*Francesco Vidoli
**Claudio Mazziotta

Disaggregazione di indicatori spaziali sintetici secondo un approccio di tipo Chow-Lin

Parole chiave: **Analisi spaziale, Disaggregazione di indicatori sintetici, Modelli di tipo Chow-Lin.**

Abstract Il presente lavoro si propone di mettere a punto una procedura statistica che consenta il passaggio di indicatori sintetici dal livello aggregato a quello disaggregato, laddove i termini aggregato/disaggregato fanno riferimento, in questo caso, alla dimensione territoriale anziché, come più frequente, a quella settoriale. Più specificamente, con riferimento alla dotazione infrastrutturale presente in Italia si è cercato di ricostruire gli indicatori a livello provinciale sulla base degli analoghi indicatori disponibili a livello regionale. A tal fine si è seguito un approccio metodologico di tipo Chow-Lin, utilizzando quali regressori rappresentativi della domanda territoriale di infrastrutture alcune variabili relative agli aspetti produttivi, demografici e di uso turistico del territorio. Il confronto tra i risultati così ottenuti (ossia gli indicatori infrastrutturali "stimati") e i dati "effettivi" (ossia gli indicatori di dotazione infrastrutturale disponibili a livello provinciale) ha consentito un'interessante e non banale verifica della coerenza tra fattori territoriali di domanda e offerta di infrastrutture in Italia.

INTRODUZIONE

Obiettivo del presente lavoro è la verifica della capacità dei modelli di disaggregazione spaziale, derivati dall'approccio di tipo Chow-Lin, di consentire la ricostruzione sufficientemente affidabile degli indici di dotazione infrastrutturale ad un certo livello territoriale, partendo dalla disponibilità degli stessi indicatori ad un livello territorialmente superiore. Ciò significa, in altri termini, verificare se il livello di dotazione infrastrutturale di determinate unità territoriali (in questo caso, delle province italiane) possa essere "spiegato" da quelli che potrebbero essere considerati i naturali fattori di generazione della domanda di infrastrutture, ossia dai fattori di natura demografica ed economica.

Più specificamente, dopo aver ottenuto, sulla base dell'approccio indicato, una stima degli indicatori provinciali di dotazione infrastrutturale, si è proceduto a confrontare tali indicatori con quelli "effettivi", disponibili a seguito delle statistiche fornite dall'ISTAT (ISTAT, 2006). Qualora tale confronto rendesse evidente una buona approssimazione tra dati effettivi e dati stimati, si potrebbe dedurre che l'offerta provinciale di infrastrutture è conforme ai fattori di generazione della corrispondente domanda; a conclusioni opposte si dovrebbe pervenire in caso di *mismatch* tra le due serie di dati.

*Francesco Vidoli, Università degli Studi Roma Tre, Dipartimento di Istituzioni pubbliche, Economia e Società (fvidoli@sose.it)
**Claudio Mazziotta, Università degli Studi Roma Tre, Dipartimento di Istituzioni pubbliche, Economia e Società (c.mazziotta@uniroma3.it)

L'APPROCCIO METODOLOGICO UTILIZZATO

Al fine di derivare gli indicatori a livello territorialmente disaggregato dai corrispondenti indicatori di livello aggregato[1], si propone un modello basato sull'approccio presentato da Chow-Lin (Chow-Lin, 1971), che si fonda, come è noto, su tre assunzioni principali:

- *structural similarity*: il modello aggregato e quello disaggregato sono strutturalmente simili, il che comporta che le relazioni tra le variabili considerate sono le stesse a livello aggregato e disaggregato, con la conseguenza che i parametri di regressione sono gli stessi nei due modelli;
- *error similarity*: gli errori spazialmente correlati presentano la stessa struttura a livello sia aggregato sia disaggregato. Ciò equivale a dire che le correlazioni spaziali non sono significativamente diverse ai due livelli;
- *reliable indicators*: le variabili assunte quali regressori presentano un buon potere predittivo a livello sia aggregato che disaggregato, ossia le misure di bontà della regressione (R^2 o F test) risultano significativamente diverse da zero.

Da rilevare che il mancato rispetto dell'assunzione 1 conduce a stime sistematicamente distorte; quello dell'assunzione 2 implica effetti di *spillover* che influenzano pesantemente le stime e il mancato rispetto dell'assunzione 3 comporta che le stime disaggregate riflettono la semplice proporzione di quelle aggregate.

Modelli di questo tipo sono stati prevalentemente utilizzati per la costruzione di serie temporali mensili o trimestrali a partire da corrispondenti serie annuali, ma più recentemente sono stati impiegati anche in alcune applicazioni di disaggregazione spaziale. In Italia, in particolare, va menzionato il lavoro di Bollino e Polinori (Bollino e Polinori, 2007), che ricostruisce il Valore Aggiunto a livello comunale in Umbria dal punto di vista della convergenza tra territori suburbani e municipalità urbane che beneficiano di un elevato livello di sviluppo, spiegata da fattori di contiguità spaziale e da meccanismi di agglomerazione.

Il modello è caratterizzato, da un lato, da una relazione funzionale tra indicatori sintetici a livello provinciale ed una serie di variabili esplicative osservabili a livello disaggregato (e, ovviamente, anche a livello aggregato), relazione sottoposta ad una prima verifica in un precedente lavoro, sempre con riferimento alla dotazione infrastrutturale (Mazziotta e Vidoli, 2009b) e, d'altro lato, da una metodologia di inferenza dei parametri incogniti a livello regionale.

Il modello si basa sull'assunzione che a livello disaggregato sia valida la relazione ecometrica (lineare) del tipo seguente:

$$y_d = \boldsymbol{X}_d \, \beta_d + \varepsilon_d \tag{1}$$

in cui: y_d rappresenta un vettore (n*1) di osservazioni degli indicatori sintetici a livello disaggregato, $\boldsymbol{X}_d$ è una matrice (n*k) di osservazioni di k variabili esplicative osservabili a livello disaggregato e n è il numero di province.

1 È bene precisare che in questo lavoro il termine "indicatore aggregato" fa riferimento ad una misura di livello territorialmente superiore (regionale, ad esempio) rispetto ad analoga misura di livello inferiore (provinciale, ad esempio), e non deve pertanto essere confuso con il termine "indicatore sintetico" (o "composto"), che rappresenta invece il risultato di un'operazione di aggregazione di indicatori semplici, considerati sempre allo stesso livello territoriale. Formalmente, la relazione tra indicatori di diverso livello territoriale può essere scritta come

$$I_a = \frac{\sum_{i=1}^{n} I_{di}}{n}, \quad \forall i \in I\,(a)$$

in cui I (a) rappresenta l'unità territoriale di livello superiore (la regione) in cui le unità di livello inferiore (le province) sono incluse.

Si assume inoltre che $\boldsymbol{C}$ sia una matrice di dimensione (n*N), in cui n è il numero delle province italiane, capace di trasformare le osservazioni disaggregate in quelle di livello superiore (indicate con N), quale che sia l'operatore di aggregazione utilizzato.
In particolare, se si adotta l'operatore somma, le stime regionali sono ottenute per somma dei corrispondenti livelli provinciali ($y_a = \sum y_d$) e il generico elemento $C_{i,j}$ sarà costruito come:

$$C_{i,j} = \begin{cases} 1, \text{ se provincia } i \in \text{regione } j \\ 0, \text{ altrimenti} \end{cases}$$

Se invece si adotta l'operatore media aritmetica, C sarà costruito come:

$$C_{i,j} = \begin{cases} \frac{1}{k}, \text{ se provincia } i \in \text{regione } j, \text{ in cui } k = \text{numero di province appartenenti alla regione } j \\ 0, \text{ altrimenti} \end{cases}$$

e le stime regionali saranno ottenute attraverso la media delle stime provinciali[2] ($y_a = E(y_d)$).

Pertanto, assumendo le ipotesi di *structural similarity* ($\beta_d = \hat{\beta}_a$), è possibile scrivere:

$$y_a = X_a \beta_d + \varepsilon_a \tag{2}$$

Sotto i seguenti vincoli aggregativi $y_a = Cy_d$, $X_a = CX_d$ e $\varepsilon_a = C\varepsilon_d$.

Recentemente, Polasek e Sellner (2008) hanno presentato un avanzamento, o meglio un'interessante generalizzazione del modello, consistente nell'introduzione di un termine[3] di autocorrelazione spaziale in un classico modello di regressione multivariata (equazione (2)). Da un punto di vista applicativo, ciò significa che il livello della variabile dipendente Y in una determinata area dipende non soltanto dalle variabili indipendenti considerate, ma anche dal livello della stessa variabile Y nelle aree circostanti.
In realtà, se si assume l'esistenza di effetti di correlazione spaziale non soltanto nei livelli di competitività tra province, ma anche e soprattutto all'interno di province molto simili, si può ipotizzare (vedi, ad esempio, Anselin, 1988) che, dati una matrice di pesi spaziali $\boldsymbol{W}_N$ e un parametro di *spatial lag* $\rho \in [0,1]$, sia verificabile a livello disaggregato una "*relazione mista auto regressiva e regressiva spaziale*":

$$y_d = \rho_d W_N y_d + X_d \beta_d + \varepsilon_d \quad \text{con} \quad \varepsilon_d \sim N[0, \sigma_d^2 I_N] \tag{3}$$

La forma ridotta dell'equazione (3) ci consente di apprezzare meglio la componente spaziale attraverso cui è stato filtrato il contributo di $\boldsymbol{X}_d$.

2 In questo caso specifico si usa la media aritmetica.

3 Per una prima introduzione alle applicazioni statistiche alla pianificazione urbana si veda: Bickman *et al.* (1998), per elaborazioni più recenti si veda. Ayuga-Téllez *et al.* (2011).

$$y_d = (I - \rho_d\ W_N)^{-1} X_d \beta_d + (I - \rho_d W_N)^{-1} \varepsilon_d \tag{4}$$

Più specificamente, questo filtro spaziale risulta essere proporzionale alla distanza tra le unità territoriali considerate; infatti, se si sviluppa in serie l'espressione $(I - \rho_d\ W_N)^{-1}$ (analogamente a quanto si fa per l'inversione della matrice di Leontief) si ottiene:

$$E\,(y_d \mid X_d) = (1 + \rho_d W_N + \rho_d^2 W_N^2 + \ldots)\, X_d\, \beta_d \tag{5}$$

La forma dell'equazione (5) consente più agevolmente di verificare che tutte le aree contigue sono coinvolte nella stima di y_d e che tale coinvolgimento viene quantificato attraverso un coefficiente di proporzionalità rispetto alla distanza (*distance decay*).

Si può a questo punto riscrivere la forma ridotta dell'equazione (4), con $R_N = (I - \rho_d\ W_N)$.

$$y_d = R_N^{-1} X_d \beta_d + R_N^{-1} \varepsilon_d \quad , \ \varepsilon_d \in N\,[0, \Sigma_d] \tag{6}$$

con la Σ_d matrice di covarianza uguale a:

$$\Sigma_d = \sigma_d^2\, (R'_N\, R_N)^{-1} \tag{7}$$

I termini incogniti dei modelli a livello disaggregato sono dunque ρ_d, β_d e la covarianza σ_d^2. Per la stima di tali incognite è possibile, nel rispetto delle ipotesi di base, sfruttare la relazione tra y e $\boldsymbol{X}$ a livello aggregato e stimare il relativo modello misto auto regressivo (vedi, ad esempio, LeSage, 1998) nella forma:

$$y_a = \rho_a\ W_N\, y_a + CX_d\, \beta_a + \varepsilon_a \ , \ \ \varepsilon_a \sim N\,[0, \Sigma_a^2\, I_N] \tag{8}$$

ottenendo così $\hat{\rho}_a$ e $\hat{\sigma}_a^2$.

Coerentemente con le assunzioni di *structural similarity* ($\rho_d = \hat{\rho}_a$, $\beta_d = \hat{\beta}_a$) e di *error similarity* ($\sigma_d^2 = \hat{\sigma}_a^2$), è possibile sostituire i parametri stimati nelle equazioni (3) e (6).

Per quanto riguarda la stima di β_a secondo il classico approccio di tipo Chow-Lin si ottiene[4]:

$$\hat{\beta}_{a,\,GLS} = (X'_a\, (C\hat{\Sigma}_d\, C')^{-1})\, X_a)^{-1}\, X'_a\, (C\hat{\Sigma}_d\, C')^{-1}\, y_a \tag{9}$$

per cui la stima di y_d a livello disaggregato può essere determinata secondo l'espressione:

$$\hat{y}_d = \underbrace{R_N^{-1} X_d\, \hat{\beta}_a}_{1°\ \text{term.}} + \underbrace{\hat{\Sigma}_d\, C'\, (C\hat{\Sigma}_d\, C')^{-1}\, (y_a - C\, \hat{R}_N^{-1}\, C'\, X_a\, \hat{\beta}_a)}_{2°\ \text{term.}} \tag{10}$$

4 Si noti che $\hat{\beta}_{a,\,GLS}$ non dipendono da $\hat{\sigma}_a^2$, ma piuttosto da $\hat{\rho}_a$.

Il primo termine dell'equazione (10) rappresenta dunque la stima *naïve* del vettore incognito y_d , mentre il secondo termine rappresenta la distribuzione dell'errore della stima a livello aggregato attraverso la "*gain projection matrix*" ***G*** (Goldberger, 1962):

$$G = \hat{\Sigma}_d C' (C\hat{\Sigma}_d C')^{-1} \tag{11}$$

Questa quantità dipende in modo cruciale dal parametro *spatial lag* $\hat{\rho}_a$ a livello aggregato; se infatti poniamo $\hat{\rho}_a = 0$, la matrice $\hat{\Sigma}_d$ diventa pari alla matrice identità e si riduce alla trasposta della *projection matrix*: $G = C' (CC')^{-1}$ ritornando al modello base di cui all'equazione (1); il parametro ρ_d e la matrice $\boldsymbol{W}_N$ permettono quindi che la 1/N parte del residuo a livello aggregato non venga assegnato in parte uguale a tutti i comuni, ma che venga filtrato attraverso la distanza spaziale tra gli stessi.

UN'APPLICAZIONE AGLI INDICATORI DI DOTAZIONE INFRASTRUTTURALE

La possibilità di applicare[5] il modello sopra illustrato richiede la disponibilità delle seguenti informazioni: i) indicatori di sintesi della dotazione infrastrutturale a livello regionale (ed anche a livello provinciale, per consentire la successiva verifica dell'accuratezza della stima); ii) variabili demografiche ed economiche correlate con i fabbisogni infrastrutturali, disponibili a livello provinciale e regionale. Gli indicatori sub i) sono costruiti attraverso un particolare metodo di sintesi di indicatori elementari (questi ultimi di fonte ISTAT) messo a punto dagli autori in un precedente lavoro (Mazziotta and Vidoli, 2009a)[6]. Le variabili sub ii) sono riportate nella Tabella 1.

Tabella 1 Fattori di generazione della domanda di infrastrutture (fonti utilizzate)

Variabile	Fonte	Anno
Prodotto interno lordo	Ist. Tagliacarne	2007
Quota di popolazione residente in comuni con più di 50 mila abitanti	ISTAT	2007
Ricettività di alberghi di categoria medio bassa per abitante	ISTAT	2006

Sulla base di tali dati si è proceduto alla stima di un modello misto auto regressivo spaziale a livello aggregato (regionale) $y_a = \rho_a W_N y_a + CX_d \beta_a + \varepsilon_a$ al fine di ottenere un buon livello di adattamento secondo l'ipotesi 3 di "*reliable indicators*". In definitiva, l'applicazione è stata impostata in ossequio a due criteri: da un lato, un criterio economico consistente nella selezione delle variabili in grado di spiegare logicamente la variabile dipendente; dall'altro, un criterio statistico, consistente nella scelta di un modello con buona capacità predittiva e soddisfacente significatività statistica.

5 L'analisi applicativa è stata sviluppata in R (*package spdep*). La codifica, scritta dagli autori, è disponibile, su richiesta, presso gli stessi.

6 Per gli indicatori elementari si veda ISTAT "*Atlante statistico territoriale delle infrastrutture*", (2008), disponibile al link http://www3.istat.it/dati/catalogo/20080805_01/. La logica che lega sviluppo territoriale e dotazione infrastrutturale poggia sull'ipotesi che dal punto di vista dell'analisi regionale (si veda, in particolare, l'approccio del potenziale di sviluppo regionale portato avanti da Biehl, 1994) le differenze che i diversi territori presentano nei rispettivi punti di partenza influenzano pesantemente le opportunità di crescita dei territori stessi. Nell'ambito di queste differenze la dotazione infrastrutturale appare di primaria importanza, per le sue caratteristiche di relativa immobilità, in grado di condizionare fortemente l'efficienza dei processi produttivi. Una dotazione elevata di capitale pubblico aumenta la produttività del sistema e riduce i costi di acquisizione da parte dei privati dei fattori della produzione (capitale privato e lavoro qualificato), rendendoli più profittevoli e dunque accrescendo la probabilità di attrarli o di mantenerli in una determinata area territoriale.

Tabella 2 Risultati[7] del modello stimato a livello regionale* ρ = 0,37, R^2 = 0,43, AIC= -26.497

Variabile	Stima	Std. Error[8]	z value	Pr(>|z|)
Prodotto interno lordo	7,893E-06	0,000	2,1024	0,0355
Quota di popolazione residente in comuni con più di 50 mila abitanti	3,916E-01	0,181	2,1587	0,0309
Ricettività di alberghi di categoria medio bassa per abitante	-1,182E-02	0,005	-2,2568	0,0240

* Vedi equazione (8)

Nella Tabella 2 sono riportati: i risultati significativi delle variabili indipendenti considerate; il valore di ρ (pari a 0.37 con p-value 0,043); la stima della bontà della regressione (R^2 uguale a 0,43). L'applicazione può dunque ritenersi soddisfacente, sia per il significato delle variabili incluse come regressori (*proxy* dello sviluppo economico, della densità demografica e dell'offerta turistica qualificata), sia per le proprietà statistiche verificate. Una volta ottenuti $\hat{\beta}_a$, le deviazioni standard stimate e il parametro $\hat{\rho}_a$, in conformità con le ipotesi di s*tructural similarity* e di *error similarity* si è proceduto alla loro immissione nell'equazione (10) al fine di ottenere l'indicatore infrastrutturale stimato a livello disaggregato.

I risultati ottenuti indicano una considerevole distanza della mappa infrastrutturale stimata dai dati "effettivi" di dotazione infrastrutturale disponibili a livello provinciale, come viene confermato sia dall'esame grafico delle due distribuzioni (Figura 1), sia dal calcolo di uno specifico indicatore di robustezza spaziale, denominato IRS (Tabella 3). Quest'ultimo è stato applicato alle due graduatorie, costruite sulla base delle posizioni occupate dalle singole province in termini di dotazione infrastrutturale, graduatorie derivanti, rispettivamente, sulla base del modello applicato e dei dati "effettivi".

Rinviando per un maggior dettaglio sugli aspetti analitici dell'indicatore ad un precedente lavoro (Mazziotta e Vidoli, 2009b), una cui sintesi è riportata nell'Appendice a questo stesso contributo, qui è sufficiente ricordare che tale indicatore è il risultato del prodotto di due matrici: la prima – matrice di contiguità, ***W*** – identifica la contiguità territoriale delle unità territoriali considerate (le province, in questo caso); la seconda – matrice di transizione, ***T*** – evidenzia le differenze tra i ranghi e indica tra quali unità territoriali tali differenze si manifestano.

Moltiplicando la matrice ***I-W*** per ***T*** si ottiene un indice (IRS, appunto), esprimibile algebricamente come:

$$IRS_{R_0,R_1} = \frac{\Sigma_{i,j} T_{i,j} (1-w_{i,j})}{Max_I}$$

che, confrontando il rango occupato da ogni provincia nelle due graduatorie considerate (quella calcolata sui dati "effettivi" e quella risultante dall'applicazione del modello) evidenzia i cambiamenti nei ranghi che hanno interessato unità spaziali non contigue. L'indice varia tra 0 e 1: il livello massimo denota la massima incoerenza tra le due graduatorie, il contrario per il livello minimo.

I risultati mostrati nella Tabella 3 e visivamente rappresentati nella Figura 1 indicano chiaramente l'esistenza di marcate differenze tra le due graduatorie (IRS pari a 0,38, con un cambiamento medio nei ranghi provinciali particolarmente elevato, circa 25). Ciò sembrerebbe doversi ascrivere non

7 Da notare che il valore di R^2 nel caso di regressione degli errori spaziali non assume lo stesso significato della procedura OLS. Si riporta anche il valore AIC, usato per confrontare modelli diversi (miglior risultato in caso di valore negativo elevato).

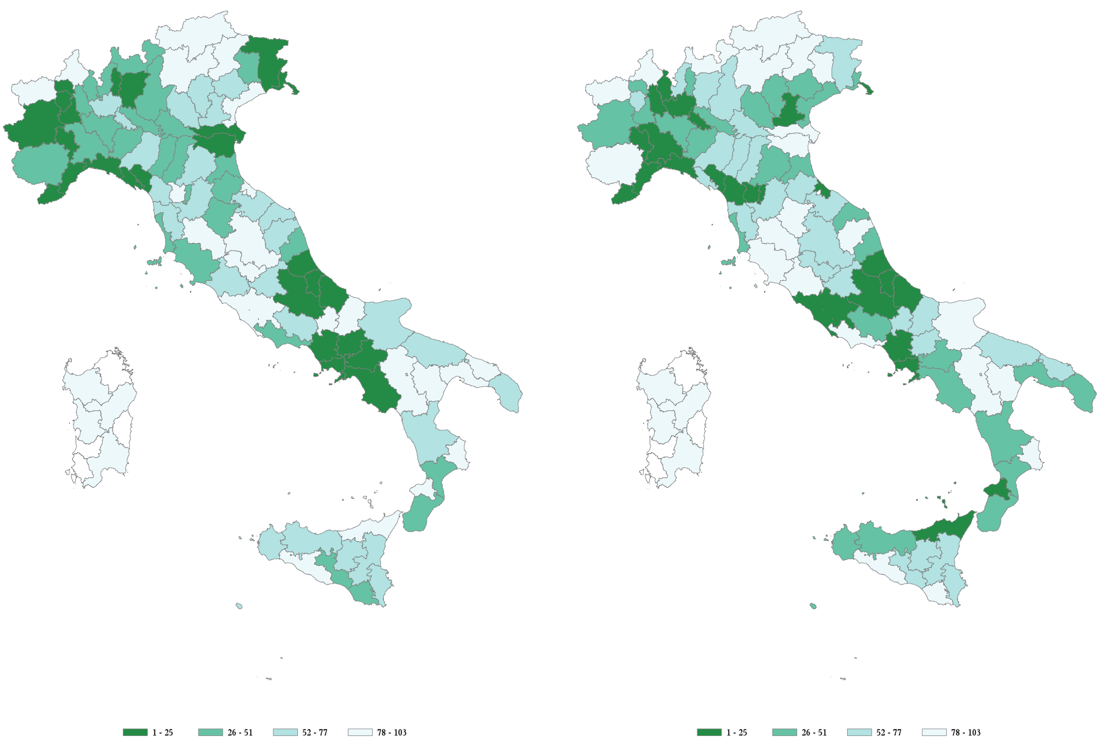
8 Numerical Hessian approximate standard errors.

tanto alla precisione del modello adottato per la disaggregazione, quanto piuttosto ad una realtà che vede una tuttora forte distanza tra domanda e offerta di infrastrutture a livello territorialmente disaggregato.

Tabella 3 Indice di robustezza spaziale (IRS) calcolato sul confronto tra indicatori "effettivi" e risultati del modello. Spostamenti nelle graduatorie delle province fuori dalle regioni di appartenenza.

IRS	N. spostamenti	Cambiamento medio di posizione nei ranghi
0,384	75	25

Figura 1 Ordinamento delle province italiane per livello di dotazione infrastrutturale: sulla base degli indicatori "effettivi" (figura a sinistra) e sulla base dei risultati del modello (figura a destra)



CONCLUSIONI

Obiettivo del lavoro qui presentato era di ricostruire la dotazione infrastrutturale (limitatamente al trasporto terrestre, nell'applicazione concretamente svolta) ad un livello di elevato dettaglio territoriale (le province italiane), contando sulla disponibilità dei medesimi indicatori ad un livello territoriale superiore (le regioni) e procedendo alla loro disaggregazione attraverso la messa a punto di un modello derivato da un approccio di tipo Chow-Lin, perfezionato sulla base di un più recente approfondimento metodologico di Polasek e Sellner.

L'inserimento nel modello di specifiche variabili di generazione della domanda di infrastrutture a livello provinciale porterebbe a ritenere che i risultati della quantificazione effettuata tendono ad approssimare i livelli di offerta infrastrutturale di cui ciascuna provincia teoricamente dovrebbe essere dotata. In effetti, l'introduzione di regressori di variabili che identificano fattori di domanda di infrastrutture conferisce al confronto tra le due graduatorie ottenute – quella basata sui dati "effettivi" e quella risultante dai dati "ricostruiti" a seguito dell'applicazione del modello – il significato di un confronto tra domanda e offerta di infrastrutture a livello territorialmente disaggregato.

Naturalmente, tale significato risulta tanto più fondato quanto più elevata è la bontà statistica del modello nella sua concreta applicazione. Sotto questo profilo, i test utilizzati mostrano che i risultati del modello possono ritenersi statisticamente soddisfacenti, tenuto conto del numero elevato di unità territoriali considerate e della quantificazione di tipo *cross-section*. Ciò non toglie che una migliore o più ampia disponibilità di variabili indipendenti possa portare significativi progressi nei risultati ottenibili dall'applicazione del modello.

Attualmente, comunque, la considerevole distanza emersa, a seguito dell'applicazione effettuata, tra livelli effettivi e livelli stimati di infrastrutturazione può essere interpretata, seppure con la cautela che si deve ad una prima sperimentazione, come la conferma del notevole divario esistente in Italia tra domanda e offerta di infrastrutture a livello territorialmente disaggregato. Questo sembra, in definitiva, essere il risultato statisticamente più evidente ed economicamente più interessante: la distribuzione territoriale della dotazione delle infrastrutture di trasporto terrestre non è coerente con i fattori "teorici" di generazione presenti a livello di province italiane.

Bibliografia

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

Ayuga-Téllez, E., Contato-Carol, M., González, C., Grande-Ortiz, M., and Velázquez, J. (2011). "*Applying Multivariate Data Analysis as Objective Method for Calculating the Location Index for Use in Urban Tree Appraisal*." J. Urban Plann. Dev., 137(3), 230–237.

Bickman L., Rog D.J, (1998), "*Handbook of applied social research methods*",eds. Sage Publications.

Biehl, D., (1994). "*The role of infrastructure in regional policy*". OECD, Working Party No. 6, Regional Development Policies, Paris.

Bollino C. A., Polinori P. (2007) *Ricostruzione del valore aggiunto su scala comunale e percorsi di crescita a livello micro-territoriale: il caso dell'Umbria*, Rivista di Scienze regionali, fascicolo 2.

Chow G. C., Lin, A. (1971) *Best linear unbiased interpolation, distribution, and extrapolation of time series by related series*, The Rev. of Economics and Statistics, 53(4): 372-375.

Goldberger A. S. (1962) *Best linear unbiased prediction in the generalized linear regression model*, American Statistical Association J., 57: 369-375.

ISTAT. (2006) *Le infrastrutture in Italia. Un'analisi provinciale della dotazione e della funzionalità*, Roma.

LeSage J. P. (1998) Spatial econometrics, Technical report, University of Toledo.

Mazziotta C., Vidoli F. (2009a) *La costruzione di un indicatore sintetico ponderato. Un'applicazione della procedura Benefit of Doubt al caso della dotazione infrastrutturale in Italia*, Italian J. of Regional Science, vol 8, n°1, Franco Angeli.

Mazziotta C, Vidoli F. (2009b) *Robustezza e stabilità spaziale di indicatori di dotazione infrastrutturale: una verifica per le province italiane*, XXX Conferenza Italiana di Scienze Regionali, Firenze.

Polasek W., Sellner R. (2008) *Spatial Chow-Lin methods: Bayesian and ML forecast comparisons*, Rimini Centre for Economic Analysis (RCEA), working paper 38-08.

APPENDICE

Costruzione di un Indicatore di Robustezza Spaziale (IRS)

Obiettivo essenziale dell'Indice di Robustezza Spaziale (IRS) qui presentato è l'analisi della stabilità dei risultati ottenuti dal confronto tra ordinamenti (graduatorie) alternativi di unità territoriali, più specificamente l'analisi della permanenza delle posizioni occupate da tali unità all'interno di più ampie partizioni territoriali.

Per chiarire il senso dell'approccio proposto, si faccia l'esempio di quattro unità territoriali, appartenenti a due aree territoriali differenti, che presentano un indicatore composito che determina un ordinamento pari a R_0:

	Unità	Rango	Area
R_0 =	A	1	Area1
	B	2	
	C	3	Area2
	D	4	

A seguito di alcune modifiche nelle ipotesi di base per la costruzione dell'indicatore o di confronti tra metodi di costruzione si possono ottenere due casi, R_1 e R_2:

	Unità	Rango	Area
R_1 =	A	2	Area1
	B	1	
	C	4	Area2
	D	3	

	Unità	Rango	Area
R_2 =	A	4	Area1
	B	2	
	C	3	Area2
	D	1	

Entrambe le situazioni presentano differenze di rango medie uguali, ma l'ordinamento R_1 appare preferibile rispetto ad R_2 in quanto spazialmente più stabile (in R_1 c'è solo scambio all'interno delle aree, mentre in R_2 lo scambio nei ranghi avviene tra aree differenti). Un indicatore di robustezza spaziale applicato ai ranghi dovrebbe, quindi, non solo mettere in evidenza le differenze medie di rango, ma evidenziare anche dove, o meglio tra quali unità territorialmente individuate, queste differenze si siano manifestate.

Per raggiungere questo scopo, sempre tenendo presente l'esempio proposto, si possono introdurre nell'analisi due matrici. La prima (W) individua l'appartenenza delle unità considerate ad un ambito territoriale più vasto (province rispetto a regioni, ad esempio), oppure individua la contiguità territoriale delle unità tra loro: a seconda dei casi si può parlare di matrici rispettivamente di "appartenenza" o "di contiguità".

		A	B	C	D
W =	A	-	1	0	0
	B	1	-	0	0
	C	0	0	-	1
	D	0	0	1	-

La seconda matrice, definibile come matrice “di transizione”, evidenzia sia le differenze nei ranghi, sia da quali unità e verso quali altre unità queste differenze si siano manifestate. Le matrici di transizione, rispettivamente tra R_0 e R_1 e tra R_0 e R_2, sono quindi:

		A	B	C	D
T_{R_0,R_1} =	A	0	1	0	0
	B	1	0	0	0
	C	0	0	0	1
	D	0	0	1	0

		A	B	C	D
T_{R_0,R_2} =	A	0	0	0	3
	B	0	1	0	0
	C	0	0	1	0
	D	3	0	0	0

La lettura di tali matrici è agevole. I numeri che vi figurano individuano l’intensità degli spostamenti avvenuti tra un’unità territoriale e l’altra a seguito delle modifiche apportate agli indicatori di sintesi: il numero 3 che si legge nella seconda matrice, ad esempio, evidenzia che tra l’ordinamento R_0 e l’ordinamento R_2 l’unità territoriale A ha perduto 3 posti, spostandosi dal rango 1 al rango 4. Inoltre, la posizione del numero in questione all’incrocio tra la riga intestata ad A e la colonna intestata a D sta a significare che lo spostamento di rango ha interessato queste due unità territoriali: A ha perso 3 posizioni a favore di D, che reciprocamente ha guadagnato le stesse 3 posizioni a svantaggio di A. Moltiplicando la matrice (I-W) per T si ottiene un indice che evidenzia i cambiamenti nei ranghi che hanno interessato unità non appartenenti allo stesso ambito territoriale (oppure che hanno interessato unità spazialmente non contigue, a seconda del tipo di matrice W utilizzato, di appartenenza o di contiguità).
L’indicatore di robustezza spaziale (IRS), in forma algebrica, può essere quindi scritto come:

$$IRS_{R_0,R_1} = \frac{\Sigma_{i,j}\, T_{i,j}\,(1\text{-}w_{i,j})}{max_I}$$

Va sottolineato che al livello massimo dell'indice (MaxI) fa riscontro la peggiore situazione dal punto di vista della coerenza tra le due graduatorie, ossia la situazione in cui l'unità i che era al primo posto nell'ordinamento R0 si ritrovi all'ultimo nell'ordinamento R1, e così via, tante volte quante sono le unità non contigue, (n-1) * (n-2) * ..., ovvero tante volte quante si ha un valore maggiore di zero nella matrice T(I-W).
Nella formulazione qui proposta per la matrice W le distanze tra le diverse unità territoriali non sono pesate, ossia queste ultime vengono registrate con uno se contigue e con zero altrimenti, indipendentemente dalla distanza più o meno grande intercorrente tra di esse. Se si volesse rimuovere questa limitazione, si potrebbe calcolare una matrice simmetrica P che rappresenti una matrice di distanza tra le unità territoriali (in termini di chilometri o di tempi di percorrenza) in modo da considerare con maggior peso le unità non contigue tra loro più lontane. Si avrebbe quindi una matrice del tipo

		A	B	C	D
	A	0	40	60	80
P =	B	40	0	30	45
	C	60	30	0	70
	D	80	45	70	0

Con questa ulteriore approssimazione l'indicatore di robustezza spaziale diventerebbe quindi:

$$IRS^{P}_{R_0,R_1} = \frac{\Sigma_{i,j}\, T_{i,j}\,(1\text{-}w_{i,j})\, p_{i,j}}{\max_{I^P}}$$

con il massimo ($\max_{I^P}$) pari a (n-1) * (n-2) * ... tante volte quante si ha un valore maggiore di zero nella matrice T_R(I-W)P.