  $H_1$ : altrimenti;
- $H_0: C_i = C_j$  per ogni  $i \neq j$ ,  $H_1$ : altrimenti.

In ogni periodo ci si attende che i prezzi delle regioni differiscano tra loro in considerazione dei diversi vantaggi che si hanno nell'abitare in una località piuttosto che in un'altra. Se si ipotizza che i miglioramenti delle regioni, che presentano ad un periodo iniziale vantaggi maggiori rispetto ad altre, si stabilizzino ad una certa soglia e che le regioni in svantaggio continuino ad adottare misure per ridurli, le differenze osservate nel periodo iniziale dovrebbero attenuarsi nel tempo. Poiché in generale maggiori vantaggi, si riflettono anche in livelli dei prezzi più elevati, si dovrebbero osservare variazioni dei prezzi via via minori per le regioni in vantaggio che potrebbero così essere recuperate dalle regioni in svantaggio. Se questo si verifica allora i segni dei coefficienti del vettore  $C_i$  dovrebbero essere negativi. In tal caso, la serie dei prezzi di ciascuna regione tende a tornare verso il proprio livello medio delle quotazioni. In caso contrario se i coefficienti  $C_i$  risultano maggiori di zero: ad aumenti dei prezzi seguono ulteriori aumenti, ovvero i tassi di variazione risulterebbero crescenti nel tempo o in modo analogo a diminuzioni seguirebbero ulteriori decrementi a tassi di decrescita crescenti nel tempo.

### La stima del modello panel

Come visto nel precedente paragrafo la specificazione del modello, suggerita dalla relazione teorica formulata da Gyourgo e Voith, richiede osservazioni bidimensionali ovvero che variano sia per unità sia nel tempo. A tal fine, partendo dai dati resi disponibili dall'Osservatorio del Mercato Immobiliare (OMI) dell'Agenzia delle Entrate, si sono costruite osservazioni in serie storica dei prezzi medi semestrali delle abitazioni riferiti a 19 regioni italiane.

L'OMI rileva semestralmente prezzi di compravendita, offerte o valori riferiti alle zone OMI<sup>1</sup> di ogni comune. Partendo da tali valori si è calcolata per i soli capoluoghi la quotazione media semestrale come media delle quotazioni relative a tutte le tipologie immobiliari residenziali. Le medie così ottenute sono state utilizzate per elaborare per ciascuna regione la media pesata delle quotazioni dei soli comuni capoluogo, utilizzando come pesi i relativi *stock* immobiliari delle abitazioni<sup>2</sup>. Tale opera-

<sup>1</sup> Per un approfondimento sulla formazione delle zone omogenee e sulla banca dati dell'Osservatorio si può consultare il "Manuale Operativo della Banca Dati dell'Osservatorio del Mercato Immobiliare, 2008".

<sup>2</sup> Il numero delle abitazioni presenti in ciascun comune è desunto dagli archivi del Catasto italiano.

zione è stata condotta per ciascun semestre a partire dal I semestre 2001 fino al II semestre 2011 per ottenere il *panel* composto da  $N = 19$  serie storiche regionali,<sup>3</sup> ciascuna costituita dalle quotazioni medie osservate per 22 semestri (I semestre 2001 - II semestre 2011). Ogni serie storica è stata deflazionata utilizzando l'indice regionale dei prezzi al consumo per l'intera collettività calcolato a livello generale comprensivo di tabacchi pubblicato dall'ISTAT. L'andamento della serie di ciascuna regione è visibile in Figura 1 che mostra prezzi caratterizzati in generale da un *trend* crescente fino al 2008 a cui segue una fase di calo. L'unica eccezione è rappresentata dal Friuli-Venezia Giulia che presenta un andamento decrescente per quasi tutto il periodo.

La variabile dipendente, come previsto dal modello, è stata costruita prendendo il logaritmo naturale del rapporto tra il prezzo al semestre  $t$  e il prezzo al semestre precedente  $t-1$  perdendo così l'osservazione al semestre iniziale (I 2001).

Per poter procedere con la stima del modello riportato nell'equazione [1] si costruiscono una serie di variabili *dummy*:

- per la stima dei coefficienti  $B_t$  si costruiscono  $T-1$  variabili *dummy* con  $T$  numero dei semestri<sup>4</sup>, che, nelle successive tabelle dei risultati, si indicano con "d\_I\_2002" fino a "d\_II\_2011";
- per la stima dei coefficienti  $A_i$  si costruiscono  $N$  variabili *dummy* che si indicano con "d\_Abr" fino a "d\_Ven";
- per stimare i coefficienti  $C_i$  si costruiscono  $N$  variabili con il ritardo della variabile prezzo relativo alla regione  $i$  al tempo  $t-1$  e si indicano con "Pt-1\_Abr" fino a "Pt-1\_Ven".

In tal modo è possibile effettuare la stima attraverso un metodo OLS sul *pooling* di dati sezionali e serie storiche.

Dalle stime riportate in appendice (Tabella 1) si nota che gli effetti fissi sono tutti statisticamente significativi ad eccezione di quelli relativi all'Abruzzo, alla Sardegna e al Friuli Venezia-Giulia. Inoltre tali coefficienti risultano molto elevati in Valle D'Aosta, Lombardia, Umbria e Molise. Per quanto riguarda i coefficienti  $C_i$  si evidenziano risultati analoghi sulla significatività. Per quanto concerne il segno, i  $C_i$  sono conformi alle aspettative e risultano tutti negativi. La componente temporale è statisticamente significativa per quasi tutti i semestri con segno sempre positivo ad eccezione del primo semestre del 2002. Questi risultati suggeriscono un impatto positivo sui tassi di crescita dei prezzi reali regionali per l'Italia.

Interessante risulta anche l'analisi della grandezza dei coefficienti  $A_i$  e  $C_i$  visualizzata nel grafico a barre di Figura 2 che mostra il valore di questi ultimi per ciascuna regione. Si nota che al crescere di  $A_i$  aumenta con segno opposto anche  $C_i$  quasi ad indicare che ad una maggiore specificità corrisponde anche una maggiore persistenza della serie. In qualche modo si può supporre che un elevato effetto fisso crei un maggiore effetto di persistenza, ovvero si accentui il processo che porta ciascuna serie a tornare verso la propria media, con la conseguenza di deviazioni accentuate rispetto alla componente di *trend* temporale globale, rappresentata dai coefficienti  $B_t$ .

### Risultati dei test

Come visto, dopo aver ottenuto le stime dei parametri è necessario condurre dei test sui coefficienti. Tale fase è infatti estremamente utile per approfondire l'analisi e trarre delle prime conclusioni sui risultati ottenuti.

Confrontando il modello completo con un modello che impone la nullità dei coefficienti  $B_t$  attraverso un test  $F$  si ottiene un *p-value* pressoché nullo che permette di concludere che esiste un effetto glo-

<sup>3</sup> Appartengono al *panel* tutte le regioni ad eccezione del Trentino Alto Adige per il quale non sono disponibili i dati relativi allo *stock*.

<sup>4</sup> Per quanto riguarda le *dummy*  $T_t$  il I semestre è stato perso per costruire la variabile dipendente.

bale temporale significativo. Tutte le serie delle variazioni dei prezzi regionali sembrano quindi subire l'influenza di un effetto nazionale legato all'evoluzione temporale. In un certo senso l'andamento nazionale trascina i prezzi di tutte le regioni innescando l'effetto "onda". Da tale risultato si può dedurre che la significatività di un effetto che influenza tutte le serie potrebbe contribuire a fare in modo che le differenze tra le quotazioni regionali si mantengono inalterate nel tempo. Tuttavia, non si è in grado di comprendere, almeno in questa fase dell'analisi, quali serie subiscono in misura maggiore tale effetto e addirittura se vi siano delle serie che non risentono in alcun modo dell'effetto.

In un passo successivo, il test sugli effetti fissi, che evidenzia le tipicità legate alle singole regioni, mostra come i coefficienti  $A_i$  siano significativamente diversi da zero e quindi non trascurabili nella spiegazione delle variazioni dei prezzi. In modo meno formale, una lettura congiunta dei risultati dei due test ora discussi permette di sostenere che se si osserva un certo tasso di apprezzamento dei prezzi a livello nazionale è difficile osservare tassi molto dissimili tra le regioni. Tuttavia, l'effetto del *trend* nazionale riassunto dai  $B_t$  è indebolito dall'importanza degli effetti specifici  $A_i$  che possono causare differenze legate a fattori che riassumono le caratteristiche proprie di ciascuna regione. Le peculiarità regionali possono essere più o meno marcate e manifestarsi in modo diverso, si può ragionevolmente sostenere che le regioni con effetti fissi più elevati e quindi con  $C_i$  più elevati mostrino una maggiore tendenza ad allontanarsi dall'andamento nazionale.

Per indagare con maggiore grado di dettaglio le similitudini e le differenze tra gli effetti fissi di ciascuna regione si sono condotti tutti i test a coppia<sup>5</sup> (*pairwise*) tra i coefficienti  $A_i$ . Nelle mappe di Figura 3 riportate in appendice sono evidenziati i risultati di tutti i confronti. La maggior parte delle regioni non mostra differenze significative con un numero elevato di regioni ma un'indagine più approfondita evidenzia che ciascuna regione mostra differenze significative con almeno cinque regioni. Merita, inoltre, particolare attenzione il fatto che esiste un gruppo di regioni formato da Lombardia, Valle d'Aosta, Umbria e Molise che presenta contemporaneamente effetti specifici non significativamente diversi tra loro ma differenze marcate con la maggior parte delle altre regioni. Si tratta di quelle regioni per le quali i coefficienti stimati  $A$  e  $C$  sono risultati i più elevati. Sarebbe quindi che tali regioni abbiano effetti specifici più influenti confermando quanto già emerso dall'analisi grafica degli effetti fissi e delle persistenze.

In un esercizio di questo tipo non si può trascurare la dimensione "temporale" che ciascuna serie storica regionale rappresenta. Il test congiunto sui coefficienti della persistenza evidenzia che per le serie, globalmente considerate, tale effetto è significativamente non nullo. Le serie hanno memoria del proprio passato. Come per gli effetti fissi anche i parametri  $C_i$  sono più evidenti per alcune regioni e meno per altre. Effettuando, anche in questo caso, tutti i possibili test a coppie si evidenziano analoghe differenze non trascurabili. Anche da questa analisi emerge la presenza del gruppo di regioni formato da Lombardia, Valle d'Aosta, Umbria e Molise rafforzando l'idea che regioni con forti effetti fissi abbiano serie storiche dei prezzi maggiormente dipendenti dal proprio passato. Ricordando che il segno del coefficiente è negativo, ciò può significare che gli *shock* esterni hanno un effetto di minore intensità sulla serie dei prezzi che fa in modo che l'aggiustamento delle variazioni verso la media della serie stessa sia più veloce. In altri termini, queste serie esibiscono andamenti diversi dal *trend* generale e sono invece più legate al proprio *pattern*.

## LA STIMA DELLE COMPONENTI PRINCIPALI

Se consideriamo i dati disponibili come una serie storica piuttosto che come un *panel*, la letteratura, in particolare quella che analizza serie storiche di tipo finanziario, suggerisce un approfondimento

<sup>5</sup> Per questo sistema di ipotesi si utilizza la statistica test *t*.

dell'analisi della struttura di covarianza tra le serie, attraverso l'utilizzo della tecnica delle componenti principali (PCA) tipica dell'analisi multidimensionale (per alcuni esempi si possono vedere ad esempio i lavori di Curto *et al.*, 2006 o Feeney & Hester, 1964). In tal modo si trascura l'autocorrelazione tra le osservazioni di ciascuna serie e si considera solo la correlazione con le serie.

L'analisi delle componenti principali<sup>6</sup> ha l'obiettivo di ridurre un numero ampio di variabili, che rappresentano il fenomeno analizzato, in un numero minore di variabili latenti ottenute come combinazione lineare di quelle originarie. La trasformazione avviene in modo tale che la prima componente spieghi la quota maggiore di varianza. Per costruzione le componenti principali sono fra loro incorrelate, sono ordinate in modo decrescente in riferimento alla varianza ed inoltre si può dimostrare che la varianza totale (somma delle varianze) si conserva nel passaggio dalle variabili osservate alle componenti principali. Un attributo peculiare di queste ultime riguarda il fatto che esse non corrispondono a caratteristiche direttamente osservabili ma vanno interpretate.

Poiché non sempre è adeguato utilizzare questo tipo di metodo, prima di procedere con l'analisi in componenti principali si verifica la fattibilità e la convenienza del metodo rispetto ai dati attraverso lo studio della matrice delle correlazioni semplici e l'applicazione dei test Kaiser-Meyer-Olkin (KMO<sup>7</sup>) e di sfericità di Bartlett. In tutti e tre i casi si sono ottenuti risultati che confermano l'esistenza di una struttura di correlazione tale da consentire l'applicazione della PCA. In particolare:

- la matrice di correlazione ha evidenziato coefficienti di correlazione molto elevati e significativi per tutte le serie con le uniche eccezioni del Friuli-Venezia Giulia e della Valle d'Aosta. Il Friuli Venezia Giulia è l'unica regione a mostrare coefficienti di correlazione negativi con tutte le regioni ed in particolare molto marcati rispetto a Liguria, Puglia e Sardegna. Come detto, il Friuli esibisce un andamento decrescente dei prezzi in quasi tutto il periodo differenziandosi molto rispetto ai *trend* di crescita elevati delle tre regioni menzionate. Inoltre non mostra correlazione significativa con altre quattro regioni. La Valle d'Aosta, invece è la regione che evidenzia le correlazioni più basse e meno significative con quelle restanti mostrando un andamento del tutto particolare. Seppure importanti dal punto di vista interpretativo, le due eccezioni possono essere trascurate ai fini della valutazione dei risultati dell'analisi di correlazione ed è possibile affermare che le regioni mostrano ampie interconnessioni con le altre. Questo suggerisce che potrebbero essere sufficienti una o due componenti per cogliere la maggior parte delle correlazioni evidenziate.
- l'indice KMO sui dati disponibili risulta di 0,74 ad evidenza di una buona correlazione supportando la convenienza di un'analisi in componenti principali;
- il test di sfericità di Bartlett applicato alla matrice dei dati mostra un *p-value* pressoché nullo che conduce al rifiuto dell'ipotesi nulla.

Conclusa l'analisi preliminare si procede con l'estrazione delle componenti principali. Per la scelta del numero di componenti da prendere in considerazione si effettua un'ispezione visiva del grafico degli autovalori mostrato in Figura 4. Risulta chiaro che gli autovalori  $\lambda$  più elevati, associati alle prime due componenti ( $\lambda_1 = 15,1$  e  $\lambda_2 = 2,43$ ), spiegano rispettivamente il 79,4% e 12,7% della variabilità. Si decide pertanto di considerare le prime due componenti per la successiva fase di interpretazione. L'analisi dei pesi associati a ciascuna regione nella prima componente, PC1, evidenzia una generale uniformità dei pesi. Infatti, la maggior parte delle regioni oscilla intorno ad un valore di 0,25. Quasi tutte le regioni contribuiscono in modo simile alla spiegazione della maggior parte della variabilità presente nei dati.

<sup>6</sup> Per un approfondimento sulla tecnica delle componenti principali si può far riferimento ad esempio al lavoro di Jolliffe (2002).

<sup>7</sup> L'indice varia tra 0 e 1 e un valore prossimo allo zero indica che l'analisi in componenti principali sarebbe inadeguata, mentre un valore vicino ad 1 indica la presenza di una struttura di correlazione che può essere colta in modo appropriato dalle componenti principali.

Quindi, la prima componente principale PC1 suggerisce la presenza di un effetto globale nazionale di breve periodo che influenza tutte le regioni. Questo risultato conferma quanto già emerso dall'analisi dei coefficienti  $B_t$  dell'analisi di regressione.

Vanno tuttavia evidenziate due eccezioni. La prima è rappresentata dal Friuli Venezia Giulia che, come già evidenziato, essendo l'unica regione ad avere quotazioni decrescenti per la maggior parte del periodo, è anche l'unica ad avere segno negativo nel peso della componente. Inoltre, la stessa PC1 mostra la presenza di un gruppo di regioni, formato da Valle d'Aosta, Lombardia e Molise, con pesi più bassi, rispetto alle altre regioni. È utile sottolineare inoltre che queste ultime sono anche quelle con effetti fissi e persistenze stimate molto elevate. Sembra quindi ragionevole ipotizzare che tali regioni risentano meno dell'effetto dell'andamento nazionale in quanto caratterizzate da importanti effetti fissi. Tali effetti rendono infatti la serie più dipendente da sé stessa e quindi meno influenzata dal *trend* nazionale.

Pur evidenziando la presenza di fenomeni locali che influenzando alcune regioni possono condurre a deviazioni rispetto all'andamento nazionale, la prima componente, che spiega quasi l'80% della varianza, conferma e rafforza l'importanza di una componente nazionale.

Di interesse però è anche l'analisi dei pesi della seconda componente principale, PC2 da cui emerge in modo più evidente la presenza di un gruppo di regioni, Lombardia, Molise, i cui pesi risultano molto elevati, a cui si aggiunge il Friuli Venezia Giulia. Lombardia e Molise hanno effetti fissi e persistenze in numero assoluto rilevanti. Sembra, quindi che la seconda componente voglia suggerire l'impossibilità di trascurare, accanto alla componente globale, le componenti locali che emergono con maggior evidenza in alcune regioni rispetto ad altre.

Tuttavia, da tale analisi non è possibile trarre conclusioni circa l'esistenza di un effetto comune alle due regioni. In altri termini, i risultati indicano che ci sono effetti locali che rendono particolari le regioni ma tale effetto non è necessariamente comune. Non si è in grado infatti di individuare la presenza di un fattore comune che solo un'analisi formale di cointegrazione potrebbe evidenziare. Il numero limitato di osservazioni disponibili per ciascuna serie non rende, infatti possibile l'utilizzo di tale tecnica di analisi.

Un modo utile per interpretare i risultati dell'analisi in componenti principali è la loro visualizzazione attraverso il biplot<sup>8</sup> (Figura 5). Quest'ultimo è una rappresentazione grafica che consente di visualizzare contemporaneamente le osservazioni e le variabili e permette di approfondire la struttura di correlazione tra le variabili, rappresentate in questo caso dalle regioni. Tale grafico indica quanto ciascuna variabile contribuisce alla spiegazione di una componente principale. Nel nostro caso, la figura rafforza attraverso l'intuizione grafica quanto emerso dall'analisi dei pesi delle due componenti. Si confermano correlazioni marcate tra alcuni gruppi di regioni, quali Calabria, Liguria, Puglia, Piemonte e Basilicata o ancora tra Emilia Romagna, Campania, Toscana e Marche. Queste regioni presentano, infatti valori simili dei pesi in entrambe le componenti. Emerge inoltre con chiarezza la distanza del Friuli Venezia Giulia dalle altre regioni. Tornando a leggere i coefficienti  $A_i$  e  $C_i$  i risultati sono confermati.

Il grafico di destra della stessa figura riporta, invece, il grafico dei *loadings* delle prime due componenti visualizzando i periodi temporali. Da questa rappresentazione, dividendo il grafico in quadranti, si possono dedurre fasi abbastanza distinte dell'andamento generale dei prezzi reali nel tempo. Dopo un primo periodo (fino all'incirca al 2002) i prezzi hanno iniziato una fase di crescita caratterizzata da tassi elevati che si sono attenuati fino a divenire negativi dal 2008.

<sup>8</sup> Poiché esistono vari modi di costruire un grafico biplot si specifica che in questo caso si è utilizzato il comando biplot disponibile nel software R che segue le linee teoriche dell'articolo originale che ha introdotto il metodo (Gabriel, 1971).

### Un tentativo di interpretazione della prima componente principale

L'analisi congiunta dei coefficienti  $B_t$  stimati nel modello di regressione e della prima componente principale suggerisce l'esistenza di un *trend* globale che sebbene sia spesso smorzato dagli effetti delle singole regioni è significativo e non trascurabile.

Un approfondimento dei fattori che influenzano l'andamento delle serie dei prezzi richiede l'implementazione di un modello per serie storiche multivariate tale da rendere possibile analisi strutturali e di politica economica. Tale studio esula dagli scopi di questo articolo per cui ai fini di un'iniziale riflessione sull'interpretazione della prima componente, dalla quale si evince l'esistenza del fattore inosservato che porta ad un co-movimento delle serie dei prezzi interregionali, si ritiene utile richiamare gli studi esistenti in letteratura. Il reddito disponibile viene descritto come uno dei fondamentali più importanti riguardante la *demand side* del mercato delle abitazioni. In particolare, viene evidenziata una forte connessione tra l'incremento del reddito personale e l'aumento dei prezzi delle abitazioni generato da un incremento della domanda delle stesse (Capozza *et al.* (2004), Kim & Rous (2012), Hwang & Quigley (2006)), e ancora nel più recente lavoro della Banca d'Italia (Nobili & Zollino 2012) sul mercato italiano. Seguendo questo risultato, sembra ragionevole poter associare la prima componente proprio con la variabile reddito disponibile. In letteratura vengono, tuttavia, sottolineati importanti legami anche con altre variabili macroeconomiche quali il tasso d'interesse, gli andamenti demografici e i tassi di occupazione. In un recente articolo (Kholodilin, 2012) sono riassunti i principali risultati ottenuti in letteratura negli studi condotti sulle determinanti dei prezzi delle abitazioni. L'autore fa una rassegna delle categorie di indicatori inseriti nei modelli tra i quali il reddito, il tasso di interesse, le variabili demografiche, il mercato del credito, il lavoro, l'offerta di terreni, l'offerta di abitazioni o ancora altre variabili non riconducibili ad una particolare categoria. Il reddito è inteso come PIL reale, reddito *pro-capite* o ancora reddito da lavoro, mentre per i tassi di interesse nella maggior parte dei modelli viene considerato il tasso di interesse reale; tra le variabili demografiche è considerata soprattutto la popolazione ma sono inserite anche variabili come il tasso di matrimoni, la proporzione di single o il numero di divorzi. Tra gli indici dei prezzi sono presi a riferimento l'inflazione o l'inflazione attesa. Tra gli indicatori non riconducibili ad una categoria specifica sono molto utilizzati i costi di costruzione, il corrispondente indice reale o ancora il valore del patrimonio abitativo.

Più in generale si ravvisa che inversioni o mutamenti del ciclo economico colpiscono aree differenti in modo relativamente simile anche in presenza di differenze strutturali tra le regioni (Oikarinen, 2007).

### IPOTESI DI CONVERGENZA NEL LUNGO PERIODO

In analogia con la teoria economica sulla crescita dei redditi e la loro convergenza proposta da Barro & Sala-i-Martin (1992) può essere ragionevole attendersi, a livello assoluto, che in un certo periodo gli ambiti territoriali che presentano livelli dei prezzi più alti mostrino tassi di apprezzamento minori così da ottenere una diminuzione delle divergenze. Nel nostro caso il coefficiente di correlazione tra il tasso di variazione complessivo delle quotazioni per l'intero periodo (2001-2011) e il livello delle stesse riferito al periodo iniziale (I semestre 2001) mostra una relazione negativa e prossima allo zero. Tuttavia, se si rimuovono Lazio e Liguria che, oltre ad avere alte quotazioni presentano tassi di crescita molto elevati, il coefficiente di correlazione risulta di -0,27. In modo meno restrittivo, se ci si riconduce al concetto di convergenza in senso relativo, si dovrebbe osservare una diminuzione nel tempo della dispersione dei valori osservati. Misurando la variabilità attraverso la deviazione *standard* per ciascun anno si nota una, seppur non accentuata, decrescita dal 2006, tuttavia, il *trend* decrescente si accentua se si escludono Lazio e Liguria.

Queste prime analisi mostrano quindi che l'ipotesi di convergenza nel lungo periodo risulta piuttosto