

CRENOS
CENTRO RICERCHE
ECONOMICHE NORD SUD
Università di Cagliari
Università di Sassari

L'ECONOMETRIA PER LE INDAGINI TERRITORIALI
Appunti metodologici e un'applicazione alla Sardegna

Simone Atzeni
Barbara Dettori
Stefano Usai

QUADERNI DI LAVORO



2004/01

Simone Atzeni
Studio Atzeni
s.atzeni@studioatzeni.it

Barbara Dettori
CRENoS
barbara.dettori@crenos.it

Stefano Usai
Università di Sassari e CRENoS
usai@uniss.it

L'ECONOMETRIA PER LE INDAGINI TERRITORIALI Appunti metodologici e un'applicazione alla Sardegna

Abstract

Il recente interesse scientifico per la *new economic geography* ha favorito il sorgere e il consolidarsi di un filone di ricerca di economia applicata che mira alla formalizzazione e misurazione delle relazioni spaziali fra fenomeni economici. Questa attività di ricerca, dapprima relegata in una nicchia accademica, oggi trova progressivamente applicazione nei corsi universitari, segno di una disciplina che suscita interesse per l'approccio innovativo all'indagine delle dinamiche sociali ed economiche.

L'econometria spaziale nasce come branca dell'econometria tradizionale con la finalità di affrontare le problematiche connesse all'impiego di dati georeferenziati, cioè collegati ad entità geografiche. Fra i vari temi assumono rilevanza particolare l'interdipendenza (o autocorrelazione) e l'eterogeneità (o eteroschedasticità) spaziale. Per essi, troppo spesso, gli elementi forniti dalla econometria tradizionale appaiono insufficienti.

Questo lavoro si propone di rendere esplicite le peculiarità proprie delle tecniche dell'econometria spaziale insieme ad alcune istruzioni d'uso per le loro applicazioni concrete.

Gennaio 2004

1. Introduzione

L'economia regionale e urbana ha avuto negli ultimi anni un notevole sviluppo teorico ed empirico, grazie a stimoli sia teorici che normativi.

Dal lato teorico si registra la rinnovata attenzione verso gli studi di Marshall (1890), complici alcune *performance* dei sistemi produttivi locali sulle economie regionali che introducono nuovi concetti rispetto alla teoria classica dei distretti industriali. Questo recupero è in parte dovuto al riconsolidato interesse da parte degli economisti per i temi della geografia, che ha portato al sorgere della cosiddetta nuova economia geografica (per un'esauritiva rassegna si veda Clark, Feldman e Gertler, 2000¹). Questo interesse è sorto e si è rinvigorito dal momento in cui a un agente economico atomistico la teoria ha progressivamente sostituito un agente che decide in dipendenza del contesto in cui agisce, in particolare degli altri agenti con le loro preferenze, utilità, strategie. Il problema dell'interazione spaziale tra agenti economici ha fatto quindi breccia nell'ortodossia economica proponendo diversi casi in cui i rapporti tra soggetti, sotto forma di norme sociali, *spillovers* da apprendimento, fenomeni emulativi danno luogo a casi di rendimenti crescenti o a cosiddette economie esterne. Queste ultime possono manifestarsi anche sotto forma di altri fattori localizzati, quali particolari dotazioni di risorse primarie, di beni intermedi e di servizi oppure la disponibilità di forza lavoro qualificata².

Accanto alla teoria, in un complicato rapporto di causa-effetto con essa, emerge il profilo normativo dello sviluppo locale. Le politiche

¹ Una nota di commento sul dibattito teorico e empirico sulla nuova economia geografica si trova in Usai (2002).

² Marshall (1890) aveva individuato tre possibili ragioni per le economie esterne: la presenza di beni e servizi intermedi, l'esistenza di un bacino di lavoratori specializzati e qualificati, e un contesto dove era possibile venire a conoscenza di particolari informazioni tecniche e tecnologiche. I fattori che danno luogo a questi fenomeni sono per loro natura immobili o possono essere trasportati da un luogo a un altro con costi diretti e indiretti molto elevati

comunitarie di accompagnamento alle dinamiche di crescita localizzata richiedono infatti un esame sempre più attento ed esperto a quei fenomeni che esplicano i propri effetti in contesti regionali e sub-regionali. Questo elemento consente di sostituire le economie di scala, inaccessibili a causa della piccola dimensione, con le economie di agglomerazione, differenti dalle prime ma pur sempre portatrici di rendimenti crescenti. Dato questo scenario, non stupisce la necessità di indagare con decisione sulle tecniche di analisi spaziale dei dati.

Sul versante teorico si è rivelata cruciale l'introduzione di una serie di innovazioni analitiche (che Krugman (1991) ama efficacemente definire *modelling tricks*, stratagemmi modellistici) che hanno consentito di affrontare in un contesto rigoroso (prettamente neoclassico) il problema dei rendimenti crescenti. Dal lato empirico-metodologico un ruolo fondamentale lo sta svolgendo, invece, l'econometria spaziale. Sorta come sottodisciplina della econometria tradizionale³ essa affronta i particolari problemi posti dalla trattazione di serie di dati riferiti a entità geografiche (Anselin, 1988). Tra questi problemi assumono rilevanza particolare l'eteroschedasticità (eterogeneità) e l'autocorrelazione (dipendenza). Si tratta, in vero, di problemi comuni a qualsiasi applicazione econometrica, ma che, nel caso di dati spaziali, possono assumere sia forme che richiedono una loro trattazione distinta, sia significati molto particolari nella lettura delle dinamiche locali. In realtà, l'eteroschedasticità spaziale, che sorge quando si utilizzano dati di unità spaziali molto diversificate per spiegare lo stesso fenomeno, può essere spesso risolta mutuando le tecniche proprie della econometria tradizionale (quella cioè delle serie storiche). Non altrettanto è, invece, possibile fare in presenza di autocorrelazione spaziale, che sorge quando il valore assunto da una variabile in un determinato luogo è correlata con il valore che la stessa variabile assume in un altro luogo o in un insieme di altri

³ L'econometria può essere quindi divisa in econometria delle serie storiche (che ha una tradizione ben più lunga e solida) e la branca della econometria dei dati spaziali che si è andata affermando dagli anni ottanta in poi.

luoghi. Proprio per questa caratteristica di multidirezionalità che caratterizza le relazioni di interdipendenza spaziale non è possibile l'applicazione delle tecniche standard per l'analisi della dipendenza temporale, che può, per sua natura, essere solo unidirezionale.

Sebbene questi problemi fossero noti, specie a coloro che si occupavano di economia urbana e regionale, la nascita della statistica spaziale può essere fatta risalire al 1948 quando Moran introduce i primi indici per lo studio della presenza di autocorrelazione spaziale. Dovevano però poi trascorrere altri trent'anni prima di giungere ad una precisa definizione della econometria spaziale e a una chiara identificazione dei suoi principali obiettivi e metodi. I primi a farlo sono stati Paelinck e Klassen (1979) con una precisa definizione del campo di analisi dell'econometria spaziale come spazio scientifico autonomo. Successivamente comparvero due altri contributi rilevanti, dovuti nell'ordine a Cliff e Ord (1981) e Upton e Fingleton (1985), che tuttavia mantennero la trattazione della dipendenza spaziale con metodi econometrici nell'ambito "angusto" dei modelli di regressione lineare. E' infine con il contributo di Anselin (1988) che l'econometria spaziale riceve dignità di branca autonoma dell'econometria, attraverso una efficace sintesi di circa dieci anni di lavori teorici ed empirici sul tema. L'econometria spaziale viene definita come l'insieme delle tecniche necessario a trattare le peculiarità determinate dalla dimensione geografica all'interno delle analisi statistiche dei modelli di scienze regionali. Lo stesso Anselin individua quattro ampie aree di interesse dell'econometria spaziale: a) la specificazione formale degli effetti spaziali nei modelli econometrici; b) la stima dei modelli che incorporano gli effetti spaziali; c) la specificazione di test e diagnostiche per la presenza di effetti spaziali; d) la previsione spaziale. L'identificazione degli obiettivi e dei principali strumenti con cui riempire la cassetta degli attrezzi dell'econometrico spaziale e la loro codifica informatica nel pacchetto econometrico *Spacestat* qualche anno dopo (nel 1992) hanno determinato una diffusione sempre maggiore delle tecniche

dell'econometria spaziale nel corso di questi ultimi anni⁴. In questo processo ha avuto un ruolo importante anche il contemporaneo diffondersi dei sistemi informativi geografici (GIS) che hanno messo a disposizione una massa crescente di dati geo-referenziati, per i quali sono necessarie tecniche specialistiche per l'elaborazione e il trattamento statistico⁵. Non è un caso, inoltre, che il mercato dei software per il trattamento delle serie territoriali e per la loro rappresentazione cartografica stia progressivamente divenendo simile alla concorrenza monopolistica: numerosi produttori, ciascuno con la propria nicchia di specializzazione (demografia, economia, dotazioni infrastrutturali, ambiente, per citare i campi di applicazione più diffusi) sebbene con elevati gradi di integrazione con i principali applicativi di matematica e statistica avanzata. Questo processo di diffusione è quindi avviato e procede rapidamente ma è, in effetti, ancora largamente incompiuto: la distanza in termini di conoscenza tra econometria spaziale e econometria tradizionale appare, infatti, appena intaccata. Sono ancora molti i lavori econometrici di economia regionale e urbana che trascurano i problemi derivanti dalla mera trasposizione dello stimatore dei minimi quadrati ordinari, tipico dell'econometria tradizionale, all'analisi inferenziale geografica. Le scuole di analisi dell'economia regionale che utilizzano tecniche di econometria spaziale in modo intensivo e ne fanno una filosofia di analisi sono poche e, sebbene questi temi siano sempre più presenti nella comunità scientifica internazionale, crescono con relativa inerzia.

⁴ A tutt'oggi Spacestat è l'unico software dedicato per i problemi di stima econometrica spaziale. La disponibilità di effettuare stime Maximum Likelihood per modelli spaziali è data anche all'interno di Matlab.

⁵ A questo proposito è utile sottolineare che Spacestat ha un'interfaccia per incrociare le sue informazioni e il suo database con ArcView uno dei più utilizzati software GIS. Come accennato, dal lato industriale proseguono parallelamente gli sforzi per arricchire gli applicativi GIS di tool per la elaborazione avanzata delle serie statistiche territoriali e l'efficace interfacciamento con i principali strumenti informatici di matematica e statistica applicata (MatLab e Gauss in questo senso rappresentano esempi importanti di un processo in atto)

In Italia, per esempio, i lavori teorici e empirici che dedicano la dovuta attenzione ai temi e alle problematiche dell'econometria spaziale sono un numero limitato. Tra i contributi teorici un importante riferimento è rappresentato dai lavori di Arbia (1996, 1999). Tra le analisi empiriche dedicate allo studio delle diverse realtà territoriali italiane (sistemi locali del lavoro, province, regioni) possono essere citati i lavori di Paci e Usai (2002) per i sistemi locali del lavoro, Pellegrini (2001) per le province, Paci e Pigliaru (2000) per le regioni italiane. Bottazzi e Peri (2002) e Moreno, Paci e Usai (2003) hanno invece utilizzato questi strumenti per il caso delle regioni europee.

Questo documento si propone come un breve e sintetico manuale per l'uso delle tecniche di econometria spaziale evidenziando, per inciso, i limiti delle tecniche econometriche standard in un contesto spaziale, e sottolineando il fatto che gli effetti spaziali possono essere considerati come un caso particolare del più generale problema della specificazione del modello e della sua stima⁶.

Nella prossima sezione si introduce il concetto di dipendenza ed eterogeneità spaziale, presentando in termini generali la nozione di connessione nello spazio, mentre nella successiva si descrivono i principali strumenti operativi per mezzo dei quali si possono ricomprendere nei lavori econometrici gli effetti spaziali, primo dei quali quello della matrice dei pesi, che rappresenta lo strumento di base per la loro formalizzazione empirica.

Nella quarta sezione vengono presentati i principali indicatori econometrici univariati che consentono di verificare e misurare il grado di interdipendenza spaziale di un fenomeno.

Nella sezione successiva vengono analizzati i problemi dei metodi di stima tradizionali dei modelli di regressione applicati ai dati

⁶ Per una trattazione più estensiva si possono consultare Vaya e Moreno (2000), ce ha fornito spunti importanti per questo lavoro, Anselin (1999) e naturalmente Anselin (1988) che rappresenta il manuale più completo e articolato su aspetti teorici e empirici riguardanti l'econometria spaziale.

spaziali, e quindi, enunciando le linee di classificazione dei modelli di econometria spaziale, sono proposte le metodologie alternative per la specificazione, verifica e stima dei modelli e alcune tecniche per la selezione della stima migliore. La settima sezione si occupa dei problemi legati all'eterogeneità e la sua possibile soluzione mediante il metodo dell'espansione spaziale e la variazione dei coefficienti random.

Il nostro contributo si conclude con un'applicazione empirica delle tecniche dell'econometria spaziale sulle determinanti del reddito per i comuni della Sardegna.

2. La dipendenza e l'eterogeneità spaziale

L'interdipendenza o autocorrelazione spaziale è la conseguenza della presenza di una relazione funzionale tra ciò che accade in un punto determinato nello spazio e ciò che accade in altri punti. In altre parole, le caratteristiche di un determinato fenomeno in una regione non sono spiegate unicamente da determinanti interne alla stessa ma anche da alcune proprie delle altre regioni, più o meno vicine. Se, per esempio i consumi in una regione crescono al crescere dei consumi nelle regioni limitrofe (e quindi dal livello del reddito in queste aree) si parlerà di autocorrelazione positiva. In questo caso sarà favorita l'omologazione nei livelli di consumo che daranno luogo a fenomeni di concentrazione spaziale. Quando invece l'autocorrelazione è negativa verrà stimolata la diversità e si formeranno configurazioni spaziali del tipo centro-periferia.

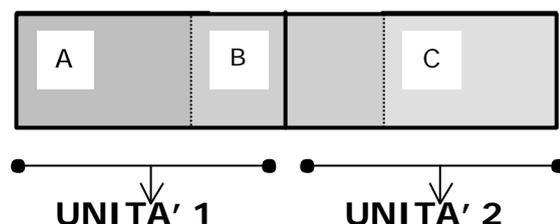
L'autocorrelazione spaziale può avere fondamentalmente due cause: 1) errori di misurazione per osservazioni riferite ad unità geografiche contigue e 2) vera e propria interazione spaziale.

I primi possono sorgere ogni qualvolta si utilizzino dati per i quali non c'è perfetta corrispondenza tra unità territoriale di analisi e estensione del fenomeno oggetto d'esame. Se, per esempio, intendiamo esaminare la relazione tra consumi e reddito attraverso una cross section di dati raccolti a livello comunale incorreremo in un errore di misurazione se il bacino di mercato si estende oltre i confini dei singoli comuni. I consumi che si realizzano in un

singolo comune potranno infatti essere influenzati dalla disponibilità di reddito dei residenti nelle aree contigue che possono muoversi da un comune all'altro per determinati acquisti. Il risultato di questo errore è che i dati di consumo mostreranno una forte interdipendenza spaziale. Tale interdipendenza non è reale, ma l'effetto di un errore nella scelta dell'unità spaziale di riferimento⁷.

Per meglio intendere questo concetto, consideriamo un semplice esempio di errore di misurazione in cui si incorre frequentemente quando i dati sono raccolti ad un livello aggregato. Come aggravante si deve aggiungere che gli errori di misurazione tendono ad estendersi oltre i confini delle unità spaziali, comportando perciò che gli errori per un'unità territoriale i sono spesso correlati agli errori in un'unità contigua j. Questo fenomeno di *spillover* spaziale degli errori di misurazione, illustrato dalla ipotetica situazione rappresentata in Figura 1, è un'ovvia causa di dipendenza spaziale, che nel contesto di una regressione portano ad un vettore di errori non-gaussiani.

Fig.1. Dipendenza spaziale e aggregazioni



Nella figura è presentato un tipico esempio del caso in cui la reale estensione spaziale delle variabili analizzate è rappresentata dalle aree A, B e C, mentre le osservazioni sono aggregate per le unità 1

⁷ Bisogna peraltro ricordare che spesso questa scelta non è libera, bensì vincolata dalla disponibilità dei dati e dalla loro qualità.

e 2. Di conseguenza la variabile osservata Y_1 sarà un'aggregazione (forzata) di Y_A e parte di Y_B , mentre Y_2 sarà un'aggregazione di Y_C e la rimanente parte di Y_B :

$$Y_1 = Y_A + \lambda Y_B$$

$$Y_2 = Y_C + (1-\lambda) Y_B$$

Un'aggregazione di questo genere sarà molto probabilmente affetta da errore nel momento della determinazione del valore del parametro λ , che è presente sia in Y_1 sia in Y_2 . Come conseguenza questi errori nella misurazione genereranno un sentiero di dipendenza spaziale.

La seconda causa di autocorrelazione, la presenza di reali fenomeni di interazione spaziale, discende dall'importanza della collocazione geografica come elemento esplicativo del comportamento umano, ed è quella che più ci interessa in quanto a essi possono essere associati relazioni di tipo economico⁸. E' il caso, per esempio, della conoscenza di determinate tecniche produttive diffusa in una data area geografica, perché là sono state introdotte e sono state sviluppate grazie all'esperienza accumulata e spesso condivisa di lavoratori, imprenditori e altri agenti operanti sul territorio. Si tratta di una fattispecie tipica della realtà del distretto industriale al cui interno si uniscono diverse imprese di territori comunali contigui formando un'entità geo-economica più ampia.

A causa della dipendenza che si crea tra fenomeni in differenti dislocazioni geografiche, ciò che si osserva in un punto è determinato, almeno in parte, da ciò che accade in un altro punto del sistema. Questo può essere formalmente espresso in un processo spaziale nel seguente modo:

$$y_i = f(y_1, y_2, \dots, y_N)$$

⁸ Altre relazioni possono essere importanti in riferimento ad aspetti sociali e culturali, per esempio. Qui ovviamente ci concentriamo sulle relazioni che hanno un riflesso e un contenuto economico.

In esso ogni osservazione della variabile y è formalmente correlata, mediante la funzione f , al valore della stessa variabile in altre unità territoriali appartenenti al sistema.

Questa semplice espressione, tuttavia, non risulta di per sé molto utile nel contesto di un lavoro empirico, dato che ne deriva un sistema indefinito, con molti più parametri (potenzialmente N^2-N) che osservazioni (N). Solo imponendo una particolare struttura alla relazione funzionale espressa dalla f , cioè esplicitando una particolare forma del processo spaziale, si possono stimare e testare empiricamente un numero limitato di caratteristiche della dipendenza spaziale. Questo procedimento rappresenta la base dei differenti approcci di specificazione e stima di modelli propri dell'econometria spaziale, che saranno discussi nelle prossime sezioni.

Il secondo tipo di effetto spaziale, l'eterogeneità, è correlato alla mancanza di stabilità nello spazio delle relazioni o dei comportamenti analizzati. Dal punto di vista della modellizzazione questo implica che le forme funzionali ed i parametri variano in ragione della loro localizzazione geografica e non sono omogenei in tutto il *data set*⁹. Poiché questa eterogeneità può essere messa in relazione direttamente con l'allocazione spaziale, viene comunemente designata con il termine di *eterogeneità spaziale*.

La letteratura delle scienze regionali e dell'economia geografica ha trattato ampiamente il tema della mancanza di uniformità nello spazio, mettendo a punto adeguate strategie che considerino nel contesto dell'analisi le peculiari caratteristiche di ogni unità territoriale.

In aggiunta alla mancanza di una stabilità strutturale nello spazio, le stesse unità territoriali risultano di per sé poco omogenee. Le

⁹ Questo accade frequentemente nei modelli econometrici stimati per un campione *cross-section* che presenta unità spaziali che differiscono notevolmente per qualche carattere, ad esempio in una regressione a livello nazionale italiano verranno incluse contemporaneamente regioni ricche settentrionali e regioni povere meridionali.

rilevazioni censuarie, ad esempio, difficilmente sono riferite ad estensioni territoriali della medesima portata, le aree urbane presentano livelli di reddito o densità insediative che possono differire notevolmente, le regioni possono mostrare disparati livelli tecnologici raggiunti. Nella misura in cui questi aspetti dell'eterogeneità si riflettono in errori di misurazione (variabili mancanti, misspecificazione funzionale etc.) possono avere come conseguenza l'eteroschedasticità.

Contrariamente a quanto accade nel caso precedentemente esposto della dipendenza spaziale, i problemi causati dall'eterogeneità possono essere risolti, nella maggior parte dei casi, attraverso le tecniche dell'econometria classica; nello specifico ci riferiamo alle tecniche di variazione dei parametri, ai coefficienti *random* e all'instabilità strutturale, metodologie facilmente adattabili per risolvere la mancanza di omogeneità delle osservazioni nello spazio. Queste tecniche saranno meglio esposte più avanti.

3. L'espressione formale degli effetti spaziali

In entrambi i casi di dipendenza spaziale descritti, la conseguenza pratica è la presenza di autocorrelazione spaziale che va o eliminata (per esempio, ricorrendo a dati riferiti alle corrette unità territoriali che "contengono" il fenomeno oggetto di studio)¹⁰ oppure, più di frequente, affrontata per mezzo di apposite metodologie. Il principale problema che emerge da un punto di vista operativo riguarda l'espressione formale della struttura della dipendenza spaziale che dev'essere incorporata in un modello. L'autocorrelazione spaziale mostra infatti caratteri che sono solo apparentemente simili a quelli propri al corrispondente fenomeno

¹⁰ Questo succede per esempio quando per gli studi riferiti ai distretti industriali è possibile riferirsi ai sistemi locali del lavoro che sono identificati in base ai dati di flussi dei lavori e quindi associano i comuni che fanno parte dello stesso mercato del lavoro che dovrebbe coincidere con l'area del distretto (si veda al proposito, Sforzi, 1991).

delle serie storiche¹¹. In quest'ultimo caso l'interdipendenza può avere un'unica direzione di causalità: gli eventi del passato possono influenzare i fatti del presente, ma non viceversa. Nel caso dell'autocorrelazione spaziale, al contrario, la relazione causale è potenzialmente bi o multidirezionale e può articolarsi in forme complesse: la *performance* di un'impresa (o di una regione) può essere influenzata non solo da un'unica impresa (regione) contigua, ma anche da un insieme di imprese (regioni) vicine e soprattutto questa stessa impresa (regione) può influenzare la *performance* di quelle che la circondano.

La conseguenza è che questa caratteristica di multidirezionalità dell'autocorrelazione spaziale rende impossibile l'uso dell'operatore dei ritardi L (Lag), caratteristico dell'analisi delle serie storiche. In effetti, il problema di esprimere formalmente la struttura della dipendenza geografica è centrale nell'econometria spaziale¹². Tale problema è stato risolto introducendo un concetto più articolato del ritardo temporale che si configura come una matrice W, detta dei pesi o dei ritardi spaziali. La matrice, nella sua forma più generale, è costruita nel modo seguente:

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1N} \\ w_{21} & 0 & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & 0 \end{bmatrix}$$

Si tratta di una matrice quadrata non stocastica i cui elementi w_{ij} riflettono l'intensità della connessione esistente tra ogni coppia di regioni i,j. Le misure di questa intensità, che necessariamente deve

¹¹ Si noti che anche ad un livello più generale, nell'econometria classica, il tema delle differenze teoriche fra *time series* e *cross section*, è stato per gran parte ignorato. Se si eccettuano alcuni cenni, prima degli anni novanta gli unici contributi interessati all'argomento sembrano quelli dovuti a Kmenta (1971) e Johnston (1984).

¹² Per una trattazione approfondita di questo aspetto si veda Anselin (1988, cap. 3).

essere non negativa e finita, possono essere diverse. Nella forma più semplice w' si basa sul concetto di contiguità binaria secondo cui la struttura delle prossimità è espressa da valori 0-1. Se due unità spaziali hanno un confine in comune, di lunghezza maggiore di zero, saranno considerati contigui e saranno contrassegnati dal valore 1. Viceversa se non sono contigui la loro accoppiata avrà valore 0.

Ipotizziamo, per esempio, che lo spazio di riferimento sia rappresentato dalla seguente griglia regolare di nove unità spaziali:

11	12	13
21	22	23
31	32	33

Se le celle vengono numerate da sinistra verso destra e dall'alto verso il basso e la prima cifra indica la riga e la seconda la colonna avremo che la corrispondente matrice dei pesi o dei ritardi spaziali avrà dimensione 9x9. Di seguito riportiamo la matrice W dove, per una migliore lettura, la prima riga e la prima colonna riportano l'indicazione delle celle rispetto alle quali viene determinata la contiguità binaria¹³.

¹³ Si noti che in questo caso è stato applicato il concetto di contiguità cosiddetto della torre, secondo il quale sono contigue quelle celle che nella griglia regolare hanno un lato in comune (per esempio la cella 11 e la cella 12). Se fosse stato applicato il criterio della regina avremmo invece dovuto considerare contigui anche quelle che hanno un angolo in comune (per esempio la cella 11 e la cella 22).

	11	12	13	21	22	23	31	32	33
11	0	1	0	1	0	0	0	0	0
12	1	0	1	0	1	0	0	0	0
13	0	1	0	0	0	0	0	0	0
21	1	0	0	0	1	0	1	0	0
22	0	1	0	1	0	1	0	1	0
23	0	0	1	0	1	0	0	0	1
31	0	0	0	1	0	0	0	1	0
32	0	0	0	0	1	0	1	0	1
33	0	0	0	0	0	1	0	1	0

La prima riga quindi ci segnala con quali celle è contigua la cella 11: il valore uno è attribuito in corrispondenza dell'incrocio con la cella 12 e la cella 21. La diagonale principale è composta di zeri, riportati in grassetto per ricordare che in questo caso non indicano assenza di contiguità bensì il riferimento alla stessa cella, w_{ii} , che per convenzione è uguale a zero¹⁴.

Il concetto di prossimità può essere più articolato di quanto espresso attraverso la contiguità binaria. Quest'ultima denuncia infatti alcuni innegabili limiti: a) non tiene conto di interazioni non reciproche a causa del suo carattere simmetrico; b) non tiene conto di casi di interazione che non siano funzione della vicinanza geografica ma di altri fenomeni; c) non distingue tra diverse tipologie di regioni vicine, per esempio rispetto alla distanza o in ragione della morfologia della zona di confine (montagna, collina, pianura) o, infine, rispetto alla estensione del confine effettivamente condiviso.

Per ovviare a quest'ultimo problema possono, in primo luogo, essere utilizzate matrici di contiguità di ordine superiore al primo.

¹⁴ La matrice W utilizzata per i ritardi spaziali è spesso standardizzata in modo che la somma degli elementi di una riga sia pari all'unità. Se cioè esistono M unità contigue, la variabile ritardata è 1/M volte la somma della relativa x_i . In termini dell'interazione spaziale questo implica una competizione tra le unità contigue: minore è il loro numero e maggiore sarà la loro influenza individuale sull'unità centrale.

Per esempio, una matrice di secondo ordine individuerà quelle aree che sono contigue alle regioni che assumono valore uno nella matrice di primo ordine. In secondo luogo, è oramai prassi consolidata associare alla matrice delle contiguità binarie anche la matrice delle distanze o altre matrici che combinano le misure di distanza con quelle di lunghezza dei confini. In questi casi quindi w_{ij} potrà essere, per esempio, l'inverso della distanza dei centri (geografici ma anche politici o amministrativi) delle unità spaziali. Tale distanza potrà essere espressa sia rispetto alla misura di distanza geografica lineare o stradale sia in termini di distanza temporale, in relazione ai tempi di viaggio. La matrice sarà quindi più complessa, perché piuttosto che valori binari (0,1) attribuirà ad ogni coppia di regioni un valore presumibilmente differente.

Per ovviare invece al primo e al secondo limite indicato sopra, si è cercato di superare il concetto di vicinanza fisica introducendo misure di "contiguità economica" tra regioni. Per esempio w_{ij} potrebbe misurare l'interscambio commerciale tra regioni (che distinguerebbe anche rapporti non reciproci) oppure il grado di affinità della specializzazione produttiva o ancora la rete delle relazioni di scambi di beni intermedi all'interno delle filiere produttive. Inoltre, le misure possono tenere conto di altre tipologie di distanze come quelle culturali, linguistiche o amministrative. In alcuni studi, per esempio, si sottolinea il fatto che, anche prescindendo dalla prossimità fisica, esistano forti interdipendenze tra le regioni europee che ospitano la capitale dello Stato. In questo caso la vicinanza avrebbe ragioni amministrative che si riflettono in strutture produttive affini perché fondate sulla specializzazione in servizi pubblici propria delle città-capitale.

Ognuno di questi indicatori ha ovviamente dei pregi e dei difetti e non esiste perciò una regola unica universalmente accettata. E' certo che la scelta vada fatta in funzione dell'obiettivo di ricerca, in modo da assicurare in ogni occasione che i pesi che compongono la matrice abbiano natura esogena rispetto all'oggetto di studio. Per questa ragione, per esempio, la matrice dei pesi costruita sulla base della matrice degli scambi tra regioni non potrà essere

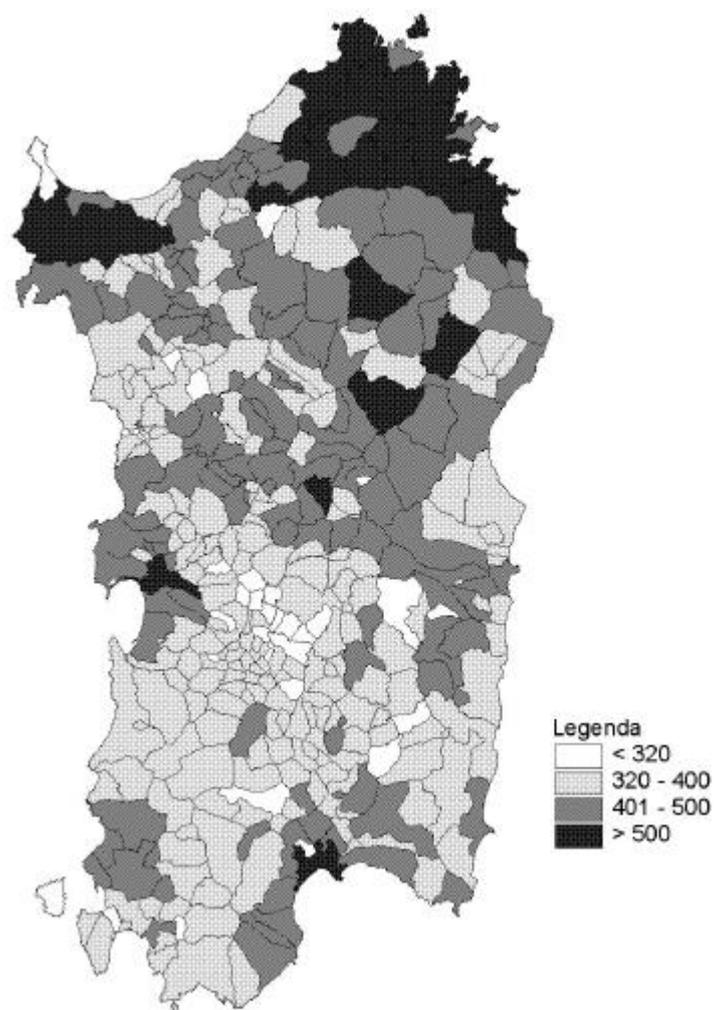
utilizzata se il nostro fine è lo studio della *performance* commerciale delle singole regioni.

4. La dipendenza spaziale univariata

La presenza di interdipendenza spaziale, è stato detto, si manifesta con fenomeni di concentrazione territoriale di valori simili (nel caso di interdipendenza positiva) o di valori differenti (nel caso di interdipendenza negativa). Una prima verifica della sua presenza può essere quindi effettuata attraverso l'esame della mappa geografica che rappresenta la distribuzione spaziale del fenomeno oggetto di studio.

Per quanto l'esame delle mappe, soprattutto se sostenuto da valide argomentazioni teoriche, possa portare a delle conclusioni ragionevoli e convincenti sulla presenza o assenza di dipendenza spaziale, è opportuno che tali conclusioni possano essere rese indipendenti dalla tecnica utilizzata per la rappresentazione (per esempio nella individuazione delle classi). Inoltre l'uso di un vero e proprio indicatore di interdipendenza si rende necessario per misurarne anche la sua intensità e consentire quindi confronti tra realtà e fenomeni differenti oppure tra stesse realtà e fenomeni in tempi diversi.

Figura 1. Autovetture immatricolate per 1000 abitanti nei comuni della Sardegna, anno 1996



Fonte: Nostre elaborazioni su dati CRENoS

L'esempio didattico proposto nella figura 1 si riferisce alla distribuzione per l'anno 1996 delle immatricolazioni di autovetture ogni 1000 abitanti nei 375 comuni della Sardegna. Dall'analisi della mappa si rileva, in prima approssimazione, una discreta interdipendenza spaziale positiva: intorno ai comuni capoluogo di Cagliari e Sassari e in particolare nell'area intorno a Olbia, in cui si registrano le maggiori frequenze assolute, emergono polarizzazioni di comuni con valori superiori alla media regionale. Concentrazioni importanti sembrano presenti anche nelle vicinanze degli altri due comuni capoluogo di provincia: Oristano e Nuoro. Si tratta di un semplice indizio, peraltro utile ed illustrativo, desunto da un esercizio di visualizzazione di informazioni statistiche. Occorre, come detto, cercare una misura più precisa della correlazione spaziale ed identificare un metodo di individuazione dell'intensità di tale correlazione.

Esiste in letteratura una batteria abbastanza nutrita di test per la verifica della presenza di autocorrelazione spaziale. Tali test possono essere utilmente suddivisi in due gruppi: indicatori globali e indicatori locali di interdipendenza spaziale.

Nel primo gruppo troviamo gli indicatori I di Moran e la G(d) di Getis e Ord, nel secondo il LISA. I primi sono abbastanza simili tra loro e per questo in questo lavoro discuteremo solo dell'I di Moran, che rappresenta il più tradizionale degli indicatori di correlazione spaziale. L'indicatore di Moran è dato dalla seguente

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

espressione:

dove x_i rappresenta la variabile che descrive il fenomeno oggetto di studio nella regione i , \bar{x} rappresenta il valore medio campionario, w_{ij} sono i pesi della matrice W e N è la dimensione del campione e $S_0 = S_i S_{ij}$

Quando N è sufficientemente grande la I di Moran standardizzata Z(I) si distribuisce come una normale standardizzata. Un valore non significativo di Z(I) quindi ci porterà a non rigettare l'ipotesi nulla di assenza di autocorrelazione, mentre un valore positivo (negativo) significativo indicherà la presenza di autocorrelazione spaziale positiva (negativa).

Nell'esempio applicativo condotto sulla serie territoriale delle immatricolazioni pro capite, come può vedersi dalla tavola sottostante¹⁵, è possibile rilevare autocorrelazione spaziale per i primi tre livelli di contiguità. Si rileva che la autocorrelazione è presente a tutti e tre i livelli essendo il valore di Z(I) sempre positivo e significativamente diverso da zero¹⁶.

Tavola 1. Test I di Moran (1°, 2° e 3° livello di contiguità)

VARIABLE	Contiguità	I	MEAN	ST.DEV.	Z Value	PROB
Immatric pro capite 1999	1	0,272	-0,002	0,032	8,661	0,001
Immatric pro capite 1999	2	0,223	-0,003	0,021	10,726	0,001
Immatric pro capite 1999	3	0,187	-0,002	0,016	11,894	0,001

Fonte: nostre elaborazioni su dati CRENoS

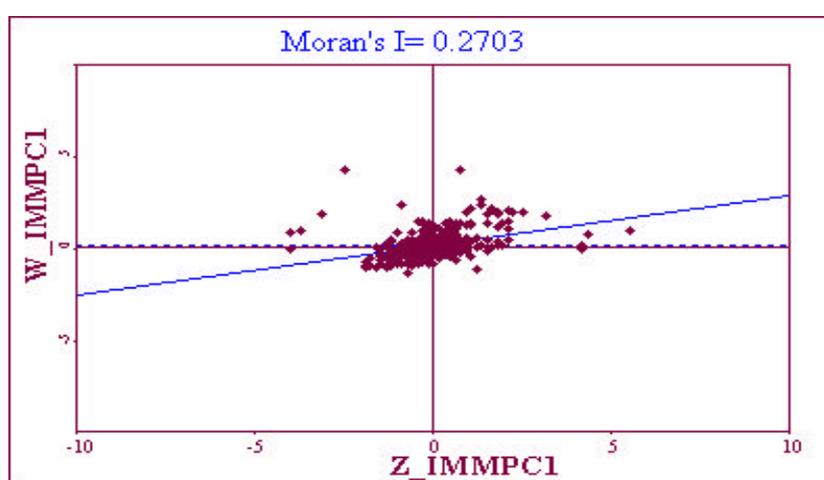
In riferimento alla statistica I di Moran è possibile associare un utile grafico che fornisce informazioni complementari e integrative. Si tratta del Moran Scatterplot che riporta in un grafico cartesiano sull'asse delle ascisse la variabile x normalizzata e su

¹⁵ Nel seguito gli elaborati statistici saranno riportati nella configurazione standard fornita da *Spacestat*.

¹⁶ Si noti che il valore di PROB, nell'ultima colonna, indica che esiste un livello di probabilità di appena lo 0.1% di commettere un errore nel rigettare l'ipotesi nulla di assenza di autocorrelazione.

quello delle ordinate il ritardo spaziale di detta variabile (Wx) anch'esso normalizzato.

Figura 2. Moran scatterplot (Immatricolazioni pro capite nei comuni della Sardegna, 1996)

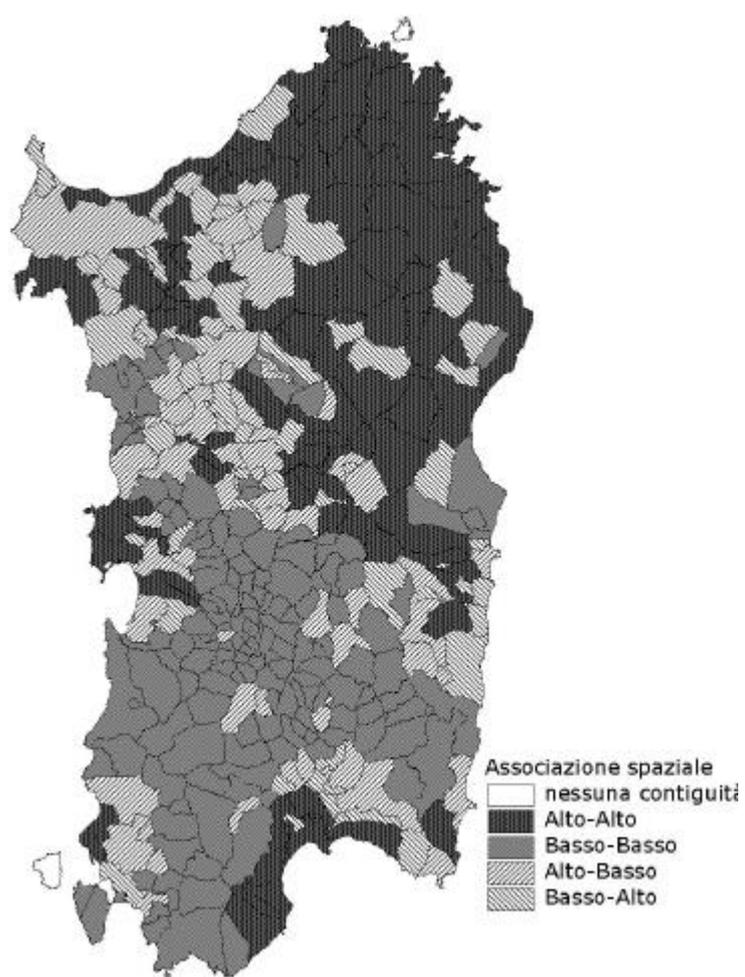


La I di Moran è rappresentata dal coefficiente angolare della relazione lineare tra le due variabili riportate sugli assi del Moran scatterplot. Se quindi i punti sono dispersi fra i quattro quadranti questo indicherà assenza di correlazione (il coefficiente angolare è zero). Se invece esiste una chiara relazione, il *Moran Scatterplot* potrà essere utilizzato per distinguere diverse tipologie di correlazione spaziale. Se, infatti, i punti sono per lo più nei quadranti NordEst e SudOvest avremo correlazione positiva e potremo distinguere fra quei casi (nel quadrante NordEst) dove la relazione si stabilisce per valori alti sia di x (la regione i) che di Wx (le regioni limitrofe a i), si parla in questo caso di relazione Alto-Alto, e quelli dove (nel quadrante SudOvest) i valori sono entrambi bassi dando luogo a

una relazione di tipo Basso-Basso. Se la concentrazione dei punti è maggiore nei due altri quadranti la correlazione è negativa. Inoltre per i punti riportati sul quadrante NordOvest avremo associati bassi valori di x e alti valori di Wx (relazione Basso-Alto) e viceversa nel quadrante SudEst (relazione Alto-Basso).

I risultati del *Moran Scatterplot* possono essere riportati su una mappa (si veda la figura 3) in modo da distinguere geograficamente le aree con le diverse tipologie di correlazione (Alto-Alto, Basso-Basso, Alto-Basso, Basso-Alto). In particolare in questo modo sarà possibile verificare se le regioni accomunate da un certo tipo di correlazione sono fra loro contigue e formano quindi un *cluster*. Il *Moran Scatterplot* ha anche l'importante funzione di mettere in evidenza i possibili casi limite (*outliers*) perché possano essere eventualmente esclusi dalla analisi se rappresentano casi anomali. La rappresentazione sulla mappa darà la possibilità di individuare esattamente le regioni anomale e di valutare se la loro posizione geografica ha una relazione con essa. Ciò sarà tanto più probabile, per esempio, se gli *outliers* sono tra loro vicini oppure sono per esempio isole o regioni di confine (dove le esternalità possono estendersi oltre i confini e non vengono colte all'interno di una analisi tradizionale con dati nazionali).

Figura 3. Moran Scatterplot Map



Fonte: nostre elaborazioni su dati CRENoS

Proseguendo nell'esercizio didattico proposto, la rappresentazione del *Moran Scatterplot* in figura 3 mostra due aree geograficamente distinte per le quali si registra, per ogni osservazione, una correlazione positiva del tipo Alto-Alto: i comuni affacciati sul Golfo di Cagliari e, più a nord, l'intera area nord orientale comprendente la Gallura e parte della Barbagia. La mappa evidenzia dunque il definirsi di due *cluster* di comuni per i quali un elevato valore delle immatricolazioni pro capite è associato ad un altrettanto elevato valore nelle unità territoriali adiacenti. Per contro si osserva nell'area centromeridionale del Campidano e nell'area a est di Cagliari (il Sarrabus – Gerrei) una macro-zona che si caratterizza per un'associazione positiva del tipo Basso-Basso, in cui si registra per ogni comune un basso valore della variabile considerata associata ad un basso valore delle osservazioni limitrofe.

Le osservazioni fino ad ora analizzate concorrono a definire la correlazione spaziale complessiva. Da queste si distinguono le aree caratterizzate da una correlazione negativa che, come era presumibile, si trovano situate nelle zone di confine tra le aree precedentemente definite. Si tratta di comuni nei quali ad un elevato (o basso) valore delle immatricolazioni pro capite corrisponde un basso (o elevato) valore delle stesse nelle zone contigue. Queste situazioni sono rilevabili nella zona intorno a Oristano, nella parte occidentale della provincia di Sassari e nella zona centro-orientale dell'Ogliastra. Coerentemente con l'impostazione della tecnica di associazione spaziale, risultano prive di intensità cromatica le unità territoriali insulari, che ovviamente non presentano osservazioni contigue, ossia i comuni di Carloforte (nell'isola di San Pietro) e di La Maddalena.

Il *Moran Scatterplot* ha il grande vantaggio di trasferire un indicatore medio valido per un insieme di regioni su un grafico che consente di esplorare meglio la sua natura distinguendo le regioni rispetto a diversi tipi di interdipendenza. In questa direzione è spesso utile associare a questo indicatore globale e alla sua rappresentazione

grafica anche un indicatore di autocorrelazione locale, in grado cioè che di misurare l'interdipendenza per ognuna delle regioni in esame. Il LISA (*Local Indicator of Spatial Association*) consente in modo efficace di associare ad ogni unità territoriale una misura del livello di associazione spaziale rispetto al suo intorno. Ovviamente si assume che la somma dei diversi LISA per ciascuna regione sia proporzionale alla corrispondente misura globale. L'indicatore solitamente utilizzato come LISA è lo stesso I di Moran calcolato a

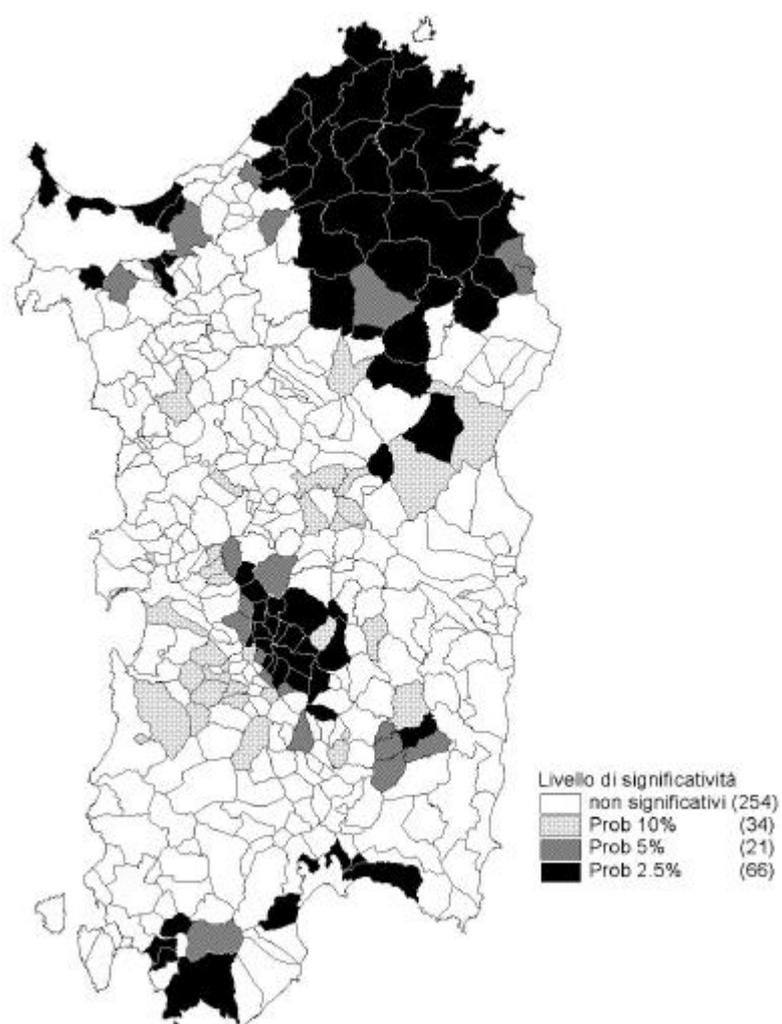
$$I_i = \frac{z_i}{\sum_j z_j^2 / N} \sum_{j \in J_i} w_{ij} z_j$$

livello locale i . L'espressione che caratterizza I_i è la seguente:

dove z_i è il valore corrispondente alla regione i della variabile già normalizzata e J è l'insieme delle regioni vicine a i . Anche in questo caso 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. Anche in questo caso la rappresentazione su una mappa può essere utile per individuare ulteriori *cluster* di regioni con correlazioni positive significative contrapposte a *cluster* di regioni con correlazioni negative significative. Potranno invece essere escluse quelle correlazioni che non raggiungeranno il livello di significatività che venivano invece riportate nella mappa del *Moran Scatterplot*.

Seguendo ancora l'esempio applicativo sulle immatricolazioni per 1.000 residenti il concetto può essere ulteriormente chiarito con l'ausilio della figura 4.

Figura 4. Indice LISA con vari livelli di significatività



Fonte: nostre elaborazioni su dati CRENoS

Seppure con diversi livelli di significatività, emergono tre agglomerazioni. Il primo, geograficamente concentrato nella porzione nord-orientale dell'isola, comprende la Gallura costiera ed i comuni interni, aree notoriamente caratterizzate da elevati livelli di benessere con buoni livelli di omogeneità. Il secondo interessa la parte occidentale della provincia di Oristano e un'altra area a cavallo tra le province di Cagliari e di Nuoro e si riferisce a una relazione positiva di tipo Basso-Basso. Infine, il terzo *cluster*, di tipo Alto-Alto, è rilevabile nei comuni di prima cintura che fanno da corona a Cagliari e trova conforto ancora una volta negli elevati livelli di reddito ed occupazione rispetto alla media regionale.

5. La dipendenza spaziale multivariata e il modello di regressione

Finora sono state presentate delle statistiche la cui funzione è quella di verificare la dipendenza spaziale per un unico fenomeno. Allarghiamo ora il nostro orizzonte di indagine per capire come studiare i casi in cui la dipendenza spaziale si estende a più fenomeni oppure alle loro componenti stocastiche. Ci riferiamo, cioè, all'applicazione dei modelli di regressione semplice e multipla corretti per tenere conto della possibile presenza di interdipendenza tra le unità spaziali.

L'interdipendenza spaziale si manifesta quando

$$\text{Cov}[y_i, y_j] = E[y_i y_j] - E[y_i] E[y_j] \neq 0 \quad \text{per } i \neq j$$

dove i e j si riferiscono a unità spaziali differenti e $y_{i(j)}$ è il valore assunto dal fenomeno oggetto di studio nella località $i(j)$. Questa covarianza diventa importante dal punto di vista dell'analisi spaziale quando la configurazione delle coppie i, j che non sono uguali a zero hanno una loro interpretazione in termini di distribuzione nello spazio o di interazione spaziale. Come in altri casi la stima della matrice della covarianza è complessa e necessita di una congettura sulla sua struttura. Tale congettura nella maggior parte dei casi è derivata attraverso la specificazione di un processo

spaziale stocastico¹⁷. Per fare questo la matrice dei pesi assume, ancora, un ruolo fondamentale perché consente di riformulare questi processi, in particolare quello autoregressivo e/o quello di medie mobili, tenendo conto della esistenza di relazioni multidirezionali e reciproche tra le unità spaziali (si veda Anselin, 1999, per un approfondimento tecnico).

Per ciò che concerne i modelli di regressione lineare la prima ovvia osservazione riguarda il metodo di stima tradizionale dei minimi quadrati ordinari (OLS). Questo metodo risulta, come nel caso delle serie storiche, inappropriato quando applicato allo studio di relazioni tra fenomeni spaziali trascurando l'esistenza di interdipendenza spaziale. Per correggere questa omissione è possibile incorporare la dipendenza spaziale in vari modi. In effetti la letteratura ha suggerito un grande numero di specificazioni dei modelli per i processi spaziali, successivamente implementati empiricamente. Questa grande varietà può apparire poco gestibile e dar l'impressione che ogni modello necessiti di un particolare contesto metodologico, ma è possibile fornire una loro riorganizzazione tassonomica, guidati dal principio che le tecniche econometriche possono essere applicate essenzialmente nella medesima maniera a differenti modelli raggruppati in base alle loro principali caratteristiche¹⁸.

Come detto in precedenza, la caratteristica principale dell'econometria spaziale è il modo in cui gli effetti spaziali vengono considerati. Chiaramente questo presuppone che lo spazio sia stato precedentemente formalizzato in qualche maniera. L'utilizzo di una matrice dei pesi spaziale rende possibile l'applicazione dei modelli spaziali a molti contesti empirici, dal momento che la dipendenza spaziale è opportunamente codificata

¹⁷ Si tratta di una relazione funzionale tra una variabile casuale localizzata in una data unità spaziale e la stessa variabile casuale in un'altra località.

¹⁸ Si noti che in questo lavoro, per ragioni di spazio, restringiamo la nostra classificazione ai modelli lineari di regressione.

nei pesi in essa espressi, e che l'eterogeneità spaziale è esplicitata nella specificazione del modello.

5.1 Tassonomia di modelli spaziali di regressione lineare.

In questa sezione presentiamo una specificazione generale che serva da quadro di riferimento per le successive modellizzazioni. Questa specificazione si riferisce alla situazione in cui le osservazioni sono disponibili per unità spaziali cross-section in un unico istante temporale. L'approccio che seguiremo ci permetterà di derivare specifici modelli mediante l'imposizione di condizioni ai parametri della formulazione generale. Il punto di partenza è la seguente espressione:

$$y = \rho W_1 y + X\beta + \varepsilon \quad (1)$$

$$\varepsilon = \lambda W_2 \varepsilon + \eta$$

con $\eta \sim N(0, O)$ e gli elementi diagonali della matrice O di covarianza degli errori $O_{ij} = h_i(z\alpha)$, dove:

β è un vettore $K \times 1$ di parametri associati alle variabili esplicative X (matrice $N \times K$);

ρ è il coefficiente della variabile dipendente ritardata spazialmente;

λ è il coefficiente in una struttura spaziale autoregressiva per il disturbo ε .

Il termine di errore η è distribuito normalmente con una matrice di covarianza diagonale O . Gli elementi diagonali consentono l'eteroschedasticità, come una funzione di $P+1$ variabili esogene Z , che includono una costante. I P parametri α sono associati con i termini non costanti, così che per $\alpha=0$ si ha $h=\rho^2$, cioè la situazione di omoschedasticità.

Le due matrici $N \times N$ W_1 e W_2 sono le matrici dei pesi spaziali (standardizzate o meno) associate rispettivamente al processo spaziale autoregressivo nella variabile dipendente e nel termine d'errore; questo consente che si possa esplicitare una differente struttura spaziale che guida i due processi.

Il modello ha $3+K+P$ parametri indeterminati. In forma vettoriale:

$$\theta = [\rho, \beta', \lambda, \rho^2, \alpha']$$

Diverse famiglie di strutture di modelli spaziali derivano quando un sottovettore del precedente è posto uguale a zero. Specificatamente le seguenti situazioni corrispondono ai modelli spaziali autoregressivi tradizionali presenti in letteratura.

1. per $\lambda = \rho = \alpha = 0$ (P+2 restrizioni)

$$y = X\beta + \varepsilon$$

il modello di regressione lineare senza effetti spaziali;

2. per $\lambda = \alpha = 0$ (P+1 restrizioni)

$$y = \rho W_1 y + X\beta + \varepsilon$$

il modello spaziale autoregressivo (Spatial LAG model).

È l'equivalente di un modello di serie temporali dove la variabile oggetto di studio y è correlata con la sua espressione ritardata. In questi casi si parla di autocorrelazione sostanziale (*substantive autocorrelation* nella terminologia di Anselin, 1988) e la sua soluzione è ovviamente l'inclusione della variabile omessa. Peraltro anche la sua semplice inclusione in un modello di stima tradizionale OLS è problematica.

3. per $\rho = \alpha = 0$ (P+1 restrizioni)

$$y = X\beta + (I - \lambda W_2)^{-1} \mu$$

il modello con disturbi spaziali autoregressivi (Spatial ERROR model), quando l'autocorrelazione è presente solo nel termine di errore.

Tale modello può essere espresso anche nel seguente modo:

$$y = \lambda W_2 + X\beta - \lambda W_2 X\beta + \mu$$

nella forma che, riprendendo la terminologia usuale per le serie storiche, è detta del fattore comune spaziale (*spatial common factor model*). Tale modello si contraddistingue per la presenza di una variabile esogena ritardata spazialmente tra i regressori, WX , e per una serie di vincoli non lineari (*common factor*) imposti sui coefficienti. In particolare, il prodotto del coefficiente autoregressivo λ con il coefficiente della regressione β deve essere uguale al coefficiente della variabile esogena WX espresso in negativo. In questo secondo caso si parla di vera e propria autocorrelazione degli errori (o *nuisance autocorrelation*). La soluzione sarà quindi l'esplicita specificazione della relazione spaziale degli errori e cioè della matrice delle covarianze $E[\varepsilon\varepsilon']$ ¹⁹.

4. per $\alpha = 0$ (P restrizioni)

$$y = \rho W_1 y + X\beta + (I - \lambda W_2)^{-1} \mu$$

il modello autoregressivo spaziale con disturbi autocorrelati, che è la combinazione dei due precedenti.

La tassonomia ora presentata si concentra principalmente sulla specificazione della dipendenza spaziale, perché, come accennato in precedenza, le forme più comuni di eterogeneità spaziale possono essere incorporata in maniera diretta mediante gli approcci che analizzeremo nelle sezioni successive di questo lavoro.

¹⁹ In questo caso, come è più frequente nelle applicazioni pratiche, la matrice delle covarianze $E(\varepsilon\varepsilon')$ è specificata con un processo stocastico autoregressivo (SAR). E' possibile peraltro utilizzare anche una specificazione per medie mobili (SMA) o anche una combinazione dei due (SARMA). Esattamente come avviene nel caso delle serie storiche.

5.2 Limiti della stima OLS nei modelli spaziali.

La dipendenza spaziale nei vari modelli autoregressivi mostra molte similitudini con la più conosciuta dipendenza temporale. Ci si potrebbe, quindi, aspettare che le proprietà degli stimatori OLS permangano valide anche nel caso di modelli con la dipendente ritardata spazialmente e/o con residui autocorrelati, ma ciò non accade. L'assenza di un'analogia diretta è da attribuirsi in primo luogo alla natura multidirezionale della dipendenza dello spazio.

Nell'econometria classica delle serie storiche, gli stimatori OLS rimangono consistenti anche in seguito all'introduzione della variabile dipendente ritardata, se il termine d'errore non è autocorrelato. Perciò, sebbene le proprietà dello stimatore risultino inficiate per piccoli campioni (poiché non più non distorto), esso può continuare ad essere utilizzato per l'inferenza asintotica su grandi campioni.

Per i modelli autoregressivi spaziali tale risultato non è valido, qualunque siano le proprietà dei residui. Diversamente dal caso delle serie storiche, infatti, il termine di ritardo spaziale Wy è correlato con i disturbi anche quando questi sono variabili casuali identicamente e indipendentemente distribuite (*i.i.d.*). In altre parole il ritardo spaziale della variabile dipendente deve essere sempre considerato una variabile endogena e la stima OLS inadeguata perché produce stimatori distorti e inconsistenti (tipici in presenza di problemi di simultaneità). Nella sezione che segue vedremo quali sono i metodi di stima alternativi utilizzabili.

Gli effetti dell'autocorrelazione spaziale dei residui sulle proprietà dello stimatore OLS sono invece più simili a quelli delle serie storiche. Le stime dei parametri saranno non distorte ma inefficienti, a causa della struttura non diagonale della matrice della varianza degli errori. Questo significa che la varianza dei parametri β sarà sovrastimata rendendo l'inferenza statistica basata su test t e sul coefficiente di determinazione R^2 non corretta. Allo stesso tempo altri test statistici come il test di Chow o il test per

l'eteroschedasticità saranno influenzati e potranno portare a conclusioni scorrette.

Si può considerare il modello spaziale come un caso particolare del modello di regressione lineare, con una particolare parametrizzazione della matrice di varianza del termine d'errore. Di conseguenza rimarranno valide le normali proprietà di OLS e GLS. Tuttavia, nel caso spaziale, la natura multidirezionale della dipendenza limita i tipi di procedure EGLS (Estimated Generalized Least Squares) che portano a delle stime consistenti. Nello specifico, questo si applica alle varie procedure EGLS in due fasi generalmente utilizzate nel caso di autocorrelazione delle serie e di eteroschedasticità. Si può dimostrare che le stime dei parametri autoregressivi ottenuti tramite OLS sono non consistenti. Conseguentemente, nel presente caso non è appropriato applicare la procedura di Cochrane-Orcutt.

6. Approcci alternativi per la stima di modelli spaziali

Dopo le considerazioni sopra esposte sull'inadeguatezza della stima OLS applicata ai modelli spaziali, presentiamo in questa sezione i metodi di stima alternativi proposti dalla letteratura. Analizzeremo in primo luogo l'approccio di massima verosimiglianza, dunque il metodo delle variabili strumentali e successivamente l'approccio robust; per quanto riguarda la dipendenza spaziale nei residui questo contributo propone l'analisi di due procedure iterative.

6.1 Stima di massima verosimiglianza (*Maximum Likelihood*)

L'inadeguatezza dello stimatore OLS ha fatto sì che l'attenzione si spostasse sulla stima di Massima Verosimiglianza come alternativa. Gli stimatori ML sono ottenuti dalla massimizzazione del logaritmo della funzione di verosimiglianza ricavata dal modello spaziale specificato sotto l'assunzione di normalità per la distribuzione degli errori. Le classiche proprietà dello stimatore ML (non distorsione, efficienza e normalità asintotica) sono valide per i modelli che includono la dipendente ritardata spazialmente,

sotto determinate condizioni²⁰. Queste riguardano l'esistenza della funzione di verosimiglianza per i valori dei parametri considerati; la derivabilità continua di questa funzione (fino al secondo e terzo ordine); l'esistenza di derivate parziali finite; l'esistenza, positività e non singolarità della matrice di covarianza.

Un ulteriore requisito, importante per la specificazione della dipendenza ed eterogeneità spaziale, è che il numero di parametri sia fisso ed indipendente dal numero delle osservazioni. Questo è per evitare il cosiddetto "*incidental parameter problem*", che rappresenta un problema comune nei modelli spazio-temporali. Ad esempio una situazione in cui ogni unità spaziale presenta un parametro che indichi le sue particolari caratteristiche, come una variabile *dummy*, non sarebbe ammissibile in un contesto in cui l'asintoticità è basata sulla dimensione *cross-section*; analogamente, risulterebbe inappropriato un modello in cui la dipendenza spaziale si accompagna ad un parametro di decadimento della distanza per ogni osservazione. Dal punto di vista operativo una situazione del genere è facilmente riconoscibile a causa della mancanza di gradi di libertà.

Per i modelli spaziali da noi considerati in questo lavoro le varie condizioni sono soddisfatte se la struttura dell'interazione spaziale, espressa congiuntamente dal coefficiente autoregressivo e dalla matrice dei pesi, non è esplosivo. Formalmente questo può essere misurato studiando le proprietà della Jacobiana associata ad ogni modello, ad esempio il determinante di $(I-\rho W)$ nella semplice formulazione autoregressiva.

Il principale problema del metodo ML sorge in relazione alla necessità di far ricorso a metodi numerici per risolvere i complessi processi di ottimizzazione non lineari. Questo approccio ha quindi un costo "computazionale" piuttosto elevato che lo rende di difficile, se non impossibile, applicazione quando si ha a che fare con un numero di osservazioni molto elevato. Inoltre il modello ML richiede necessariamente l'uso di un *software* realizzato

²⁰ Per una dimostrazione formale si rimanda al lavoro di Bates e Whites (1985).

appositamente per l'analisi delle serie spaziali (attualmente solo *SpaceStat* e *Matlab* sono disponibili sul mercato).

Il modello lineare con effetti spaziali nel termine di errore può essere considerato come un caso particolare del più generale modello con una matrice di covarianza degli errori non scalare.

La matrice di covarianza è generalmente non conosciuta, ed espressa in termini di un numero finito di parametri. Nel modello spaziale questi sono: il parametro autoregressivo, di decadimento della distanza, o i coefficienti associati alle variabili che specificano la forma dell'eteroschedasticità.

Quando si conosce la matrice di varianza degli errori, gli usuali stimatori Generalized Least Squares (GLS) sono BLUE (Best Linear Unbiased Estimator), così come quelli ML. Tuttavia, nella più realistica condizione in cui i parametri della varianza degli errori si basano su delle stime, solo lo stimatore ML presenta le proprietà desiderabili di consistenza, efficienza e normalità asintotiche. Inoltre, nella maggior parte dei casi, le stime per i parametri regolari del modello (i β della notazione comune) sono anche non distorti.

6.1.1 Test delle ipotesi per stime ML

Dato il diffuso utilizzo dell'approccio ML nella stima dei modelli spaziali, anche la maggior parte dei test delle ipotesi per i parametri di questi modelli si basano su considerazioni asintotiche²¹. In letteratura sono stati analizzati principalmente tre test: il Wald (W), il Likelihood Ratio (LR) e il Lagrange Multiplier (LM).

La maggior parte dei test delle ipotesi possono essere espressi come test sulle funzioni dei parametri del modello:

$$H_0: g(\theta) = 0$$

²¹ Bisogna constatare, però, che, al contrario di ciò che è avvenuto per la stima ML, ampiamente trattata dalla letteratura sull'econometria spaziale, i vari approcci per i test non hanno ricevuto la dovuta attenzione, specie in riferimento al loro utilizzo per piccoli campioni.

$$H_1: g(\theta) \neq 0$$

Dove g è una funzione matriciale (lineare o meno) degli elementi del vettore dei parametri θ . Come caso generale g include test sulla significatività dei singoli coefficienti, semplificandosi in un vettore con 1 in corrispondenza del coefficiente considerato, e 0 negli altri casi.

Nel modello spaziale generale ci si concentra sulla significatività di entrambi i processi autoregressivi (con $H_0: \rho = 0$, o $H_0: \lambda = 0$), sui parametri regressivi β , e sulla presenza di eteroschedasticità (con $H_0: \alpha = 0$).

In generale i tre test si basano su diverse misure della differenza tra una stima non ristretta e una stima che soddisfa le restrizioni imposte dall'ipotesi nulla, cioè una stima ristretta. Per esempio se si suddivide un vettore di parametri θ in $\theta' = [\theta_1' \mid \theta_2']$, l'ipotesi nulla può assumere la forma:

$$H_0: \theta_1 = 0$$

Una stima ristretta (θ_R) consiste nello stimare θ_2 , con tutti i parametri in θ_1 nulli, la stima non ristretta è invece il normale vettore completo θ . I test si basano sulla misura della differenza tra le due. Intuitivamente, se questa misura è troppo elevata non si può assumere che le restrizioni siano valide, per cui si rifiuta l'ipotesi nulla.

La maniera in cui i test misurano la differenza tra la stima ristretta e non ristretta ha implicazioni per il tipo di stima che è necessario effettuare: per il test Wald è necessario il modello completo, cioè sono utilizzati i parametri non ristretti; per il test LM il punto di partenza sono i coefficienti non ristretti, e dunque il modello è stimato in una specificazione più semplice con l'imposizione dell'ipotesi nulla; per il test LR, invece, sono necessarie entrambe le specificazioni.

La procedura più conosciuta è il test Wald, espresso nella sua forma più generale come:

$$W = g' [G'VG]^{-1}g$$

dove g è un vettore $q \times 1$ dei valori ottenuti da una stima ristretta ML,

G è una matrice $(3+K+P) \times q$ di derivate parziali per i parametri stimati,

V è la matrice della varianza asintotica stimata, di dimensioni $3+K+P$.

Il test LR è basato sulla differenza fra il *log-likelihood* per il modello generale e quello ristretto:

$$LR = 2 [L(\theta) - L(\theta_r)]$$

Il test LM è basato su un approccio di ottimizzazione, cioè è calcolato sulle condizioni del primo ordine per la massimizzazione di una funzione lagrangiana sul *log-likelihood*:

$$f = L(\theta) + \eta' g(\theta)$$

dove f è la lagrangiana,

L è il log-likelihood,

η è un vettore di moltiplicatori di Lagrange che corrisponde alle restrizioni q .

Tutti i tre test presentati sono asintoticamente distribuiti come un χ^2 con q gradi di libertà, dove q rappresenta il numero di restrizioni. È da notare, però, che per piccoli campioni mostrano valori differenti, che possono complicare la loro interpretazione. In generale è valida la seguente relazione:

$$W \geq LR \geq LM$$

per la quale un utilizzo non critico del test di Wald porta ad un rifiuto più frequente dell'ipotesi nulla, mentre il test LM porta ad una sua accettazione.

Per evitare alcuni problemi associati alla stima ML²² possono essere utilizzate diverse alternative.

6.2 Stima con variabili strumentali (IV)

Sotto vari punti di vista, il metodo di stima con variabili strumentali presenta proprietà asintotiche simili all'approccio ML, con il vantaggio che la sua implementazione numerica è molto più semplice. Nonostante queste proprietà desiderabili, tuttavia, la sua applicazione ai modelli spaziali è stata esigua.

Come detto, l'inadeguatezza dello stimatore OLS nei modelli spaziali autoregressivi è dovuta alla correlazione tra la variabile spaziale e la variabile d'errore a prescindere dalle proprietà di quest'ultima. Questa situazione è simile alla stima di parametri in un sistema d'equazioni simultanee, dove la dipendenza tra le variabili endogene e i termini d'errore è la causa della mancanza di consistenza degli OLS.

Intuitivamente, l'approccio IV è basato sull'esistenza di un set di strumenti Q , che sono fortemente correlati con le variabili originarie $Z = [y_L, X]$, ma non correlate asintoticamente col termine d'errore.

Questo approccio può dunque trovare applicazione al caso dei modelli con variabile dipendente ritardata spazialmente, se si può approntare un appropriato set di strumenti, con il vantaggio che la stima può essere effettuata con l'utilizzo dei pacchetti econometrici standard²³.

Le proprietà dello stimatore IV non sono solamente asintotiche. Per contro, però, sebbene consistente, spesso non è lo stimatore più efficiente: l'efficienza dipende in maniera determinante sulla scelta appropriata degli strumenti. È da considerare, inoltre, che la

²² In particolare ci riferiamo alla complessità numerica dell'ottimizzazione non lineare e alle condizioni restrittive dei parametri.

²³ Sebbene il caso qui analizzato comprenda la variabile ritardata spazialmente come unica causa dei problemi, l'approccio IV è applicabile anche ad una formulazione più generale, ad esempio al caso in cui alcune delle altre variabili esplicative sono variabili stocastiche o endogene in un sistema di equazioni.

normalità asintotica non si estende alla distribuzione di un campione finito, così che l'inferenza sugli effettivi *data set* empirici può essere problematica. Infine la determinazione degli strumenti dal punto di vista empirico può rappresentare un problema: nella teorica essa dovrebbe essere basata su precise ipotesi, mentre nella pratica spesso si presentano poche indicazioni da seguire.

Idealmente gli strumenti dovrebbero essere fortemente correlati alle variabili originarie e non correlati asintoticamente con i termini d'errore. Sebbene non esista una regola generale per la verifica, il primo requisito può essere facilmente verificato ad esempio tramite il calcolo della semplice correlazione. La scelta del numero degli strumenti può essere guidata da tecniche simili: mentre asintoticamente il numero degli strumenti non è importante, per campioni finiti alcuni limiti pratici sono imposti da problemi di multicollinearità e riguardano i gradi di libertà.

Il requisito di non correlazione asintotica tra gli strumenti ed i termini d'errore è più difficile da stimare. In alcuni semplici casi nei quali gli strumenti consistono in valori stimati la mancanza di correlazione può essere mostrata analiticamente. Quando il problema è più complesso possono essere utilizzati diversi test per verificare l'esogeneità degli strumenti, come ad esempio l'approccio Wu-Hausman²⁴.

Anselin (1980) propone diverse soluzioni per il modello generale di autoregressione spaziale, ad esempio l'utilizzo del *predicted value* ritardato spazialmente di una regressione della variabile dipendente sui regressori non spaziali, oppure dei ritardi spaziali delle variabili esogene del modello, sebbene questo possa comportare problemi di multicollinearità. In generale, quasi ogni combinazione delle variabili esogene, ritardate spazialmente o no, da risultati soddisfacenti.

²⁴ Dato che questi test possono essere applicati al caso spaziale in maniera analoga a quello classico, in questa sede non verranno trattati in maniera esplicita.

I vari test di esogeneità approntati in letteratura possono anche essere utilizzati per verificare la serietà del problema determinato dalla presenza della variabile dipendente ritardata spazialmente. In alcune situazioni, infatti, questo effetto non giustifica l'utilizzo di un approccio IV o ML. Sebbene consistente, la stima IV per piccoli campioni può essere inferiore a quella OLS in termini di errore quadratico medio. Un compromesso tra le stime OLS, distorte ma con un errore quadratico medio potenzialmente minore, e le stime IV, consistenti, si ottiene con i cosiddetti stimatori ristretti, delle *k*-classi. In termini generali questi sono costruiti come media ponderata dei due stimatori, nella quale i pesi sono da ricercarsi in un *range* ottimale. In alcuni casi questi stimatori mostrano una migliore *performance* e l'applicazione di questo approccio in un contesto spaziale dovrebbe essere automatica, anche se sulle sue proprietà si deve ancora indagare.

6.3 L'approccio *robust* ai processi spaziali

Al contrario della stima IV, l'approccio robusto è stato spesso utilizzato per le analisi spaziali. Nello specifico analizziamo una particolare tecnica robusta, il *bootstrap*, per l'inferenza statistica nei modelli di dipendenza spaziale.

In generale la tecnica di *bootstrap*, e il relativo *jackknife*, sono esempi di tecniche di ricampionamento, il cui principio è di utilizzare la casualità presente in data set ricampionati creati artificialmente, come base per l'inferenza statistica. Questo porta ad ottenere stime dei parametri e misure della distorsione e della varianza alternative, e alla costruzione di livelli di pseudo-significatività e di intervalli di confidenza.

Nel caso del *bootstrap* si effettuano le stime per uno svariato numero di data set, ottenuti dal ricampionamento *random* con ripetizione delle osservazioni originarie. Di conseguenza la stessa osservazione può essere inclusa più di una volta nel pseudo- *data*

set, oppure neanche considerata²⁵. Di conseguenza gli pseudo dati sono assunti indipendenti e distribuiti normalmente.

La statistica che interessa è poi calcolata per ogni replicazione degli pseudo dati, e la sua distribuzione di frequenza è utilizzata per ricavare la stima dei parametri, della distorsione e della varianza. Si considerino per esempio R ripetizioni di un campione casuale da un data set x, che determinano gli pseudo-data set x_1, x_2, \dots, x_R . Per ognuno di essi si ottiene la stima dei parametri che interessano $\theta = f(x)$ (una qualsivoglia statistica espressa in funzione delle osservazioni), e quindi $\theta_1, \theta_2, \dots, \theta_R$. La stima *bootstrap* per il parametro sarà:

$$\theta_B = (1/R) \sum_r \theta_r$$

cioè la media della stima su tutte le ripetizioni. La varianza ad esso associata è misurata come:

$$\text{var} [\theta_B] = [1/(R-1)] \sum_r (\theta_r - \theta_B)^2$$

La differenza tra la stima *bootstrap* per tutte le ripetizioni e la stima dei parametri del data set iniziale, fornisce una misura della distorsione di quest'ultima.

In uno fra i vari approcci che conducono ad una stima *bootstrap* proposti dalla letteratura il ricampionamento è basato su un set di residui ottenuti al primo stadio da una regressione, ad esempio mediante OLS. In termini formali il modello di stima sull'universo è:

$$\begin{aligned} y &= X\beta + e && \text{con residui OLS:} \\ e &= y - X\hat{\beta} \end{aligned}$$

dove $\hat{\beta}$ è il vettore della stima OLS. Per i residui si assume una funzione empirica di distribuzione che assegna un'eguale probabilità ad ognuno, cioè ogni residuo avrà

²⁵ Lo schema di campionamento evidenziato assegna un'eguale probabilità ad ogni osservazione, cioè $1/N$ se si considerano N osservazioni.

probabilità $1/N$ (con N osservazioni) di essere riutilizzato nello schema di ricampionamento²⁶.

Si costruisce una ripetizione *bootstrap* da un set di residui estratti casualmente (con ripetizione) in combinazione con i parametri stimati nel primo stadio.

$$y_r = e_r + Xb$$

dove: e_r è il vettore dei residui ricampionati,

b è la stima del primo stadio

X è la matrice delle osservazioni sulle variabili esogene fisse.

Una stima per β di questo pseudo-*data set*, che indichiamo con b_r , si ottiene con lo stesso metodo applicato all'intero universo. Questo processo è ripetuto un elevato numero di volte (R) per generare una distribuzione di frequenze empirica di b_r . La media di questa distribuzione empirica rappresenta la stima *bootstrap* per β .²⁷

6.4 Stima in presenza di errori autocorrelati spazialmente

La dipendenza spaziale nei termini d'errore in una regressione lineare comporta una matrice di covarianza degli errori non sferica $s^2O(?)$. Di conseguenza la varianza associata alle stime OLS per i coefficienti del modello non presenterà la forma usuale $s^2(X'X)^{-1}$ ma sarà una funzione complessa del parametro $?$. Conseguentemente, sebbene le stime OLS permangano non distorte, l'inferenza basata sulla stima della varianza può essere fuorviante. Se la covarianza degli errori è conosciuta, la *Best Linear Unbiased Estimate* è la *Aitken Generalized Least Squares* (AGLS). Questo implica che sia la struttura della dipendenza spaziale (la W) così come il coefficiente associato (?) siano noti, ma nella pratica questo non si verifica, tranne che nelle simulazioni artificiali Monte Carlo.

²⁶ Sebbene sia ben noto che i residui OLS non siano indipendenti, li si assume tali ai fini del ricampionamento.

²⁷ Per quanto riguarda la il metodo di ricampionamento e la scelta dello stimatore da utilizzare nella fase iniziale e nella stima degli pseudo dataset creati si rimanda ad Anselin (1988).

Il metodo Estimated GLS si basa su stime consistenti dei parametri di O , in sostituzione dei valori sconosciuti della popolazione. La natura asintotica dei risultati della stima EGLS è un punto spesso ignorato nei lavori empirici. Nei campioni finiti la distribuzione della stima non è ben definita, né la EGLS è necessariamente superiore alla OLS in termini dell'errore quadratico medio. Inoltre le proprietà della stima EGLS sono sensibili ad una corretta specificazione di $O(?)$, che nei modelli spaziali è determinata essenzialmente dalla scelta della matrice dei pesi.

6.4.1 Due procedure iterative

In analogia alle conosciute procedure di Cochrane-Orcutt (1949) e Durbin (1960) per gli errori autocorrelati delle serie storiche, sono stati proposti due approcci simili per il modello spaziale ad opera di Hordijk (1974), Bartels (1979) e Anselin (1980).

L'analogo spaziale della metodologia Cochrane-Orcutt, consiste in tre fasi: si applica la stima OLS sul modello lineare. I residui sono quindi adoperati per ottenere una stima del coefficiente autoregressivo spaziale ρ , che successivamente è sostituito nella procedura EGLS. Questo è ripetuto sino all'ottenimento della convergenza.

L'approccio Durbin è basato sull'equivalenza formale del modello ad una specificazione spaziale autoregressiva, in analogia a quanto suggerito da Durbin per il caso delle serie storiche. Formalmente:

$$y = X\beta + (I - \rho W)^{-1}\mu, \quad \text{o:} \quad y = \rho Wy + X\beta + \mu$$

dove: μ è un termine d'errore con caratteristiche desiderabili.

Una stima LM o un approccio IV produrranno stime consistenti per ρ e β . La procedura di stima può essere interrotta a questo punto, o portata avanti con un'ulteriore EGLS che si basa sulla stima di ρ . Le stime ML o IV hanno le consuete proprietà asintotiche, mentre per piccoli campioni non sono disponibili i

risultati esatti. Il principale vantaggio dell'approccio ML è che consente l'applicazione dei test Wald, Likelihood Ratio e Lagrange Multiplier.

6.5 Alcune note pratiche

Infine alcuni accenni alla strategia di selezione del modello "migliore" una volta che è stata accertata la presenza di autocorrelazione spaziale e si è proceduto con la stima ML o con un'altra tecnica capace di ovviare ai problemi della stima OLS.

La prima guida per una buona selezione del modello è sempre la teoria. In particolare se questa consente di esprimere una forma ridotta, che contenga al suo interno la dipendenza spaziale sia in riferimento a variabili endogene sia esogene, la migliore strategia è quella di verificare direttamente la significatività dei ritardi spaziali inseriti nel modello.

Se invece non esiste un modello teorico a priori e si procede attraverso tecniche di *data mining* una possibile strategia è quella di trasferire la tecnica di sovrapparametrizzazione o *general to specific*, dal contesto temporale dell'econometria tradizionale a quello spaziale. Questo metodo prevede l'inserimento di un insieme di variabili che possano essere correlate temporalmente (spazialmente) a diversi livelli di ritardo (contiguità). Se la matrice dei pesi è quella delle contiguità binarie introdurremo variabili ritardate anche per livelli superiori al primo. Se invece si tratta di una matrice delle distanze ciò non sarà necessario.

Questo metodo ha quindi tre fasi:

- 1) individuazione dell'insieme S di variabili che per ragioni teoriche o in ragione della osservazione empirica potrebbero essere correlate spazialmente con il fenomeno in esame.
- 2) Stima del modello generale con l'insieme S incluso con il metodo appropriato (OLS se non è inclusa la variabile endogena tra i regressori, ML o IV se invece è inclusa)
- 3) Test della significatività individuale delle variabili che fanno parte di S. Il processo va avanti se non tutte le variabili sono significative attraverso la loro eliminazione (una alla volta partendo

da quella con il t-statistico più basso) fino a giungere a una specificazione che include solo variabili significative.

Anche in questo caso non mancano i problemi legati soprattutto al fatto che la sequenza di eliminazione condiziona i risultati ottenuti. Il pregio sta nel fatto che viene comunque ridotta al minimo la discrezionalità del ricercatore.

Infine si noti che, nell'ambito delle tecniche di stima spaziali, il coefficiente di determinazione R^2 perde il suo significato originale di quota della varianza spiegata. Al suo posto può essere utilizzato il criterio di Akaike che penalizza quelle specificazioni che contengono un numero maggiore di variabili senza che vi sia un vantaggio in termini di precisione della stima.

6. L'eterogeneità spaziale

Come detto in precedenza, molti tratti caratteristici dell'eterogeneità sono affrontati mediante l'uso delle tecniche dell'econometria standard. In questa sezione, pertanto, ci concentriamo su quegli aspetti che sono peculiari alla dimensione spaziale, introducendo il metodo dell'espansione spaziale dei parametri e la variazione dei coefficienti *random*. In molte situazioni, la base per la specificazione della forma particolare dell'eterogeneità nei modelli spaziali può essere derivata dalla teoria che può fornire un'interpretazione sulle caratteristiche dei *data set* spaziali che sono probabili cause dell'eterogeneità, ed indicare variabili importanti che determinino la sua forma. Un'ulteriore difficoltà nell'analisi spaziale è che la misspecificazione del modello e gli errori di misurazione che possono comportare eteroschedasticità (come il problema della scelta dell'unità spaziale d'osservazione) sono anche frequenti cause di autocorrelazione spaziale.

7.1 Espansione spaziale dei parametri

Da un punto di vista econometrico, il metodo dell'espansione spaziale può essere considerato una variazione sistematica dei

coefficienti in un modello di regressione. L'eterogeneità del fenomeno analizzato viene riflessa nei valori dei parametri che differiscono per ogni osservazione. Questa differenziazione è espressa come funzione di un certo numero di variabili ausiliarie che portano ad una formulazione del modello più complessa. Nella terminologia del metodo dell'espansione la semplice specificazione originaria omogenea è chiamata modello iniziale, mentre la formulazione eterogenea complessa prende il nome di modello finale.

Nelle prime applicazioni del metodo dell'espansione le variabili ausiliarie consistevano in polinomi "Trend Surface", rappresentati dalle coordinate dell'allocatione delle osservazioni, da cui il termine espansione spaziale. Più recentemente si è giunti a forme più complesse d'espressione del *trend surface* in termini delle sue componenti principali ortogonali.

Le proprietà del metodo dell'espansione spaziale possono essere illustrate formalmente, in termini di una regressione semplice con una variabile esplicativa. Il modello iniziale è:

$$y = \beta_0 + \beta_1 X + e$$

dove β_0 e β_1 sono i coefficienti di regressione
 X è il vettore di osservazioni sulla variabile esplicativa.

L'eterogeneità si riflette nella mancanza di stabilità dei parametri sulle unità d'osservazione. Questa è incorporata nel modello assumendo ogni parametro individuale (o un sottoinsieme dei parametri) come funzione esatta di un numero finito di variabili di espansione, ad esempio z_1 e z_2 :

$$\beta_1 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2$$

La sostituzione del parametro espanso β_1 nella formulazione originaria porta al modello finale:

$$y = \beta_0 + (\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2) x + e$$

o:

$$y = \beta_0 + \beta_1 x + \beta_2 (z_1 x) + \beta_3 (z_2 x) + e$$

Se il modello finale espanso rappresenta la specificazione corretta, la stima dei parametri del modello iniziale è distorta a causa del problema della variabile omessa.

In una tipica applicazione empirica del metodo dell'espansione, il grado dell'appropriato polinomio *trend surface* è determinato da una serie di regressioni a stadi (*stepwise*). Sebbene spesso ignorato nei lavori empirici, i problemi associati alla ricerca di questo tipo di specificazione ad hoc sono seri, e tendono ad invalidare il contesto probabilistico formale sul quale si basa l'inferenza. Inoltre, tendenzialmente le variabili espansive mostrano un alto grado di multicollinearità, il che diminuisce la precisione della stima.

Riassumendo, il metodo dell'espansione spaziale fornisce un metodo semplice ed interessante per considerare l'eterogeneità nei coefficienti nel modello di regressione, fornendo inoltre un'esplicita rappresentazione spaziale dell'instabilità nei parametri. Tuttavia la sua implementazione dev'essere effettuata con cautela, specialmente quando non esistono forti motivazioni a priori che guidino nella scelta delle variabili d'espansione.

Nei lavori empirici l'esistenza di una relazione esatta tra i coefficienti e la loro espansione spaziale risulta essere un'assunzione forte. Perciò la ricerca del giusto grado del polinomio d'espansione e l'uso delle componenti principali implicano l'esistenza di un termine d'errore stocastico. Questo porta alla seguente espressione formale:

$$\beta_1 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \mu$$

dove μ è un termine d'errore stocastico distribuito normalmente e non autocorrelato, con varianza (σ_μ^2).

Sostituendo nel modello iniziale:

$$y = \beta_0 + \beta_1 x + \beta_2 (z_1 x) + \beta_3 (z_2 x) + e$$

che rappresenta un modello con errori eteroschedastici, dove con:

$$e = \mu x + e$$

si è indicato il nuovo termine d'errore.

In presenza di errori non sferici²⁸ nel modello iniziale, la varianza dell'errore nel modello terminale risulta essere una funzione delle variabili esplicative x . Di conseguenza l'inferenza sui parametri del modello finale che ignora l'eteroschedasticità dei termini d'errore può essere fuorviante. Ciò è importante in questo contesto, poiché la significatività dei parametri β_1 e β_2 determina la valutazione dell'instabilità strutturale.

7.2 Variazione dei coefficienti random

Questo approccio è stato proposto da Arora e Brown (1977) come *soluzione* al problema dell'autocorrelazione spaziale. Si tratta in effetti di una sopravvalutazione di queste tecniche che non risolvono l'autocorrelazione ma sono adatte a contemplare l'eterogeneità nella sua forma generale che può avere, anche se non necessariamente, una peculiare interpretazione in termini spaziali. Nelle applicazioni empiriche in cui non sono disponibili variabili che esprimano in maniera ovvia la specifica forma della variazione spaziale nei coefficienti di regressione, un approccio è il modello dei coefficienti *random* di Hildreth - Houck (1968) dove i coefficienti di osservazione individuale sono interpretati come una estrazione casuale da una distribuzione multivariata. Formalmente per ogni osservazione i si ha:

$$y_i = x_i' \beta_i$$

dove x_i' è un vettore riga di variabili esplicative
 β_i è un vettore $K \times 1$ di coefficienti determinati
come:

$$\beta_i = \beta + \mu_i$$

il cui termine d'errore μ è distribuito secondo una normale k -variata. Questa distribuzione ha media 0 e matrice di covarianza S ,

²⁸ Errori non sferici sono quelli che violano le ipotesi classiche del modello econometrico.

che può essere diagonale nel caso particolare di assenza di covarianza tra gli errori dei parametri individuali. Sostituendo si ottiene l'usuale forma con termine di errore eteroschedastico ϵ_i :

$$y_i = x_i' \beta + \epsilon_i$$

dove $\epsilon_i = x_i' \mu_i$ e $E[\epsilon_i] = 0$; $V[\epsilon_i] = x_i' S x_i$

I parametri del modello possono essere stimati con l'approccio ML o con diverse procedure iterative EGLS. Nelle situazioni operative con un modello scarsamente specificato può insorgere un problema se la stima degli elementi di S non portano ad una matrice di covarianza definita positivamente. La funzione per la variazione dei coefficienti random può essere definita prima della stima di una specificazione più complessa, per mezzo di test per questa forma particolare di eteroschedasticità, come il test Breusch-Pagan.

Chiaramente il modello dei coefficienti *random* espresso in questa forma rappresenta un caso particolare di eteroschedasticità, e per questo motivo non si può affermare che esso risolva il problema della dipendenza spaziale. Per incorporare quest'ultima è necessaria una specificazione più complessa. Ad esempio si può separare la casualità dell'intercetta da quella degli altri coefficienti, escludendo quindi l'interazione tra l'errore della costante e degli altri elementi casuali. La dipendenza spaziale sarà allora riflessa nel termine di errore dell'intercetta nella sua forma usuale, ad esempio come processo autoregressivo.

Formalmente il modello può essere espresso come:

$$y_i = a + x_i' \beta_i + e_i$$

dove: e_i è il disturbo associato all'intercetta.

Il modello generale per l'errore può essere espresso come:

$$\epsilon_i = e_i + \eta_i$$

con $e_i = \eta_i W_i + F_i$

7. Applicazioni empiriche

Per completare il discorso sulle differenti specificazioni dei modelli di econometria spaziale, in questa sezione presentiamo una rassegna empirica di differenti metodi di stima e test analizzati da un punto di vista formale nelle sezioni precedenti.

Effettueremo l'analisi di dati *cross-section* mediante un semplice modello sulle determinanti del reddito per i comuni della Sardegna, per il quale saranno applicate le metodologie presentate precedentemente.

Il modello che sarà utilizzato per l'analisi dei dati è una semplice espressione lineare nella quale il reddito pro capite nei comuni della Sardegna (REDD) è espresso in funzione della densità della popolazione (DENS), della quota di addetti nel settore della manifattura e dei servizi (INDSER), della quota di possessori di titolo di studio superiore (diploma o laurea) sul totale della popolazione (LAUDIP), del tasso di disoccupazione (TDIS), e dei consumi energetici per usi produttivi (CONS). Le osservazioni si riferiscono ai 375 comuni della Sardegna, e i dati sono relativi al 1991; le variabili sono state analizzate previa loro standardizzazione. Ovviamente ci si attende che tra non tutte le variabili vi sia una correlazione positiva, ipotizzando che il tasso di disoccupazione presenti un coefficiente di segno negativo.

L'equazione da stimare può essere così rappresentata:

$$\text{REDD} = \alpha_1 + \alpha_2 \text{DENS} + \alpha_3 \text{INDSER} + \alpha_4 \text{LAUDIP} + \alpha_5 \text{TDIS} + \alpha_6 \text{CONS}$$

Vogliamo sottolineare che questo modello è utilizzato essenzialmente per illustrare i diversi effetti spaziali, e quindi ha finalità prettamente metodologiche; non va dunque interpretato come un contributo significativo alla comprensione delle determinanti territoriali del reddito. Lo strumento informatico utilizzato è *SpaceStat*, cui si è fatto cenno nella prima parte di questo contributo.

Gli effetti spaziali sono incorporati in duplice maniera: da una parte la dipendenza spaziale per mezzo della matrice dei pesi,

dall'altra l'eterogeneità nella forma dell'eteroschedasticità, della variazione dei parametri e del cambiamento strutturale spaziale. Per quanto riguarda il primo punto abbiamo considerato la contiguità di primo ordine tra i comuni con il criterio *rook*, mediante una matrice dei pesi standardizzata per riga. La variazione spaziale dei parametri è formulata nel metodo dell'espansione come funzione delle coordinate dei centroidi. Una modificazione spaziale della struttura è introdotta distinguendo tra comuni costieri e non costieri, espressa mediante una variabile *dummy*.

8.1 Regressione OLS con diagnostica per la dipendenza spaziale

L'utilizzo della stima OLS per la stima del modello enunciato in precedenza conduce ai risultati riportati nella Tavola 2:

Tav. 2 – Regressione OLS

VARIABLE	COEFF	S.D.	t-value	Prob
CONSTANT	4.60353E-07	0.032883	0.000014	0.999989
DENS	0.195	0.0368826	5.287039	0.000000
INDSER	0.408968	0.0382827	10.682859	0.000000
LAUDIP	0.326128	0.043615	7.477419	0.000000
TDIS	-0.120293	0.0352612	-3.411496	0.000718
CONS	0.0618778	0.0330605	1.871654	0.062045

R2 = 0.6010	SIG-SQ = 0.405484
R2-adj= 0.5956	SIG-SQ(ML) = 0.398996
LIK = -359.826	AIC = 731.653
	SC = 755.214

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Koenker-Bassett test	5	2.421028	0.788339

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	55.290539	0.000037

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.196622	6.338968	0.000000
LM (error)	1	36.998597	0.000000
Robust LM (error)	1	9.048145	0.002630
LM (lag)	1	33.186807	0.000000
Robust LM (lag)	1	5.236355	0.022120
LM (SARMA)	2	42.234952	0.000000

I coefficienti stimati sono fortemente significativi, tranne per il caso della variabile CONS, che raggiunge una significatività solo per un livello di significatività del 10%. La regressione ha inoltre un buon potere esplicativo in base all' R^2 e all' R^2 aggiustato che raggiungono il valore 0.60. Sono riportate anche due statistiche per la varianza dei residui, la prima non distorta e la seconda di massima verosimiglianza. Poiché quest'ultima è ottenuta dividendo la somma dei quadrati dei residui per il numero totale delle osservazioni, invece che dei gradi di libertà, sarà necessariamente minore della prima.

Nella tavola sono riportate le statistiche sull'eteroschedasticità. Poiché non si è specificata una variabile che controlli questo problema, il test è stato condotto per la variazione di coefficienti random. Poiché il modello include 5 variabili esplicative (escludendo la costante), queste rappresentano i gradi di libertà per il test Koenker-Basset che, come si nota, non è significativo. Questo però non è confermato dal test di White che, con un valore maggiore di 55 è altamente significativo con $Prob \ll 0.01$, evidenziando potenziali problemi non mostrati dalla precedente statistica. Un punto da tenere a mente in tutte le situazioni in cui sono presenti sia eteroschedasticità che dipendenza spaziale, è che i test per l'eteroschedasticità sono molto sensibili alla presenza di dipendenza spaziale: in altre parole, mentre questi test segnalano eteroschedasticità, in realtà il problema è quello dell'autocorrelazione spaziale.

Per quanto riguarda la diagnostica spaziale, il tradizionale test per la presenza di autocorrelazione nel termine di errore è basato sull' I di Moran calcolato per i residui, che nel nostro caso mostra una $Prob$ nulla, ad indicare termini di errore fortemente correlati. Tutti gli altri test sono altamente significativi, evidenziando la difficoltà ad individuare quale può essere la corretta specificazione alternativa, se il modello con dipendente ritardata spazialmente o quello che include il termine autoregressivo degli errori.

8.2 Stima ML del modello spaziale autoregressivo (LAG model)

Nella Tavola 3 sono indicati i risultati della stima del modello autoregressivo che include la variabile W_REDD ritardata spazialmente, necessaria in quanto il test LM nella tavola precedente da forti indicazioni sulla possibile presenza di autocorrelazione spaziale nella variabile dipendente.

Tav. 3 – Stima ML del LAG model

VARIABLE	COEFF	S.D.	z-value	Prob
W_REDD	0.259257	0.0496233	5.224508	0.000000
CONSTANT	-0.00969662	0.0312768	-0.310026	0.756541
DENS	0.14416	0.0357863	4.028361	0.000056
INDSER	0.358534	0.0374635	9.570213	0.000000
LAUDIP	0.338656	0.0414109	8.177943	0.000000
TDIS	-0.0962558	0.0337521	-2.851844	0.004347
CONS	0.0538593	0.0313948	1.715545	0.086245

R2 = 0.6248	SIG-SQ = 0.365361	Sq. Corr. = 0.6347
LIK = -345.724	AIC = 705.448	SC = 732.936

DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	ALUE	PROB
Breusch-Pagan test	5	72.228909	0.000000
Spatial B-P test	5	72.228941	0.000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
SPATIAL LAG DEPENDENCE FOR WEIGHTS MATRIX FULL			
TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	28.205012	0.000000

LAGRANGE MULTIPLIER TEST ON SPATIAL ERROR DEPENDENCE					
WEIGHT	STAND	ZERO	DF	VALUE	PROB
FULL	yes	yes	1	4.834910	0.027889

I risultati della stima ML sono riportati assieme allo pseudo R^2 e alla correlazione quadratica tra valori osservati e stimati, che però hanno un valore limitato e non devono essere utilizzati per effettuare una comparazione con i risultati della stima OLS. Invece le misure basate sull'approccio della verosimiglianza (LIK, AI, e SC) sono utili per un paragone delle due regressioni. Da queste risulta che la stima migliora nettamente con l'introduzione del ritardo spaziale nel modello, come indicato dall'aumento del *log-likelihood* (rispettivamente da -359.8 per l'OLS a -345.7) e dalla diminuzione sia nell'AIC che nel SC. Questo miglioramento poteva comunque essere intuito dato che il coefficiente autoregressivo risulta fortemente significativo.

I coefficienti per le altre variabili della regressione sono simili ai precedenti, evidenziando esclusivamente un aumento della significatività per CONS.

Anche in questo caso è riportata la diagnostica per l'eteroschedasticità, che mostra che questo problema permane e ancora necessita di essere affrontato.

A seguire, nella parte della diagnostica spaziale, è indicato con 28.2 il valore del test LR sul coefficiente autoregressivo, che ha $\text{Prob} \ll 0.01$ ed è quindi, come detto poco sopra, fortemente significativo. Da ultimo analizziamo il test LM per l'autocorrelazione dei residui, che, nonostante mostri un aumento della Prob fino a 0.03, permane significativo, ad indicare che la dipendenza spaziale non è stata completamente eliminata con l'introduzione del termine autoregressivo.

8.3 Stima ML del modello con dipendenza spaziale nei residui (ERROR model)

Analizziamo ora i risultati della stima del modello che introduce la dipendenza spaziale nel termine di errore.

Tav. 4 – Risultati della stima ML per lo Spatial error model

VARIABLE	COEFF	S.D.	z-value	Prob
CONSTANT	0.00754328	0.0537518	0.140335	0.888395
DENS	0.141817	0.0389142	3.644367	0.000268
INDSER	0.373006	0.0386347	9.654697	0.000000
LAUDIP	0.363321	0.0416313	8.727098	0.000000
TDIS	-0.108225	0.0335183	-3.228848	0.001243
CONS	0.0512506	0.0302363	1.695001	0.090075
LAMBDA	0.434449	0.0648882	6.695346	0.000000

R2 = 0.5502 SIG-SQ = 0.350052 Sq. Corr. = 0.5984

LIK = -342.525 AIC = 697.050 SC = 720.611

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	66.752106	0.000000
Spatial B-P test	5	66.752276	0.000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHTS MATRIX FULL

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	34.602803	0.000000

TEST ON COMMON FACTOR HYPOTHESIS

TEST	DF	VALUE	PROB
Likelihood Ratio Test	5	7.054553	0.216615
Wald Test	5	7.145407	0.210049

LAGRANGE MULTIPLIER TEST ON SPATIAL LAG
DEPENDENCE

WEIGHT	STAND	ZERO	DF	VALUE	PROB
FULL	yes	yes	1	0.611594	0.434188

Anche in questo caso per effettuare una comparazione tra risultati ottenuti con il metodo ML e OLS bisogna comparare i valori delle tre statistiche *likelihood*, che mostrano un aumento del valore e il test LIK e una diminuzione dell'AIC e SC, indicando un miglioramento della stima, confermato dall'altissima significatività del coefficiente λ della dipendenza del termine di errore. Questa stima risulta migliore anche di quella del lag model, come poteva essere dedotto dalle indicazioni del test LM del modello standard (LM error > LM lag).

I test per l'eteroschedasticità continuano a segnalare la presenza di questo problema nella regressione.

Nella parte successiva il test LR sul coefficiente autoregressivo conferma la sua elevata significatività. I successivi test sui fattori comuni, LR e Wald, hanno invece una Prob > 10%, segnalando che non si verifica una inconsistenza della specificazione del modello (se entrambi fossero stati significativi avremmo dovuto interpretare inappropriato l'error model).

8.4 Stima dello Spatial Durbin model

Come affermato precedentemente, un metodo alternativo per la stima di un modello con errori autocorrelati è rappresentato dal metodo Durbin. Questo equivale all'applicazione delle tecniche ML in un modello che include, oltre alla dipendente, tutte le variabili esplicative ritardate spazialmente.

I risultati sono riportati nella Tavola 5.

Tav. 5 – Stima dello Spatial Durbin model

VARIABLE	COEFF	S.D.	z-value	Prob
CONSTANT	0.00694642	0.0308248	0.225352	0.821706
DENS	0.118059	0.04018	2.938256	0.003301
INDSER	0.357833	0.0392248	9.122602	0.000000
LAUDIP	0.366096	0.0417811	8.762251	0.000000
TDIS	-0.103846	0.0336086	-3.089875	0.002002
CONS	0.0519988	0.0305784	1.700511	0.089035
W_REDD	0.401123	0.0665543	6.027005	0.000000
W_DENS	0.0586807	0.0575662	1.019361	0.308032
W_INDSER	-0.062173	0.071311	-0.871856	0.383287
W_LAUDIP	-0.221854	0.0717256	-3.093091	0.001981
W_TDIS	0.0315346	0.0631503	0.499358	0.617527
W_CONS	0.02022	0.0735823	0.274795	0.783474

R2 = 0.6300	Sq. Corr. = 0.6546	SIG-SQ	0.345652
LIK = -338.998	AIC = 701.995	SC = 749.118	

DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	10	94.070721	0.000000
Spatial B-P test	10	94.070875	0.000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
SPATIAL LAG DEPENDENCE FOR WEIGHTS MATRIX FULL			
TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	30.185143	0.000000

LAGRANGE MULTIPLIER TEST ON SPATIAL ERROR DEPENDENCE					
WEIGHT	STAND ZERO	DF	VALUE	PROB	
FULL	yes	yes	1	2.705476	0.100004

Mentre il coefficiente ρ è evidentemente molto significativo, quello relativo al lag delle variabili esplicative presenta una $\text{Prob} < 0.05$ solo per LAUDIP. Inoltre molti dei segni sono opposti a quelli attesi.

L'eteroschedasticità continua a persistere.

La parte più interessante dei risultati è fornita invece dal test LM sulla correlazione spaziale degli errori, che indica che il problema della dipendenza spaziale è stato positivamente risolto mediante l'applicazione di questo metodo di stima.

8.5 Il metodo dell'espansione spaziale

Uno dei modi in cui l'eterogeneità spaziale può essere incorporata in un modello è quello della espansione spaziale. Per il nostro esercizio didattico sono state introdotte 10 nuove variabili, costruite moltiplicando le esplicative originarie per le coordinate X e Y del centroide relativo alla rispettiva osservazione²⁹. Questa espansione lineare è solo una delle possibili specificazioni, ma poiché il modello presentato in questo lavoro ha solamente propositi illustrativi, non riteniamo necessaria una formulazione più complessa.

I risultati sono riportati nella Tavola seguente.

²⁹ Il centroide rappresenta il punto baricentrico di una regione, le cui coordinate possono essere facilmente calcolate mediante il *software ArcView*.

Tav. 6 – Il metodo dell'espansione spaziale

SPATIAL EXPANSION OF COEFFICIENTS-
FOLLOWING EXPANSION VARIABLES X AND Y COORD

VARIABLE	COEFF	S.D.	t-value	Prob
CONSTANT	0.0340377	0.0389818	0.873170	0.383154
DENS	-4.24897	5.93648	-0.715739	0.474618
INDSER	-0.599706	3.09296	-0.193894	0.846369
LAUDIP	-1.55435	3.54255	-0.438765	0.661096
TDIS	0.780781	2.95514	0.264211	0.791769
CONS	-26.5078	13.657	-1.940968	0.053045
A_DENS	1.47417E-06	3.47288E-06	0.424480	0.671470
B_DENS	8.54661E-07	1.14161E-06	0.748644	0.454562
A_INDSER	2.26339E-06	1.35496E-06	1.670452	0.095702
B_INDSER	-3.52451E-08	7.08887E-07	-0.049719	0.960374
A_LAUDIP	-1.78403E-06	1.45542E-06	-1.225789	0.221082
B_LAUDIP	6.25526E-07	7.74158E-07	0.808009	0.419621
A_TDIS	1.4975E-06	1.25665E-06	1.191661	0.234181
B_TDIS	-3.7012E-07	6.46675E-07	-0.572343	0.567448
A_CONS	1.23453E-06	4.14707E-06	0.297688	0.766113
B_CONS	5.99425E-06	3.36069E-06	1.783635	0.075328

R2 = 0.6176 SIG-SQ = 0.399463
R2-adj = 0.6016 SIG-SQ(ML) = 0.382419
LIK = -351.870 AIC = 735.739SC = 798.570

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER

1469.464044

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	28119.496270	0.000000

TEST ON EXPANSION OF COEFFICIENTS

TEST	DF	VALUE	PROB
F-test	10 359	1.556222	0.118044

Tav. 6 Cont.

EXPANSION OF INDIVIDUAL COEFFICIENTS

TEST	DF	VALUE	PROB
DENS	2 359	0.287335	0.750432
INDSER	2 359	1.431656	0.240274
LAUDIP	2 359	1.127006	0.325145
TDIS	2 359	0.894492	0.409724
CONS	2 359	2.746280	0.065515

DIAGNOSTICS FOR HETEROSKEDASTICITY

LINEAR SPECIFICATION USING VARIABLES

CONSTANT DENS INDSER LAUDIP TDIS

CONS

TEST	DF	VALUE	PROB
Koenker-Bassett test	5	2.514772	0.774269

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.178700	5.997733	0.000000
Lagrange Multiplier (error)	1	30.561184	0.000000
Robust LM (error)	1	6.655161	0.009887
Kelejian-Robinson (error)	16	94.767705	0.000000
Lagrange Multiplier (lag)	1	30.150829	0.000000
Robust LM (lag)	1	6.244807	0.012456
Lagrange Multiplier (SARMA)	2	36.805990	0.000000

L'introduzione di nuove variabili ha una scarsa influenza sulla bontà della regressione, come mostrato da un incremento solo marginale dell' R^2 ed R^2 aggiustato, questo anche a causa della non significatività dei coefficienti delle nuove esplicative, siano esse riferite alla coordinata X o alla Y. I coefficienti non espansi sono anch'essi non significativi, e mostrano un cambiamento di segno. È da sottolineare che una tale specificazione avrebbe sicuramente

mostrato effetti differenti se applicata, ad esempio, al contesto nazionale, in cui una differenziazione nord-sud ha un preciso significato anche dal punto di vista economico.

Nonostante quanto detto il modello sembra però risolvere positivamente il problema dell'eteroschedasticità, dato che il test di Koenker-Bassett non è più significativo.

Un'ultima considerazione riguarda la misura in cui un'espansione spaziale come quella effettuata riesce a contenere gli altri effetti spaziali. Nell'ultimo riquadro, dove è riportata la diagnostica spaziale, tutti i test portano però a concludere che la dipendenza spaziale non è stata eliminata in questa specificazione.

8.6 Analisi della stabilità strutturale in un Error model

Una differente prospettiva per affrontare il problema dell'eterogeneità spaziale è la considerazione della stabilità strutturale. Nel modello qui utilizzato questo è effettuato mediante due diverse regressioni, di cui una per i comuni costieri, e l'altra per quelli interni. Tale scelta, sicuramente arbitraria, non lo è però più di quella di considerare separatamente i comuni settentrionali e quelli meridionali, dato che nessuna indicazione ci è pervenuta in questo senso dalla significatività dei coefficienti delle variabili espansive lungo la coordinata Y.

La stima del cosiddetto modello non ristretto, nel quale i coefficienti per i diversi raggruppamenti di comuni possono assumere valori differenti, è effettuata con il metodo ML nella specificazione dell'Error model³⁰. I risultati sono riportati nella Tavola 7.

³⁰ Risultati non dissimili si ottengono anche con la stima OLS

Tav. 7 – Stima dell'Error model nella specificazione non ristretta

COMUNI INTERNI				
VARIABLE	COEFF	S.D.	z-value	Prob
CONSTANT	-0.0510449	0.0422595	-1.207891	0.227089
DENS	0.334895	0.0807686	4.146357	0.000034
INDSER	0.365319	0.0434602	8.405829	0.000000
LAUDIP	0.240643	0.0495987	4.851797	0.000001
TDIS	-0.160342	0.0371244	-4.319040	0.000016
CONS	0.255738	0.173156	1.476922	0.139696
LAMBDA	0.149325	0.0839009	1.779775	0.075113

DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	5	62.582458	0.000000
Spatial B-P test	5	62.582782	0.000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
SPATIAL ERROR DEPENDENCE FOR WEIGHTS MATRIX INT			
TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	2.695491	0.100632

TEST ON COMMON FACTOR HYPOTHESIS			
TEST	DF	VALUE	PROB
Likelihood Ratio Test	5	5.385959	0.370613
Wald Test	5	5.361474	0.373375

LAGRANGE MULTIPLIER TEST ON SPATIAL LAG DEPENDENCE					
WEIGHT	STAND	ZERO	DF	VALUE	PROB
INT	yes	yes	1	0.001760	0.966534

Tav.7 Cont.

COMUNI COSTIERI				
VARIABLE	COEFF	S.D.	z-value	Prob
CONSTANT	0.164703	0.13755	1.197402	0.231150
DENS	0.0417998	0.0300925	1.389046	0.164819
INDSER	0.295407	0.0638897	4.623710	0.000004
LAUDIP	0.513977	0.0543046	9.464700	0.000000
TDIS	0.0379384	0.0581895	0.651979	0.514415
CONS	0.0312213	0.0206159	1.514428	0.129917
LAMBDA	0.654218	0.0757494	8.636606	0.000000

DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	5	6.063909	0.300050	
Spatial B-P test	5	6.063927	0.300048	

DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL ERROR DEPENDENCE FOR WEIGHTS MATRIX COST				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	34.348233	0.000000	

TEST ON COMMON FACTOR HYPOTHESIS				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	5	5.002256	0.415605	
Wald Test	5	3.855045	0.570469	

LAGRANGE MULTIPLIER TEST ON SPATIAL LAG DEPENDENCE					
WEIGHT	STAND	ZERO	DF	VALUE	PROB
COST	yes	yes	1	0.474877	0.490752

Le stime mostrano risultati abbastanza differenti rispetto al modello ristretto, riferito cioè all'intero campione utilizzato. Prima le variabili DENS e TDIS presentavano coefficienti significativi (il primo positivo e il secondo negativo), mentre ora sono significativi solo per i comuni interni. Il coefficiente λ perde di valore nella prima specificazione, mentre nella seconda è maggiore di quello del modello ristretto.

Test per la stabilità strutturale sono l'LR e Wald. Sia per i comuni interni che per quelli costieri presentano valori della Prob >> 0.05, che escludono l'inconsistenza delle due specificazioni. Di conseguenza, quando si considera la presenza di dipendenza spaziale, è confermato l'idea di una instabilità strutturale nel modello ristretto.

Note bibliografiche

Anselin L. (1980), *Estimation Methods for Spatial Autoregressive Structures*. Ithaca, NY: Cornell University, Regional Science Dissertationa and Monograph Series #8.

Anselin L. (1988), *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht/Boston/London.

Anselin L. (1999), *Spatial Econometrics*, Bruton College, University of Texas at Dallas, mimeo

Arora S. e Brown M. (1977), *Alternative Approaches to Spatial Autocorrelation: An Improvement over Current Practice*. International Regional Science Review, 2, pp. 67-78.

Arbia G. (1996), *Modelling the spatial pattern and the diffusion of individual firms*, Quaderni di Base, N. 1/96, Università' Commerciale "Luigi Bocconi"

Arbia G. (1999), *A method-of moment procedure for estimating the spatial correlogram with sample data combined with remotely sensed information*, Atti SCO'99I, Venezia

Atzeni S. e Usai S. (2002), *Econometria Spaziale. Note teoriche ed empiriche*. In Note di Lavoro CRENoS – Atlantis.

Bartels C. (1979), *Operational Statistical Methods for Analysing Spatial Data*. In Ketellapper, pp. 5–50, Boston: Martinus Nijhoff.

Bates C. e Whites H. (1985), *A Unified Theory of Consistent Estimation for Parametric Models*. Econometric Theory, 1, pp.151-178.

Besag J. (1974), *Spatial Interaction and the Statistical Analysis of Lattice Systems*. Journal of the Royal Statistical Society B, 36, pp. 192-236.

Bottazzi L. e Peri G. (2000), *Innovation, Demand and Knowledge Spillovers: Theory and Evidence from European Regions*, CEPR WP. N. 2279

Bottazzi L. e Peri G. (2001) *Innovation and Spillovers: Evidence from European Regions*, CESifo Working Paper, #340, March 2001

Bottazzi L. e Peri G. (2002) *Innovation and Spillovers in Regions: Evidence from European Patent Data* IGIER Working Paper Series, N°215, June 2002

Burridge P. (1980), *On the Cliff – Ord Test for Spatial Correlation*. Journal of the Royal Statistical Society B, 42, pp. 107-108.

Clark G.L, Feldman M.P. e Gertler M.S. (2000), *The Oxford Handbook of Economic Geography*, Oxford University Press

Cliff A. e Ord J. (1973), *Spatial Autocorrelation*. London: Pion.

Cliff A. e Ord J. (1981), *Spatial Processes, Models and Applications*, London: Pion

Cochrane D. e Orcutt G. (1949), *Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms*. Journal of the American Statistical Association, 44, pp. 32 – 61.

Durbin J. (1960), *Estimation of Parameters in Time – Series Regression Models* Journal of the Royal Statistical Society B, 22, pp. 139-153.

Getis A. e Ord J. (1992), *The Analysis of Spatial Association by use of Distance Statistics*, Geographical Analysis, 17, 81-88

Hildreth C. e Houck J. (1968), *Some Estimators for a Linear Model with Random Coefficients*. Journal of the American Statistical Association, 63, pp. 584-595.

Hordijk L. (1974), *Spatial Correlation in the Disturbance of a Linear Interregional Model*. *Regional Science and Urban Economics*, 4, pp. 117-140.

Johnston J. (1984), *Econometric Methods*, New York, Mac Graw Hill

Kelejian H.H. e Robinson D.P. (1993), A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model, *Papers in Regional Science*, 72, 297-312.

Kmenta J. (1971), *Elements of Econometrics*, New York, Mac Millan

Krugman P. (1991), *Economic Geography and Trade*. Cambridge, MA: MIT Press.

Maciocco G., Pittaluga P. e Usai S. (a cura di) (2003), *La Nuova Economia Geografica: note teoriche e risvolti empirici*. Territori e Progetto, Franco Angeli, Milano

Marshall A. (1890), *Principles of Economics*, London, MacMillan.

Moran P. (1948), *The Interpretation of Statistical Map*, *Journal of Royal Statistical Society B*, 10, 243 – 251

Moreno R. e Vayà E. (2000), *La Utilidad de la Econometría Espacial en el Ambito de la Ciencia Regional*, Barcellona, FEDEA, Documentos de Trabajo 2000-13

Moreno R. e Vayà E. (2000), *Técnicas econométricas para el tratamiento de datos espaciales: La Econometría Espacial*. Edicions de la Universitat de Barcelona, UB 44 Manuals.

Moreno R., Paci R. e Usai S. (2003), *Spatial Distribution of Innovation Activity. The Case of European Regions*, *Contributi di Ricerca CRENoS 03/10*, Cagliari

Paci R. e Usai S. (2001), *Externalities and local economic growth in manufacturing industries*, Contributi di Ricerca CRENoS 01/13, Cagliari

Paci R. e Pigliaru F. (2000), *Externalities, knowledge spillovers and the spatial distribution of innovation*, Contributi di Ricerca CRENoS 00/2, Cagliari

Paci R. e Pigliaru F. (2001), *Technological Diffusion, Spatial Spillovers And Regional Convergence In Europe*, Contributi di Ricerca CRENoS 01/1, Cagliari

Paelinck J. e Klassen L. (1979), *Spatial Econometrics*, Saxon House, Farnborough.

Sforzi F., Openshaw S e Wymer C. (1997), *La procedura di identificazione dei sistemi locali del lavoro*, in I Sistemi Locali del Lavoro in Italia 1991, Roma, ISTAT, Argomenti n. 30, pp 235 – 242

Tobler W. (1979), *Cellular Geography*. In *Philosophy in Geography*, edited by S. Gale and G. Olsson, pp. 379-86. Dordrecht: Reidel.

Upton G. e Fingleton B. (1985), *Spatial Data Analysis by Example*, New York, Wiley

Usai S. (2003), *La Nuova Economia Geografica: note teoriche e risvolti empirici*, Università di Sassari e CRENoS, mimeo.

Whittle P. (1954), *On Stationary Processes in the Plane*. *Biometrika*, 41, pp. 434-449.