

Rapporto n. 241

Il modello Tucker3 per l'analisi dei bilanci dei comuni italiani

Simona C. Minotti, Mirko Signorelli

Settembre 2013

Dipartimento di Statistica e Metodi Quantitativi

Università degli Studi di Milano Bicocca

Via Bicocca degli Arcimboldi 8 - 20126 Milano - Italia

Tel +39/02/64483102/3 - Fax +39/2/64483105

Segreteria di redazione: Andrea Bertolini

IL MODELLO TUCKER3 PER L'ANALISI DEI BILANCI DEI COMUNI ITALIANI

Simona C. Minotti - Mirko Signorelli

Dipartimento di Statistica e Metodi Quantitativi
Università degli Studi di Milano-Bicocca
simona.minotti@unimib.it, m.signorelli6@campus.unimib.it

Abstract

Il presente articolo affronta la tematica dell'analisi di un array contenente gli indici di bilancio rilevati su un campione di comuni italiani. In apertura si pongono in evidenza le ragioni per cui l'impiego del modello Tucker3 rappresenti lo strumento metodologicamente corretto per sintetizzare le informazioni contenute nell'array, mostrando in seguito come tale modello rappresenti una naturale generalizzazione dell'analisi delle componenti principali a due vie al caso three-way. Ampio spazio è dedicato, in seguito, alla stima di tre distinti modelli Tucker3, attraverso i quali vengono poste in evidenza le principali dimensioni che caratterizzano la gestione dell'ente (disponibilità di risorse pro-capite, autonomia e virtuosità) e le differenze geografiche e demografiche che caratterizzano i comuni.

Keywords: analisi delle componenti principali a tre vie, analisi di bilancio per indici, enti locali, modelli multiway, Tucker3.

1 Introduzione

L'analisi statistica degli array tridimensionali¹ viene spesso condotta applicando tecniche statistiche multivariate ad una matrice bidimensionale ottenuta pretrattando l'array per mezzo di procedure (flattering² o unfolding³) che semplificano la struttura tridimensionale dei dati, comprimendone una delle tre dimensioni (tipicamente quella temporale); questa semplificazione, impropria in quanto non tiene conto della struttura three-way dell'array e comporta una perdita di informazione relativa alla struttura dei dati, può essere evitata facendo ricorso ai modelli multiway [12], che generalizzano alcune tecniche statistiche multivariate al caso multiway.

L'introduzione dei modelli multiway risale al 1963, ma per molti anni il loro impiego è stato confinato alla psicomетria e alla chemiometria. In tempi più recenti sono comparse applicazioni di tali modelli in ambito econometrico e statistico-economico: si vedano, ad esempio, [1], [6] e [13].

¹Un array tridimensionale $\mathcal{X} = \{x_{ijk} : i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K\}$ è una struttura dei dati che raccoglie i valori di J variabili rilevate su I unità statistiche in K occasioni.

²Il flattering consiste nella sostituzione, per ciascuna coppia (i, j) , della media dei K valori \bar{x}_{ij} , al posto delle K condizioni x_{ijk} .

³L'unfolding è la giustapposizione, orizzontale o verticale, delle K slices verticali, in modo da trasformare l'array in una matrice $I \times J \times K$ o $I \times JK$.

Nel presente articolo verranno presentate alcune applicazioni a dati di bilancio del modello Tucker3 [15], che costituisce il modello più generale di analisi delle componenti principali a tre vie. L'approccio impiegato è quello proprio dell'analisi di bilancio per indici, tecnica sviluppatasi in ambito aziendale e solo successivamente estesa allo studio dei bilanci degli enti locali⁴. Scopo delle applicazioni è sintetizzare le informazioni presenti nei bilanci dei comuni italiani nel periodo 2003-2010: dal momento che i dati in nostro possesso sono organizzati in un array tridimensionale (comuni \times indici di bilancio \times anni), l'applicazione del modello Tucker3 in luogo dell'analisi delle componenti principali classica (two-way) risponde all'esigenza di analizzare propriamente e più sinteticamente l'array. Cercheremo di dare risposta ai seguenti interrogativi: quali sono le principali differenze che caratterizzano la gestione degli enti locali in Italia, sia dal lato delle entrate che da quello delle spese? Nell'analisi di tali differenze è più rilevante la dimensione geografica o quella demografica? Per quanto attiene alla collocazione geografica dei comuni, sono maggiori le differenze fra un comune collocato in una regione a statuto ordinario ed uno posto in una regione a statuto speciale oppure fra un comune del nord ed uno del sud?

Il presente lavoro costituisce una novità nell'ambito dell'analisi dei bilanci dei comuni italiani in quanto si propone di applicare una tecnica statistica avanzata a dati in genere analizzati solo con strumenti tradizionali. Ciò è stato reso possibile anche grazie alla disponibilità di una nuova fonte informativa, la banca dati del progetto "Open Bilanci" [14], che raccoglie i bilanci dei comuni italiani e li rende pubblici in formato open data. Più in generale, inoltre, l'intenzione è quella di mostrare l'utilità di un approccio multiway all'analisi di un insieme di bilanci, indipendentemente dalla loro natura: il modello Tucker3 che qui presentiamo, infatti, è suscettibile di applicazione a qualsiasi situazione reale nella quale siano stati collezionati una molteplicità di indici di bilancio misurati per più anni su un collettivo di imprese, enti locali, banche o fondazioni e si abbia la necessità di fornire una visione sintetica delle numerose informazioni disponibili. Una situazione di questo genere è stata affrontata con un approccio multiway da Andolina e Vassiliadis [1], ma con alcune importanti differenze: dal punto di vista metodologico, infatti, essi fanno riferimento all'impiego dell'analisi fattoriale a tre vie (pur non esplicitando ipotesi distributive), mentre qui si propone un approccio esplorativo fondato sull'analisi delle componenti principali a tre vie. Si intende infine porre in evidenza la coerenza dei modelli Tucker3 stimati mettendoli in relazione con l'analisi delle componenti principali a due vie.

La trattazione è così articolata: dopo aver riassunto brevemente gli aspetti salienti del modello Tucker3 si procede alla descrizione degli indici di bilancio impiegati e delle caratteristiche del campione selezionato; la sezione successiva mostra come sia possibile rilevare una struttura di correlazione fra indici di bilancio e componenti principali persistente nel tempo e come risulti quindi ragionevole e parsimonioso impiegare un unico modello multiway; infine vengono presentati i risultati di tre differenti modelli multiway, dapprima analizzando separatamente le differenze geografiche e demografiche e quindi considerandole congiuntamente.

2 Il modello Tucker3

Il modello Tucker3 [15] consente di generalizzare al caso tridimensionale il metodo delle componenti principali, estraendo rispettivamente $P < I$ componenti per le unità statistiche, $Q < J$ componenti per le variabili e $R < K$ componenti per le condizioni.

In linguaggio matriciale, il metodo Tucker3 opera la fattorizzazione dell'array tridimensionale \mathcal{X} in tre matrici di componenti A , B e C legate tra loro dal full core array \mathcal{G} , di dimensione $P \times Q \times R$. Indicata con \mathbb{X}

⁴Il metodo dell'analisi di bilancio per indici si fonda sulla costruzione di rapporti ("ratios") che fanno ricorso in genere ad aggregati desumibili dal bilancio dell'impresa o dell'ente locale e, talvolta, ad alcune informazioni extracontabili (come, ad esempio, il numero di abitanti di un comune).

la matricizzazione di dimensione $I \times JK$ estratta da \mathcal{X} e con \mathbb{G} la matricizzazione $P \times QR$ del core array \mathcal{G} , si procede all'estrazione di tre matrici di componenti A, B e C tali che risulti

$$\mathbb{X} = A\mathbb{G}(C' \otimes B') + E,$$

ovvero

$$x_{ijk} = \hat{x}_{ijk} + e_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{prq} a_{ip} b_{jq} c_{kr} + e_{ijk},$$

dove i coefficienti a_{ip} , b_{jq} , c_{kr} rappresentano i punteggi riferiti rispettivamente alla p-esima componente delle osservazioni, alla q-esima componente delle variabili e alla k-esima componente delle condizioni, mentre gli elementi g_{prq} del core array esprimono il peso delle diverse combinazioni di componenti nella determinazione dell'immagine strutturale \mathcal{X} , così che la variabilità spiegata dalla combinazione di componenti $a_p b_q c_r$ è data dal quadrato del corrispondente elemento di \mathcal{G} , g_{pqr}^2 .

La varianza spiegata complessivamente dal modello è data dalla somma delle varianze spiegate da ciascuna combinazione di componenti ed è perciò pari a

$$\sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}^2.$$

La stima di x_{ijk} secondo il modello Tucker3 è data da

$$\hat{x}_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{prq} a_{ip} b_{jq} c_{kr}$$

e l'array $\hat{\mathcal{X}}$ contenente le stime \hat{x}_{ijk} è detto immagine strutturale di \mathcal{X} .

3 La definizione del campione

Il campione estratto è composto da 205 comuni italiani aventi popolazione superiore ai 5000 abitanti, stratificato per area geografica e dimensione demografica.

Gli indici di bilancio considerati fanno riferimento alla parte corrente del conto del bilancio; per la definizione di tali indici si veda [2].

Tra gli indici considerati, alcuni analizzano la composizione delle entrate correnti (indici di autonomia finanziaria, impositiva e tariffaria) o il loro ammontare pro-capite (pressione finanziaria e tributaria, entrate correnti pro-capite, trasferimenti pro-capite, intervento erariale, intervento regionale), altri riguardano le spese correnti (rigidità della spesa corrente, spese correnti pro-capite, spese per il personale pro-capite), altri ancora considerano congiuntamente entrate e spese correnti (rigidità della gestione corrente, rapporto fra entrate e spese correnti, rapporto fra entrate proprie e spese correnti, rapporto fra entrate da trasferimenti e spese correnti).

Il raggruppamento per area geografica impiegato, che di seguito indicheremo con la denominazione "Geo-SO/SS", distingue 5 macroaree, che raggruppano i comuni sulla base della regione di appartenenza tenendo conto sia della collocazione geografica che dello statuto della regione; le macroaree sono perciò: comuni del nord di regioni a statuto speciale (Valle D'Aosta, Friuli Venezia Giulia, Trentino Alto Adige), comuni del nord di regioni a statuto ordinario (Piemonte, Lombardia, Liguria, Veneto, Emilia Romagna),

Metodo	Statistica	Geo-SO/SS	Geo-NUTS1
Analisi discriminante canonica	η_1^2	90,6%	71,8%
Analisi discriminante parametrica	Accuratezza	78,2%	70,7%
Modello logistico multinomiale	Accuratezza	85,7%	82,7%

Tabella 1: Confronto del potere discriminante delle partizioni Geo-SO/SS e Geo-NUTS-1; η_1^2 indica il quadrato del coefficiente di correlazione canonica.

comuni di regioni del centro - tutte a statuto ordinario - (Toscana, Umbria, Marche, Lazio), comuni di regioni del sud - tutte a statuto ordinario - (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria) e comuni di regioni delle isole - aventi tutte statuto speciale - (Sardegna e Sicilia). La scelta di impiegare questa particolare partizione è motivata non solo dalla necessità di ricorrere ad una partizione parsimoniosa e in analogia con le analisi di [5], ma anche dal fatto che l'impiego di tre diverse tecniche di valutazione della bontà della partizione (analisi discriminante canonica, analisi discriminante parametrica e regressione logistica) hanno mostrato l'elevato potere discriminante ad essa associato, sensibilmente superiore rispetto a quello della tradizionale partizione NUTS1 (si confrontino le statistiche riportate nella Tabella 1).

4 Approccio two-way

Prima di analizzare i modelli di analisi delle componenti principali a tre vie è opportuno riflettere sui risultati che è possibile ottenere con l'analisi delle componenti a due vie effettuata sulle singole slices annuali.

L'idea che ci guida in questa riflessione è che nel caso in cui la struttura di correlazioni che emerge dai diversi modelli two-way si presentasse ripetutamente nei diversi anni, sarà ragionevole attendersi che il modello Tucker3 riproduca una struttura di correlazioni sostanzialmente analoga a quella emersa nelle analisi two-way: se ciò si verifica, avremo la conferma della bontà del Tucker3 come generalizzazione a tre vie delle componenti principali.

Per portare avanti questo ragionamento riportiamo quindi i grafici (Figure 1 e 2) che mostrano le correlazioni fra gli indici di bilancio e le prime due componenti principali estratte per ciascun anno. Consideriamo soltanto le prime due componenti non solo per ragioni grafiche, ma anche perché le componenti dei modelli Tucker3 che estrarremo in seguito nella dimensione delle variabili saranno esattamente due, sicché sarà possibile effettuare un confronto diretto tra i risultati dell'analisi two-way e three-way.

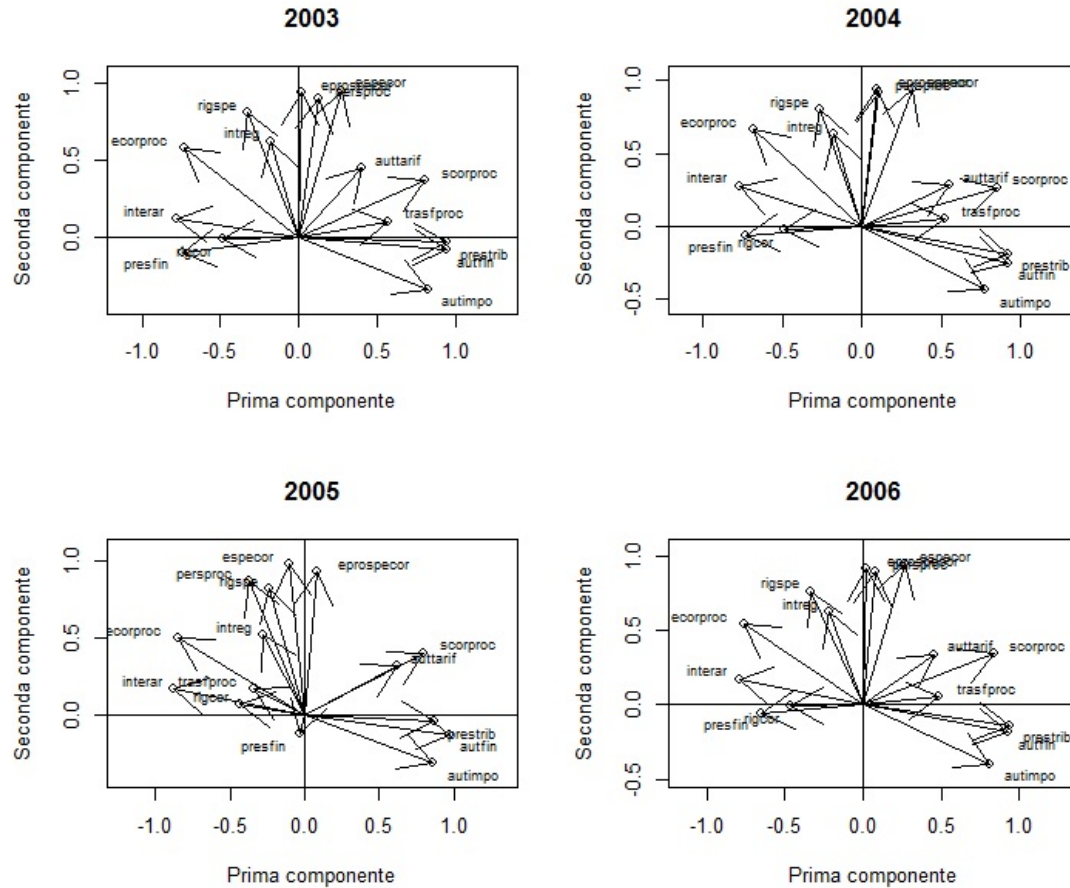


Figura 1: Cerchi delle correlazioni per gli anni 2003-2006.

A seconda dell'anno, la quota di variabilità spiegata complessivamente dalle prime due componenti varia dal 61,7% al 68,1%. Dall'analisi delle correlazioni fra componenti principali e indici di bilancio emerge che:

1. la struttura delle correlazioni è pressoché identica negli anni 2003, 2004, 2006 e 2007;
2. negli anni 2008, 2009 e 2010 la struttura delle correlazioni è la stessa che si osserva nel 2003, ma avviene un'inversione delle componenti: le variabili che erano spiegate dalla prima componente ora sono spiegate dalla seconda, quelle che erano spiegate dalla seconda ora sono spiegate dalla prima;
3. dal 2007 in poi si osserva che la prima e la seconda componente spiegano una percentuale di varianza simile. Conseguentemente, l'inversione degli assi di cui si è detto al punto precedente non è un fenomeno che stravolge il sistema di correlazioni ed è possibile ritenere che i pattern individuati ai punti 1 e 2 siano sostanzialmente similari;
4. il risultato del 2005 è anomalo in quanto dissimile da quello degli altri 7 anni.

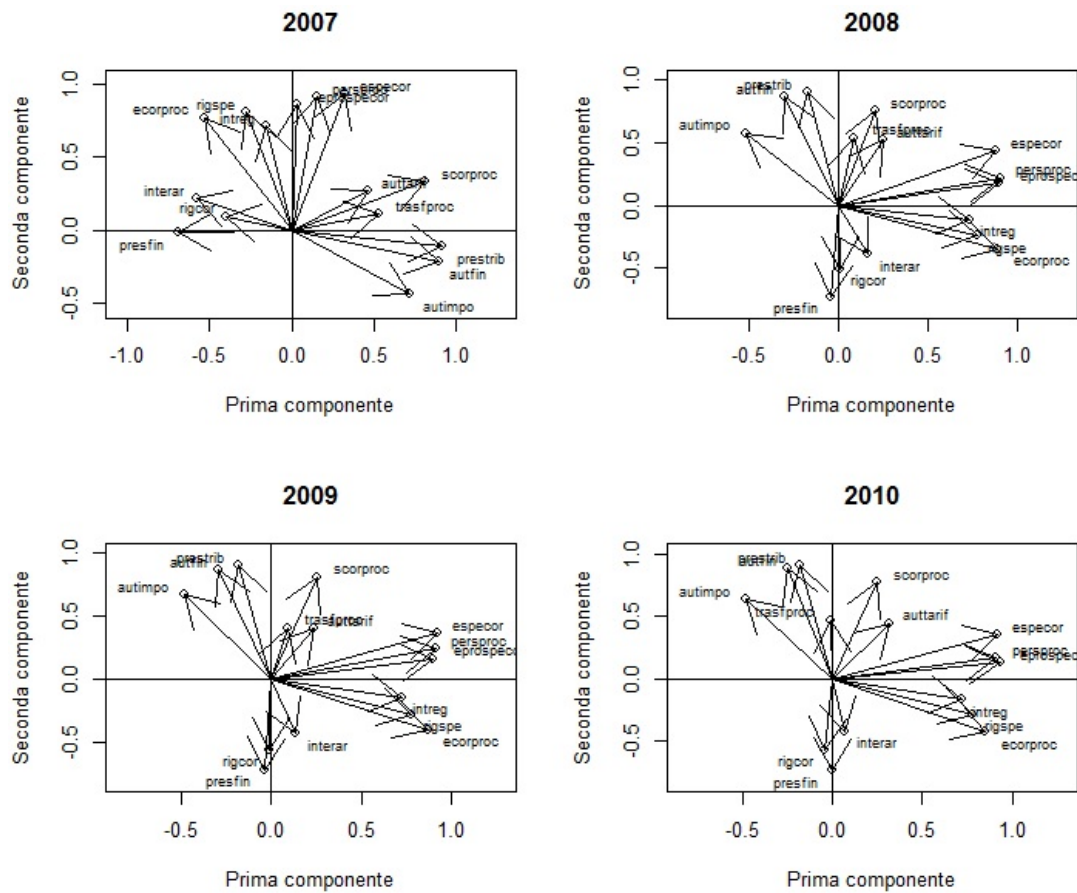


Figura 2: Cerchi delle correlazioni per gli anni 2007-2010.

Osserviamo in conclusione che l'analisi two-way mostra la persistenza di un preciso schema di correlazione sottostante agli indici, schema che si manifesta nello stesso modo per 7 degli 8 anni considerati.

Ci si attende quindi che i modelli Tucker3 stimati di seguito riproducano, in quanto modelli sintetici che operano sull'insieme delle 8 slices annuali, tale struttura di correlazione.

Anticipiamo al lettore che tale considerazione risulterà confermata dalle analisi three-way, seppur con qualche lieve differenza, come si può evincere dal confronto delle Figure 1 e 2 con le corrispondenti Figure 3, 8 e 13 nel caso three-way.

5 Applicazioni multiway

In questa sezione presentiamo i risultati che emergono dalla stima di tre diversi modelli Tucker3, stimati rispettivamente su array contenenti il dato mediano degli indici per area geografica, classe di ampiezza demografica e secondo entrambi i criteri.

La scelta di stimare il modello a partire dal dato mediano per area geografica (o, analogamente, per classe di ampiezza demografica), in luogo di un modello stimato sui dati delle singole città, è motivata da diverse considerazioni. In primo luogo, scopo della presente applicazione è mostrare le principali differenze geografiche e di dimensione demografica esistenti fra i comuni italiani: per ottenere dei coefficienti del modello direttamente associati a tali aree è necessario che i dati in input siano riferiti direttamente a tali aree, e non ai comuni (campionati) che le compongono. Questa scelta risulta inoltre ragionevole se si tiene conto del fatto che i dati analizzati appartengono ad un campione e non all'intera popolazione: il dato campionario mediano per area di un indice può perciò considerarsi, per mezzo della procedura di campionamento, una stima del valore dell'indice per l'area corrispondente.

Prima di procedere alla stima, la matrice dei dati è stata sottoposta a centratura attraverso le dimensioni A e C e normalizzazione entro la dimensione B [10].

Le soluzioni che presentiamo sono state ruotate per mezzo della rotazione Joint Orthomax proposta da Kiers [7], che ruota sia le matrici di componenti A, B e C che il full core array \mathcal{G} , operando un compromesso nella ricerca di strutture semplici tanto nelle matrici di componenti quanto nel core array.

5.1 Modello per l'analisi delle differenze territoriali

Il primo modello che proponiamo, che per ragioni di brevità denomineremo ModGeo, è stimato a partire da un array con dimensioni $I=5$, $J=15$, $K=8$, contenente i dati mediani dal 2003 al 2010 degli indici di bilancio nelle 5 aree geografiche.

La scelta del numero di componenti da estrarre nelle diverse dimensioni si basa su un algoritmo di stima proposto da Kiers e Kienderen [11] che consente di paragonare il fit di modelli con un diverso numero di componenti nelle diverse dimensioni. Sulla base di tale criterio è stato scelto il modello che presenta 2 dimensioni in ciascuna componente (sicché di seguito $P=Q=R=2$) ed un fit (misurato in termini di varianza spiegata dal modello) del 64,75%. Si tratta quindi di un modello che ha un fit elevato ed allo stesso tempo parsimonioso in termini di numero di componenti.

I coefficienti della matrice B sono riportati in Tabella 2 e rappresentati in Figura 3. Alla prima componente è possibile associare le variabili che rappresentano le entrate e le spese correnti pro-capite, gli interventi erariale e regionale, le spese per il personale pro-capite e il livello di pressione finanziaria; alla seconda i livelli di autonomia finanziaria, impositiva e tariffaria, i rapporti fra entrate e spese correnti e fra entrate proprie e spese correnti; sono inoltre correlati ad entrambe le dimensioni gli indici di rigidità della spesa e della gestione corrente, nonché il livello di pressione tributaria e dei trasferimenti pro-capite. Attraverso i coefficienti del core array (si veda Tabella 3) si possono calcolare le percentuali di varianza spiegata del modello associate alle due componenti: rispettivamente 38,8% e 61,2%. In termini sintetici si può pertanto affermare che la prima componente è correlata positivamente con la disponibilità di risorse pro-capite che l'ente locale è in grado di reperire - indipendentemente dalla modalità con cui le reperisce - mentre la seconda coglie sia il grado di autonomia dell'ente, e quindi la sua indipendenza dai trasferimenti, che la virtuosità della gestione in termini di (buon) sopravanzo delle entrate rispetto alle spese.

	B1	B2
autfin	-0.17	0.36
autimpo	-0.16	0.36
autturif	-0.04	0.37
especor	-0.05	0.37
eprospecor	-0.05	0.31
rigspe	0.13	0.30
rigcor	0.14	0.30
trasfproc	0.16	0.26
prestrib	0.17	0.27
presfin	0.30	0.09
ecorproc	0.28	0.14
scorproc	0.41	-0.03
persproc	0.43	-0.04
interar	0.37	0.05
intreg	0.43	-0.13

Tabella 2: ModGeo. Matrice B.

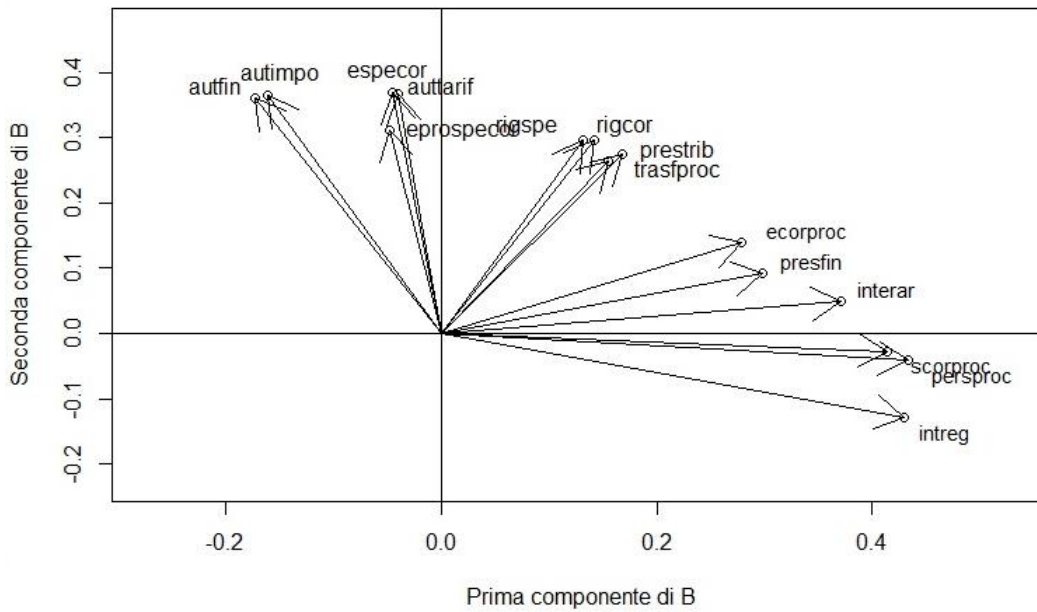


Figura 3: ModGeo. Rappresentazione grafica della matrice B in coordinate normali.

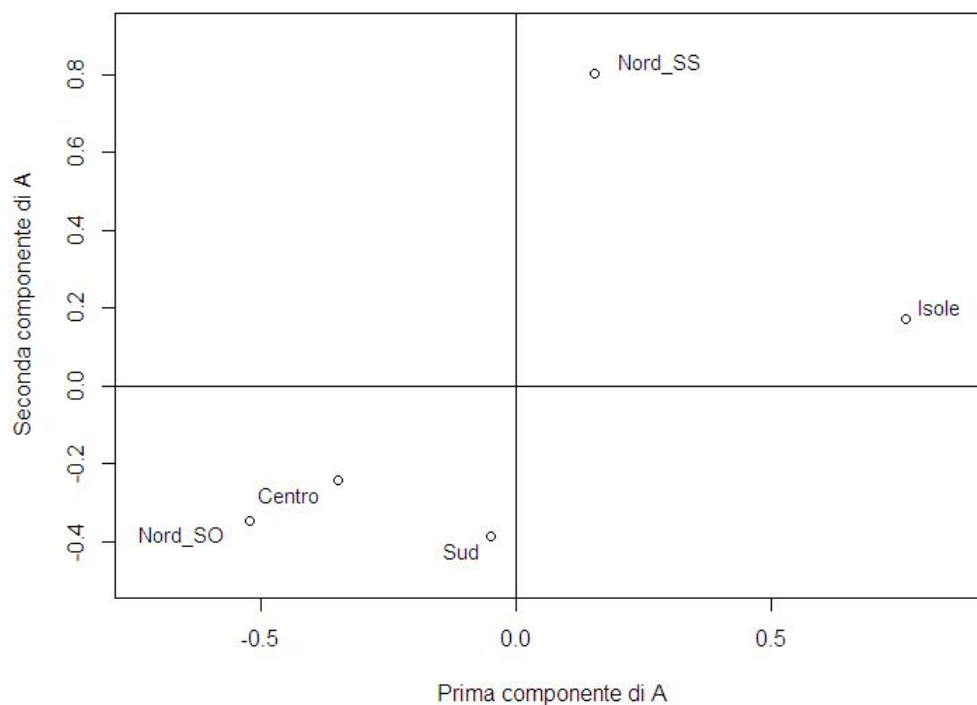


Figura 4: ModGeo. Grafico dei punteggi della matrice A.

L'analisi della matrice A, effettuata per via grafica (si veda Figura 4), consente di osservare come tanto la prima quanto la seconda componente riferite alle unità distinguono principalmente le aree a statuto speciale da quelle a statuto ordinario; in particolare la prima componente - che spiega il 13,7% della varianza spiegata dal modello - enfatizza le peculiarità dei comuni insulari, mentre la seconda - che spiega la maggior parte della varianza spiegata dal modello: l'86,3% - quelle dei comuni di regioni del nord a statuto speciale.

L'analisi della matrice C, rappresentata in Figura 5, non consente infine di ravvisare alcun trend temporale - risultato che appare coerente con l'analisi delle statistiche descrittive per anno rispetto agli indici di bilancio (non riportata in questo lavoro), il cui andamento non mostra un trend temporale chiaro ma, piuttosto, variazioni annuali che risultano difficili da interpretare in termini di andamento generale.

L'analisi del core array riportato in Tabella 3 pone in evidenza come la combinazione di componenti più forte sia il risultato dell'interazione fra A2, B2 e C2 sicché molto forte è la differenziazione in termini di livelli di autonomia e "virtuosità" dell'ente fra i comuni del nord di regioni a statuto speciale e i comuni di tutte le regioni a statuto ordinario; altra combinazione di componenti particolarmente pesante nel modello è fra A2, B1 e C2, sicché anche la differenza in termini di disponibilità di risorse pro-capite è più marcata fra comuni del nord di regioni a statuto speciale e comuni di regioni a statuto ordinario che non fra questi ultimi e quelli delle isole.

Per sintetizzare, quindi, il modello qui proposto mostra come le due componenti principali che consentono di riassumere gli indici di bilancio analizzati sintetizzano da un lato l'ammontare delle risorse medio per cittadino di cui l'ente dispone e dall'altro i livelli di autonomia e virtuosità che ne caratterizzano la gestione; rispetto ad ambedue gli aspetti emerge una marcata dicotomia fra comuni di regioni a statuto speciale e di

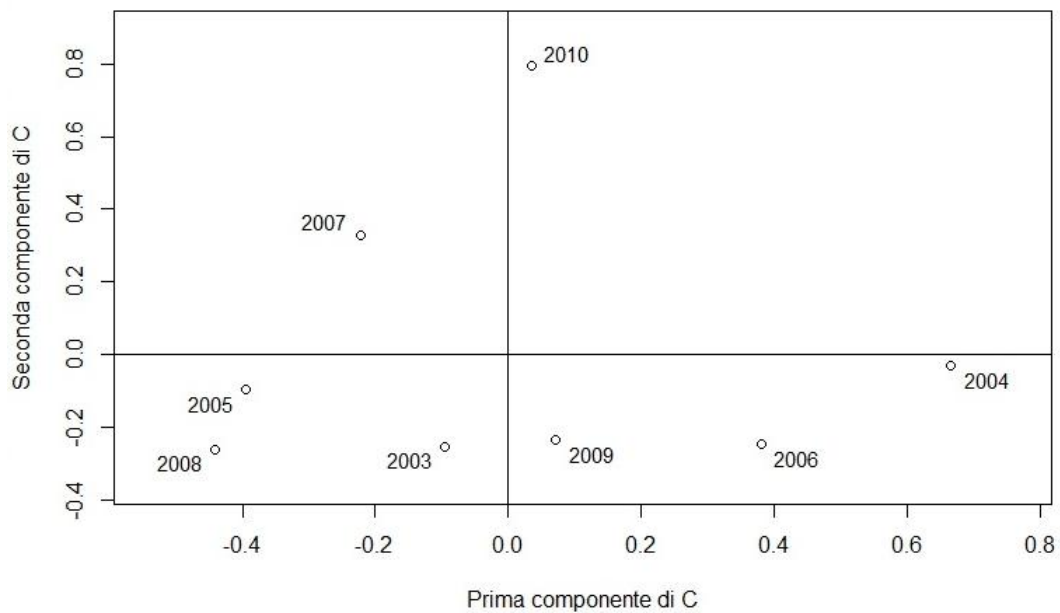


Figura 5: ModGeo. Grafico dei punteggi della matrice C.

	B1xC1	B2xC1	B1xC2	B2xC2
A1	-0.41	3.23	6.51	0.44
A2	9.92	-6.86	3.11	13.43

Tabella 3: ModGeo. Core array.

regioni a statuto ordinario e tale differenza è estremamente marcata per quanto riguarda i comuni di regioni settentrionali a statuto speciale.

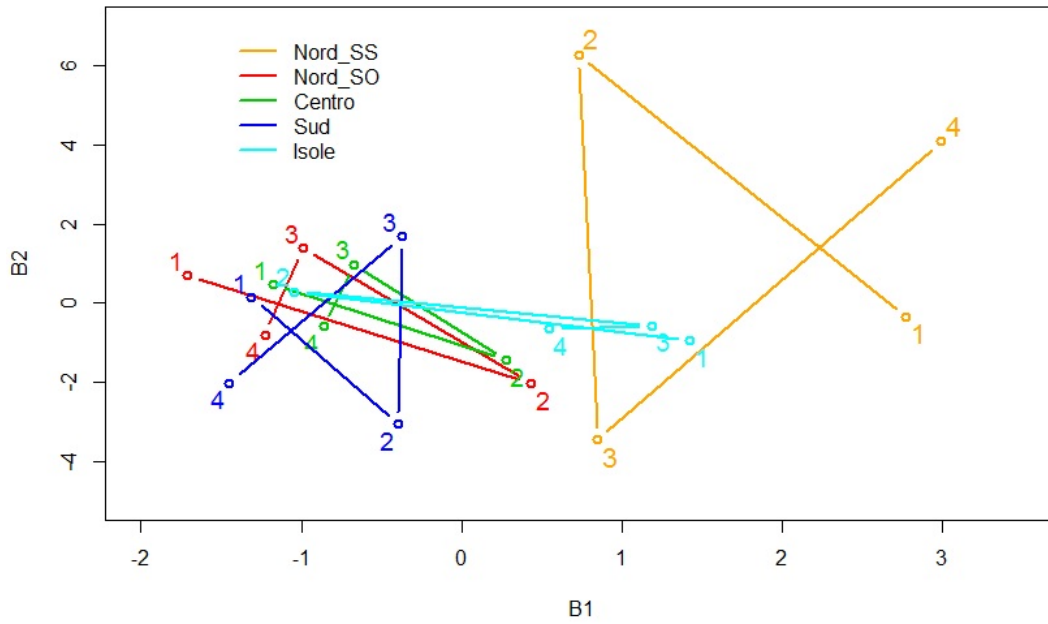


Figura 6: ModGeo. Trajectory plot per gli anni dal 2003 (nr. 1) al 2006 (nr. 4).

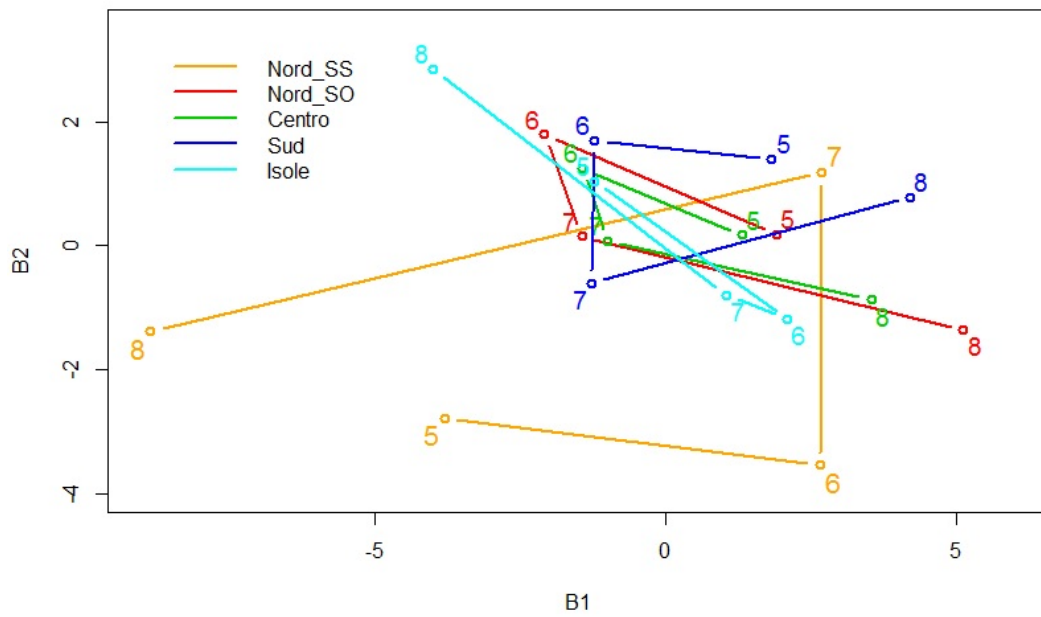


Figura 7: ModGeo. Trajectory plot per gli anni dal 2007 (nr. 5) al 2010 (nr. 8).

Interessante, infine, è l'analisi dei trajectory plot, che rappresentano su un piano cartesiano i coefficienti che è possibile ricavare dal modello per ciascuna coppia area geografica - anno [8]. La Figura 6 riporta tali coefficienti per gli anni dal 2003 al 2006, la Figura 7 quelli dal 2007 al 2010. Dall'analisi di entrambi i grafici emergono alcune evidenze:

- la sostanziale sovrapposizione che caratterizza i comuni del centro e del nord di regioni a statuto ordinario e il loro andamento simile, seppur non perfettamente sovrapposto, a quello dei comuni del sud;
- l'andamento molto diverso da un lato dei coefficienti associati ai comuni insulari e dall'altro di quello dei comuni del nord a statuto speciale - in quest'ultimo caso la traiettoria è particolarmente estesa, elemento che fa pensare che per questi comuni la variabilità temporale degli indici sia particolarmente accentuata rispetto all'andamento generale.

5.2 Modello per l'analisi delle differenze relative alla dimensione demografica

Il secondo modello che analizziamo, di seguito denominato ModDem, prende in considerazione i dati medi per classe di ampiezza demografica degli indici di bilancio nel periodo 2003-2010, così che le dimensioni dell'array risultano $I=4$, $J=15$ e $K=8$. Anche in questo caso abbiamo estratto 2 componenti per ciascuna dimensione ($P=Q=R=2$); il fit del modello associato è ancora una volta elevato, pari al 64,8%.

	B1	B2
autfin	0.21	-0.16
autimpo	0.23	-0.18
auttarif	0.40	-0.03
especor	0.41	0.00
eprospecor	0.41	-0.02
rigspe	0.40	0.07
rigcor	0.37	0.09
trasfproc	0.21	-0.00
prestrib	0.18	0.03
presfin	-0.07	0.38
ecorproc	-0.08	0.38
scorproc	0.02	0.43
persproc	0.15	0.37
interar	0.07	0.43
intreg	-0.01	0.36

Tabella 4: ModDem. Matrice B.

La lettura della Tabella 4 e l'analisi della Figura 8 permettono di osservare immediatamente come la struttura di correlazioni nel sottospazio definito dalle due componenti estratte sia simile a quella del modello

precedente: la prima componente - che spiega il 36,8% della varianza spiegata dal modello - coglie l'autonomia nella realizzazione delle entrate e la capacità di ottenere un rapporto elevato fra entrate e spese correnti; la seconda - che spiega il 63,2% della varianza spiegata dal modello - interpreta la disponibilità di risorse per abitante. Non si ha in questo caso un set di indici correlato sia con la prima che la seconda componente, in quanto anche gli indici di rigidità, il valore dei trasferimenti pro-capite e la pressione finanziaria risultano correlati prevalentemente con la prima componente.

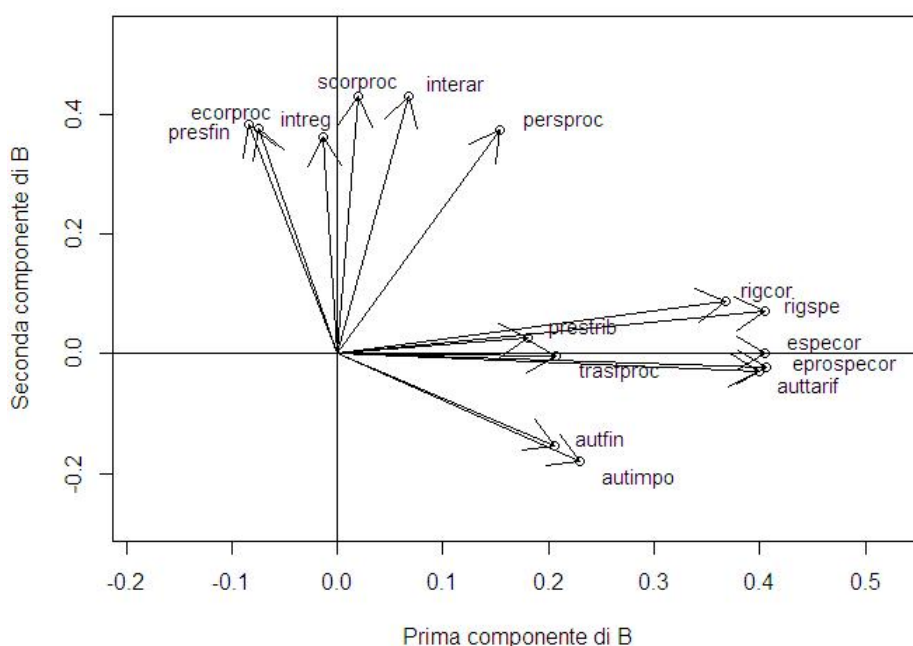


Figura 8: ModDem. Rappresentazione grafica della matrice B in coordinate normali.

I punteggi che emergono dalla matrice A (Figura 9) mostrano come la prima componente relativa alle unità statistiche operi un ordinamento crescente rispetto alla dimensione demografica, mentre la seconda distingue più genericamente i comuni al di sotto dei 50000 abitanti da quelli più grandi. Si osservi come in questo caso la prima componente spieghi ben il 94,8% della varianza spiegata, mentre poco rilevante risulti la seconda componente (5,2%).

L'analisi della matrice C (fig. 10) non consente ancora una volta di individuare alcun trend temporale rilevante.

Il core array, infine, mostra come le combinazioni di componenti più importanti nel modello siano fra le componenti A1, B2 e C1, A1, B1 e C2 e A1, B2 e C2 (vedasi Tabella 5). Conseguentemente, tanto la componente B1 quanto la componente B2 sono prevalentemente legate alla componente A1, sicché le differenze negli indici sono particolarmente marcate proprio per quanto concerne la componente A1, che meglio coglie la dimensione demografica (in quanto ordina in senso crescente le classi di ampiezza demografica). Ciò significa che al crescere della dimensione del comune è in genere maggiore l'ammontare delle risorse pro-capite disponibili, ma sono minori tanto l'autonomia dell'ente quanto la virtuosità della sua gestione.

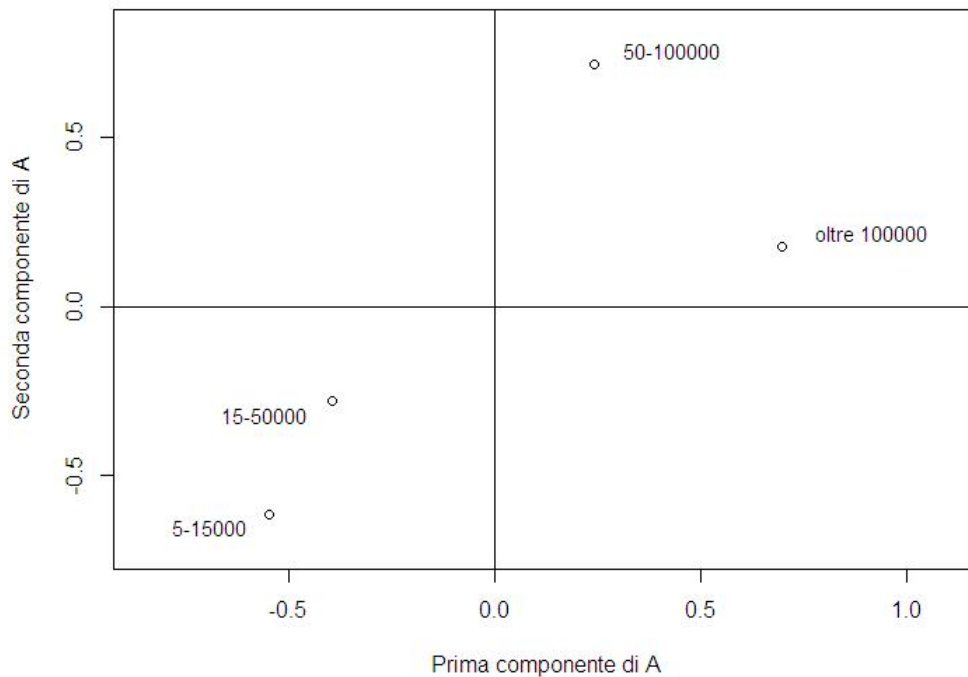


Figura 9: ModDem. Grafico dei punteggi della matrice A.

	B1xC1	B2xC1	B1xC2	B2xC2
A1	-3.77	12.30	-9.54	-6.19
A2	-1.15	-1.44	-2.83	2.20

Tabella 5: ModDem. Core array.

L'analisi dei trajectory plot (Figure 11 e 12), infine, consente di distinguere i comuni con meno di 50000 abitanti da quelli più grandi: tanto nel primo quanto nel secondo grafico la traiettoria dei comuni con un numero di abitanti compreso tra 5000 e 15000 è sostanzialmente identica per verso, dimensione e forma a quella dei comuni con numero di abitanti fra 15000 e 50000; dall'altro lato, le due classi che raggruppano i comuni più grandi hanno traiettorie di verso opposto a queste ultime, con l'aggiunta che la traiettoria dei comuni con oltre 100000 abitanti è più estesa di quella dei comuni con numero di abitanti fra i 50000 e i 100000 (segno, questo, di maggiore variabilità nel tempo).

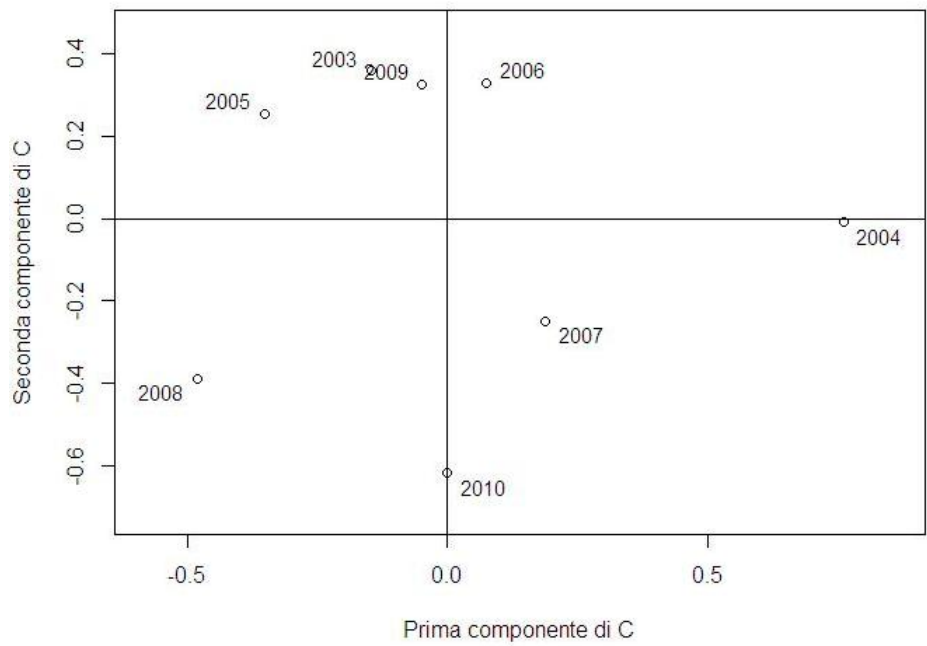


Figura 10: ModDem. Grafico dei punteggi della matrice C.

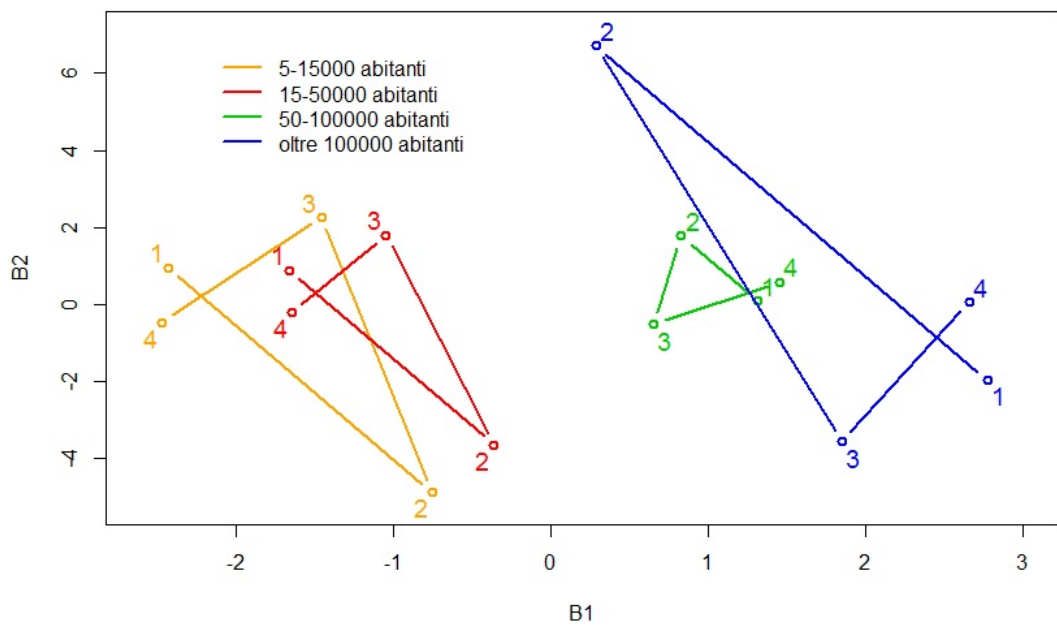


Figura 11: ModDem. Trajectory plot per gli anni dal 2003 (nr. 1) al 2006 (nr. 4).

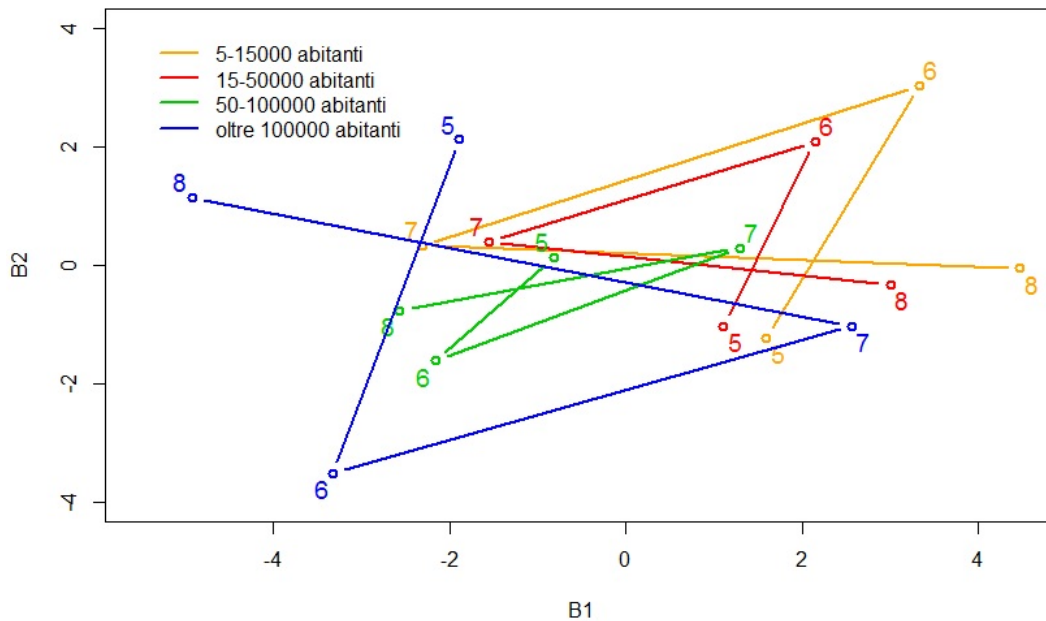


Figura 12: ModDem. Trajectory plot per gli anni dal 2007 (nr. 5) al 2010 (nr. 8).

5.3 Un modello più generale

Nei paragrafi precedenti abbiamo presentato due modelli Tucker3 che analizzano separatamente l'aspetto geografico (ModGeo) e la dimensione demografica (ModDem) dei comuni; proponiamo ora un modello più generale, che considera congiuntamente i due aspetti. Il modello, denominato ModGeoDem, è perciò stimato a partire da un array avente dimensioni $I=20$, $J=15$ e $K=8$, dove le 20 unità statistiche sono in questo caso tutte le possibili coppie area geografica - classe di ampiezza demografica; per ciascuna coppia si sono stimati, per ogni anno, i valori mediani dei diversi indici di bilancio impiegati.

Anche in questo caso, per favorire la confrontabilità con i modelli presentati in precedenza, abbiamo estratto 2 componenti in ciascuna dimensione ($P=Q=R=2$), conseguendo un fit leggermente inferiore (ma comunque buono per un modello three-way), pari al 52,17%.

	B1	B2
autfin	0.32	-0.16
autimpo	0.31	-0.16
auttarif	0.37	-0.03
especor	0.36	-0.03
eprospecor	0.38	-0.05
rigspe	0.34	0.08
rigcor	0.33	0.08
trasfproc	0.26	0.13
prestrib	0.28	0.13
presfin	0.09	0.34
ecorproc	0.03	0.37
scorproc	-0.10	0.42
persproc	0.04	0.40
interar	0.07	0.41
intreg	-0.07	0.38

Tabella 6: ModGeoDem. Matrice B.

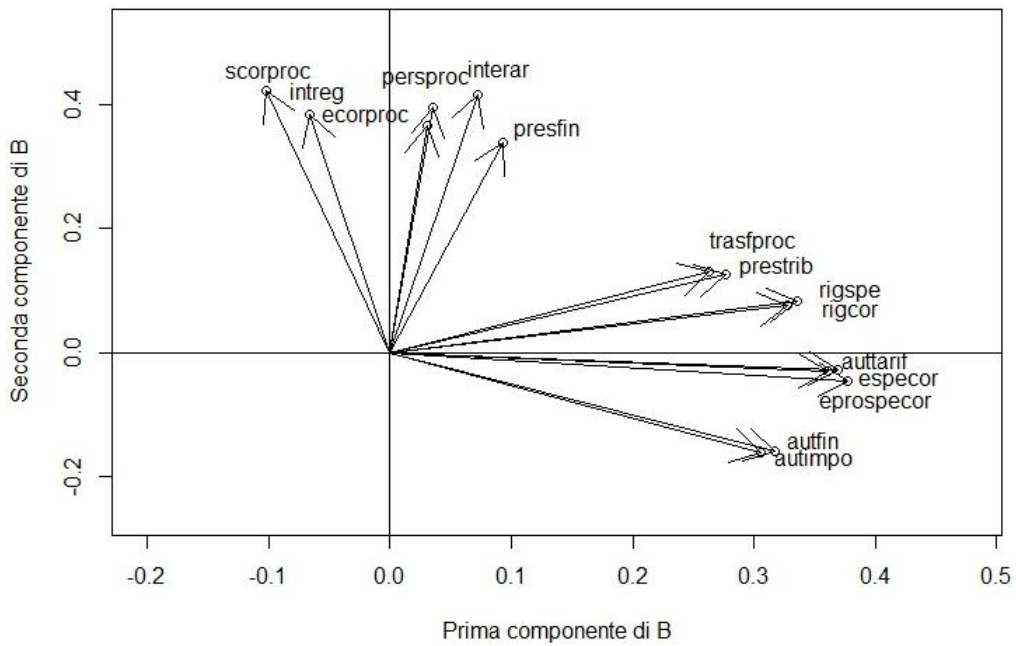


Figura 13: ModGeoDem. Rappresentazione grafica della matrice B in coordinate normali.

Il significato delle componenti della matrice B, riportata in Tabella 6 e rappresentata in Figura 13, è analogo a quello esposto nella precedente sezione (5.2): la prima componente - che in questo caso spiega il 64,8% della varianza spiegata dal modello - coglie l'autonomia nella realizzazione delle entrate e la capacità di ottenere un rapporto elevato fra entrate e spese correnti, la seconda - cui è associato il restante 35,2% di variabilità - interpreta la disponibilità di risorse per abitante.

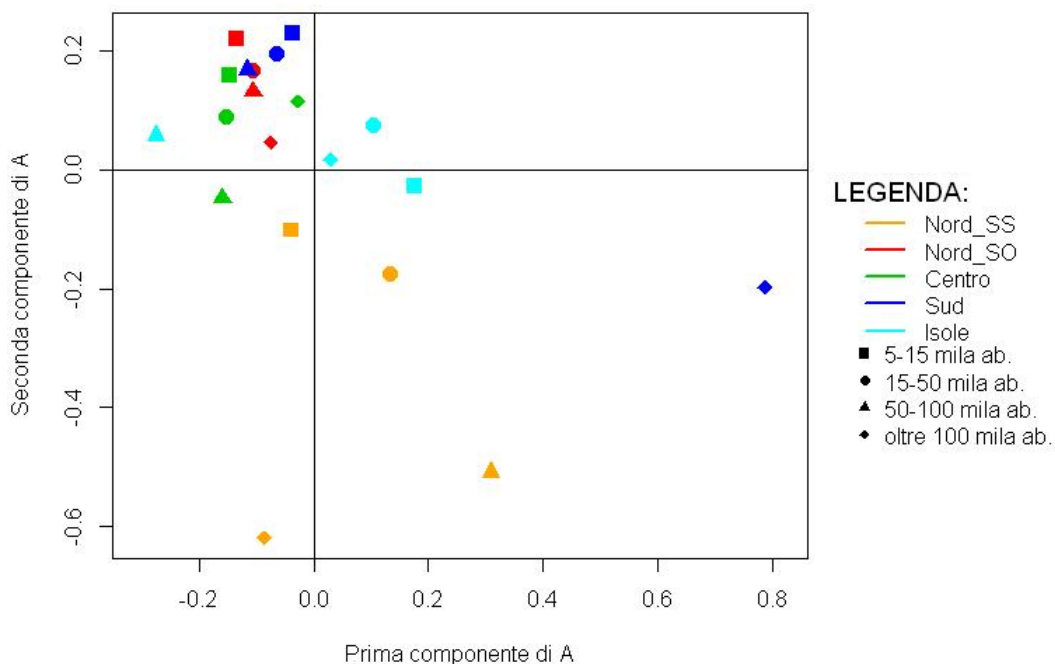


Figura 14: ModGeoDem. Grafico dei punteggi della matrice A.

L'analisi della matrice A deve ora tener conto congiuntamente tanto dell'area geografica quanto della classe di popolazione: graficamente si può osservare (vedasi Figura 14) come la disposizione dei punti nei quadranti sia orientata essenzialmente dall'area geografica; ciò ci porta ad osservare che, sulla base di questo modello, l'aspetto prevalente di differenziazione fra i comuni sembra essere l'area geografica, mentre il numero di abitanti che popolano il comune sembra avere un peso decisamente minore.

Ancora una volta, la differenziazione avviene sulla base dell'appartenenza del comune ad una regione a statuto speciale oppure ordinario; ciò si può notare dal fatto che quasi tutti i punti associati a comuni di regioni a statuto ordinario si collocano nel secondo quadrante, mentre quelli delle aree a statuto speciale si collocano altrove nel grafico. Nel caso delle regioni a statuto ordinario, inoltre, si conferma una sostanziale omogeneità fra nord, centro e sud, mentre fra le due aree a statuto speciale - nord e isole - la differenziazione (marcata dalla seconda componente) è ancora una volta più netta. Si aggiunga a ciò il fatto che la prima componente spieghi il 9,5% della variabilità del modello contro il 90,5% della seconda, ovvero risulti confermato come la differenziazione più forte sia quella dei comuni del nord a statuto speciale, mentre quella dei comuni insulari sia meno marcata. Il modello in esame riproduce quindi, in estrema sintesi, i risultati di ModGeo (par. 5.1).

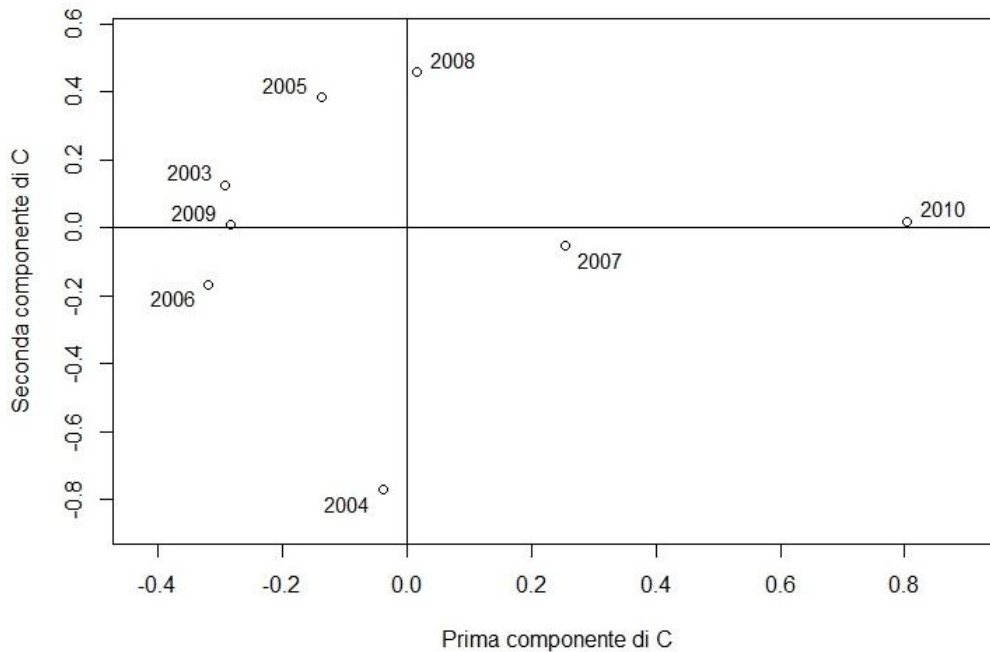


Figura 15: ModGeoDem. Grafico dei punteggi della matrice C.

	B1xC1	B2xC1	B1xC2	B2xC2
A1	-5.62	7.93	0.24	-4.95
A2	-27.00	-8.43	-7.12	16.81

Tabella 7: ModGeoDem. Core array.

La Figura 15, infine, mostra ancora una volta l'assenza di un trend temporale nei vari anni considerati.

6 Conclusioni

Nel presente articolo sono state presentate alcune applicazioni del modello Tucker3 allo studio dei bilanci dei comuni italiani.

Dal punto di vista metodologico, esso propone un approccio innovativo nel campo dell'analisi di bilancio per indici, mostrando come l'impiego dei modelli multiway possa costituire un valido strumento di sintesi in situazioni in cui si debbano analizzare più bilanci annuali di un collettivo di imprese o enti locali. Il ricorso al modello Tucker3, infatti, non solo si configura come la soluzione metodologicamente corretta allo studio di un array, ma consente altresì di condensare in un unico modello i risultati che si potrebbero ottenere con più analisi two-way separate anno per anno.

Nella sezione 4, in particolare, si è posto in evidenza il fatto che il Tucker3 è una generalizzazione coerente dell'analisi delle componenti principali a due vie mostrando l'analogia sostanziale fra i grafici

delle correlazioni fra indici di bilancio e componenti principali nel caso two-way e i corrispondenti grafici nel caso three-way.

Un aspetto distintivo dell'articolo è che esso presenta la prima applicazione di modelli multiway ai dati di bilancio dei comuni. La stima dei modelli ModGeo, ModDem e ModGeoDem ha posto in evidenza alcuni risultati particolarmente significativi:

1. tutti i modelli hanno consentito di individuare nella gestione corrente dell'ente due dimensioni principali, una indicante la disponibilità di risorse pro-capite a disposizione del comune e l'altra l'autonomia dell'ente e la sua capacità di ottenere un buon sopravanzo delle entrate sulle spese;
2. dal punto di vista delle differenze territoriali, si è riscontrato come la distinzione più rilevante sia l'appartenenza del comune ad una regione a statuto ordinario oppure speciale, mentre la direttiva nord-sud rappresenta uno strumento interpretativo secondario;
3. per quanto attiene la dimensione demografica, l'evidenza empirica suggerisce che al crescere del numero di abitanti del comune siano maggiori le risorse pro-capite disponibili e minori tanto l'autonomia quanto la virtuosità dell'ente;
4. ciò che differenzia maggiormente i comuni è l'area geografica, mentre la dimensione demografica è un aspetto di differenziazione meno importante.

Un limite del presente articolo è che i modelli impiegati non hanno consentito di individuare un trend temporale rilevante. Al riguardo, giova ricordare che i modelli multiway non tengono conto in modo esplicito della dimensione temporale ma, come rilevato da Kroonenberg [12], se nei dati è insito un andamento temporale rilevante, esso dovrebbe esser posto in evidenza dal valore delle componenti della matrice C. Nel caso in esame, tuttavia, non si è ravvisato alcun trend evidente e tale risultato appare coerente con l'andamento poco sistematico nel tempo delle statistiche descrittive degli indici di bilancio per area geografica e dimensione demografica.

Riferimenti bibliografici

- [1] Andolina P., Vassiliadis E. (2001), *Indicatori di bilancio e dinamiche produttive delle imprese manifatturiere italiane. Un'analisi "three-way"*, Studi e note di economia, 1, 145-168.
- [2] Fontana F., Rossi M. (2004), *Le analisi di bilancio nell'ente locale*, Giuffré.
- [3] Giordani P., Kiers H. A. L., Del Ferraro M. A. (2013), *Three-way component analysis using the R package ThreeWay*, Journal of Statistical Software, accettato per la pubblicazione.
- [4] Henrion R. (1994), *N-way principal component analysis. Theory, algorithms and applications*, Chemometrics and intelligent laboratory systems, 25, 1-23.
- [5] Ires Piemonte, IRPET, SRM, Eupolis Lombardia, IPRES (2012), *La finanza locale in Italia. Rapporto 2011*, Franco Angeli.
- [6] Lacangellera M., Liberati C., Mariani P. (2011), *Banking services evaluation: a dynamic analysis*, Journal of Applied Quantitative Methods, 6 (4), 3-13.
- [7] Kiers H. A. L. (1998), *Joint orthomax rotation of the core and component matrices resulting from Three-mode principal component analysis*, Journal of Classification, 15, 245-263.

- [8] Kiers H. A. L. (2000), *Some procedures for displaying results from three-way methods*, Journal of Chemometrics, 14, 151-170.
- [9] Kiers H. A. L. (1997), *Three-mode orthomax rotation*, Psychometrika, 62, 579-598.
- [10] Kiers H. A. L. (2000), *Towards a standardized notation and terminology in multiway analysis*, Journal of Chemometrics, 14, 105-122.
- [11] Kiers H. A. L., der Kienderen A. (2003), *A fast method for choosing the numbers of components in Tucker3 analysis*, British Journal of Mathematical and Statistical Psychology, 56, 119-125.
- [12] Kroonenberg P. M. (2008), *Applied Multiway Data Analysis*, Wiley.
- [13] Palazzi P. (1997), *Dinamica della struttura economica mondiale e i suoi effetti sulle relazioni nord-sud. Un'analisi empirica*, Giappicchelli.
- [14] Sito della banca dati del progetto "Open Bilanci": www.openbilanci.it.
- [15] Tucker L. (1966), *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31, 279-311.