

Il benessere economico e la sua evoluzione nelle regioni italiane: un'analisi attraverso indicatori

ADRIANA MONTE, GABRIELLA SCHOIER

Abstract

In questo studio si utilizzano alcuni degli indicatori proposti nel progetto Benessere equo e sostenibile (Bes). Questo progetto è stato sviluppato dall'Istat per misurare "un benessere equo e sostenibile" ed è il risultato dello sviluppo a livello internazionale dell'idea di sostenibilità relativa al benessere economico. Sono stati scelti diciotto indicatori appartenenti a tre delle dodici dimensioni in cui è articolato il Bes (Istruzione e formazione, Lavoro e conciliazione tempi di vita, Benessere economico) per descrivere il benessere economico nelle regioni italiane prima e dopo la crisi del 2008. In particolare, vengono utilizzati gli indicatori disponibili a livello regionale per ciascun anno nel periodo esaminato (2005-2015). Un'analisi cluster viene applicata per identificare gruppi di regioni italiane simili sulla base degli indicatori scelti, vengono poi analizzate le dinamiche del benessere economico sulla base dei cambiamenti nella composizione dei cluster.

1. Introduzione

Negli anni più recenti l'interesse per la ricerca di indicatori capaci di descrivere i vari aspetti della vita economico sociale di un paese al fine di poter monitorare l'evoluzione del progresso economico sociale e del benessere, ha portato alla nascita e allo sviluppo di numerosi progetti.

L'utilizzo del Pil come misura del progresso economico di un paese è stato criticato da vari studiosi che hanno posto l'accento su come la crescita economica, misurata con il Pil, non vada necessariamente di pari passo con il progresso socio-economico. Il Pil infatti non tiene conto, ad esempio, delle esternalità negative sociali e ambientali e nemmeno degli aspetti distributivi e di disegualianza del reddito. Nel 2007 l'Organizzazione per la Cooperazione e lo Sviluppo Economico (Ocse) al Forum mondiale di Istanbul ha affermato la necessità di individuare indicatori per misurare il progresso delle società¹. In ambito Europeo, la conferenza *Beyond the Gdp* (2007) si è posta l'obiettivo tra gli altri di decidere come misurare i concetti di progresso, ricchezza e benessere². Stiglitz et al. (2009) indicavano come fosse arrivato il momento di spostare il centro dell'attenzione dalla misurazione della produzione economica alla misurazione del benessere delle persone, individuando tre aspetti fondamentali: condizioni materiali, qualità della vita e sostenibilità (si veda anche Balducci et al., 2015). Da allora l'interesse per questi aspetti è stato crescente³ ed ha visto lo svilupparsi di molte iniziative tra le quali ricordiamo ad esempio il *Canadian Index of Wellbeing*, il *Measures of Australia's Progress*, il *Gross National Happiness Index in Buthan*. In ambito europeo ne è testimonianza l'iniziativa *Beyond the Gdp*⁴, nata sulla scia della conferenza del 2007 e lo sviluppo di iniziative da parte dei vari paesi europei, ricordiamo ad esempio nel Regno Unito il programma *Measuring National Well-being*⁵.

Il principale contributo italiano⁶ nell'ambito dell'iniziativa *Beyond the Gdp* è il progetto Benessere equo e sostenibile (Bes) che nasce nel 2010 dalla collaborazione tra il Consiglio Nazionale dell'Economia e del Lavoro (CNEL) e l'Istituto Nazionale di Statistica (Istat) con l'obiettivo di sviluppare un sistema di indica-

¹ <https://www.oecd.org/newsroom/38883774.pdf>

² http://ec.europa.eu/environment/beyond_gdp/2007_conference_en.html

³ Si veda Istat (2016), *Bes 2016, Il Benessere Equo e Sostenibile in Italia*, <http://www.istat.it/it/archivio/194029>

⁴ http://ec.europa.eu/environment/beyond_gdp/index_en.html

⁵ <http://www.ons.gov.uk/ons/guide-method/user-guidance/well-being/index.html>

⁶ Si veda anche Unioncamere, 2013.

tori per misurare gli aspetti fondamentali secondo la linea proposta da Stiglitz et al. (2009).

Il primo rapporto sul Benessere equo e sostenibile (Istat, 2013), basato su informazioni che arrivano al dettaglio regionale, viene pubblicato nel 2013 e da allora annualmente. Ad esso si sono affiancate iniziative per la misurazione del benessere equo e sostenibile a un maggiore dettaglio territoriale (Bes delle province e Urbes)⁷. La rilevanza riconosciuta a questo approccio alla misurazione del benessere individuale e sociale è testimoniata dal fatto che alcuni degli indicatori del Bes sono stati inclusi tra gli strumenti di programmazione e valutazione della politica economica nazionale, affiancando in questo il tradizionale Pil⁸. Inoltre indicatori del Bes sono rientrati anche nell'Agenda 2030 (Agenda globale per lo sviluppo sostenibile) delle Nazioni Unite approvata nel 2015, che individua 17 obiettivi di sviluppo sostenibile, misurati con oltre 200 indicatori⁹.

Molti sono stati i lavori che hanno cercato di cogliere la situazione economico sociale e il suo sviluppo nel nostro paese, sintetizzando le informazioni contenute negli indicatori Bes attraverso varie analisi multivariate (si veda ad esempio Chelli et al. 2015; Monte et al., 2017) e anche con la costruzione di indicatori di sintesi (si veda ad esempio Mazziotta et al., 2013; Chelli et al., 2016; Ciommi et al., 2017).

In questo lavoro ci si pone l'obiettivo di individuare i cluster di benessere economico delle regioni italiane e di studiare la loro evoluzione nel tempo per descrivere e interpretare i cambiamenti avvenuti nel panorama territoriale italiano nel periodo 2005- 2015, focalizzando in particolare l'attenzione sul periodo precedente e quello successivo alla crisi del 2008. Gli indicatori utilizzati per misurare il benessere economico sono selezionati dal database Bes dell'Istat e si riferiscono al periodo 2005-2015. Questo periodo è stato scelto per monitorare le transizioni dei diversi paesi da un cluster all'altro prima e dopo la crisi del 2008.

Dopo una descrizione degli indicatori utilizzati nel lavoro (paragrafo 2) e un esame alla loro evoluzione nel panorama nazionale durante il decennio 2005-2015 (paragrafo 3), viene descritta dal punto di vista teorico la metodologia utilizzata per l'analisi dei cluster e per monitorare la loro evoluzione (paragrafo 4). I risultati ottenuti vengono riportati e commentati nel paragrafo 5. Seguono le conclusioni.

⁷ <http://www.istat.it/it/benessere-e-sostenibilita>

⁸ http://www.mef.gov.it/inevidenza/article_0276.html

⁹ <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>

2. Il Bes e il benessere economico

Il sistema di indicatori individuato nell'ambito del Bes nasce da un approccio multidimensionale "per misurare il benessere equo e sostenibile che integra l'indicatore dell'attività economica, il Pil, con le fondamentali dimensioni del benessere e con misure di disegualianza e sostenibilità economica"¹⁰. Tali indicatori sono articolati in 12 dimensioni (domini) e cioè *Salute, Istruzione e formazione, Lavoro e conciliazione tempi di vita, Benessere economico, Relazioni sociali, Politica e istituzioni, Sicurezza, Benessere soggettivo, Paesaggio e patrimonio culturale, Ambiente, Ricerca e innovazione, Qualità dei servizi*. Una Commissione scientifica di esperti ha avuto il compito di individuare gli indicatori statistici più adeguati alla misurazione delle dimensioni citate.

Uno di questi domini viene chiamato *Benessere economico* e viene definito nell'ambito del Bes come "il mezzo attraverso il quale un individuo riesce ad avere e sostenere un determinato standard di vita. Un'analisi del benessere economico fa riferimento al reddito, alla ricchezza, alla capacità di consumo, ma anche ad alcune dimensioni di benessere materiale che tali strumenti permettono di acquisire (condizioni abitative, possesso di beni durevoli, ecc.)" (Istat, 2016).

Nel presente lavoro vengono considerati anche altri due domini del Bes congiuntamente al *Benessere economico* e cioè il *Lavoro e conciliazione tempi di vita*, sulla base della considerazione che gli indicatori utilizzati per la sua misurazione sono tendenzialmente capaci di darci indicazioni sulla possibilità di mantenere il benessere economico nel tempo, e l'*Istruzione e formazione* misurata da indicatori che possono dare un'indicazione della capacità del sistema di preparare all'immissione nel mercato del lavoro e di fornire una formazione continua. Queste ultime due dimensioni del Bes integrano così il concetto di benessere economico di un territorio, tentando di estenderlo anche a considerare la capacità futura di poterlo mantenere.

Gli indicatori relativi alle tre dimensioni individuate (*Istruzione e formazione, Lavoro e conciliazione tempi di vita, Benessere economico*) sono 35, non tutti sono disponibili a livello regionale, alcuni vengono rilevati a partire da anni successivi alla crisi del 2008, altri hanno una cadenza superiore all'anno. Nella Tabella 1 vengono riportate le numerosità degli indicatori per ciascuna delle dodici dimensioni del Bes. Come si vede il numero di indicatori disponibili a livello regionale per tutto il periodo in esame è più limitato; per quanto riguarda in particolare i tre domini considerati in questo studio si tratta di 18 dei 35 indicatori individuati.

¹⁰ Si veda <http://www.istat.it/it/benessere-e-sostenibilit%C3%A0/misure-del-benessere>

Nella Tabella 2 sono elencati gli indicatori che sono stati utilizzati nel presente lavoro. Si tratta di 5 degli 11 indicatori del dominio *Istruzione e formazione* (si veda Tabella 1), 6 dei 14 indicatori del dominio *Lavoro e conciliazione tempi di vita* e 7 degli 11 del *Benessere economico*. Gli indicatori sono stati scelti sulla base della loro disponibilità a livello regionale per tutti gli anni del periodo oggetto di studio. Nel Bes vi sono infatti alcuni indicatori che vengono rilevati per macroaree (Nord, Centro, Sud), altri che presentano una cadenza superiore all'anno ed altri che seppure disponibili per regione, sono stati introdotti successivamente (si veda Istat, 2016).

TABELLA 1. Domini del Bes secondo numero di indicatori

Dominio	Numero di indicatori	Numero di indicatori con cadenza annuale disponibili dal 2005 al 2015 per le regioni
1. <i>Salute</i>	14	5
2. <i>Istruzione e formazione</i>	11	5
3. <i>Lavoro e conciliazione tempi di vita</i>	14	6
4. <i>Benessere economico</i>	10	7
5. <i>Relazioni sociali</i>	9	5
6. <i>Politica e istituzioni</i>	12	-
7. <i>Sicurezza</i>	11	1
8. <i>Benessere soggettivo</i>	4	1
9. <i>Paesaggio e patrimonio culturale</i>	12	1
10. <i>Ambiente</i>	15	2
11. <i>Ricerca e innovazione</i>	7	2
12. <i>Qualità dei servizi</i>	11	2

TABELLA 2. Elenco e descrizione degli indicatori Bes utilizzati in questo studio, secondo dimensione Bes di appartenenza

Dominio Bes	Indicatore	Descrizione
Istruzione e formazione	Dip – Persone con almeno il diploma	Percentuale di persone di 25-64 anni che hanno completato almeno la scuola secondaria di II grado (scuola media superiore)
	Univ – Persone che hanno conseguito un titolo universitario	Percentuale di persone di 30-34 anni che hanno conseguito un titolo universitario
	Prec – Uscita precoce dal sistema di istruzione e formazione	Percentuale di persone di 18-24 anni che hanno conseguito solo la licenza media e non sono inseriti in un programma di formazione
	Neet – Giovani che non lavorano e non studiano (Neet)	Percentuale di persone di 15-29 anni né occupate né inserite in un percorso di istruzione o formazione (Neet)
	Form – Partecipazione alla formazione continua	Percentuale di persone di 25-64 anni che hanno partecipato ad attività di istruzione e formazione nelle 4 settimane precedenti l'intervista
Lavoro e conciliazione tempi di vita	Occ – Tasso di occupazione 20-64 anni	Percentuale di occupati di 20-64 anni sulla popolazione di 20-64 anni
	Nolav – Tasso di mancata partecipazione al lavoro	Percentuale di disoccupati di 15-74 anni + forze di lavoro potenziali di 15-74 anni che non cercano lavoro ma sono disponibili a lavorare sul totale delle forze di lavoro 15-74 anni + forze di lavoro potenziali 15-74 anni che non cercano lavoro ma sono disponibili a lavorare
	Oltre5 – Percentuale di occupati in lavori a termine da almeno 5 anni	Percentuale di dipendenti a tempo determinato e collaboratori che hanno iniziato l'attuale lavoro da almeno 5 anni
	Sovra – Incidenza di occupati sovraistruiti	Percentuale di occupati che possiedono un titolo di studio superiore a quello maggiormente posseduto per svolgere quella professione sul totale degli occupati
	Donne – Rapporto tra tasso di occupazione delle donne di 25-49 anni con figli in età prescolare e delle donne senza figli	Tasso di occupazione delle donne di 25-49 anni con almeno un figlio in età 0-5 anni sul tasso di occupazione delle donne di 25-49 anni senza figli
	Invpt – Quota di part-time involontario	Percentuale di occupati che dichiarano di svolgere un lavoro a tempo parziale perché non ne hanno trovato uno a tempo pieno sul totale degli occupati

Dominio Bes	Indicatore	Descrizione
Benessere economico	Red - Reddito medio disponibile pro-capite	Rapporto tra il reddito disponibile delle famiglie consumatrici e il numero totale di persone residenti (<i>in euro</i>)
	Dis - Indice di disuguaglianza del reddito disponibile	Rapporto tra il reddito equivalente totale ricevuto dal 20% della popolazione con il più alto reddito e quello ricevuto dal 20% della popolazione con il più basso reddito
	Rpov - Indice di rischio di povertà	Percentuale di persone a rischio di povertà con un reddito equivalente inferiore o pari al 60% del reddito mediano sul totale delle persone residenti
	Depr - Indice di grave deprivazione materiale	Percentuale di persone che vivono in famiglie con grave deprivazione materiale povertà
	Ab - Indice di bassa qualità dell'abitazione	Persone che vivono in situazioni di sovraffollamento abitativo, in abitazioni prive di alcuni servizi e con problemi strutturali (<i>valori percentuali</i>)
	Econ - Indice di grande difficoltà economica	Quota di persone in famiglie che dichiarano di arrivare con grande difficoltà alla fine del mese
	Lav - Molto bassa intensità lavorativa	Incidenza di persone che vivono in famiglie dove le persone in età lavorativa (tra i 18 e i 59 anni, con l'esclusione degli studenti 18-24) nell'anno precedente, hanno lavorato per meno del 20 per cento del loro potenziale (con esclusione delle famiglie composte soltanto da minori, da studenti di età inferiore a 25 anni e da persone di 60 anni o più)

3. L'evoluzione temporale degli indicatori Bes utilizzati

Una sintesi del comportamento degli indicatori nel tempo viene rappresentata nei box plot in appendice (Figure 3, 4 e 5), in cui vengono riportate per ogni anno dal 2005 al 2015 le distribuzioni dei diversi indicatori. In particolare per quanto riguarda il dominio *Istruzione e Formazione*, tre dei cinque indicatori migliorano, aumenta infatti la percentuale di persone con almeno il diploma (**Dip**) e anche quella delle persone che hanno conseguito un titolo universitario (**Univ**), mentre diminuiscono le uscite precoci dal sistema di istruzione e formazione (**Prec**). La partecipazione alla formazione continua (**Form**) aumenta negli anni più recenti, ma con una maggiore dispersione della distribuzione a prova del fatto che le regioni si sono comportate in modo più diversificato rispetto agli anni precedenti. I giovani che non lavorano e non studiano (**Neet**) aumentano invece in particolare dal 2009, anno successivo all'inizio della crisi finanziaria (Figure 3).

Il dominio *Lavoro e conciliazione tempi di vita* viene misurato con i sei indicatori riportati in Figura 4. L'andamento dell'indicatore **Occ** (Tasso di occupazione della popolazione in età 20-64 anni), i cui valori mediani dal 2009 tendono leggermente a scendere, rivela un aumento della variabilità nel tasso di occupazione tra le regioni, ma con un costante aumento dell'incidenza degli occupati sovraistrutti (**Sovra**). Dal 2009 tende ad aumentare anche il tasso di mancata partecipazione al lavoro della popolazione in età 15-74 (**Nolav**) in modo maggiormente differenziato tra le regioni. Mentre la percentuale di occupati in lavori a termine da almeno 5 anni (**Oltre5**) presenta un andamento pressoché costante nei valori mediani, l'indicatore che risente maggiormente della crisi del 2008 è **Invpt** (quota di part time involontario) che cresce costantemente non solo nei valori mediani, ma anche in termini di variabilità tra regioni.

Per quanto riguarda il dominio *Benessere economico* (Figura 5) vi è un generale peggioramento nel tempo. Dal 2009 aumenta il valore mediano dell'indice di grave deprivazione materiale (**Depr**), dell'indice di rischio povertà (**Rpov**) e dell'indice di bassa qualità dell'abitazione (**Ab**) ed anche la loro variabilità tra le regioni. La disuguaglianza del reddito disponibile (**Dis**) e la percentuale di persone con meno di 60 anni che vivono in famiglie a intensità lavorativa molto bassa (**Lav**) pur presentando valori mediani pressoché stabili nel tempo, presentano una maggiore dispersione e asimmetria. Il reddito disponibile pro-capite (**Red**) rimane praticamente invariato dal 2009.

4. La metodologia: la *cluster analysis* e la *cluster transition*

4.1 LA CLUSTER ANALYSIS

Esistono diversi metodi statistici per la classificazione delle unità in gruppi omogenei; essi possono essere suddivisi in due grandi categorie: *supervised classification* e *unsupervised classification*. Nel primo caso si hanno a priori unità osservate che appartengono a due o più popolazioni differenti e di ognuna si conoscono i valori delle variabili considerate. Lo scopo di questo tipo di analisi è l'assegnazione di ulteriori unità alla popolazione di appartenenza, minimizzando la probabilità di errore di attribuzione. Al contrario i metodi del tipo *unsupervised classification* sono tipicamente esplorativi e consistono nella ricerca, nelle osservazioni *n*-dimensionali, di gruppi di unità tra loro simili, non sapendo a priori se tali gruppi omogenei esistono effettivamente nel dataset (Zani et al., 2007). La classificazione non supervisionata, o *cluster analysis*, ha quindi l'obiettivo di riconoscere dei gruppi che si caratterizzano per un'elevata omogeneità all'interno e per un'elevata eterogeneità tra di essi.

Gli ambiti di applicazione della *cluster analysis* sono molto vari, essa è utilizzabile per: ridurre i dati in forma grafica (per evidenziare le più importanti informazioni rilevate oppure per presentare i risultati di analisi multivariate); generare ipotesi di ricerca (prima di provare un qualsiasi modello di analisi sui dati rilevati è utile individuare le connessioni reali tra le entità e intuire in base a queste i modelli presenti nei dati); individuare gruppi di unità con caratteristiche distintive che, nell'insieme, facciano percepire la fisionomia del sistema sociale osservato; costruire sistemi di classificazione automatica; stratificare popolazioni da sottoporre a campionamento.

Gli algoritmi di *clustering* suddividono i dati in un certo numero di cluster (o gruppi, sottoinsiemi, categorie). Non esiste una definizione univoca per queste procedure, anche se molti studiosi convergono nel riconoscere un cluster nel caso di omogeneità all'interno di ogni gruppo e di eterogeneità tra i diversi gruppi.

Si supponga di considerare un dataset di $N = \{1, 2, \dots, n\}$ unità e di disporre per ognuna le rilevazioni su p variabili; queste informazioni vengono inserite nella matrice dei dati \mathbf{X} di dimensioni $(n \times p)$, il *partitional clustering* consiste nel ricercare una partizione di \mathbf{X} in K gruppi ($K \leq N$), $C = \{C_1, \dots, C_k\}$ tali che:

1. $C_i \neq \emptyset$ per $i = 1, \dots, K$;
2. $\bigcup_{i=1}^K C_i = \mathbf{X}$;
3. $C_i \cap C_j = \emptyset$ con $i, j = 1, \dots, K, i \neq j$.

Come si nota dai vincoli imposti, ogni osservazione appartiene ad un unico cluster tuttavia è possibile allentare questa restrizione e supporre che un'osservazione appartenga a tutti i cluster con un certo grado di appartenenza, $u_{i,j} \in [0,1]$, che rappresenta il coefficiente di appartenenza della j -esima osservazione all' i -esimo cluster in questo caso si parla di *fuzzy clustering* di cui non si tratta nel presente lavoro.

La *cluster analysis* tradizionale consiste in quattro semplici fasi strettamente collegate tra loro; come si vede in Figura 1, la procedura può richiedere una serie di tentativi e di ripetizioni dei vari passaggi che vengono di seguito sintetizzati: scelta delle variabili; scelta dell'algoritmo di *clustering*; validazione dei cluster; interpretazione dei risultati.

Esistono diversi algoritmi di *clustering* che possono essere classificati secondo il seguente schema: metodi gerarchici (agglomerativi, divisivi), metodi non gerarchici, metodi basati sull'errore quadratico, metodi basati sui modelli mistura, metodi basati sulla teoria dei grafi e altri metodi.

Si supponga di partire da un dataset di $N = \{1, 2, \dots, n\}$ unità e di possedere per ognuna le rilevazioni per p variabili. Si consideri la matrice dei dati \mathbf{X} di dimensioni $(n \times p)$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

dove x_{ih} rappresenta l'osservazione della variabile h sulla unità i , con $h = 1, \dots, p$; $i = 1, \dots, n$.

Secondo i tradizionali metodi di *clustering*, per individuare dei gruppi di unità omogenei è fondamentale ricavare per ogni coppia di elementi degli indici di prossimità; grazie a questi è possibile raggruppare le n unità in g sottoinsiemi e "ridurre le dimensioni" dello spazio \mathbb{R}^n .

Un indice di prossimità tra due generiche unità statistiche u_i e u_j è definito come funzione dei vettori riga della matrice dei dati \mathbf{X} :

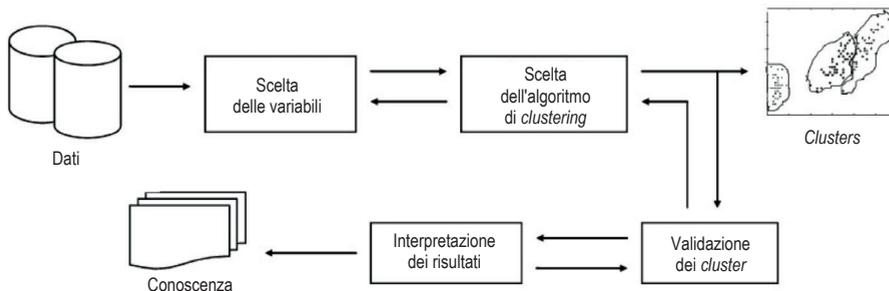
$$IP_{ij} = f(\mathbf{x}'_i, \mathbf{x}'_j) \quad i, j = 1, 2, \dots, n$$

Gli indici di prossimità vengono abitualmente distinti tra indici di dissimilarità (ai quali appartengono le distanze), applicati solitamente a fenomeni quantitativi e indici di similarità, utilizzati per fenomeni qualitativi (Zani et al., 2007).

Esistono diversi tipi di distanze, tra queste la distanza euclidea è indubbiamente quella più conosciuta. Si definisce distanza euclidea tra due unità statistiche i e j la norma euclidea della differenza tra i rispettivi vettori:

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \left[\sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2}$$

FIGURA 1. Le fasi di una procedura di *clustering*



FONTE: Xu, 2005

Un altro tipo di distanza che, a differenza della precedente, tiene conto della relazione tra le variabili è la distanza di Mahalanobis. Si dice distanza di Mahalanobis tra i vettori \mathbf{x}_i e \mathbf{x}_j l'espressione:

$${}_M d_{ij} = c \left[(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \right]^{1/2}$$

dove \mathbf{S} è la matrice di covarianza e $c = \sqrt{\frac{n}{n-1}}$ è un fattore correttivo (introdotto in quanto la matrice di covarianza \mathbf{S} è stimata in base alle n osservazioni) che tende ad 1 quando $n \rightarrow +\infty$.

4.1.2 I metodi gerarchici

I metodi gerarchici aggregativi consentono di ottenere una famiglia di partizioni delle n unità statistiche partendo da quella in cui tutte le unità sono distinte, per arrivare a quella in cui tutte sono riunite in modo aggregativo; sono quindi dei metodi che si sviluppano per fasi ordinate, in modo che ad ogni passo vengano uniti i due gruppi (o le due unità nella prima fase) scelti a seconda della misura di distanza (o di similarità) precedentemente definita¹².

Gran parte dei metodi gerarchici partono dalla matrice di distanze \mathbf{D} calcolata per le n unità statistiche. In questo caso la procedura generale per il raggruppamento delle unità è il seguente.

FASE 1: si individuano nella matrice \mathbf{D} le due unità tra loro più simili (in pratica quelle con minor distanza¹³) e si aggregano. Questo è il primo gruppo di unità. Si ottiene una partizione con $(n - 1)$ gruppi, di cui $(n - 2)$ costituiti da singole unità e l'altro formato da due unità.

FASE 2: si ricalcolano le distanze del gruppo ottenuto dagli altri gruppi (alcuni saranno costituiti da una sola unità), ottenendo una nuova matrice delle distanze con dimensioni diminuite di uno.

FASE 3: si individua nella nuova matrice delle distanze la coppia di gruppi (o unità) con minore distanza, unendoli in un solo gruppo.

FASE 4: si ripetono la fase 2 e la fase 3 fino a che tutte le unità sono riunite in un unico cluster.

¹¹ Questa tipologia di metodo viene anche definita *bottom up*, poiché parte dalle singole unità statistiche (dal basso) e procede ad unirle in gruppi sempre più grandi.

¹² Se si parte da una matrice di indici di similarità non si guarda alla minore distanza ma alla "maggiore similarità".

La differenza tra i diversi metodi gerarchici consiste solamente nel diverso criterio utilizzato per calcolare la distanza tra due gruppi di unità. Si supponga di avere due cluster C_1 e C_2 formati rispettivamente da n_1 e n_2 unità: sono possibili diverse definizioni di distanza tra i due gruppi, che identificano altrettanti metodi gerarchici. Di seguito si riportano i metodi più noti.

Nel metodo del legame singolo (o del vicino più prossimo) la distanza tra i due gruppi è definita come il minimo delle $n_1 n_2$ distanze tra ciascuna delle unità di un gruppo e ciascuna delle unità dell'altro gruppo: $d(C_1, C_2) = \min (d_{rs})$ con $r \in C_1, s \in C_2$.

In pratica con questa definizione, ad ogni passo si valuta la distanza tra due cluster attraverso la distanza dei punti più vicini.

Al contrario, nel metodo del legame completo (o del vicino più lontano) la distanza tra i due gruppi è definita come il massimo delle $n_1 n_2$ distanze tra ciascuna delle unità di un gruppo e ciascuna delle unità dell'altro gruppo:

$$d(C_1, C_2) = \max (d_{rs}) \text{ con } r \in C_1, s \in C_2$$

Con questo metodo tutte le distanze tra le unità di C_1 e le unità di C_2 sono minori (o uguali) alla distanza definita dal criterio.

Infine il metodo del legame medio tra i gruppi, secondo il quale la distanza tra due cluster viene definita tramite la media aritmetica delle $n_1 n_2$ distanze tra ciascuna delle unità di un gruppo e ciascuna delle unità dell'altro gruppo:

$$d(C_1, C_2) = \frac{1}{n_1 n_2} \sum_r \sum_s d_{rs} \text{ con } r \in C_1, s \in C_2.$$

I metodi visti fino a questo punto richiedono solamente la conoscenza della matrice delle distanze. Vi sono altri metodi gerarchici che utilizzano anche la matrice dei dati di partenza.

Il metodo del centroide prevede che la distanza tra due cluster C_1 e C_2 venga calcolata come la distanza tra due centroidi \bar{x}_1 e \bar{x}_2 : $d(C_1, C_2) = d(\bar{x}_1, \bar{x}_2)$.

Il centroide iniziale non è altro che il vettore che contiene i valori medi delle p variabili per le unità incluse nel gruppo. Al passo successivo, il centroide del nuovo cluster potrà essere calcolato come una media aritmetica dei centroidi dei due gruppi iniziali: $\text{centroide}(C_1 \cup C_2) = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n_1 + n_2}$.

Il metodo del centroide presenta alcune analogie con il metodo del legame medio: in quest'ultimo si considera la media delle distanze tra le unità dell'uno e dell'altro gruppo, mentre nel metodo del centroide si individua prima il "centro" di ogni gruppo e si misura poi la distanza tra essi.

Nel metodo di Ward (o della minima devianza) non è richiesto il calcolo preliminare di una matrice delle distanze, ma si definisce esplicitamente una

funzione obiettivo. Poiché, come già ribadito, lo scopo della classificazione è quello di ottenere gruppi con la maggiore coesione interna, si considera la scomposizione della devianza totale (indicata con la lettera T) delle p variabili in devianza nei gruppi (*Within*, indicata con la lettera W) e devianza tra i gruppi (*Between*, indicata con la lettera B).

Se si suppongono g gruppi si definisce $T = \sum_{s=1}^p \sum_{i=1}^{n_i} (x_{is} - \bar{x}_{s,l})^2$ la devianza totale delle p variabili, ottenuta come somma delle devianze delle singole variabili rispetto alla corrispondente media generale \bar{x}_s . Questa devianza può essere così scomposta:

$$T = W + B = \sum_{s=1}^p \sum_{i=1}^{n_i} (\bar{x}_{s,l} - \bar{x}_s)^2 + \sum_{s=1}^p \sum_{i=1}^{n_i} n_i (\bar{x}_{s,l} - \bar{x}_s)^2.$$

Il primo termine rappresenta la somma delle devianze di gruppo, il secondo la somma, calcolata su tutte le variabili, delle devianze ponderate delle medie di gruppo rispetto alla corrispondente media generale.

Questa procedura aggrega ad ogni passo i cluster aventi la minima devianza nei gruppi (*Within*), cioè i cluster più omogenei al loro interno.

In generale è possibile rappresentare graficamente le varie partizioni che si ottengono nelle varie fasi di un metodo gerarchico mediante un albero n -dimensionale che viene definito dendrogramma. Questo diagramma ad albero permette di visualizzare i gruppi ottenuti ad ogni stadio dell'operazione di *clustering*.

4.2 LA CLUSTER TRANSITION

Lo studio del clustering evolutivo è diventato un importante tema di ricerca, soprattutto nell'ultimo decennio. La necessità di descrivere e comprendere il comportamento di un determinato fenomeno nel tempo ha portato a nuovi metodi incentrati sull'evoluzione temporale. La cluster analysis è ampiamente utilizzata in ambito economico per il raggruppamento di unità statistiche. Il raggruppamento è stato principalmente applicato a dati statici in modo che la struttura possa descrivere bene la situazione in un istante di tempo (si veda, ad esempio Berkhin, 2006) ma altrettanto importanti si sono rivelate le applicazioni della cluster analysis alle dinamiche delle strutture dei raggruppamenti. Il nostro lavoro si colloca all'interno di questo secondo filone di ricerca. Per cogliere le trasformazioni che si verificano nella struttura dei cluster tra istanti diversi di tempo, è stata considerata la seguente tassonomia formata da cinque schemi tassonomici (vedi Oliveira et al., 2012):

- Nascita: emerge un nuovo cluster

- Morte: un cluster scompare
- Divisione: un cluster si suddivide in due o più cluster
- Fusione: due o più cluster si fondono in un unico cluster
- Sopravvivenza: un cluster che non subisce nessuna delle transizioni di cui sopra

Queste transizioni sono esterne, poiché sono legate a cambiamenti nell'intero clustering e rappresentano i cambiamenti dei cluster in due istanti. Per essere in grado di rilevare le transizioni, è necessario scoprire quali cluster al tempo t corrispondono a cluster precedentemente trovati in $t-1$; per questo motivo è fondamentale eseguire la mappatura dei cluster in un dato intervallo di tempo, per scoprire le corrispondenze dei cluster. Nel presente lavoro abbiamo usato il metodo indicato da Spiliopoulou et al. (2006): il processo di mappatura è limitato da una soglia predefinita o soglia di sopravvivenza τ di un cluster che assume il minimo pari a 0,5 cioè un cluster è ancora esistente se almeno il 50% degli oggetti appartenenti al cluster al tempo t rimane nello stesso cluster al tempo $t + 1$ (Spiliopoulou et al., 2006).

5. Applicazione e risultati

In questo paragrafo presentiamo i dati e i risultati dell'applicazione. Descriviamo l'evoluzione dei cluster di benessere economico, confrontando la situazione prima e dopo la crisi del 2008.

I dati

Per l'analisi è stato utilizzato il database dell'Istat che contiene gli indicatori utilizzati nel Bes, con riferimento alle 20 regioni italiane.

Nell'analisi preliminare sono state calcolate le matrici di correlazione per ciascun anno dal 2005 al 2015. Gli indicatori individuati dal Bes per ciascuno dei domini considerati in questa analisi, presentano in alcuni casi delle correlazioni elevate sia all'interno del dominio stesso (in particolare nel caso del dominio *Benessere economico*) sia con indicatori che appartengono agli altri due domini (*Istruzione e formazione professionale*, *Lavoro e conciliazione tempi di vita*). In appendice in particolare sono riportate le correlazioni per il 2005 e il 2015 (Tabelle 8 e 9). Le variabili che presentano correlazioni elevatissime ($>0,975$) sono riportate nella Tabella 3; solo la correlazione tra il tasso di mancata partecipazione al lavoro (**Nolav**) e il tasso di occupazione (**Occ**) riguarda indicatori che appartengono allo stesso dominio (*Lavoro e conciliazione tempi di vita*), mentre

le altre sono correlazioni tra indicatori appartenenti a domini diversi. La percentuale di correlazioni elevate in ciascun anno è non trascurabile come si può leggere in Tabella 4.

TABELLA 3. Correlazioni >0,975 in almeno uno degli anni considerati (2005-2015)

	Corr(Nolav, Neet)	Corr(Nolav, Occ)	Corr(Neet, Occ)	Corr(Neet, Rpov)	Corr(Rpov, Nolav)	Corr(Rpov, Occ)
2005	0,989	-0.989	-0.985	0,979	0,984	0,977
2006	0,983	-0.989	-0.980	0,966	0,978	0,978
2007	0,983	-0.992	-0.979	0,963	0,978	0,976
2008	0.985	-0.993	-0.984	0.973	0.983	0.982
2009	0.977	-0.991	-0.969	0.923	0.961	0.965
2010	0,989	-0.991	-0.972	0,950	0,965	0,969
2011	0,986	-0.994	-0.972	0,945	0,959	0,960
2012	0,988	-0.994	-0.984	0,958	0,963	0,967
2013	0,986	-0.995	-0.979	0,951	0,966	0,966
2014	0,986	-0.993	-0.971	0,944	0,957	0,962
2015	0.980	-0.994	-0.971	0.946	0.962	0.967

TABELLA 4. Percentuale di correlazioni elevate tra gli indicatori per ciascun anno

	Percentuale di correlazioni >0,85	Percentuale di correlazioni >0,90
2005	19,6%	10,5%
2006	18,3%	13,1%
2007	21,6%	15,7%
2008	19,0%	15,0%
2009	17,0%	13,1%
2010	18,3%	11,8%
2011	20,9%	13,7%
2012	11,8%	9,8%
2013	15,0%	13,1%
2014	11,8%	9,8%
2015	11,1%	9,2%

L'individuazione dei clusters

In questo lavoro vengono utilizzati algoritmi di cluster gerarchica applicati agli indicatori standardizzati. Viene utilizzata la distanza di Mahalanobis, adatta ai casi in cui vi siano valori elevati nella matrice delle correlazioni. I risultati presentati sono stati ottenuti con il metodo di Ward.

La scelta sul numero dei cluster si è basata su tre criteri¹³: il CCC, lo pseudo F e il t quadrato. Tutti e tre i criteri hanno dato lo stesso risultato e cioè di considerare un numero di cluster pari a cinque nel 2009 e pari a quattro in tutti gli altri anni.

Risultati e discussione

In questa sezione si riportano i risultati dell'analisi, in particolare i valori descrittivi dei cluster vengono riportati per il primo e l'ultimo anno (2005 e 2015) e per l'anno immediatamente successivo alla crisi del 2008; la Figura 2 mostra le variazioni dei cluster e le transizioni nel tempo.

Le Tabelle 5, 6 e 7 mostrano per ciascun cluster i valori medi delle variabili nel 2005, nel 2009 e nel 2015. Nelle tabelle vengono utilizzati dei colori per permettere di cogliere la performance globale del cluster. Il cluster migliore, è quello in cui prevalgono le celle verdi, mentre il peggiore è quello in cui prevalgono le celle rosse. Le celle gialle si riferiscono invece a valori medi di cluster che sono peggiori rispetto alla media dei 20 valori regionali e le celle grigie individuano valori medi di cluster migliori rispetto a tale media.

Nel 2005 (Tabella 5) ci sono quattro cluster riportati in ordine di performance degli indicatori:

- Cluster-1/05 - Emilia Romagna, Friuli Venezia Giulia, Liguria, Lombardia, Marche, Piemonte, Toscana, Trentino Alto Adige, Valle d'Aosta, Veneto;
- Cluster-3/05 - Abruzzo, Lazio, Molise, Umbria;
- Cluster-4/05 - Basilicata, Puglia, Sardegna;
- Cluster-2/05 - Calabria, Campania, Sicilia.

Nel 2005 si individua un cluster più numeroso, costituito da tutte le regioni del Nord Italia insieme con Toscana ed Emilia Romagna, in cui gli indicatori relativi al *Lavoro e conciliazione tempi di vita* e al *Benessere economico* presentano i valori migliori dei quattro raggruppamenti, mentre il Cluster-3/05, formato da quattro regioni dell'Italia centrale si contraddistingue per la miglior condizione del dominio *Istruzione e Formazione*. Gli ultimi due cluster sono entrambi formati da tre regioni regioni dell'Italia meridionale ed insulare; in questi due cluster gli

¹³ I tre criteri sono presenti nel linguaggio di programmazione SAS versione 9.4 utilizzato per le elaborazioni.

indicatori presentano sempre valori peggiori della media. Il Friuli Venezia Giulia (FVG) appartiene al cluster “migliore”, ma se si confrontano i valori degli indicatori con quelli medi di cluster, si nota come il tasso di occupazione (**Occ**) sia minore di quello medio del cluster e come quasi tutti gli indicatori del dominio Benessere economico, tranne l'indice di bassa qualità dell'abitazione (**Ab**) e l'indice di disuguaglianza del reddito disponibile (**Dis**) mostrino delle performance peggiori della media di cluster.

I cluster individuati nel 2009 sono cinque (Tabella 6), sono anch'essi riportati in ordine di performance e sono così composti:

- Cluster-1/09 - Emilia Romagna, Friuli Venezia Giulia, Liguria, Marche, Toscana, Umbria, Veneto;
- Cluster-5/09 - Lombardia, Piemonte, Trentino Alto Adige, Valle d'Aosta;
- Cluster-3/09 - Abruzzo, Lazio, Molise;
- Cluster-4/09 - Basilicata, Calabria;
- Cluster-2/09 - Puglia, Sardegna, Campania, Sicilia.

La Tabella 7 riporta invece la situazione alla fine del periodo considerato nell'analisi e cioè nel 2015. Vengono individuati quattro cluster, la cui composizione è la seguente:

- Cluster-1/15 - Emilia Romagna, Friuli Venezia Giulia, Liguria, Lombardia, Piemonte, Toscana, Trentino Alto Adige, Valle d'Aosta, Veneto;
- Cluster-3/15 - Abruzzo, Lazio, Marche, Umbria;
- Cluster-2/15 - Basilicata, Molise;
- Cluster-4/15 - Calabria, Campania, Puglia, Sardegna, Sicilia.

Durante il periodo analizzato (2005-2015) i cluster si sono modificati ed evoluti e negli ultimi anni si sono stabilizzati nelle unità che li compongono, in particolare modo quelli del Centro Nord Italia. Come si può leggere nella Tabella 7, nel 2015 il cluster con i valori peggiori degli indicatori è costituito da cinque regioni dell'Italia meridionale e insulare che sembrano allontanarsi di più dalla media di quanto non accadeva all'inizio del periodo (Cluster-2/05 e Cluster-4/15). Il Friuli Venezia Giulia si colloca sempre nel cluster con le prestazioni migliori, ma vi è un numero maggiore di indicatori che hanno valori peggiori rispetto alla media di cluster. Tra questi si evidenzia ad esempio l'indicatore **Neet** (giovani che non lavorano e non studiano) che è aumentato tra il 2005 e il 2015 e che ha valori più elevati in Friuli Venezia Giulia rispetto alla media del cluster di appartenenza. Se, per fare un altro esempio, si considera l'indice di grave deprivazione materiale (**Depr**), questo indicatore è peggiorato in Friuli Venezia Giulia tra il 2005 e il 2015, ma il suo valore è migliore della media di cluster.

TABELLA 5. Valori medi degli indicatori per cluster, 2005

Indicatore	Cluster-1/05 n=10	Cluster-2/05 n=3	Cluster-3/05 n=4	Cluster-4/05 n=3	Media	Friuli Venezia Giulia
Dip - Persone con almeno il diploma	52,13	44,19	56,46	43,22	50,47	53,14
Form - Partecipazione alla formazione continua	5,90	5,32	7,08	5,53	5,99	6,66
Neet - Giovani che non lavorano e non studiano	11,92	31,85	16,99	26,87	18,17	11,41
Prec - Uscita precoce dal sistema di istruzione e formazione	18,78	24,64	14,99	25,92	19,97	16,20
Univ - Persone che hanno conseguito un titolo universitario	17,61	13,75	20,55	12,79	16,90	19,32
Donne - Rapporto tra tasso di occupazione delle donne di 25-49 anni con figli in età prescolare e delle donne senza figli	79,61	71,04	81,64	66,54	76,77	84,58
Invpt - Quota di part-time involontario	3,58	7,73	5,47	6,06	4,95	3,37
Nolav - Tasso di mancata partecipazione al lavoro	7,44	31,10	15,42	25,36	15,27	7,00
Occ - Tasso di occupazione 20-64 anni	68,37	48,40	61,16	52,32	61,53	66,49
Oltre5 - Percentuale di occupati in lavori a termine da almeno 5 anni	16,92	32,90	20,63	24,71	21,23	15,67
Sovra - Incidenza di occupati sovrastrutturati	16,24	15,75	20,84	15,22	16,93	16,46
Ab - Indice di bassa qualità dell'abitazione	5,74	12,99	6,90	8,26	7,44	4,39
Depr - Indice di grave deprivazione materiale	2,79	16,51	4,22	10,27	6,25	2,92
Dis - Indice di disuguaglianza del reddito disponibile	4,42	6,40	4,67	5,06	4,86	4,09
Econ - Indice di grande difficoltà economica	9,52	23,97	12,08	24,24	14,41	9,71
Lav - Molto bassa intensità lavorativa	7,31	19,94	8,07	13,78	10,32	9,17
Red - Reddito medio disponibile pro-capite	19433,67	12047,02	16152,22	12575,05	16640,59	18770,07
Rpov - Indice di rischio di povertà	10,18	37,77	18,04	28,99	18,71	10,47

Nota: I valori medi degli indicatori per ciascun cluster sono comparati con le medie delle 20 regioni.: verde si riferisce ai valori migliori, rosso ai peggiori, giallo a valori peggiori della media e grigio a valori migliori della media.

TABELLA 6. Valori medi degli indicatori per cluster, 2009

Domitio	Indicatori	Cluster-1/09 n=7	Cluster-2/09 n=4	Cluster-3/09 n=3	Cluster-4/09 n=2	Cluster-4/09 n=2	Media	
<i>Istruzione e formazione</i>	Dip - Persone con almeno il diploma	58,77	44,70	58,50	52,27	55,31	54,57	
	Form - Partecipazione alla formazione continua	6,58	5,32	6,60	6,14	5,88	6,15	
	Neet - Giovani che non lavorano e non studiano	13,64	29,96	18,16	25,90	13,66	18,81	
	Prec - Uscita precoce dal sistema di istruzione e formazione	14,60	24,32	13,91	14,58	19,43	17,40	
	Univ - Persone che hanno conseguito un titolo universitario	20,61	13,99	23,00	21,19	18,07	19,19	
	Donne - Rapporto tra tasso di occupazione delle donne di 25-49 anni con figli in età prescolare e delle donne senza figli	84,58	66,37	78,65	72,32	77,16	77,34	
	Invpt - Quota di part-time involontario	5,85	8,72	6,93	7,78	4,92	6,59	
	Nolav - Tasso di mancata partecipazione al lavoro	9,67	29,90	16,86	29,56	7,83	16,41	
	Occ - Tasso di occupazione 20-64 anni	68,50	48,90	60,15	49,93	70,47	61,86	
	Oltre5 - Percentuale di occupati in lavori a termine da almeno 5 anni	14,15	21,81	22,13	32,18	17,96	19,45	
<i>Lavoro e conciliazione tempi di vita</i>	Sovra - Incidenza di occupati sovrastruiti	22,32	17,61	23,59	21,87	16,41	20,34	
	Ab - Indice di bassa qualità dell'abitazione	5,65	9,47	8,65	11,29	5,80	7,46	
	Depr - Indice di grave deprivazione materiale	3,60	13,26	5,08	10,51	4,19	6,56	
	Dis - Indice di disuguaglianza del reddito disponibile	4,22	5,50	4,83	5,10	4,59	4,73	
	Econ - Indice di grande difficoltà economica	11,46	25,15	15,57	23,05	8,67	15,42	
	Lav - Molto bassa intensità lavorativa	6,14	14,96	8,91	15,44	4,93	9,01	
	Red - Reddito medio disponibile pro-capite	19438,86	13249,77	16702,89	12776,91	20992,59	17435,20	
	Rpov - Indice di rischio di povertà	10,46	30,66	22,24	34,10	9,51	18,44	
	<i>Benessere economico</i>							

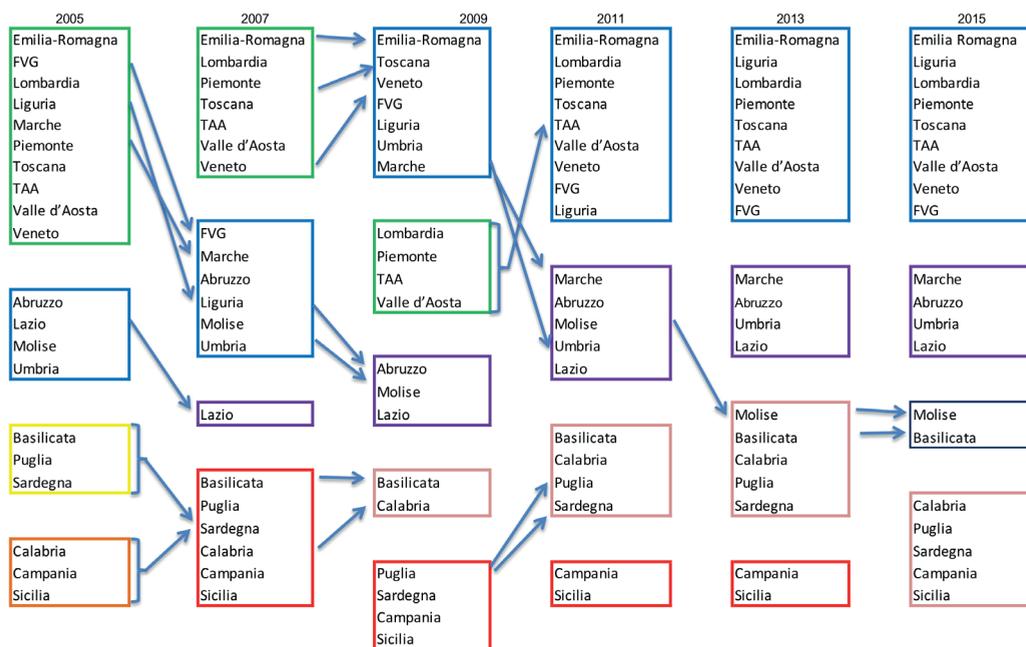
Nota: I valori medi degli indicatori per ciascun cluster sono comparati con le medie delle 20 regioni: verde si riferisce ai valori migliori, rosso ai peggiori, giallo a valori peggiori della media e grigio a valori migliori della media.

TABELLA 7. Valori medi degli indicatori per cluster, 2015

Dominio	Indicatori	Cluster-1/15 n=9	Cluster-2/15 n=2	Cluster-3/15 n=4	Cluster-4/15 n=5	Media	Frituli Venezia Giulia
<i>Istruzione e formazione</i>	Dip - Persone con almeno il diploma	63,12	59,26	66,26	50,72	60,26	65,46
	Form - Partecipazione alla formazione continua	8,57	6,84	7,81	5,87	7,57	10,34
	Neet - Giovani che non lavorano e non studiano	18,30	26,88	22,74	35,87	24,44	18,47
	Prec - Uscita precoce dal sistema di istruzione e formazione	11,85	10,19	10,92	19,78	13,48	6,93
<i>Lavoro e conciliazione tempi di vita</i>	Univ - Persone che hanno conseguito un titolo universitario	27,31	27,62	29,28	19,62	25,81	26,86
	Donne - Rapporto tra tasso di occupazione delle donne di 25-49 anni con figli in età prescolare e delle donne senza figli	82,91	75,81	86,12	73,87	80,58	78,15
	Invpt - Quota di part-time involontario	10,12	11,43	12,96	14,61	11,94	9,53
	Nolav - Tasso di mancata partecipazione al lavoro	13,36	29,32	19,06	38,65	22,42	14,15
	Occ - Tasso di occupazione 20-64 anni	69,57	53,18	63,99	45,81	60,88	68,06
	Oltre5 - Percentuale di occupati in lavori a termine da almeno 5 anni	17,26	23,37	18,04	25,23	20,02	14,90
	Sovra - Incidenza di occupati sovrastruiti	22,72	26,23	27,87	21,79	23,87	23,78
	Ab - Indice di bassa qualità dell'abitazione	7,83	7,60	12,63	11,69	9,73	5,40
	Depr - Indice di grave deprivazione materiale	6,95	11,95	9,86	20,16	11,34	4,86
	Dis - Indice di disuguaglianza del reddito disponibile	4,40	4,72	5,31	6,39	5,11	3,89
<i>Benessere economico</i>	Econ - Indice di grande difficoltà economica	10,88	12,83	13,59	24,84	15,11	10,13
	Lav - Molto bassa intensità lavorativa	6,51	12,75	10,79	20,26	11,43	7,78
	Red - Reddito medio disponibile pro-capite	20489,21	13581,51	17542,77	13163,05	17377,61	19743,98
	Rpov - Indice di rischio di povertà	10,30	27,57	18,64	33,50	19,50	8,24

Nota: I valori medi degli indicatori per ciascun cluster sono comparati con le medie delle 20 regioni: verde si riferisce ai valori migliori, rosso ai peggiori, giallo a valori peggiori della media e grigio a valori migliori della media.

FIGURA 2. Transizione dei cluster e loro composizione



La Figura 2 presenta una visualizzazione dell'evoluzione dei cluster nel tempo considerando una soglia τ pari a 0,5. Vengono rappresentati solo alcuni anni: 2005 e 2007 (anno precedente la crisi), 2009 (anno successivo), 2011, 2013, 2015. Le frecce blu indicano le transizioni delle varie regioni da un cluster all'altro nel tempo. Ogni freccia corrisponde a una regione, tranne nei casi in cui rappresenta la fusione di un cluster con un altro, dando origine ad uno nuovo (è il caso di due cluster del 2005, il giallo e l'arancio, che si fondono nel cluster rosso nel 2007) o il passaggio di tutte le regioni di un cluster ad uno già esistente (è il caso del cluster verde del 2009 che confluisce interamente nel cluster blu del 2011). Il maggior numero di frecce si ha tra il 2007 e il 2009 (sette regioni si muovono e si forma anche un nuovo cluster – il cluster rosa) e dal 2009 al 2011, in cui le transizioni sono otto (quattro regioni cambiano cluster ed il cluster verde confluisce interamente nel cluster blu), a testimonianza del fatto che vi è una maggior instabilità generale nei valori degli indicatori e quindi nella composizione dei cluster successiva alla crisi del 2008. La situazione si stabilizza dal 2011, come si può vedere dalla scarsa presenza di frecce. Nella figura ogni cluster viene

individuato da un colore specifico nell'anno 2005. Lo stesso colore negli anni successivi indica la sopravvivenza del cluster, ciò si verifica se almeno il 50% delle unità rimane nello stesso cluster. Questo è il caso ad esempio dei cluster blu, rosso, rosa. Un cluster può fondersi con un altro come nel caso del cluster verde nel 2009 che si fonde con il cluster azzurro e questo è anche il caso dei due cluster giallo e arancione del 2005 che si fondono nel cluster rosso nel 2007. Inoltre può nascere un cluster come nel caso del cluster viola (2007), del cluster rosa (2009) e del cluster blu scuro (2015).

6. Conclusioni

Misurare il benessere dei membri di una collettività e disporre di indicatori in grado di coglierne le variazioni nei vari aspetti multidimensionali del concetto è importante per la pubblica amministrazione e per gli operatori economici e i decisori politici, in quanto non solo consente di conoscere meglio una collettività e le sue esigenze e quindi permette un uso efficiente delle risorse, ma anche può essere una guida alla progettazione e alla implementazione di politiche e di servizi che tendano a migliorare il benessere dei cittadini stessi. Avere a disposizione indicatori a livello locale (regionale, provinciale, urbano) consente di conoscere le differenze territoriali che in un indicatore nazionale non si possono leggere. Un'analisi su indicatori a livello territoriale subnazionale consente quindi di conoscere le esigenze differenziate delle varie aree che compongono il paese.

In questo lavoro affrontiamo il problema del monitoraggio dell'evoluzione dei cluster di benessere economico in Italia. La metodologia utilizzata per l'identificazione dei cluster di benessere economico e la successiva osservazione della loro evoluzione hanno dato risultati soddisfacenti. In un analogo lavoro fatto sugli stati Europei, si registrava chiaramente un numero maggiore di transizioni e trasformazioni nella struttura dei cluster nei periodi successivi alle crisi (2008 e 2010), segnale evidente della diversità dei comportamenti dei singoli Stati (le differenti politiche nazionali) come reazione alla crisi stessa (Monte et al., 2017).

Nel caso invece qui considerato delle regioni italiane, questo aspetto non emerge in modo così evidente probabilmente perché di fronte ad una crisi di carattere internazionale gli andamenti regionali sono fortemente influenzati dalle politiche centrali. Quello che si coglie, invece, è la variabilità degli adattamenti

della struttura produttiva locale e delle politiche economiche regionali adottate durante la crisi. Nel lavoro si sono visualizzate le transizioni delle regioni italiane da un cluster all'altro nel tempo, permettendo così di cogliere in modo immediato i cambiamenti e la loro entità. Gli indicatori utilizzati in questo contesto per misurare il benessere economico sono riferiti al livello regionale; l'uso di indicatori a un livello territoriale più dettagliato potrebbe consentire una migliore rappresentazione dell'evoluzione reale del benessere economico.

Appendice

TABELLA 8. Correlazioni tra gli indicatori utilizzati ordinati secondo dimensione Bes di appartenenza, anno 2005

	Istruzione e formazione					Lavoro e conciliazione tempi di vita						Benessere economico						
	Dip	Form	Neet	Prec	Univ	Donne	Invpt	Nolav	Occ	Oltre5	Sovra	Ab	Depr	Dis	Econ	Lav	Red	Rpov
Dip	1,000																	
Form	0,613	1,000																
Neet	-0,669	-0,338	1,000															
Prec	-0,848	-0,594	0,582	1,000														
Univ	0,834	0,487	-0,614	-0,760	1,000													
Donne	0,507	0,193	-0,592	-0,595	0,619	1,000												
Invpt	-0,348	-0,071	0,849	0,365	-0,369	-0,527	1,000											
Nolav	-0,617	-0,252	0,989	0,499	-0,572	-0,556	0,861	1,000										
Occ	0,616	0,262	-0,985	-0,494	0,570	0,560	-0,838	-0,989	1,000									
Oltre5	-0,397	-0,175	0,685	0,235	-0,481	-0,223	0,576	0,713	-0,664	1,000								
Sovra	0,676	0,376	-0,124	-0,606	0,655	0,470	0,128	-0,070	0,056	-0,174	1,000							
Ab	-0,419	-0,182	0,807	0,488	-0,487	-0,502	0,731	0,794	-0,758	0,664	-0,074	1,000						
Depr	-0,681	-0,373	0,951	0,624	-0,660	-0,524	0,821	0,935	-0,912	0,750	-0,187	0,864	1,000					
Dis	-0,392	-0,280	0,806	0,512	-0,461	-0,609	0,855	0,777	-0,746	0,612	-0,131	0,845	0,847	1,000				
Econ	-0,676	-0,341	0,917	0,662	-0,636	-0,663	0,826	0,898	-0,886	0,550	-0,119	0,732	0,900	0,727	1,000			
Lav	-0,610	-0,399	0,925	0,588	-0,621	-0,530	0,836	0,892	-0,888	0,705	-0,139	0,845	0,962	0,864	0,873	1,000		
Red	0,627	0,217	-0,914	-0,432	0,551	0,448	-0,744	-0,942	0,944	-0,620	0,008	-0,666	-0,847	-0,560	-0,860	-0,777	1,000	
Rpov	-0,578	-0,286	0,979	0,459	-0,559	-0,518	0,835	0,984	-0,977	0,717	-0,041	0,792	0,937	0,798	0,866	0,895	-0,917	1,000

NOTA: i colori si riferiscono alle diverse dimensioni Bes cui appartengono gli indicatori

TABELLA 9. Correlazioni tra gli indicatori utilizzati ordinati secondo dimensione Bes di appartenenza, anno 2015

	Istruzione e formazione					Lavoro e conciliazione tempi di vita						Benessere economico						
	Dip	Form	Neet	Prec	Univ	Donne	Invpt	Nolav	Occ	Oltre5	Sovra	Ab	Depr	Dis	Econ	Lav	Red	Rpov
Dip	1,000																	
Form	0,719	1,000																
Neet	-0,793	-0,794	1,000															
Prec	-0,801	-0,539	0,714	1,000														
Univ	0,851	0,641	-0,708	-0,751	1,000													
Donne	0,512	0,237	-0,536	-0,304	0,264	1,000												
Invpt	-0,498	-0,538	0,747	0,643	-0,436	-0,225	1,000											
Nolav	-0,802	0,766	0,980	0,655	-0,669	-0,629	0,705	1,000										
Occ	0,788	0,779	-0,971	-0,611	0,674	0,590	-0,677	-0,994	1,000									
Oltre5	-0,402	-0,443	0,686	0,485	-0,336	-0,501	0,394	0,664	-0,650	1,000								
Sovra	0,492	-0,043	-0,060	-0,475	0,508	0,372	0,074	-0,073	0,048	-0,035	1,000							
Ab	-0,101	-0,339	0,465	0,193	-0,187	0,023	0,278	0,418	-0,431	0,364	0,324	1,000						
Depr	-0,797	-0,744	0,838	0,706	-0,726	-0,409	0,690	0,827	-0,818	0,565	-0,194	0,270	1,000					
Dis	-0,508	-0,577	0,744	0,721	-0,511	-0,257	0,900	0,697	-0,671	0,504	-0,054	0,246	0,750	1,000				
Econ	-0,791	-0,728	0,850	0,792	-0,741	-0,386	0,648	0,811	-0,807	0,512	-0,227	0,378	0,826	0,686	1,000			
Lav	-0,775	-0,731	0,920	0,728	-0,731	-0,422	0,833	0,909	-0,900	0,567	-0,094	0,339	0,886	0,839	0,821	1,000		
Red	0,731	0,752	-0,912	-0,490	0,595	0,513	-0,639	-0,947	0,958	-0,569	-0,109	-0,431	-0,767	-0,562	-0,707	-0,867	1,000	
Rpov	-0,701	-0,775	0,946	0,607	-0,612	-0,515	0,766	0,962	-0,967	0,656	0,026	0,441	0,836	0,784	0,786	0,936	-0,928	1,000

NOTA: i colori si riferiscono alle diverse dimensioni Bes cui appartengono gli indicatori

FIGURA 3. Distribuzioni degli indicatori per anno; dimensione Istruzione e Formazione

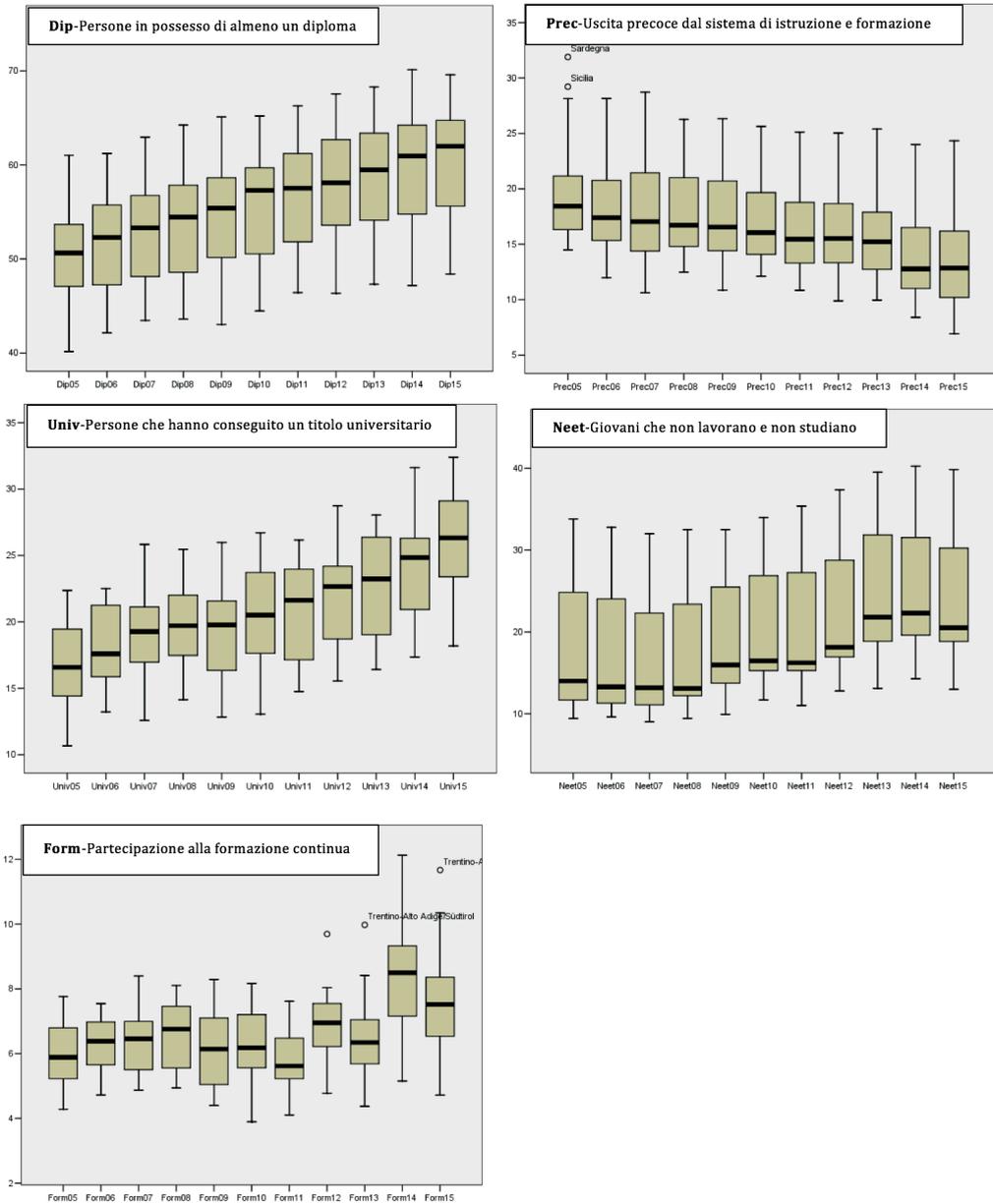


FIGURA 4. Distribuzioni degli indicatori per anno; dimensione Lavoro e conciliazione tempi di vita

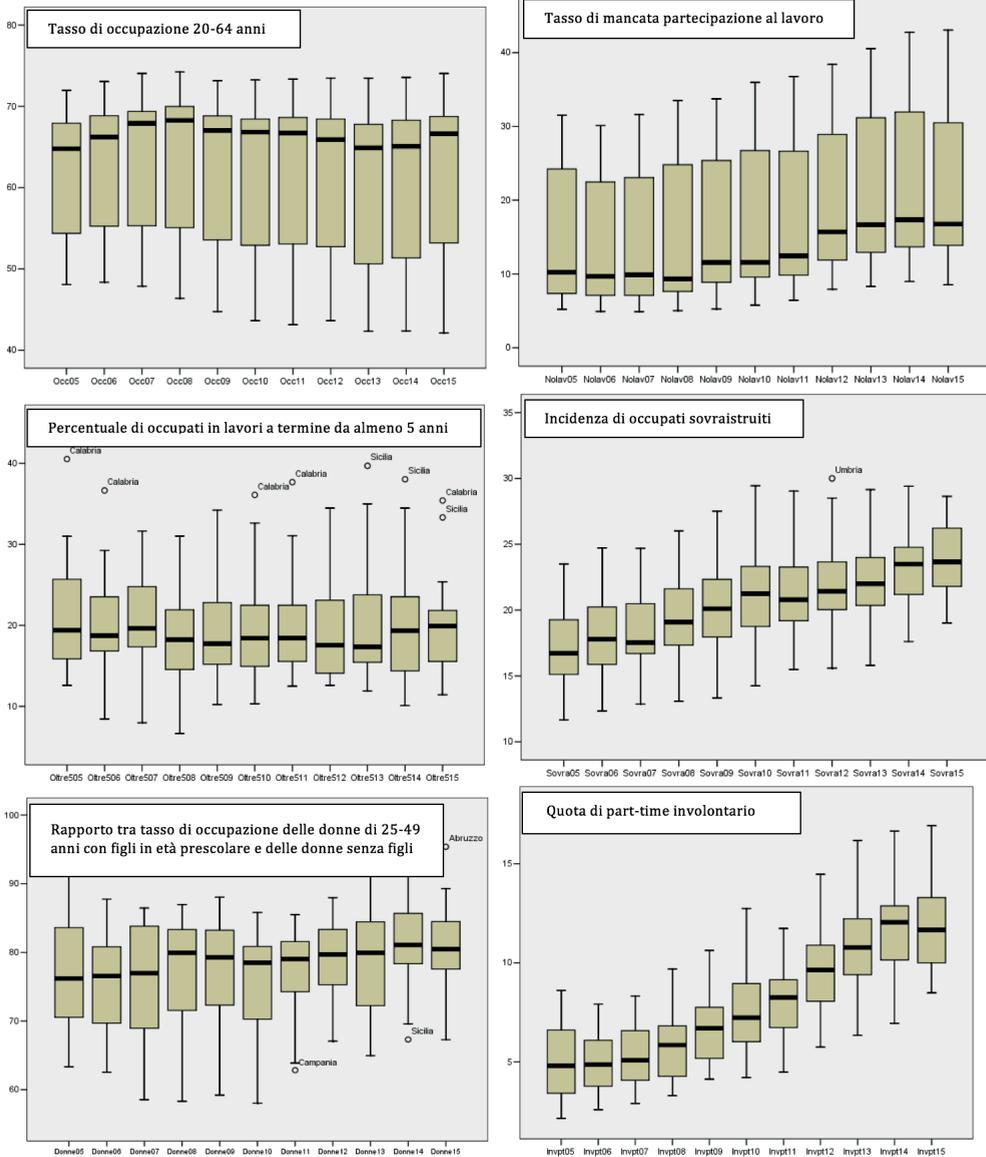
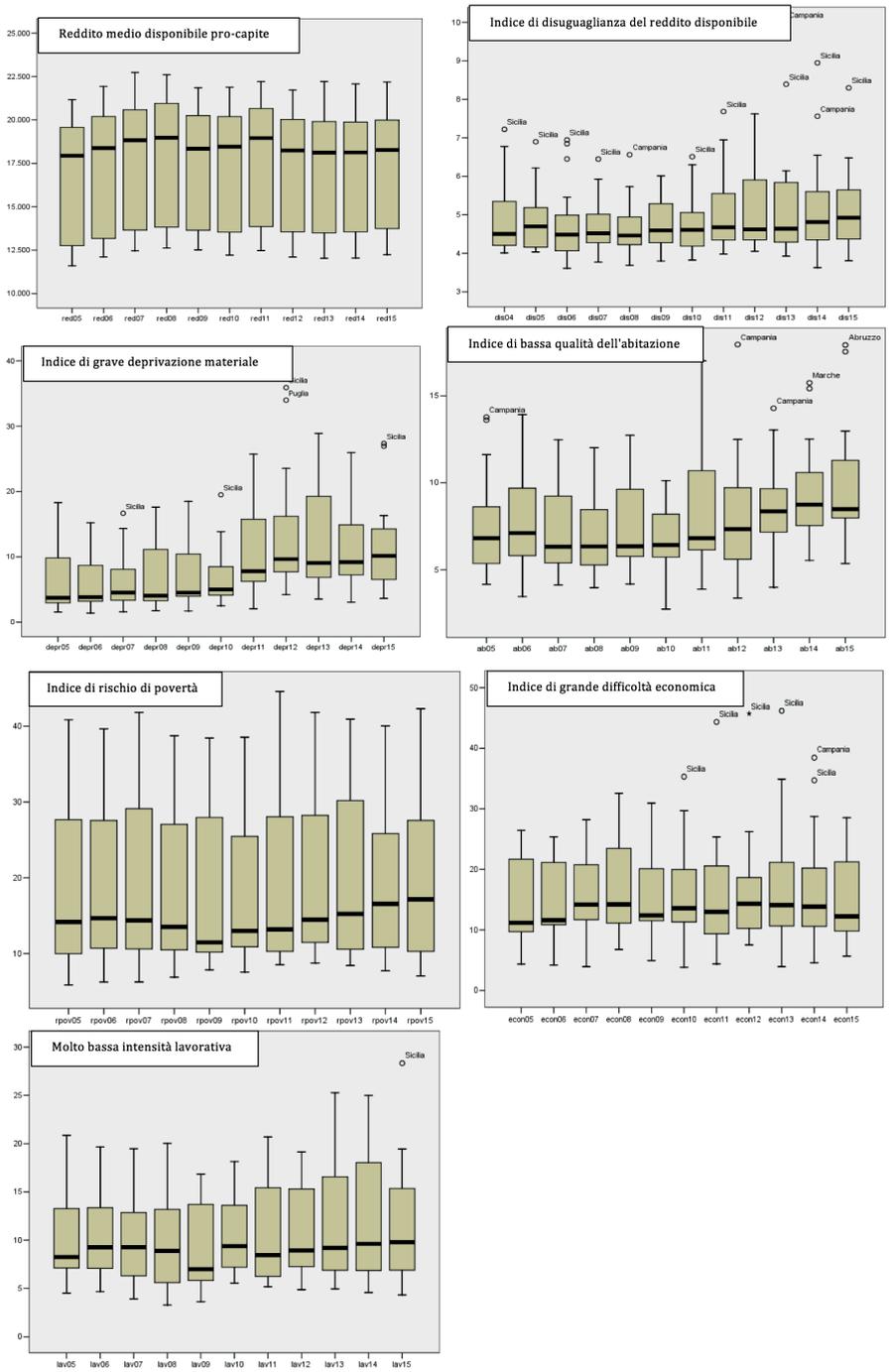


FIGURA 5. Distribuzioni degli indicatori per anno; dimensione Benessere economico



Bibliografia

- Balducci F., Ciommi M., Gigliarano C., Chelli F.M., Gallegati M. (2015), *Effetti collaterali della crescita economica. Una valutazione dello sviluppo sostenibile nelle regioni italiane*, Giappichelli Editore, Torino.
- Berkhin P. (2006), A survey of clustering data mining techniques, in Kogan J., Nicholas C., Teboulle M. (Eds) *Grouping Multidimensional Data. Recent Advances in Clustering*, Berlin Heidelberg: Springer, pp. 25-71.
- Chelli F.M., Ciommi M., Emili A., Gigliarano C., Taralli S. (2015), Comparing equitable and sustainable well-being (Bes) across the Italian Provinces. A factor analysis-based approach, *Rivista Italiana di Economia, Demografia e Statistica*, vol. LXIX, n. 3, pp. 61-72.
- Chelli F.M., Ciommi M., Emili A., Gigliarano C., Taralli S. (2016), Assessing the equitable and sustainable well-being of the Italian Provinces, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 24, pp. 39-62.
- Ciommi M., Gigliarano C., Emili A., Taralli S., Chelli F.M. (2017), A new class of composite indicators for measuring well-being at the local level: An application to the Equitable and Sustainable Well-being (BES) of the Italian Provinces, *Ecological Indicators*, vol. 76, pp.281-296.
- Istat (2013), *Bes 2013: Il Benessere Equo e Sostenibile in Italia*, <http://www.istat.it/it/archivio/84348>
- Istat (2016), *Bes 2016: Il Benessere Equo e Sostenibile in Italia*, <http://www.istat.it/it/archivio/194029>.
- Mazziotta, M., Pareto, A. (2013), A non-compensatory composite index for measuring well-being over time, *Cogito. Multidisciplinary Research Journal*, vol. 5, pp. 93-104.
- Monte A., Schoier G., Zucchet I. (2017), An evaluation of the dynamics of economic well-being in Europe: a statistical analysis, *Rivista Italiana di Economia, Demografia e Statistica*, vol. LXXI, n. 1, pp. 101-112.
- Oliveira M., Gama J. (2012), A framework to monitor clusters' evolution applied to economy and finance problems, *Intelligent Data Analysis*, vol. 16, pp. 93-111.
- Spiliopoulou M., Ntoutsis I., Theodoridis Y., Schult R. (2006), MONIC – Modeling and monitoring cluster transitions, in KDD, pp. 706-711.
- Stiglitz J. E., Sen A., Fitoussi J. P. (2009), Report by the commission on the measurement of economic performance and social progress. Paris: Commission on the Measurement of Economic Performance and Social Progress.
- Unioncamere del veneto – Area studi e ricerche (a cura di) (2013), *Oltre il Pil 2013. La geografia del benessere nelle regioni italiane e nelle città metropolitane*, <http://www.oltreilpil.it/rapporti>.
- Zani, S., Cerioli, A. (2007) *Analisi dei dati e data mining per le decisioni aziendali*, Giuffrè Editore, Milano.
- Xu, R., Wunsch, D. (2005) Survey of Clustering Algorithms, *IEEE transactions on neural networks* 16(3), pp.645-678.