

Come il Bes Può Cogliere i Flussi Migratori Nazionali

Una Analisi Spaziale

Federico Grazi

28 Marzo 2023

1 Introduzione

Per decenni l'economia classica si è fondata sul dogma per cui il consumo, la capacità di spesa e il PIL fossero indicatori adeguati per la misura del benessere e della felicità delle persone, tuttavia, con l'accelerazione dei ritmi di produzione e di consumo si stanno pian piano sgretolando le convenzioni che attribuivano alla felicità l'uguaglianza con il concetto di ricchezza. Nelle parole del giornalista Tiziano Terzi:

"Oggi l'economia è fatta per costringere tanta gente a lavorare a ritmi spaventosi per produrre delle cose perlopiù inutili, che altri lavorano a ritmi spaventosi, per poter comprare, perché questo è ciò che dà soldi alle società multinazionali, alle grandi aziende, ma non dà felicità alla gente."

L'idea di fondo in molti studi sociali è la necessità di poter carpire altri aspetti della quotidianità che non possono essere ridotto alla ricchezza materiale, ma che comunque possano valutare il benessere.

In Italia, dal 2013, l'Istat produce annualmente il *rapporto sul Benessere Equo e Sostenibile* (BES), ovvero un rapporto che misura variabili oggettive e soggettive su 12 Domini che impattano la vita in Italia. Come descritto sul sito dell'Istat, il BES è

"[...] una linea di ricerca, un processo che assume come punto di partenza la multidimensionalità del benessere e, attraverso l'analisi di un ampio set di indicatori, descrive l'insieme degli aspetti che concorrono alla qualità della vita dei cittadini."

Questo progetto ha come scopo di valutare quanto e che tipo di relazione possa sussistere fra questi indicatori e la propensione alla migrazione; ciò che si vuole approfondire è quali aspetti della vita impattino sulla decisione di cambiare provincia. Questa analisi verrà effettuata tenendo in considerazione la dimensione spaziale. I dati verranno trattati per le unità spaziali NUTS3, come definite dal sito ufficiale dell'Eurostat; saranno quindi valutati gli impatti delle province limitrofe nella decisione di trasferirsi e, in generale, nell'intera analisi.

2 Descrizione del Dataset

I dataset sono stati recuperati appieno dal sito dell'Istat, sia il dataset riguardante gli indicatori "BES per le province" sia i dati riguardanti ai cambi di residenza. Infatti, per misurare i flussi migratori, sarà utilizzato il dato di cambi di residenza che sono avvenuti attraverso l'anagrafe; questo dato è fortemente distorto in quanto non vengono considerati studenti fuori sede; lavoratori non regolari; lavoratori in affitto; etc etc.. Si può capire come i risultati di questo progetto non possano essere generalizzabili alla popolazione intera, ma solamente ad un sottogruppo che, per motivi di qualunque genere, si rechina all'anagrafe per cambiare residenza. Tuttavia, ritenendo il cambio di residenza come una pratica impegnativa e che richiede un profondo ragionamento sulle caratteristiche della provincia in cui si va ad abitare, il dataset può essere lo stesso considerato adeguato per la nostra analisi - ad esempio, uno studente fuori sede difficilmente prenderà in considerazione la disponibilità sanitaria quando decide dove andare a studiare.

Tutti i dati sono relativi al 2017, in quanto anno con maggior rilevazione di indicatori del BES. Per alcuni valori mancanti è stato necessario interpolare i dati usando la serie storica dal 2013 al 2019. La variabile dipendente sarà la differenza fra emigrati e immigrati in una provincia, che sarà positiva se quella provincia acquista residenti e negativa se perde residenti.

3 Metodologia

Per svolgere un'analisi spaziale sui dati verrà usato un modello di regressione spaziale auto-regressive (SAR) di tipo:

$$y = \rho W y^* + X\beta + \varepsilon, \quad (1)$$

$$y = (I - \rho W)^{-1} X\beta + u, \quad (2)$$

$$\text{con } u = (I - \rho W)^{-1} \varepsilon,$$

dove W è la matrice dei pesi spaziali che vengono associati alle varie unità spaziali. Il modello SAR riesce a carpire gli impatti spaziali attraverso la serie infinita $\sum_i^\infty \rho^i W^i$ in modo che gli effetti spaziali più vicini siano presi in maggior considerazione rispetto a quelli più lontani.

Un semplice modello su tutti e 68 indicatori risulterebbe troppo oneroso, per cui verrà applicata una PCA ad ogni dominio (Ambiente, Sicurezza, Qualità dei servizi, Lavoro e conciliazione dei tempi di vita, Sicurezza, Istruzione e Formazione, Politica e istituzioni e Benessere Economico) in modo da ridurre gli indicatori ad un numero ragionevole per poterli maneggiare in una semplice regressione lineare.

I valori della PCA verranno trasformati attraverso la standardizzazione *Mazziotta-Pareto index* in questo modo:

$$r_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} 60 + 70 \quad (3)$$

Ogni indicatore piuttosto di essere una variabile con media 0, risulterà una variabile con media 100 più facilmente confrontabile.

Infine verranno confrontati tre tipi di Cluster Analysis per poi condizionare il modello ai vari cluster ed analizzare in modo più specifico l'effetto spaziale in relazione ai cluster generati.

4 Analisi Descrittiva

Gli indicatori del BES riguardano 111 province con dati per ogni provincia; tuttavia per alcuni indicatori non sono stati rilevati i valori, per cui andrà tenuto in considerazione quest'aspetto. Nella Tabella 1 sono mostrati quante province hanno valori mancanti per ogni dominio.

DOMINIO	n
Ambiente	25
Sicurezza	15
Lavoro e conciliazione dei tempi di vita	12
Benessere economico	10
Innovazione, ricerca e creatività	5
Istruzione e formazione	4
Qualità dei servizi	2
Politica e istituzioni	1

Tabella 1: Valori NA per ogni dominio

Per quanto interessante, non sarà possibile portare avanti nell'analisi il dominio *Ambiente* in quanto il numero di valori mancanti è troppo grande e non permette uno studio completo per molte province.

4.1 BES

Gli indicatori del Bes riguardano sia misurazioni di carattere oggettivo che, una volta aggregati, possano rispecchiare misurazioni di carattere soggettivo. Un esempio di indicatore è in Figura 1, dove viene rappresentato l'indicatore *Reddito Pro Capite* e la medesima variabile calcolata attraverso il lag spaziale, ottenuto attraverso il pacchetto *spdep*. Da notare come la variabile 'laggata' abbia un range più piccolo rispetto alla variabile originale.

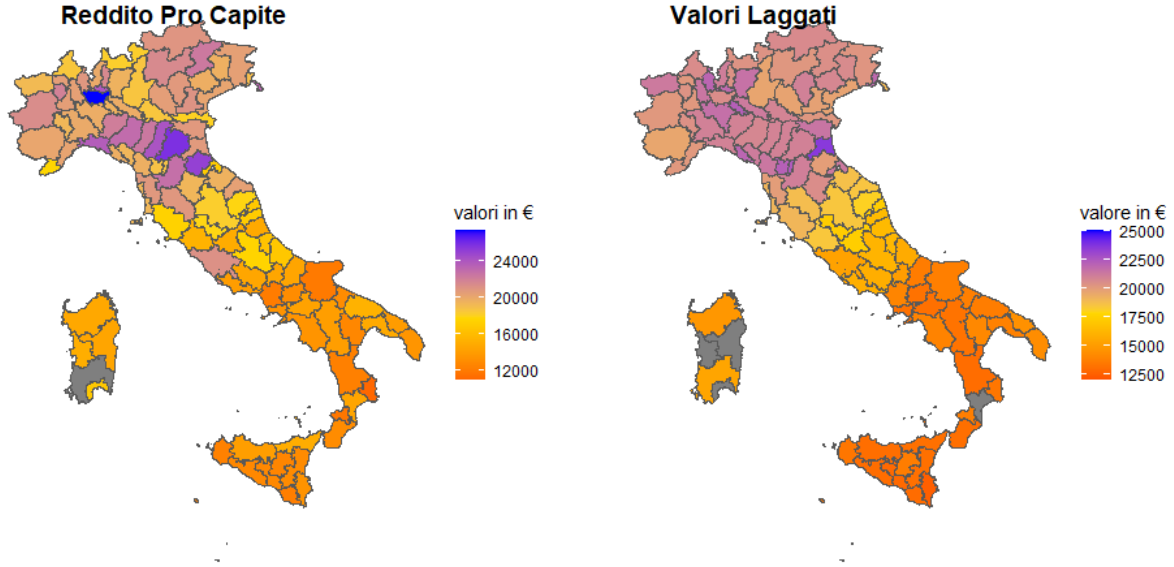


Figura 1: Variabile *Reddito Pro Capite*

Come è solito, vi è una chiara distinzione, che traspare in maniera netta nei valori laggati, fra nord, centro e sud. Nella sottostante Tabella 2 possiamo vedere alcuni degli indicatori e i loro relativi minimi e massimi.

Dominio	Indicatore	Valore	Città
Ambiente	PM 2.5	-0.09	Pordenone, F-C, Rimini, Pisa
		15.91	Padova
Ambiente	Verde urbano	-0.81	Vercelli
		946.49	Matera
Istruzione e formazione	Passaggio all'Università	-0.09	Brescia
		14.31	Isernia
Lavoro e conciliazione dei tempi di vita	Retribuzione Giornaliera	-0.05	Macerata
		9.55	Lecco
Politica e istituzioni	Amministratori Donne	-0.10	Crotone
		8.00	Bologna
Salute	Mortalità infantile	-0.07	Genova, Verona, Massa-Carrara
		3.83	Barletta-Andria-Trani, Messina
Sicurezza	Delitti	-0.85	Teramo
		269.65	Rimini
Sicurezza	Mortalità stradale	-0.03	Avellino
		10.57	Foggia

Tabella 2: Minimi e massimi per alcuni indicatori del BES

Come si può vedere Padova è da sola in cima alla lista della città più inquinata da pm 2.5, mentre Pordenone, Forlì-Cesena, Rimini e Pisa sono a parimerito per le meno inquinate.

4.2 Principal Components Analysis

Usando l'analisi delle Componenti Principali, si riesce a ridurre notevolmente il numero di variabili sino ad una decina, riportate nella Tabella 4. Per scegliere il numero di componenti principali è stato utilizzato un criterio non eccessivamente restrittivo; per la maggior parte delle PC la varianza spiegata doveva essere maggiore di il 50%.

Dominio	PC	Var Spiegata	Var Cumulata
Istruzione e Formazione	istruiti	60%	60%
Qualità dei Servizi	disponibilità sanitaria	35%	35%
Qualità dei Servizi	differenziazione sanitaria	22%	57%
Benessere Economico	ricchezza	79%	79%
Lavoro	qualità del lavoro	76%	76%
Politica e Istituzioni	elezioni o giovani	48%	48%
Politica e Istituzioni	amm donne giovani	35%	83%
Sicurezza	delitti	44%	44%
Sicurezza	morti passive	29%	73%
Salute	vita longeva	44%	44%
Salute	mortalità avanzata	19%	63%

Tabella 3: Varianze per ogni PC

Come si può notare nella Tabella 3, quasi tutte le PC riescono a spiegare la maggior parte della varianza solamente da sole. Nel caso ci fosse stato bisogno di aggiungere una seconda PC si può notare che la percentuale di varianza spiegata fosse ugualmente particolarmente alta. Una volta ottenute le componenti, sono state trasformate attraverso la standardizzazione utilizzata per costruire il Mazzetta-Pareto index, in modo tale da avere indicatori che variassero intorno a 100 piuttosto che attorno a 0. Ciò ci restituirà una regressione in cui è più chiaro interpretare gli effetti marginali. Per avere una visione di come si comportano le variabili costruite fra di loro, possiamo analizzare le correlazioni, visualizzate nella Figura 2. Interessante come la variabile "ricchezza" sia significativamente correlata con la variabile "vita longeva" e inversamente correlata a "morti passive". Ciò indica in modo alquanto distinto come la ricchezza di una provincia influenzi lo stile di vita. Possiamo anche notare come "morti passive" abbia un outlier che assume un valore assai maggiore alla media (Foggia). Nella maggior parte dei casi le distribuzioni delle variabili è unimodale, a parte

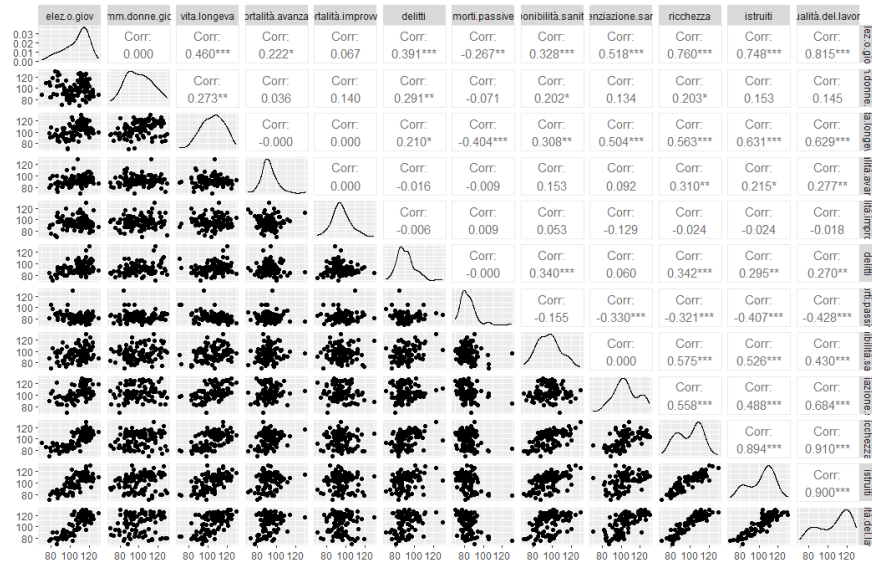


Figura 2: Distribuzioni e Correlazioni fra le PC create

ricchezza, istruzione e qualità del lavoro, ad indicare una divisione in, almeno, due categorie nette all'interno della variabile. Questa divisione può essere meglio osservata nella Figura 3, in cui sono meglio evidenziate le due curve che determinano i due intervalli. Ciò, visualizzato anche in Figura 3b, è un forte indice di disparità territoriale fra il Nord e il Sud del paese.

DOMINIO	Indicatori PCA	Indicatori Positivi	Indicatori Positivi
Lavoro	qualità.del.lavoro	Retribuzione Giornaliera, Tasso di Occupazione Giovanile, Tasso di Occupazione	Tasso di Mancata Partecipazione al Mondo del Lavoro, Tasso di Mancata Partecipazione Giovanile al Mondo del Lavoro
Politica	amm.donne.giov	Amministratori Donne, Amministratori Giovani	NA
Politica	elez.o.giov	Partecipazione alle Elezioni	Amministratori Giovani
Benessere Economico	ricchezza	Patrimonio Per Capite, Pensione PC, Reddito PC	Pensionati con reddito pensionistico di basso importo, Tasso di ingresso in sofferenza dei prestiti bancari
Istruzione e Formazione	istruiti	Bambini aderenti a servizi per l'Infanzia, Diplomatici, Laureati, Partecipazione alla formazione continua	Analfabeti, Anumerici, NEET, Partecipazione al Sistema Scolastico
Qualità dei servizi	differenziazione.sanitaria	Servizio di raccolta differenziata	Emigrazione Ospedaliera, Irregolarità del servizio elettrico, Medici Specialisti, Posti letto per specialità
Qualità dei servizi	disponibilità.sanitaria	Medici Specialisti, Posti letto per specialità, Posti letto negli ospedali	Emigrazione Ospedaliera, Irregolarità del servizio elettrico
Sicurezza	Delitti	Altri delitti, Delitti	Mortalità Stradali, Omicidi volontari
Sicurezza	morti.passive	Altri Delitti, Morti Stradali, Omicidi volontari	NA
Salute	vita.longeva	Mortalità Demografica, Speranza di Vita	Mortalità evitabile, Mortalità infantile, Mortalità per tumori
Salute	mortalità avanzata	Mortalità per incidenti, Mortalità demografica, Mortalità per tumori	Mortalità infantile

Tabella 4: Indicatori PCA per Dominio

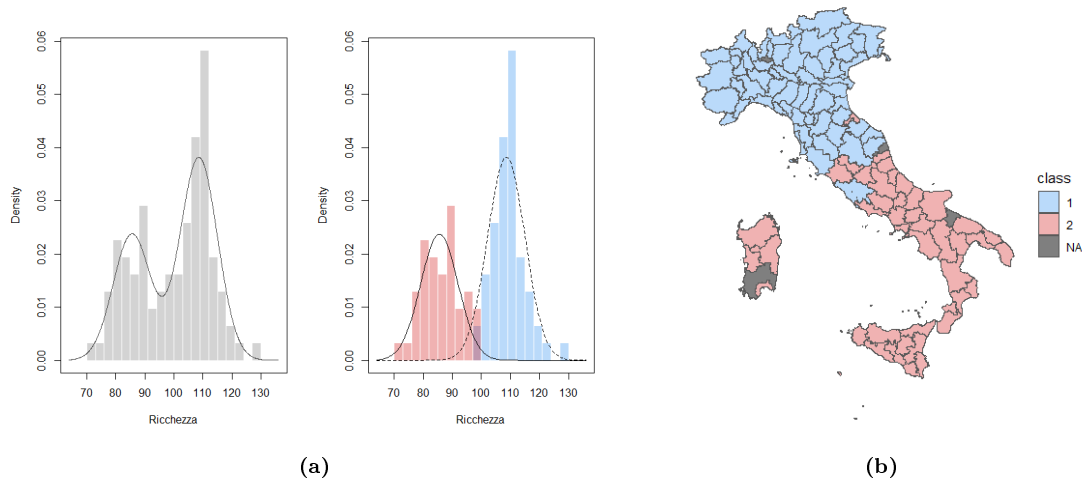


Figura 3: Distribuzione Bimodale della variabile Ricchezza

Utilizzando l'indicatore 'ricchezza' e 'vita longeva' appena citati, si può osservare come la distribuzione laggata in modo spaziale abbia delle aree ben definite, come mostra la Figura 4

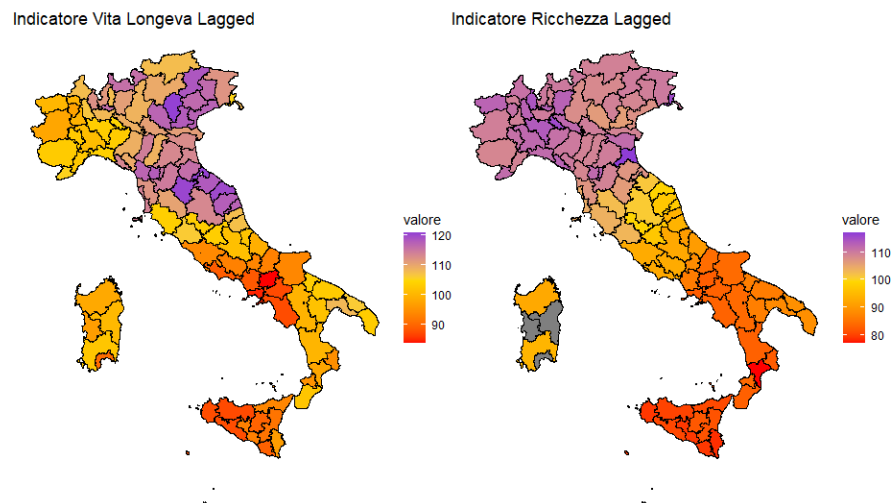


Figura 4: Distribuzioni laggate delle due variabili

Per *vita.longeva* vi è un contrasto fra costa Adriatica e il mar Tirreno nel Sud; la prima dev'essere dovuta all'alto tenore di vita e il basso livello di mortalità per tumori in quella zona, che, al contrario, ha valori massimi in città Campane (Napoli e Caserta).

4.3 Dati sui Cambiamenti di Residenza

I dati sugli spostamenti di residenza tengono conto solamente degli spostamenti da una provincia ad un'altra, non sono considerati gli spostamenti interni in quanto possono essere determinati da motivi personali piuttosto che ad una ragionata scelta in base al territorio e alle opportunità offerte dal luogo di destinazione, che è ciò che riguarda questo progetto.

Nella Tabella 5 sono riportate le 5 città con il rapporto maggiore di nuovi residenti. Come si può notare, i risultati non sono così scontati: mancano le maggiori città italiane (Roma, Milano, Napoli) ma è presente Bologna, presumibilmente influenzato dalla forte presenza universitaria nella città. Le altre città sono città

Provincia	% Nuovi Residenti
Lodi	3.31%
Monza e della Brianza	3.26%
Asti	2.87%
Pavia	2.85%
Bologna	2.80%

Tabella 5: Immigrazioni sui Residenti totali

ricche del nord che attirano residenti in maniera diversa, quali possono essere maggior offerta di lavoro specializzata, livelli di vita maggiori e condizioni più agiate.

Questa analisi considererà solamente la differenza fra i nuovi residenti e chi ha cambiato residenza, ovvero:

$$\text{Differenza} = \# \text{ Nuovi Residenti} - \# \text{ Residenti Persi}$$

che risulta positivo se vi sono più nuovi residenti e la città ha conseguentemente acquistato più residenti di quanti ne ha persi, ed è negativo nel caso contrario.

5 Analisi di Regressione Spaziale

In questa sezione verrà analizzata e confrontata l'effetto spaziale delle province limitrofe contro un semplice modello lineare. In una prima sezione verranno analizzati modelli non condizionanti e conseguentemente sarà condizionato il modello rispetto ad un appropriato modello di clustering utilizzando un numero ristretto di variabili selezionate

5.1 Selezione del Modello

Al fine di avere un modello significativo dovremmo tener conto quali variabili considerare e decidere quanta influenza dovrà avere la dimensione spaziale.

Scelta dei Vicini Verrà utilizzata la tecnica di *K-Near Neighbor* per determinare quali province possano essere considerate "limitrofe" e possano influenzare la scelta di cambiare residenza. Per decidere il numero k si userà un criterio molto soggettivo: considerando che le province considerate sono 105, il numero di province per regione è circa 5, per cui per tener conto di tre regioni sarà usato 15 come numero di vicini da utilizzare. Ciò produrrà una mappa di abbastanza intricata, come si può visualizzare nella Figura 5a, tuttavia questa situazione genererà una analisi accurata e completa del fenomeno.

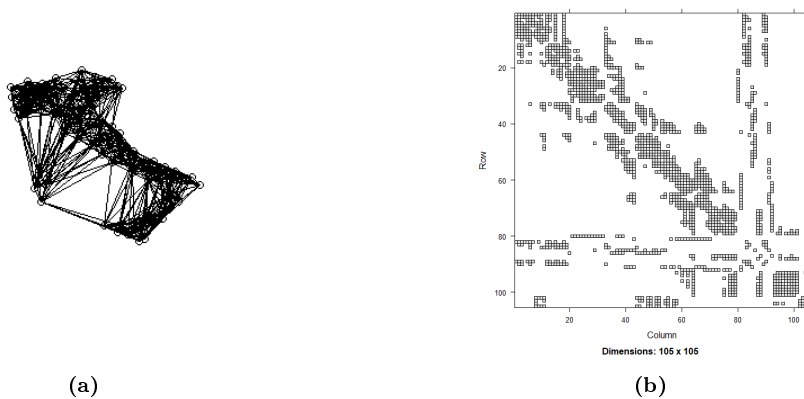


Figura 5: Relazione Spaziale fra le Province e Pesi Spaziali

Da questa lista di tipo **knn** verrà poi generata la matrice di pesi spaziali, visualizzata nella Figura 5b, una matrice *Sparse*, ovvero che ha molti elementi uguali a 0.

Scelta dei Regressori Poiché una ideale regressione su tutte le componenti principali generate risulterebbe di difficile interpretazione e assai onerosa da trattare, piuttosto che utilizzare il potenziale effetto di causalità che esse hanno.. Allora, per scegliere il numero ottimale verrà utilizzato il metodo della 10 Fold Cross Validation utilizzando come *leap method* quello Stepwise, per garantirci maggior flessibilità nella scelta del modello. Nella Tabella 6 possiamo vedere i risultati di due statistiche riguardanti 8 modelli generati dalla Cross Validation: sia il RMSE sia il MAE indicano 5 regressori come valori che minimizzano la differenza coi valori reali.

n	RMSE	MAE
3	2587.72	1677.49
4	2511.31	1636.82
5	2355.71	1580.50
6	2471.54	1659.96
7	2550.65	1664.00
8	2627.32	1796.75
9	2495.22	1709.36
10	2553.87	1744.02

Tabella 6: Statistiche per modelli da 3 a 10 regressori

I 5 regressori selezionati dalla *Stepwise Selection* sono *elez.o.giov*, *istruiti*, *ricchezza*, *delitti* e *vita.longeva*; l'unico dominio che è stato escluso dalla regressione è *Qualità dei servizi*, mentre per ogni altro dominio è stata scelta la prima componente principale. Per quanto questo metodo di valutazione della bontà non necessariamente possa essere esteso anche ad un modello di tipo spaziale, utilizzeremo i risultati ottenuti anche nella regressione di tipo SAR.

5.2 Modello OLS

Un primo modello lineare stimato attraverso OLS (quindi una PCR) sarà usato per valutare l'impatto dei regressori sulla nostra variabile dipendente; la pretesa di questo modello è di valutare l'effetto causale che vi è fra i 5 regressori e i cambiamenti di residenza.

Va tenuto conto che questo modello si tratta di una regressione sulle componenti principali, poi ritrasformate, per cui gli effetti diretti e indiretti dovranno essere ri-trasformati prima di poterli rapportare alle nostre variabili originali presenti nel BES.

Il modello OLS, come indicato nella Tabella 7, ha valori positivi per ricchezza e istruiti, seppur solo il primo sia significativo. Ciò vuol dire, considerando come sono costruiti gli indicatori, valori alti di patrimonio per capite, reddito per capite, di laureati e di diplomati fanno aumentare la differenza dei residenti, ovvero che alti valori sono considerati come "attrattivi" nei confronti di abitanti di altre province.

Interessante è il segno dell'indicatore *elez.o.giov* che, nuovamente per costruzione, suggerisce un impatto della partecipazione elettorale negativa e un impatto della presenza di una amministrazione giovane positiva; questo trend va quindi a premiare province giovani e con un ricambio generazionale nella politica.

5.3 Modello SAR

Introducendo i pesi spaziali W come sono nell'Equazione (2), ricaveremo un modello di tipo Auto-Regressivo Spaziale la cui significatività può essere valutata attraverso la significatività del coefficiente ρ .

Come possiamo vedere nella seconda colonna della Tabella 7 li rho è positivo e significativo all'1%, per cui l'introduzione delle dipendenze spaziale è significativa e migliora il modello sia nell'AIC, nell'RMSE e nel MAE, per cui il nostro modello è più accurato. L'effetto principale che il modello SAR introduce è una riduzione del coefficiente della componente principale *ricchezza* che non risulta più significativa; l'effetto con maggiore significatività rimane al dominio politico e che restituisce una particolare attenzione da parte di nuovi residenti di come sia amministrata la provincia in questione e quelle attorno alla propria. Va notato come il coefficiente ρ sia positivo, implicando che gli effetti indiretti delle quindici province limitrofe abbiano il medesimo segno e impatto minore rispetto a quello restituito dalla propria provincia. Ciò implica che se una provincia ha valori bassi in *ricchezza*, supponendo che le province vicine siano in condizioni simili, esse

aumentino la probabilità di avere la differenza citata nella Sezione 4.3 negativa, piuttosto che suggerire un effetto inverso; per cui si potrebbe generare una situazione del tipo *rich gets richer*, ovvero che le province ricche, siano influenzate e influenzino le province limitrofe per attirare residenti.

	OLS	SAR
ricchezza	150*	67
elez.o.giov	-143***	-140***
vita.longeva	-23	4
delitti	-68*	-10
istruiti	90	86
rho		0.578***
R^2	0.256	0.325
AIC	1993.6	1979.3
RMSE	3033	2770
MAE	1889	1595
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		

Tabella 7: Coefficienti per i modello OLS e SAR

Nella Figura 6 si possono osservare i diversi impatti che i regressori hanno per ogni provincia. Si nota come la ricchezza e l'istruzione siano valutate maggiormente al Nord, mentre la presenza di delitti ha un forte impatto nelle regioni del Mezzogiorno.

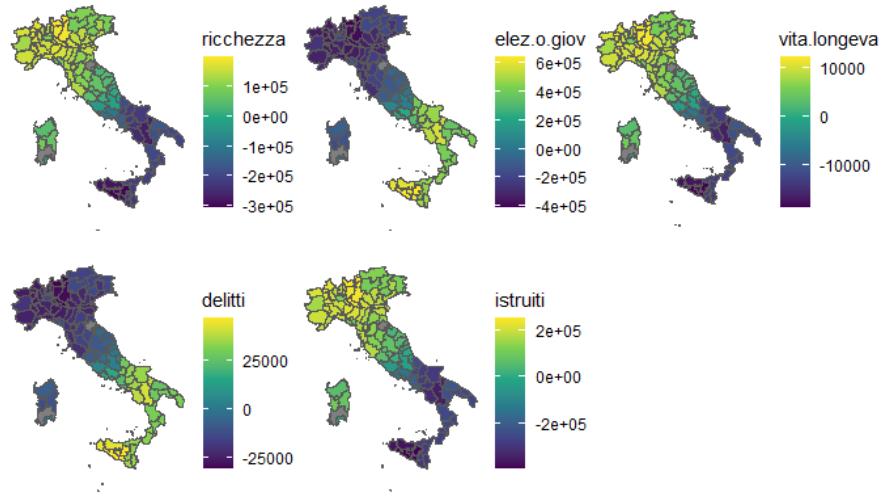


Figura 6: Impatti delle Componenti Principali per Provincia

Questa regressione non condizionata raffigura una chiara distinzione all'interno del paese, in cui gli effetti si differenziano sulla base della latitudine. Un approfondimento per diversi cluster potrebbe restituire una più chiara immagine di come sia eterogenea l'Italia nei flussi di cambi di residenza.

6 Clustering Spaziale

In questa sezione verranno confrontati due metodi di clustering utilizzati poi per condizionare il modello SAR appena ricavato. I due metodi affrontati in questo progetto sono il Clustering Gerarchico con e senza vincoli spaziali; questi si basano sul pacchetto `hclust` e `hclustgeo`, ed entrambi considerano come in input una matrice di distanze relative ai valori, calcolate attraverso la distanza euclidea, e restituiscono come output un dendrogramma in cui le fusioni di cluster avvengono attraverso il metodo di Ward. La differenza fra il primo

e il secondo, vi è poiché il secondo tiene conto anche di una seconda matrice di distanze geografiche quando si uniscono i cluster.

6.1 Numero Ottimale di Cluster

Prima di procedere alla applicazione delle funzioni, va definito il numero ottimale di cluster; in modo superficiale si potrebbero discernere tre cluster definiti in base alla comune conoscenza della Penisola: Nord, Centro e Sud. Tuttavia due indici di bontà per la scelta del numero di cluster ottimali, Dunn e Silhouette, ci restituiscono 7 come numero ottimale. La Figura 7 indica che 7 può essere un numero adeguato; infatti, mentre l'indice di Dunn va massimizzato, l'indice di Silhouette va minimizzato, e i valori che entrambi assumono per il numero 7 rispettano questi due criteri.

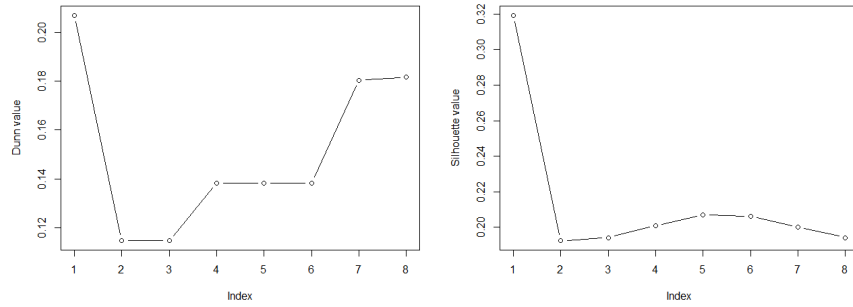


Figura 7: Indici di Dunn e Silhouette

6.2 H-clust Senza Vincoli Spaziali

Utilizzando 7 come il numero ottimale di cluster, con il metodo di Ward su una matrice di distanze calcolate sulle variabili PCA ci verrà restituito il seguente dendrogramma.

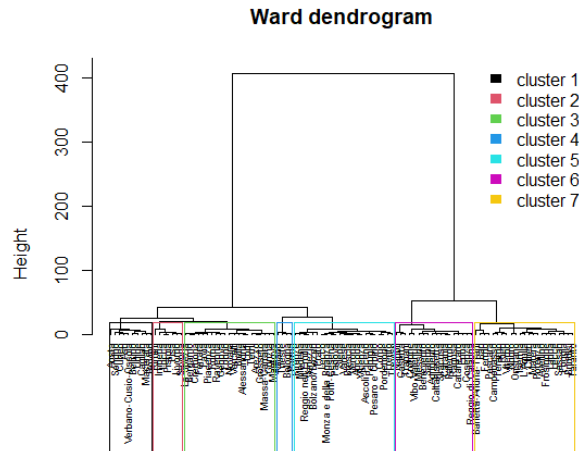


Figura 8: Dendrogramma

Da questa partizione otterremo 7 cluster che sono distinti dai valori che si trovano nella Figura 9 e sono distribuiti come riportato nella Figura 10a. Il primo cluster riguarda principalmente la Sicilia e parte del

Mezzogiorno che viene bagnato dal Mar Tirreno si caratterizza da valori mediamente sotto la media e per avere il valore più basso per *istruiti*. Il sesto e il settimo cluster sono simili per molte variabili ma si distinguono per valori diversi di *istruiti*; nel sesto cluster vi sono le province metropolitane di Bologna e di Milano, ad esempio.

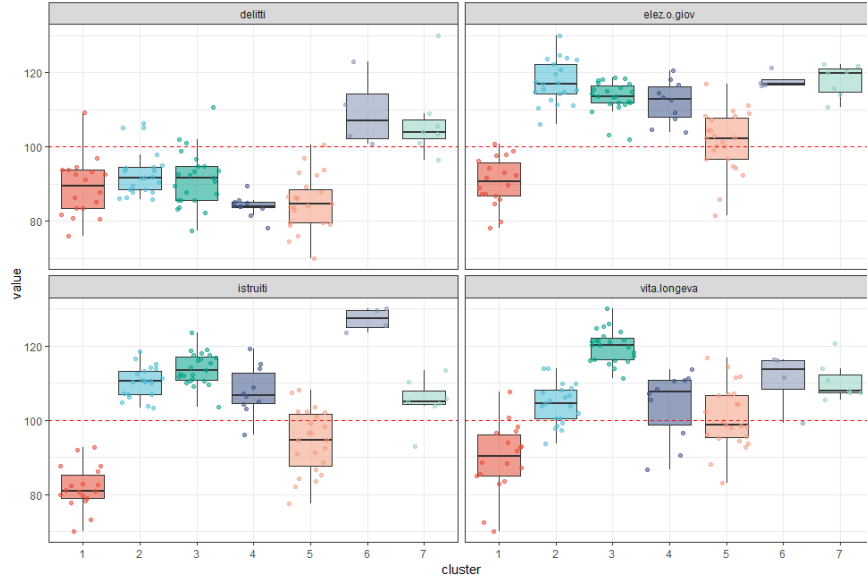
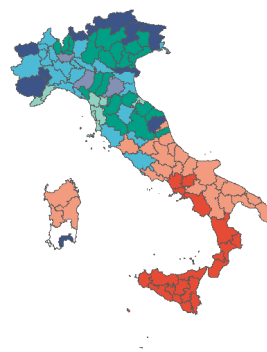
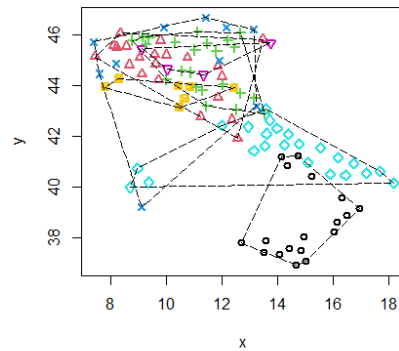


Figura 9: Valori per i Cluster

Nonostante non sono stati introdotti vincoli spaziali in questo modello di clustering, i risultati ottenuti sono abbastanza soddisfacenti; innanzitutto vi sono due cluster al Sud che comprendono l'intero Mezzogiorno in maniera omogenea. La situazione risulta essere assai diversa al Nord, dove i cluster sono geograficamente frammentati e difficilmente individuabili nello spazio. Queste osservazioni fanno sì che non risulti efficace per condizionare rispetto a cluster geografici.



(a) h-clust senza vincoli



(b) Confini dei cluster

Figura 10: Distribuzione dei Cluster Gerarchici

6.3 H-clust Con Vincoli Spaziali

L'utilizzo di vincoli riguardanti la distanza che vi è fra i centroidi geografici di ogni provincia permette di avere cluster più omogenei a livello geografico, come si vede nella Figura 11, dove si possono notare 7 cluster

ben delimitati.

Mentre i due cluster al Mezzogiorno sono simili a quelli già ottenuti, adesso vi è maggior distinzione fra Centro, Nord, Nord-Est e Nord-Ovest, che prima non veniva risaltata.

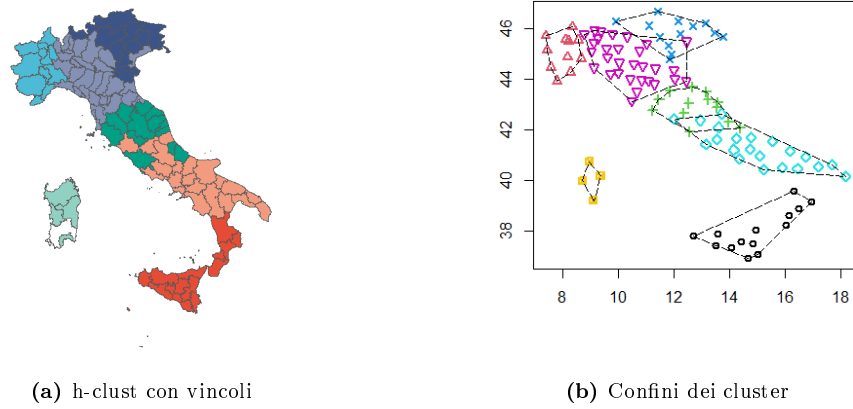


Figura 11: Distribuzione dei Cluster Gerarchici con Vincoli

Come possiamo notare dai valori medi dei cluster nella Figura 12 i cluster riguardanti il Meridione sono quelli meno istruiti, con poca partecipazione alle elezioni e *vita.longeva* più bassa. Si osserva, inoltre, come il cluster del Nord-Ovest abbia un valore molto alto nella partecipazione elettorale e alto livello di istruzione.

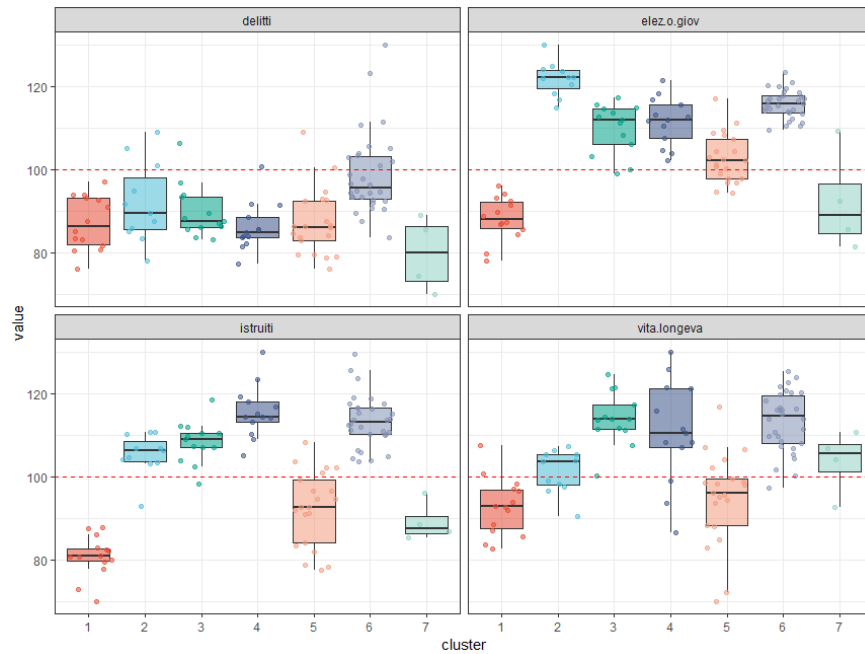


Figura 12: Valori per i Cluster con Vincoli

Un'ulteriore osservazione è che la numerosità dei cluster è distribuita in maniera più equa fra cluster, con il solo cluster 'Sardegna' che ha 4 unità e, che sarà conseguentemente esclusa dal condizionamento successivo per la numerosità troppo ridotta.

6.4 Modello SAR Condizionato

Condizionando quindi ai 6 cluster rimanenti, i risultati ottenuti si possono osservare nella Tabella 8.

	SAR
clust_1	-6883.8
clust_2	8517.6
clust_3	-20589.62
clust_4	-46903.59***
clust_5	5242.46
clust_6	-25149.03
clust_1:ricchezza	-94.97
clust_2:ricchezza	368.07
clust_3:ricchezza	195.46
clust_4:ricchezza	275.32***
clust_5:ricchezza	-35.45
clust_6:ricchezza	195.51
clust_1: elez.o.giov	-29.5
clust_2: elez.o.giov	-325.41.
clust_3: elez.o.giov	-174.16
clust_4: elez.o.giov	60.89
clust_5: elez.o.giov	115.73.
clust_6: elez.o.giov	-8.22
clust_1: vita.longeva	19.76
clust_2: vita.longeva	-105.91
clust_3: vita.longeva	-75.41
clust_4: vita.longeva	72.43*
clust_5: vita.longeva	68.78.
clust_6: vita.longeva	70.9
clust_1: delitti	0.91
clust_2: delitti	156.63*
clust_3: delitti	113.92
clust_4: delitti	66.85**
clust_5: delitti	-304.24***
clust_1: istruiti	91.76
clust_2: istruiti	-64.48
clust_3: istruiti	143.5
clust_4: istruiti	-22.44
clust_5: istruiti	127.13*
clust_1: mortalità.avanzata	52.28
clust_2: mortalità.avanzata	-51.28
clust_3: mortalità.avanzata	42.97
clust_4: mortalità.avanzata	-3.15
clust_5: mortalità.avanzata	-93.96.
rho	-0.95***
R^2	0.138
AIC	1926.4
RMSE	1549.6
MAE	1169.2

Tabella 8: Coefficienti per Modello Condizionato

Come si può osservare il coefficiente auto-regressivo ρ è altamente significativo ed è diventato negativo. Ciò indica che, analizzando per cluster, l'effetto delle province limitrofe è inverso. Ulteriori osservazioni indicano

che il modello non riesce ad adattarsi meglio ai dati in quanto l' \bar{R}^2 diminuisce, diventando anche minore del modello OLS normale; gli altri valori per la valutazione della bontà di adattamento del modello indicano che si riescono meglio a spiegare i dati e a fittare in maniera più precisa i valori, ciò dovuto alla segmentazione dei dati.

Di particolare interesse è come vari coefficienti del quarto cluster, quello del Nord Italia, siano significativi almeno al 5%; la costante è negativa, tuttavia il coefficiente di *ricchezza* è positivo e alto, come lo è per *vita.longeva*. Da notare anche come *delitti* sia significativo e negativo nel cluster 5 (Mezzogiorno senza Isole), ad indicare una grave situazione a livello di giustizia.

In generale si nota come alcuni cluster abbiano trend opposti riguardo ai coefficienti rispetto ad altri.

7 Conclusione

Seppur vi siano vari motivi per cui una persona si decida a cambiare residenza, questa analisi ha portato alla nostra attenzione come vi sia una divisione all'interno della nostra nazione che è distinta e delineata da alcuni aspetti della quotidianità che possono essere colti dal BES. L'utilizzo delle Componenti Principali ha permesso di analizzare in maniera sintetica gli aspetti che i numerosi indicatori rilevati dall'Istat ritraggono a proposito della situazione Italiana.