istat
working
papers

N.12
2023

Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella Rilevazione Oros

Indices of ratio variables: property analysis, decomposition, and case study from Istat's Oros Survey

Marco Lattanzio

Istat
Istituto Nazionale
di Statistica

istat working papers

N.12
2023

Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella Rilevazione Oros

Indices of ratio variables: property analysis, decomposition, and case study from Istat's Oros Survey

Marco Lattanzio

Istat

Istat Working Papers

Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella Rilevazione Oros

(Indices of ratio variables: property analysis, decomposition, and case study from Istat's Oros Survey)

N. 12/2023

ISBN 978-88-458-2134-9

Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella Rilevazione Oros

(Indices of ratio variables: property analysis, decomposition, and case study from Istat's Oros Survey)

Marco Lattanzio[1]

Sommario

Il lavoro intende mostrare i risultati di uno studio finalizzato a offrire un quadro teorico e analitico sulle proprietà dell'indice semplice a base fissa calcolato su una variabile rapporto e le caratteristiche connesse al suo utilizzo. Con lo scopo di offrire spunti per spiegarne la dinamica, l'indice generale viene scomposto nelle singole componenti rispetto a una qualsiasi classificazione delle unità, nonché analizzato rispetto alle dinamiche di numeratore e denominatore. Le stesse scomposizioni vengono esplicitate sui tassi di variazione dell'indice. Nel documento, a titolo esemplificativo, vengono considerati gli indici trimestrali delle retribuzioni per Ula prodotti dalla Rilevazione Oros. Infine, viene mostrata una formula generale per calcolare la relazione di scomposizione delle variazioni tendenziali, una volta nota la relazione di scomposizione degli indici.

Parole chiave: Indici semplici, tassi di variazione, variabili rapporto, scomposizione, congiuntura.

Abstract

This work shows the results of a study for defining a theoretical and analytical framework on the fixed base index numbers calculated on ratio variables, focussing on some of their properties to be considered when used. To provide hints to explain its dynamic, the general index is broken down into individual components with respect to any possible classification of units, and analysed in relation to the dynamics of its numerator and denominator. The same breakdowns are spelled out for the calculation of the related growth rates. As case study, the quarterly indexes of wages per Fte calculated by the Oros survey are considered. Furthermore, a general formula for decomposing growth rates based on the relation between indexes is presented.

Keywords: Fixed base index numbers, growth rates, ratio variables, decomposition, short-term statistics.

1 Marco Lattanzio (marco.lattanzio@istat.it), Istituto Nazionale di Statistica - Istat.

I punti di vista espressi sono quelli dell'autore e non riflettono necessariamente le opinioni ufficiali dell'Istituto Nazionale di Statistica - Istat.

L'autore ringrazia i revisori anonimi (almeno due per ogni lavoro, coinvolti su base volontaria e gratuita, con un approccio di tipo double - anonymised) per i loro commenti e suggerimenti, che hanno migliorato la qualità di questo Istat working papers N. 12/2023.

Indice

1. Introduzione

Nell'ambito della produzione di indicatori congiunturali dell'Istat, i numeri indice sono generalmente calcolati secondo una metodologia condivisa e diffusi seguendo un predefinito schema che contempla la classificazione delle unità statistiche in insiemi disgiunti definiti con lo scopo di conseguire gli obiettivi conoscitivi della rilevazione. Ad esempio nell'ambito delle indagini economiche, le unità statistiche sono classificate in base all'attività economica svolta, utilizzando attualmente la classificazione Ateco 2007 (versione italiana della NACE rev.2; vedi Istat-2009), la quale contempla insiemi di classificazione più fine quali le divisioni Ateco e aggregazioni di esse quali le sezioni e i macro settori. Pertanto in tale contesto, considerando che gli indicatori sono calcolati generalmente quali numeri indice a base fissa, risulta spesso necessario, soprattutto quando si osservano andamenti particolari o anomali delle serie oggetto di analisi, da una parte risalire alle cause che hanno generato tali valori, dall'altra conoscere il contributo fornito dai singoli sotto settori componenti il generico aggregato di cui si osserva la dinamica, per acquisire sufficienti informazioni, anche quantitative, per una corretta analisi delle serie storiche. Per quanto riguarda il primo obiettivo, la ricerca di una relazione matematica di scomposizione dell'indice generale può servire per impostare un problema di *editing selettivo*, ovvero fornire un'indicazione quantitativa del contributo di ogni singola unità statistica all'indice o al tasso di variazione e fornirne un ordinamento, in modo da selezionare il sottoinsieme di unità da sottoporre al successivo ed eventuale processo di correzione. Per quanto riguarda il secondo obiettivo, volto all'analisi delle serie storiche, si vuole fornire un'indicazione quantitativa del contributo all'indice o al tasso di variazione relativo alle sotto aggregazioni componenti un aggregato generale, anche per le quali sono generalmente diffusi indici e tassi di variazione. Anche in questo caso il problema da risolvere è quello di ricavare, quando non nota, una relazione matematica che leghi gli indici delle sotto aggregazioni all'indice generale, che rappresentano in generale sottoinsiemi di unità statistiche disgiunti. Il problema ulteriore è quello poi di ricavare una relazione anche per i tassi di variazione, generalmente differente da quella per gli indici. Tra le varie tipologie di indici, il lavoro focalizza l'attenzione sugli indici calcolati su variabili rapporto, portando ad esempio il contesto dell'Indagine congiunturale Oros, che fornisce indicatori a cadenza trimestrale sull'andamento del costo del lavoro per Ula e delle relative componenti quali retribuzioni e oneri sociali diffuse per sezioni e macro settori Ateco. In tale contesto non risulta utile solamente analizzare il contributo dei sotto-settori componenti l'aggregato, ma anche calcolare una misura economicamente interpretabile degli effetti prodotti dalle variabili a numeratore e denominatore dell'indice (e sui tassi di variazione), cercando di scinderne il singolo contributo. Il risultato ottenuto, è una generalizzazione dell'applicazione studiata nei lavori di Keller (2009) prima e dell'Istat (2010) poi, nei quali, per gli indicatori relativi alle retribuzioni per dipendente si scorpora dal tasso di variazione il cosiddetto "effetto di ricomposizione dell'occupazione", risultante dalla classificazione dei dipendenti per qualifica professionale. Nel lavoro qui presentato l'operazione di scomposizione è applicata al contesto dell'Indagine Oros, con lo scopo di analizzare il contributo dell'*effetto di ricomposizione* risultante dall'analisi della variazione della struttura occupazionale tra i singoli settori Ateco nella definizione della variazione di uno specifico macro settore, offrendo spunti di riflessione interessanti sulle dinamiche osservate oltre a permettere di giustificare alcune incongruenze riscontrate nel quadro delle statistiche diffuse. Nell'ottica di fornire una panoramica sugli approcci e i contesti di applicazione di problemi di scomposizione delle formule per il calcolo degli indici e dei tassi di variazione, va osservato innanzitutto che nel caso in cui la metodologia di calcolo sia relativa a un indice complesso, quale ad esempio l'indice di *Laspeyres*, metodologia utilizzata in molte delle indagini congiunturali (Eurostat-2006; Predetti-2006), la relazione di scomposizione dell'indice è nota. L'indice di *Laspeyres* relativo all'aggregato generale è per definizione una media ponderata di sotto indici, sia che essi siano indici elementari, sia che essi siano indici di sotto-aggregazioni, e gli indici relativi a uno specifico aggregato vengono ottenuti per aggregazioni successive degli indici di livello inferiore della classificazione, con pesi di ponderazione fissi all'anno base. Per quanto il tasso di variazione dell'indice di *Laspeyres*, anche nella sua versione con concatenamento utilizzata nell'ambito del calcolo dell'indice dei prezzi al consumo, la relativa formula di scomposizione viene ricavata nel lavoro di Brunetti (2010). Un approccio noto al problema di scomposizione dei tassi di variazione è quello offerto dall'analisi *shift-share*. Questa tecnica, riprendendo dal testo di Biffignandi (1993), ha l'obiettivo di "*... studiare l'evoluzione temporale di aggregati economici territoriali tentando di separare il contributo di diversi fattori di sviluppo, costruiti e interpretati secondo criteri di logica economica. La tradizionale impostazione della tecnica shift-share spiega la variazione intervenuta in un determinato periodo di tempo in un fenomeno economico rilevato su diversi settori e diverse aree territoriali come attribuibile a tre componenti: nazionale o tendenziale, mix settoriale nazionale e locale*"[2]. Pertanto si noti che, in generale, la tecnica presuppone l'osservazione di una variabile economica secondo una classificazione a due stadi (ad esempio, settoriale e territoriale oppure solo settoriale o solo territoriale a due stadi come ad esempio la classificazione in nazione-regione-provincia) e che è applicabile ad "aggregati", ovvero variabili espresse come totali e non rapporti di aggregati, caso di studio di questo lavoro. Tentativi di scomposizione (secondo un approccio shift-share) di tassi di variazione relativi a variabili rapporto sono presentati nel lavoro di Keller (2009) e un'applicazione di quest'ultimo è presente nel lavoro dell'Istat (2010), in cui l'approccio di scomposizione di Keller viene applicato agli indicatori dell'Indagine mensile sull'Occupazione, le retribuzioni e l'orario di lavoro nelle Grandi Imprese (GI). Nel lavoro di Keller si ottiene una formula di scomposizione del tasso di variazione delle retribuzioni per dipendente secondo le qualifiche. Malgrado l'approccio venga configurato come derivante dall'applicazione della tecnica *shift-share* sembrerebbe che manchino degli elementi per consentire tale configurazione secondo l'approccio originariamente esplicitato in Biffignandi (1993). Di fatto, mentre per la scomposizione dei tassi di variazione di variabili rapporto presente nel lavoro di Keller, come anche quella descritta in questo lavoro, si ottengono tre componenti indipendentemente dalla stratificazione utilizzata nella classificazione delle unità statistiche, nella tecnica *shift-share* la scomposizione in tre componenti è unicamente riconducibile alla classificazione a due stadi.

2 Per la formulazione algebrica delle componenti si rimanda al testo Biffignandi-1993.

Utilizzando la classificazione a uno stadio (il settore di attività economica o la qualifica), la scomposizione porterebbe solo a due componenti. Partendo da queste considerazioni, l'obiettivo di questo lavoro è quello di offrire un approccio alternativo per ricavare la formula di scomposizione dei tassi di variazione e degli indici calcolati su variabili rapporto relative a unità statistiche classificate in insiemi disgiunti (ad esempio la classificazione Ateco) secondo una qualsiasi classificazione delle unità di analisi, permettendo di ottenere una dimostrazione della formula utilizzata nei lavori di Keller (2009) e dell'Istat (2010), la quale rappresenta un'applicazione della formula generale ricavata in questo lavoro e, in seconda battuta, fornire una formulazione generale applicabile alle diverse tipologie di indice (per cui non solo per gli indici semplici di variabili rapporto) per la scomposizione dei tassi di variazione una volta nota la formula di scomposizione dei relativi indici. Il lavoro è organizzato nel modo seguente: nel secondo paragrafo vengono ricavate le formule di scomposizione dell'indice semplice calcolato su variabili rapporto, dei relativi tassi di variazione utilizzando un approccio algebrico di tipo *top-down*, per cui, partendo dalla formula applicata all'aggregato generale si cerca di ottenere una relazione nota che considera gli indici di livello più fine. Nel paragrafo 3 la formula ricavata viene applicata per l'analisi dell'indice e delle variazioni tendenziali delle retribuzioni per Ula calcolate nel contesto dell'Indagine Oros per la scomposizione settoriale, fornendo anche utili elementi per un'analisi delle serie storiche in cui venga scorporato, dalla variazione generale, l'effetto di ricomposizione dell'occupazione. Inoltre vengono discusse alcune criticità e offerti spunti di riflessione, e infine mostrata un'applicazione per impostare un problema di *editing selettivo*, utilizzato nell'ambito dell'Indagine Oros. Nel paragrafo 4 vengono delineate le conclusioni ottenute nel lavoro. Infine nell'Appendice A vengono richiamate alcune definizioni utili per le dimostrazioni presentate nel lavoro, e nell'Appendice B presentata una formula generale di scomposizione dei tassi di variazione con approccio analitico, utilizzando l'espansione in serie di Taylor per funzioni in più variabili, con la relativa dimostrazione.

2. Calcolo delle scomposizioni per indici semplici di variabili rapporto e i relativi tassi di variazione

Per semplicità espositiva le definizioni relative a indici e tassi di variazione utilizzate nel documento sono inserite nell'Appendice A. L'obiettivo del lavoro presentato nei paragrafi seguenti è quello di ottenere relazioni generali di scomposizione degli indici e dei tassi di variazione calcolati su variabili rapporto, sia come funzione degli indici e dei tassi di variazione delle variabili a numeratore e denominatore, sia come funzione degli indici e dei tassi di variazione relative alle componenti di uno specifico aggregato.

2.1 Scomposizione numeratore-denominatore dell'indice semplice e del tasso di variazione di variabili rapporto

Risulta di facile verifica che l'indice semplice di una variabile rapporto X è uguale al rapporto degli indici rispettivamente della variabile a numeratore N e a denominatore D:

$$I_t^X = \frac{N_t/D_t}{N_b/D_b} = \frac{N_t/N_b}{D_t/D_b} = \frac{I_t^N}{I_t^D} \quad (3)$$

Scomponiamo ora il tasso di variazione percentuale di passo h nei tassi di variazione di numeratore e denominatore:

$$V_h(I_t^X) = \left(\frac{X_{t+h}}{X_t} - 1\right)100 = \left(\frac{N_{t+h}/D_{t+h}}{N_t/D_t} - 1\right)100 =$$

$$= \left(\frac{N_{t+h}/N_t}{D_{t+h}/D_t} - 1\right)100 = \left(\frac{\frac{V_h(N_t)}{100}+1}{\frac{V_h(D_t)}{100}+1} - 1\right)100 =$$

$$= \frac{\left(\frac{V_h(N_t)}{100}+1-1-\frac{V_h(D)}{100}\right)}{1+\frac{V_h(D_t)}{100}}100 = \frac{V_h(N_t)-V_h(D_t)}{1+\frac{V_h(D_t)}{100}} = \frac{V_h(I_t^N)-V_h(I_t^D)}{1+\frac{V_h(I_t^D)}{100}} \quad (4)$$

Nei passaggi algebrici è stata utilizzata la definizione di tasso di variazione presentata nell'Appendice A.

2.2 Scomposizione dell'indice semplice e del tasso di variazione di variabili rapporto negli indici e nei tassi di variazione delle componenti di un aggregato

In questo paragrafo l'obiettivo è quello di ottenere relazioni generali di scomposizione degli indici e dei tassi di variazione calcolati su variabili rapporto per uno specifico aggregato, come funzione degli indici e dei tassi di variazione calcolati su sotto-aggregati generici. In formula, l'obiettivo è quello di ottenere relazioni f e g del tipo:

$$\boldsymbol{I_t^X = f\left(I_{1,t}^X, \dots, I_{n,t}^X\right)} \quad (1)$$

$$\boldsymbol{V_h\left(I_t^X\right) = g\left[V_h\left(I_{1,t}^X\right), \dots, V_h\left(I_{n,t}^X\right)\right]} \quad (2)$$

Come esempio noto, nel caso dell'indice di *Laspeyres* la f è già nota per definizione, ed è pari alla media ponderata con pesi fissi all'anno base degli indici $\boldsymbol{I}^{1}_{t}, \dots, \boldsymbol{I}^{n}_{t}$.

$$\boldsymbol{L_t = f\left(I_{1,t}, \dots, I_{n,t}\right) = \sum_{i=1}^{n} I_{i,t}\pi_p^i}$$

La scomposizione g, già nota (Brunetti 2010), viene ottenuta anche come applicazione della formula con approccio analitico nell'Appendice B di questo lavoro, ed è la seguente:

$$\boldsymbol{V_h(L_t) = g\left[V_h\left(I_{1,t}\right), \dots, V_h\left(I_{n,t}\right)\right] = \sum_{i=1}^{n} V_h\left(I_{i,t}\right)\frac{I_{i,t}\pi_p^i}{\sum_i I_{i,t}\pi_p^i} = \sum_{i=1}^{n} V_h\left(I_{i,t}\right)\frac{I_{i,t}\pi_p^i}{L_t}}$$

La formula di scomposizione sarà utile per esplicitare il contributo di ogni singolo aggregato componente, analizzare il contributo della variabile denominatore separatamente da quello relativo alla variabile rapporto, e, infine, verificare alcune proprietà tipiche degli indici utilizzati nell'ambito delle indagini congiunturali. Di norma, nella maggior parte degli indicatori congiunturali la metodologia consiste nel calcolo di indici semplici a un certo livello di dettaglio fine della classificazione utilizzata, definiti indici elementari, e nell'ottenimento degli indici relativi ad aggregati superiori tramite aggregazioni successive seguendo uno schema di ponderazione con pesi fissi (si pensi ad esempio all'indice di *Laspeyres*). Per queste tipologie di indice sono verificate due importanti proprietà: la proprietà di additività e la proprietà di coerenza interna, in particolare quest'ultima garantisce che l'indice relativo a un aggregato sarà contenuto nell'intervallo costituito dal valore minimo e massimo degli indici degli aggregati componenti. La stessa proprietà è verificata dai tassi di variazione. Per quanto riguarda l'indice semplice di variabile rapporto, il problema consiste nel verificare che, nonostante non sia calcolato seguendo una metodologia ad aggregazioni successive, tipica delle indagini congiunturali, è possibile, risolvendo un problema di scomposizione dell'indice generale risalire a una relazione di tipo additivo simile alla metodologia *Laspeyres*. L'indice generale del settore S, $I^{X}_{S,t}$, può essere scomposto utilizzando un approccio *top down* (effettuando una disaggregazione per passi successivi) in modo da ricavare una relazione tra l'indice generale e gli indici relativi a n sotto insiemi di unità statistiche $S_1, \dots, S_n$. Si cercherà di ottenere una relazione del tipo:

$$I_{S,t}^X = f\left(I_{s_1,t}^X, I_{s_2,t}^X, \dots, I_{s_n,t}^X\right) \quad (5)$$

Saremo interessati inoltre a verificare le proprietà di additività e coerenza interna dell'indice. In particolare l'indice aggregato verificherà la proprietà di coerenza interna se è possibile ricondurre il suo calcolo a una media ponderata di sotto-indici, ricavando un sistema di pesi, $\pi_1, \ldots, \pi_n$, che sommano a 1. Le relazioni da verificare sono le seguenti:

$$\begin{aligned} &i)\ I^X_{S,t} = \sum_{i=1}^n I^X_{s_i,t}\,\pi_i \\ &ii)\ \sum_{i=1}^n \pi_i = 1 \end{aligned} \qquad (6)$$

Vedremo che l'indice semplice di variabili rapporto non verifica la proprietà di coerenza interna, ma soltanto la condizione di additività (*6 i*). Per quanto riguarda gli indici semplici di variabili espresse come totali (non variabili rapporto, ipotizzando di utilizzare l'indice semplice per ogni livello di dettaglio e non uno schema di aggregazione per passi successivi) la coerenza interna e l'additività sono facilmente verificabili, e la dimostrazione viene fornita in Appendice A.

Di altrettanto interesse può risultare una relazione del tipo:

$$I^X_{S,t} = f\left(I^N_{s_1,t}, I^N_{s_2,t}, \ldots, I^N_{s_n,t}, I^D_{s_1,t}, I^D_{s_2,t}, \ldots, I^D_{s_n,t}\right) \quad (7)$$

nella quale si mette in relazione l'indice generale della variabile rapporto con i sotto indici degli *n* sottoinsiemi calcolati sia per la variabile numeratore *N* sia per la variabile denominatore *D*. Il calcolo di quest'ultima relazione è più semplice e viene presentato per primo.

$$I^X_{S,t} = \frac{X_{S,t}}{X_{S,b}} = \frac{N_{S,t}/D_{S,t}}{N_{S,b}/D_{S,b}} = \frac{\sum_i N_{s_i,t} / \sum_i D_{s_i,t}}{\sum_i N_{s_i,b} / \sum_i D_{s_i,b}} = \frac{\sum_i \frac{N_{s_i,t}}{N_{s_i,b}} N_{s_i,b} \Big/ \sum_i \frac{D_{s_i,t}}{D_{s_i,b}} D_{s_i,b}}{\sum_i N_{s_i,b} / \sum_i D_{s_i,b}} =$$

$$= \frac{\sum_i \frac{N_{s_i,t}}{N_{s_i,b}} \pi^N_{s_i,b}}{\sum_i \frac{D_{s_i,t}}{D_{s_i,b}} \pi^D_{s_i,b}} \quad (8)$$

.

definendo con $\pi^N_{Si,b} = \frac{N_{s_i,b}}{\sum_i N_{s_i,b}}$ e $\pi^D_{Si,b} = \frac{D_{s_i,b}}{\sum_i D_{s_i,b}}$, i quali rappresentano i pesi rispettivamente della variabile *N* e della variabile *D* misurate sulle unità appartenenti al sotto settore *i*-esimo Si rispetto a S_i calcolato all'anno base *b*. La (8) si può quindi scrivere come:

$$I_{S,t}^{X} = \frac{\sum_i I_{s_i,t}^{N}\, \pi_{s_i,b}^{N}}{\sum_i I_{s_i,t}^{D}\, \pi_{s_i,b}^{D}} = f\left(I_{s_1,t}^{N}, I_{s_2,t}^{N}, \dots, I_{s_n,t}^{N}, I_{s_1,t}^{D}, I_{s_2,t}^{D}, \dots, I_{s_n,t}^{D}\right) \quad (9)$$

ovvero come un rapporto di medie ponderate.

Passiamo a ricavare la relazione (5) riprendendo i passaggi effettuati nella (8):

$$I_{S,t}^{X} = \frac{\sum_i \frac{N_{s_i,t}}{D_{s_i,t}} \frac{D_{s_i,t}}{\sum_i D_{s_i,t}}}{\sum_i N_{s_i,b} \big/ \sum_i D_{s_i,b}} =$$

$$= \frac{\sum_i \frac{N_{s_i,t}}{D_{s_i,t}} \Big/ \frac{N_{s_i,b}}{D_{s_i,b}} \quad D_{s_i,t} \Big/ \sum_i D_{s_i,t} \frac{N_{s_i,b}}{D_{s_i,b}}}{\sum_i N_{s_i,b} \big/ \sum_i D_{s_i,b}} = \sum_i I_{s_i,t}^{X} \, \frac{D_{s_i,t} \big/ \sum_i D_{s_i,t}}{D_{s_i,b} \big/ \sum_i D_{s_i,b}} \frac{N_{s_i,b}}{\sum_i N_{s_i,b}}$$

Nell'ultima relazione si può definire:

$$I_{s_i,t}^{\pi_D} = \frac{D_{s_i,t} \big/ \sum_i D_{s_i,t}}{D_{s_i,b} \big/ \sum_i D_{s_i,b}}$$

che rappresenta l'indice della composizione della variabile denominatore *D* misurato sulle unità appartenenti al sotto settore *i*-esimo S_i rispetto all'aggregato *S*.

Abbiamo quindi:

$$I_{S,t}^{X} = \sum_i I_{s_i,t}^{X}\, I_{s_i,t}^{\pi_D}\, \pi_{s_i,b}^{N} = f\left(I_{s_1,t}^{X}, I_{s_2,t}^{X}, \dots, I_{s_n,t}^{X}, I_{s_1,t}^{\pi_D}, I_{s_2,t}^{\pi_D}, \dots, I_{s_n,t}^{\pi_D}\right) \quad (10)$$

Analizzando la relazione appena descritta vediamo che l'indice generale non dipende solo dagli indici della variabile rapporto dei sotto-settori ma anche da altrettanti indici (variabili rispetto a *t*) che misurano la dinamica della variabile a denominatore tramite una sua funzione, la composizione rispetto ai macro settori. La relazione è di tipo additivo in quanto la condizione (*6 i*) è verificata ma non verifica la proprietà di coerenza interna in quanto non è verificata la condizione *9 ii*. Infatti i coefficienti di ponderazione, variabili nel tempo, $\pi_i = I^{\pi D}_{Si,t}\, \pi^{N}_{Si,b}$ non sommano a 1.

Ora vogliamo ricavare la relazione *g*, ovvero la scomposizione per i tassi di variazione degli indici. Per semplicità, trattandosi di passaggi algebrici simili, utilizzeremo l'approccio seguito nella (10):

$$V_h\left(I_{S,t}^X\right) = \left(\sum_i \frac{\frac{N_{s_i,t+h}}{D_{s_i,t+h}}}{\frac{N_{s_i,t}}{D_{s_i,t}}} \frac{N_{s_i,t}}{D_{s_i,t}} \frac{\frac{D_{s_i,t+h}}{\sum_i D_{s_i,t+h}}}{\frac{\sum_i N_{s_i,t}}{\sum_i D_{s_i,t}}} - 1\right) 100 =$$

$$= \left(\sum_i \frac{\frac{N_{s_i,t+h}}{D_{s_i,t+h}}}{\frac{N_{s_i,t}}{D_{s_i,t}}} \frac{N_{s_i,t}}{\sum_i N_{s_i,t}} \frac{\frac{D_{s_i,t+h}}{\sum_i D_{s_i,t+h}}}{\frac{D_{s_i,t}}{\sum_i D_{s_i,t}}} - 1\right) 100 =$$

$$= \left(\sum_i \left(1 + \frac{V_h\left(X_{s_i,t}\right)}{100}\right) \frac{N_{s_i,t}}{\sum_i N_{s_i,t}} \frac{\frac{D_{s_i,t+h}}{\sum_i D_{s_i,t+h}}}{\frac{D_{s_i,t}}{\sum_i D_{s_i,t}}} - 1\right) 100 =$$

$$= \left(\sum_i \left(1 + \frac{V_h\left(X_{s_i,t}\right)}{100}\right) \frac{N_{s_i,t}}{\sum_i N_{s_i,t}} \left(1 + \frac{V_h\left(\pi_{s_i,t}^D\right)}{100}\right) - 1\right) 100 =$$

$$= \left(\sum_i \left(1 + \frac{V_h\left(I_{s_i,t}^X\right)}{100}\right) \frac{N_{s_i,t}}{\sum_i N_{s_i,t}} \left(1 + \frac{V_h\left(I_{s_i,t}^{\pi_D}\right)}{100}\right) - 1\right) 100 =$$

$$= \left(\sum_i \left(1 + \frac{V_h\left(I_{s_i,t}^X\right)}{100}\right) \pi_{s_i,t}^N \left(1 + \frac{V_h\left(I_{s_i,t}^{\pi_D}\right)}{100}\right) - 1\right) 100 =$$

$$= \left(\sum_i \left(1 + \frac{V_h\left(I_{s_i,t}^X\right)}{100} + \frac{V_h\left(I_{s_i,t}^{\pi_D}\right)}{100} + \frac{V_h\left(I_{s_i,t}^X\right) V_h\left(I_{s_i,t}^{\pi_D}\right)}{100^2}\right) \pi_{s_i,t}^N - 1\right) 100 =$$

$$= \left(\sum_i \pi_{s_i,t}^N + \sum_i \left(\frac{V_h\left(I_{s_i,t}^X\right)}{100} + \frac{V_h\left(I_{s_i,t}^{\pi_D}\right)}{100} + \frac{V_h\left(I_{s_i,t}^X\right) V_h\left(I_{s_i,t}^{\pi_D}\right)}{100^2}\right) \pi_{s_i,t}^N - 1\right) 100 =$$

$$= \sum_i \left(V_h\left(I_{s_i,t}^X\right) + V_h\left(I_{s_i,t}^{\pi_D}\right) + \frac{V_h\left(I_{s_i,t}^X\right) V_h\left(I_{s_i,t}^{\pi_D}\right)}{100}\right) \pi_{s_i,t}^N =$$

$$= \sum_i V_h\left(I_{s_i,t}^X\right) \pi_{s_i,t}^N + \sum_i V_h\left(I_{s_i,t}^{\pi_D}\right) \pi_{s_i,t}^N + \sum_i \frac{V_h\left(I_{s_i,t}^X\right) V_h\left(I_{s_i,t}^{\pi_D}\right)}{100} \pi_{s_i,t}^N =$$

$$= g\left(V_h\left(I_{s_1,t}^X\right), \dots, V_h\left(I_{s_n,t}^X\right), V_h\left(I_{s_1,t}^{\pi_D}\right), \dots, V_h\left(I_{s_n,t}^{\pi_D}\right)\right) \quad (11)$$

dove si è posto $\left(\frac{\frac{D_{s_i,t+h}}{\sum_i D_{s_i,t+h}}}{\frac{D_{s_i,t}}{\sum_i D_{s_i,t}}} - 1\right) 100 = V_h\,(\pi^D{}_{Si,t}) = V_h\,(I^{\pi D}{}_{Si,t})$, termine che rappresenta la variazione di passo *h* della composizione della variabile denominatore *D* (o del suo indice) nel settore s_i rispetto all'aggregato *S*.

Nell'Appendice B viene esposto e dimostrato un risultato generale con il quale viene ricavata la relazione (11) utilizzando un approccio analitico che utilizza l'espansione in serie di Taylor per funzioni in più variabili. Il risultato è generalizzato e viene applicato a varie tipologie di indici (ad esempio indici di *Laspeyres*).

Le formule sulla scomposizione di indici e variazioni sviluppate precedentemente fanno emergere inoltre alcune interessanti caratteristiche dell'indice semplice di variabili rapporto:

- la variazione dell'indice semplice di variabili rapporto non è una media ponderata di variazioni degli indici delle variabili rapporto degli aggregati componenti, ma è pari alla somma di tre medie ponderate: una prima media delle variazioni delle variabili rapporto, una seconda media delle variazioni della composizione della variabile denominatore e, infine, una terza media ponderata del prodotto tra le due variazioni precedentemente descritte, diviso 100. Quest'ultimo termine $\sum_i \frac{V_h\left(X_{s_i,t}\right)V_h\left(\pi^D_{s_i,t}\right)}{100}\,\pi^N_{s_i,t}$, risulta trascurabile se i termini a numeratore sono minori di 1;
- la dinamica generale risente sia della dinamica della variabile al numeratore sia di quella della variabile al denominatore. In particolare, essa dipende dalla dinamica della variabile rapporto osservata sui sotto-indici ma anche dalla dinamica di una funzione del denominatore, ovvero la composizione tra i settori osservati. Questa "composizione" della variabile *D* tiene conto della dinamica del peso della variabile denominatore calcolata sulle unità dell'insieme *i*-esimo rispetto al totale dell'aggregato su cui si calcola l'indice generale. L'impatto di questa componente, definita *effetto di ricomposizione*, è diverso a seconda della scomposizione dell'aggregato *S* che si considera;
- dalla scomposizione ne consegue che l'indice non gode, in generale, della proprietà di coerenza interna rispetto ai sotto-indici, per cui si può verificare che il valore dell'indice generale è inferiore al minimo oppure superiore al massimo degli indici dei sotto settori (casi definiti *fuori range*);
- a queste relazioni seguono delle implicazioni sull'utilizzo di un approccio indiretto alla destagionalizzazione. L'eventuale destagionalizzazione delle serie dei sotto-indici e una successiva loro aggregazione per ottenere l'indice aggregato destagionalizzato non è consigliabile, poiché richiederebbe anche la destagionalizzazione dell'indice della composizione della variabile *D*;
- la scomposizione appena mostrata è molto utile per la ricerca delle cause di valori anomali all'interno dei sotto-insiemi di classificazione, in quanto definisce la formula (cfr. la penultima riga della dimostrazione (11)), per il calcolo del contributo delle singole componenti sul tasso di variazione complessivo.

Le formule ricavate precedentemente sono generalizzate rispetto ad altri tipi di classificazione della popolazione osservata (ne sono un esempio la classificazione dei lavoratori per qualifica, orario lavorativo, appartenenza a imprese grandi o medio piccole, eccetera). Nel prossimo paragrafo verranno applicati i risultati di scomposizione ottenuti in questo lavoro ai dati dell'Indagine Oros.

3. Il caso della Rilevazione Oros: l'indice del costo del lavoro per Ula

3.1 Elementi di novità rispetto all'applicazione sull'indicatore delle retribuzioni nelle grandi imprese

Nel lavoro dell'Istat (2010) viene illustrata una applicazione della formula (11) nell'analisi dell'indicatore sulle retribuzioni per dipendente calcolato nel contesto dell'Indagine mensile sull'Occupazione, l'Orario di Lavoro e le Retribuzioni nelle Grandi Imprese (GI), volto a calcolare l'effetto di ricomposizione della variabile occupazionale a seguito delle ricomposizioni nel tempo delle strutture occupazionali per qualifica professionale. Il lavoro ripropone l'approccio presente in Keller (2009).

Nel successivo paragrafo verrà applicata la formula di scomposizione (11) per l'analisi delle serie storiche dei tassi di variazione delle Retribuzioni per Ula nel totale delle imprese appartenenti al settore privato B-S, analizzando i dati dell'Indagine Oros relativi a un intervallo temporale che arriva fino al secondo trimestre 2016, focalizzando l'attenzione sulla ricomposizione settoriale nel tempo delle Unità di Lavoro a tempo pieno (Ula) tra i settori appartenenti a uno specifico macro settore. Inoltre, verrà utilizzata la formula di scomposizione per indagare su eventuali anomalie riscontrate sui dati e sulla presenza di valori cosiddetti *fuori range*, caratteristica generalmente poco gradita dagli utilizzatori degli indicatori.

3.2 Obiettivi conoscitivi dell'Indagine Oros e metodologia

La Rilevazione Oros, acronimo di Occupazione, Retribuzioni e Oneri sociali ha come obiettivo la produzione di indicatori congiunturali sulla dinamica del costo del lavoro e delle relative componenti quali retribuzioni e oneri sociali per tutti i dipendenti dei settori da B a S della classificazione Ateco relativi a imprese e istituzioni private. A partire da giugno 2015 la rilevazione diffonde anche indici sulle posizioni lavorative dipendenti.

L'indagine rappresenta un caso peculiare in Istat di sfruttamento intensivo di dati amministrativi per fini congiunturali, producendo indici trimestrali calcolati sulla base di dati amministrativi di fonte Inps (Istat 2019).

Gli indici Oros sul costo del lavoro totale unitario e sulle sue due componenti relative a retribuzioni e oneri sociali, sono calcolati a partire da variabili rapporto, aventi a numeratore la componente di costo del lavoro e a denominatore l'input di lavoro che ha prodotto il reddito a numeratore[3].

La variabile rapporto sarà pari alla generica variabile di Costo del lavoro per Ula, indicata con il simbolo CU_t e i relativi numeratore e denominatore verranno indicati rispettivamente con il simbolo C_t e con il simbolo U_t.

Gli indici del costo del lavoro per Ula vengono prodotti e diffusi per sezione della classificazione Ateco 2007 e per macro settori. Gli indici relativi sia a uno specifico settore, sia ai rispettivi aggregati, vengono calcolati determinando la variabile rapporto

3 Per maggiori dettagli sul calcolo delle singole componenti dell'indice cfr. Istat 2019.

a partire dai rispettivi totali che, per i macro settori, sono ottenuti per somma dei valori dei settori componenti.

La motivazione principale che ha consentito questa scelta sta nel fatto che l'indagine è praticamente censuaria sia nelle stime finali, sia nelle stime provvisorie[4]: la scelta di rilevare l'insieme delle posizioni lavorative trimestralmente senza fissare una struttura di riferimento all'anno base, consente di cogliere interamente gli effetti della demografia d'impresa e della variazione della composizione occupazionale nell'intervallo temporale di analisi.

Gli indicatori così costruiti vanno a definire un sistema di indici a due dimensioni di aggregazione: una aggregazione di tipo *settoriale* e una aggregazione definita *costo del lavoro*, conseguente alla relazione che lega il costo del lavoro come somma delle componenti retribuzioni e oneri sociali.

Mentre per le aggregazioni *costo del lavoro* degli indici delle variabili *pro capite* e per l'aggregazione settoriale della variabile occupazione il sistema di indici gode di una coerenza interna giustificata dalla riconducibilità delle formule a quelle degli indici calcolati su totali (scomposizione generale mostrata in A5, per la dimostrazione cfr. Istat (2019,143)), per quanto riguarda l'aggregazione *settoriale* delle variabili delle componenti di costo del lavoro *pro capite* questa coerenza non è mantenuta proprio per la natura della metodologia utilizzata per il calcolo degli indici, la quale consiste nell'uso dell'indice semplice della variabile rapporto, in cui numeratore e denominatore sono variabili nel tempo. Utilizzando la (10) è possibile ricavare la relazione esistente tra l'indice generale dell'aggregato S e gli indici dei sotto settori $S_1, \dots, S_n$:

$$I_{S,t}^{CU} = \sum_i I_{s_i,t}^{CU}\, I_{s_i,t}^{\pi_U}\, \pi_{i,b}^{C} = f\left(I_{s_1,t}^{CU}, I_{s_2,t}^{CU}, \dots, I_{s_n,t}^{CU}, I_{s_1,t}^{\pi_U}, I_{s_2,t}^{\pi_U}, \dots, I_{s_n,t}^{\pi_U}\right) \quad (12)$$

Per quanto riguarda i tassi di variazione, viene utilizzata la formula (11) non vista come scomposizione dei singoli contributi dei sotto-settori ma nel modo seguente:

$$V_h\left(CU_{S,t}\right) = \sum_{i=1}^{n} V_h\left(CU_{s_i,t}\right) \pi_{s_i,t}^{C} + \sum_{i=1}^{n} V_h\left(\pi_{s_i,t}^{U}\right) \pi_{s_i,t}^{C} + \sum_{i=1}^{n} \frac{V_h\left(CU_{s_i,t}\right) V_h\left(\pi_{s_i,t}^{U}\right)}{100}\, \pi_{s_i,t}^{C} =$$

$$= M_1 + M_2 + M_3 \quad (13)$$

In cui il tasso di variazione è pari alla somma di tre medie M_k.

3.3 Alcuni risultati dell'analisi di scomposizione sugli indici di retribuzioni per Ula di Oros

Di seguito vengono mostrati alcuni risultati dell'analisi di scomposizione relativamente a una delle componenti di costo del lavoro prodotti dalla Rilevazione Oros, ossia le retribuzioni (R) per Ula (U), valutando la dinamica nei settori dell'Industria e dei Servizi di Mercato separatamente. Per facilitare la lettura dei risultati, l'analisi viene effettuata non sugli indici ma sulle variazioni tendenziali, la cui regola di scomposizione, come evidenziato nella relazione (13) si può esprimere come somma di tre componenti: la componente *effetto retribuzioni* ottenuta

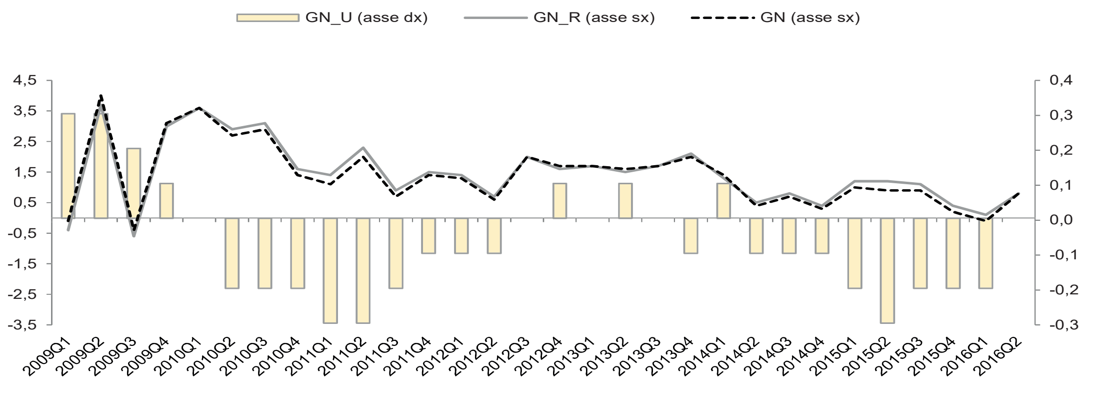
4 Nelle stime provvisorie i dati disponibili coprono circa il 95% dell'occupazione dipendente della popolazione osservata.

da una media ponderata delle retribuzioni per Ula dei sotto-settori (indicata in seguito con il suffisso R accanto al nome dell'aggregato), la componente dell'*effetto ricomposizione* (con suffisso U e nell'asse secondario del grafico), ottenuta dalla media ponderata delle variazioni del peso occupazionale dei sotto-settori, espresso in termini di Ula, e una componente residua che risulterà trascurabile utilizzando i dati a disposizione. Sono presentati due grafici (Figura 3.1 e 3.2) che rappresentano le serie delle componenti e della variazione finale e una Tavola (Tavola 3.1) che contiene dati relativi alle componenti e ai tassi di variazione tendenziali del peso occupazionale dei sotto-settori che concorrono all'aggregazione, al fine di interpretare alcuni effetti che si notano sulla dinamica.

Figura 3.1 – Retribuzioni per Ula nell'Industria: "effetto retribuzioni" ed "effetto ricomposizione"
(Serie storica I trimestre 2009 - II trimestre 2016) (a)

Fonte: Istat, Elaborazioni su dati dell'Indagine Oros, Retribuzioni per Ula nell'Industria
(a) Tassi di variazione tendenziali BF, effetto retribuzioni BFR, effetto ricomposizione BFU, serie storica trimestri Q1:2009-Q2:2016.

Figura 3.2 – Retribuzioni per Ula nei Servizi di Mercato: "effetto retribuzioni" ed "effetto ricomposizione"
(Serie storica I trimestre 2009 - II trimestre 2016) (a)



Fonte: Istat, Elaborazioni su dati dell'Indagine Oros, Retribuzioni per Ula nell'Industria
(a) Tassi di variazione tendenziali GN; effetto retribuzioni GNR, effetto ricomposizione GNU.

Nella Figura 3.1 e nella Tavola 3.1 è rappresentata la serie storica delle retribuzioni per Ula nel settore dell'industria e le relative serie componenti BFR e serie BFU derivanti dalla scomposizione dell'aggregato nei sotto-settori industria in senso stretto e costruzioni (BF=BE + F).

È possibile osservare nel 2012 e nel 2013 un effetto ricomposizione rilevante se confrontato con la serie storica, con valori che oscillano tra 0,3 e 0,4 punti percentuali, a tal punto da causare nei trimestri Q2:2012, Q3:2012 e Q4:2013 la presenza di *fuori range* settoriali di valore rispettivamente pari a 0,2, 0,3 e 0,3 punti percentuali. Il calo dell'occupazione che si osserva nel settore delle costruzioni è stato "veloce" e persistente tanto da comportare una ricomposizione occupazionale a favore dell'industria in senso stretto, la cui occupazione ha tenuto maggiormente. La ricomposizione occupazionale osservata non è conseguente a un travaso di occupazione da un settore all'altro, ma al naturale effetto di ricomposizione della distribuzione dei dipendenti del settore dell'industria a seguito del licenziamento dei lavoratori del settore delle costruzioni. Si ricordi che il valore dell'effetto ricomposizione in formula (11) è il risultato di una media delle variazioni del peso occupazionale dei sotto-settori ponderate per il monte retributivo che, in media in serie storica, è pari all' 83% per BE e al 17% per F. Nella fase di crisi, il peso delle costruzioni in termini di Ula nel settore dell'industria è diminuito del 6,6% tra il trimestre Q4:2013 e il trimestre Q4:2012, mentre il peso dell'industria in senso stretto è cresciuto dell'1,8%. Ne risulta infine un effetto ricomposizione di 0,4, seguendo lo schema di ponderazione nella formula (11). Il valore così ottenuto, sommandosi a quello dell'effetto retribuzioni pari al 2,2%, restituisce un valore finale di 2,6%, fuori dall'intervallo definito dalla variazione del settore BE pari a 2,3% e a quello delle costruzioni pari a 1,6%. Il *fuori range* è la principale evidenza sui dati della mancanza della proprietà di coerenza interna negli indici semplici di variabili *pro capite* e può essere causa di una scorretta e difficile interpretazione da parte degli utilizzatori dei dati, per i quali i risultati di questo lavoro sono offerti come strumento per l'interpretazione.

Nella Figura 3.2 e nella Tavola 3.1 viene rappresentata la serie storica delle variazioni tendenziali delle retribuzioni per Ula nel settore dei servizi di mercato, serie GN, e le componenti (facendo riferimento alla formula (13)), serie GNR e serie GNU. Nel caso specifico si è analizzata la scomposizione dell'aggregato GN come somma delle sezioni da G a N, e conseguentemente l'effetto ricomposizione è il risultato della redistribuzione della variabile occupazionale tra le sezioni indicate. Analizzando le serie storiche, l'effetto ricomposizione maggiore che si osserva è di 0,3 e 0,2 in valore assoluto, valori osservabili nel periodo 2010-2011 e 2015 inizio 2016. Essi sono il risultato di una media ponderata delle variazioni del peso occupazionale dei settori G – N. I settori G, H, J e K sono i settori aventi il peso in termini di monte retributivo a t maggiore (insieme costituiscono il 73% in media del monte retributivo dell'aggregato), in cui si osservano valori in media in serie storica pari rispettivamente a 30%, 18%, 10% e 15%. In questo esempio i valori più alti di effetto ricomposizione sono trainati da questi settori, in particolare dal settore G avente il maggiore peso. Nell'esempio riportato, nel settore dei servizi, non si osservano valori *fuori range*.

Tavola 3.1 - Serie storica delle componenti "effetto retribuzioni" ed "effetto ricomposizione"
(Serie storica I trimestre 2010-II trimestre 2016; Elementi della scomposizione e tassi di variazione tendenziali)

TRIMESTRI	BF = BE + F							GN = G + H + I + J + K + L + M + N											
	BF_R (a)	BF_U (b)	BF_{RU} (c)	BF (d)	π^U (e) BE	π^U (e) F	U (f) F	GN_R (a)	GN_U (b)	GN_{RU} (c)	GN (d)	π^U (e) G	H	I	J	K	L	M	N
2010Q1	4,1	0,1	0	**4,2**	0,6	-2,0	-9,1	3,6	0	0	**3,6**	0,3	-0,5	1,9	-0,2	0,9	0,2	-4,2	0,5
2010Q2	4,7	0,1	0	**4,8**	0,9	-2,7	-5,9	2,9	-0,2	0	**2,7**	-0,2	0,0	1,3	-1,4	-0,7	1,1	-4,4	3,3
2010Q3	3,5	0,1	0	**3,6**	0,9	-2,7	-5,2	3,1	-0,2	0	**2,9**	-0,4	-0,2	1,6	-2,0	-1,1	1,0	-4,1	3,9
2010Q4	2,5	0,2	0	**2,7**	1,0	-3,2	-5,4	1,6	-0,2	0	**1,4**	-0,2	-0,2	1,1	-2,1	-0,6	0,2	-3,8	3,9
2011Q1	4,3	0,1	0	**4,4**	0,7	-2,3	-2,6	1,4	-0,3	0	**1,1**	-0,6	-0,2	-0,5	-2,6	-1,9	-2,3	0,4	5,0
2011Q2	4,1	0,1	0	**4,2**	1,0	-3,2	-4	2,3	-0,3	0	**2**	-0,6	-0,6	-0,1	-2,5	-1,7	-2,0	0,2	5,0
2011Q3	2,2	0,2	0	**2,4**	1,1	-3,6	-4,7	0,9	-0,2	0	**0,7**	-0,7	-0,4	0,3	-1,1	-1,6	-1,8	-0,1	3,9
2011Q4	2,1	0,2	0	**2,3**	1,1	-3,5	-5,2	1,5	-0,1	0	**1,4**	-0,5	-0,5	0,4	-0,5	-1,1	-3,2	0,3	2,6
2012Q1	3,3	0,2	0	**3,5**	1,6	-5,5	-8,9	1,4	-0,1	0	**1,3**	-0,2	-0,6	1,7	0,9	-0,5	-1,0	-0,6	0,2
2012Q2	1,1	0,2	0	**1,3**	1,6	-5,3	-9,2	0,7	-0,1	0	**0,6**	-0,4	-0,3	3,2	0,4	0,0	-1,0	-0,1	-1,5
2012Q3	2,7	0,3	0	**3**	1,7	-5,7	-10,7	2	0	0	**2**	-0,4	-0,1	2,3	0,2	0,7	-0,5	0,4	-2,0
2012Q4	3,5	0,3	0	**3,8**	1,5	-5,2	-10,7	1,6	0,1	0	**1,7**	-0,7	-0,1	1,4	1,7	0,3	1,3	1,5	-1,5
2013Q1	1,3	0,3	0	**1,6**	1,7	-6,3	-11,4	1,7	0	0	**1,7**	-0,8	0,1	-0,2	0,3	-0,2	-0,4	0,8	1,6
2013Q2	2,4	0,2	0	**2,6**	1,6	-5,9	-10,8	1,5	0,1	0	**1,6**	-1,0	0,1	-3,0	1,7	-0,2	0,4	1,4	3,4
2013Q3	2,4	0,3	0	**2,7**	1,7	-6,1	-10	1,7	0	0	**1,7**	-1,1	-0,1	-2,2	1,8	-0,9	1,5	0,8	4,2
2013Q4	2,2	0,4	0	**2,6**	1,8	-6,6	-9,4	2,1	-0,1	0	**2**	-1,2	0,4	-2,4	0,8	-0,6	-0,6	-0,4	4,5
2014Q1	2,1	0,2	0	**2,3**	1,4	-5,8	-8	1,3	0,1	0	**1,4**	-1,4	0,1	-2,9	0,8	0,5	11,9	0,9	3,4
2014Q2	1,7	0,2	0	**1,9**	1,6	-6,2	-7,8	0,5	-0,1	0	**0,4**	-1,6	0,2	-0,4	-0,5	-0,1	13,7	0,1	3,2
2014Q3	1,9	0,2	0	**2,1**	1,4	-5,6	-6,7	0,8	-0,1	0	**0,7**	-1,4	0,1	-0,9	-0,3	-0,5	13,9	-0,3	3,7
2014Q4	1,5	0,2	0	**1,7**	1,1	-4,5	-5,5	0,4	-0,1	0	**0,3**	-1,6	0,0	-0,6	0,1	-1,0	12,3	0,2	3,6
2015Q1	2,2	0,1	0	**2,3**	0,6	-2,5	-3,6	1,2	-0,2	0	**1**	-1,8	-0,1	-0,2	0,4	-1,7	9,4	0,9	4,0
2015Q2	1,7	0	0	**1,7**	0,2	-0,8	-0,9	1,2	-0,3	0	**0,9**	-2,0	-0,5	0,6	1,8	-3,5	8,1	0,7	4,5
2015Q3	1,6	0	0	**1,6**	0,2	-0,7	-1,9	1,1	-0,2	0	**0,9**	-1,4	-1,2	0,9	0,5	-1,9	7,9	1,1	3,1
2015Q4	1,6	0	0	**1,6**	-0,1	0,4	0,3	0,4	-0,2	0	**0,2**	-0,8	-1,5	2,4	0,7	-2,4	10,0	1,1	1,4
2016Q1	-0,1	0	0	**-0,1**	-0,1	0,6	2,2	0,1	-0,2	0	**-0,1**	-0,1	-2,0	4,9	1,4	-3,2	9,0	2,0	-1,6
2016Q2	0,9	0,1	0	**1**	0,4	-1,8	-1,2	0,8	0	0	**0,8**	0,5	-1,9	0,7	0,6	-1,7	7,6	2,7	-0,9

Fonte:Istat, Elaborazioni su dati dell'Indagine Oros, Retribuzioni per Ula nell'Industria e nei Servizi di Mercato
(a) Effetto retribuzioni (media ponderata di variazioni tendenziali delle Retribuzioni per Ula dei sotto-settori).
(b) Effetto composizione (media ponderata di variazioni tendenziali della composizione occupazionale tra i sotto-settori).
(c) Effetto misto trascurabile.
(d) Tasso di variazione tendenziale delle retribuzioni per Ula.
(e) Tasso di variazione tendenziale della composizione occupazionale delle Ula (quota).
(f) Tasso di variazione tendenziale delle Ula (livello).

Analizzando le serie storiche e considerando i risultati ottenuti per gli indici semplici di variabili rapporto all'interno della Rilevazione Oros si può osservare come in molti casi la proprietà di coerenza interna viene mantenuta, anche se per definizione dell'indice non dovrebbe essere garantita. Questo accade in quanto, oltre al termine residuo M_3 della (13), anche il termine dell'*effetto ricomposizione*, M_2, osservato nella scomposizione tra macro settore e sezioni Ateco, risulta il più delle volte vicino a 0 a causa dell'inerzia persistente nella struttura occupazionale osservata nella serie storica, semplificando la variazione del costo del lavoro *pro capite* alla media ponderata $M_1 = \sum_{i=1}^{n} V_h\left(CU_{s_i,t}\right) \pi_{s_i,t}^{C}$.

Diversamente in casi in cui la ricomposizione occupazionale diventa rilevante questa proprietà viene meno, aumentando l'impatto dell'effetto composizione sulla variazione finale.

Il significato economico di un effetto ricomposizione basso consiste nell'avere una struttura occupazionale che non varia tra gli istanti temporali messi a confronto. Infatti, analizzando la scomposizione in (13) si possono interpretare le prime due medie M_1 e M_2

rispettivamente nel contributo alla variazione di retribuzione *pro capite* calcolato a struttura occupazionale fissa e al contributo dovuto alla variazione di struttura occupazionale calcolata a retribuzione *pro capite* fissata. In questo modo è possibile isolare l'effetto "prezzo" relativo alla sola remunerazione dall'effetto quantità relativo alla composizione dei lavoratori, utile anche ai fini analitici.

3.4 Criticità e spunti di riflessione

Con lo scopo di comprendere l'origine della dinamica di indici semplici calcolati su variabili rapporto, le scomposizioni mostrate hanno fatto emergere come una delle principali problematiche nell'utilizzo di questa metodologia sia l'assenza della proprietà di coerenza interna degli indici, con tutte le possibili conseguenze che tale aspetto può implicare soprattutto nelle situazioni di cui si osservano cambiamenti nella struttura di base.

Alcune possibili alternative all'uso dell'indice semplice su variabili rapporto possono essere:

- calcolo di un indice semplice di valore, applicato alla variabile "monte retribuzioni".

  Il vantaggio sarebbe sicuramente la garanzia di una struttura di pesi con cui aggregare i sotto-indici nell'indice generale e conseguente garanzia di coerenza interna. Un indice di questo tipo non permetterebbe, tuttavia, di scindere l'effetto del valore in effetto prezzo e quantità. Utilizzando questa formulazione si avrebbe un'indicazione sull'evoluzione della retribuzione complessiva e non del contributo unitario dell'input di lavoro (misurato in Ula) a generare reddito da lavoro:

  $$I_{S,t}^{R} = \frac{R_{S,t}}{R_{S,b}} = \frac{\sum_i R_{s_i,t}}{\sum_i R_{s_i,b}} = \sum_i \frac{R_{s_i,t}}{R_{s_i,b}} \frac{R_{s_i,b}}{\sum_i R_{s_i,b}} = \sum_i I_{s_i,t}^{R} \, \pi_{s_i,b}^{R} \quad (14)$$

  Nel grafico seguente sono mostrate le serie delle variazioni tendenziali delle retribuzioni per Ula prodotte in Oros per il totale economia B-N (serie V4(RU)) e, insieme, i valori relativi alle variazioni dell'indice del monte retributivo (indice di valore ipotizzato come alternativa all'indice del *pro capite*, serie V4(R)) e dell'indice del monte Ula (serie V4(U)). Utilizzando un'approssimazione della formula (4), il tasso di variazione dell'indice delle retribuzioni per Ula è pari alla differenza dei tassi di variazione dell'indice del monte retributivo e del monte Ula (per semplicità di letture il termine V4(U) è stato inserito nel grafico con segno invertito).

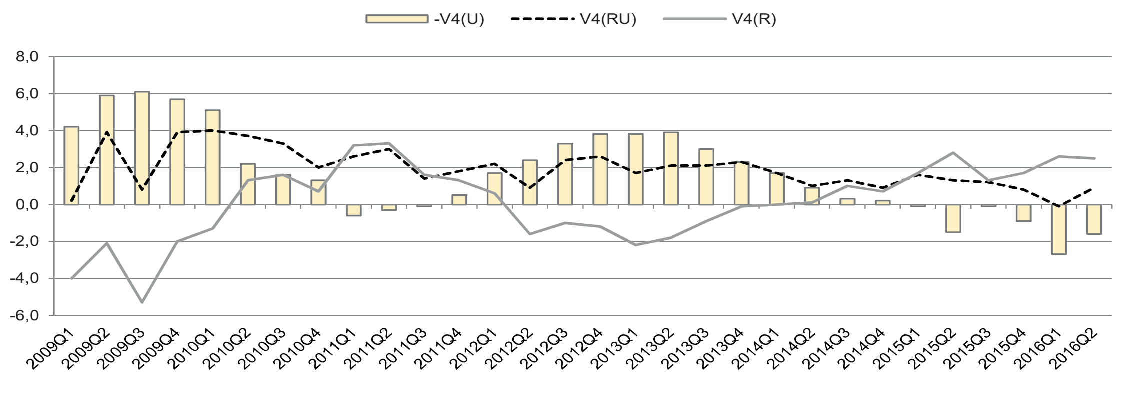
  $$\boldsymbol{V_h\left(I_t^{RU}\right) \approx V_h\left(I_t^{R}\right) - V_h\left(I_t^{U}\right)} \quad (15)$$

  Si noti come, nei trimestri in cui l'occupazione varia molto, i due indicatori di valore e di prezzo si discostino fortemente tanto da dare indicazioni diverse in termini di segno di crescita.

- Utilizzo di un indice di tipo *Laspeyres*. È possibile definire valore, prezzo e quantità a livello elementare?

  Da un punto di vista metodologico, un approccio di questo tipo viene applicato nel calcolo del *Labour Cost Index* (LCI), in cui si utilizza uno schema di ponderazione di tipo *Laspeyres* concatenato per calcolare gli indici relativi ai macro settori. Gli indici elementari sono gli indici semplici di variabili rapporto calcolati a livello di sezione Ateco (Istat 2019). Pertanto l'effetto ricomposizione osservato tra le sezioni del macro settore aggregato è nullo per definizione, e resta valorizzato ma non direttamente osservabile, l'effetto ricomposizione che si osserva tra le aggregazioni al di sotto della sezione Ateco.

Figura 3.3 – Scomposizione numeratore/denominatore applicata all'indice delle Retribuzioni per Ula
(Serie storica I trimestre 2009 - II trimestre 2016) (a)



Fonte: Istat, Elaborazioni su dati dell'Indagine Oros, Retribuzioni per Ula nel totale economia B-N (solo settore privato)
(a) Tassi di variazione tendenziali.

3.5 Altri esempi di scomposizione: scomposizione micro e impatto della mancata risposta totale

Come descritto nel paragrafo 1 il problema di scomposizione del tasso di variazione può essere utile per impostare un problema di *editing selettivo*, nel caso in cui la variabile *target* scelta come variabile da monitorare è proprio il tasso di variazione. Questo tipo di applicazione viene utilizzata nel contesto dell'Indagine Oros. Sia l'aggregato S l'unione di sottoinsiemi, ognuno costituito da una singola unità. L'indice sarà il risultato dell'aggregazione dei singoli micro indici. Analoga argomentazione vale anche per i tassi di variazione micro. È necessaria una precisazione: la scomposizione dei tassi di variazione in micro contributi è ben definita se e solo se si può osservare per ogni unità statistica il valore relativo al trimestre tendenziale o congiunturale. Questo presuppone che non ci siamo demografia di impresa ma che tutte le unità osservate facciano parte di un *panel* chiuso alle entrate e alle uscite. In questo modo infatti, è possibile calcolare l'indice e il tasso di variazione della variabile rapporto per ogni unità eliminando i casi in cui o il dato al tempo $t + h$ o il dato al tempo t siano pari a zero per motivi legati a nascita o cessazione dell'impresa. Supponiamo, inoltre, che il dato al tempo t non sia affetto da mancata risposta. La mancata risposta nel contesto dell'Indagine Oros ha una precisa configurazione. I dati sono calcolati come medie mensili dei dati relativi al trimestre osservato e non possono essere parziali, nel senso che il *record* relativo all'impresa può essere è vuoto oppure pieno, con tutte le variabili valorizzate, ammettendo il solo caso di mancata risposta totale, relativa sia alla variabile numeratore sia denominatore. Nel caso in cui siano presenti mancate risposte totali nel mese, esse influenzeranno sia numeratore sia denominatore delle variabili rapporto oggetto di analisi. Supponendo come esempio un caso di mancata risposta in un mese per una singola unità statistica, e di poter stimare la media mensile calcolata sui tre mesi, con la media mensile calcolata sui due mesi di riposta (ad esempio nei trimestri in cui non vengono erogate tredicesima o quattordicesima o i premi di qualunque natura), si può essere indotti a sottostimare l'effetto della mancata risposta sull'indice generale. Tuttavia, analizzando le relazioni (10) e (11), considerando l'evidenza per cui la formula di scomposizione non è una media ponderata delle sole variabili pro-capite, e considerando quindi l'influenza sull'*effetto di ricomposizione*, il quale è funzione del livello dell'occupazione e non di un valore *pro capite*, esso verrebbe sottostimato in caso di mancata risposta. L'eventuale revisione del dato finalizzata a incorporare risposte ritardatarie giunte in periodi successivi sarebbe affetta da questo "*errore di revisione*".

Se l'impresa *j*-esima risulta influente sulla composizione occupazionale dell'aggregato, sia in termini di variazione della composizione occupazionale sia in termini di monte costo del lavoro osservato al tempo *t*, l'*effetto ricomposizione* che ne deriva va a influire sulla variazione della variabile rapporto oggetto di studio calcolata sull'aggregato osservato.

$$V_h\left(I_{S,t}^X\right) = \sum_{i \in S}\left(V_h\left(I_{i,t}^X\right) + V_h\left(I_{i,t}^{\pi_D}\right) + \frac{V_h\left(I_{i,t}^X\right)V_h\left(I_{i\,,t}^{\pi_D}\right)}{100}\right)\pi_{i,t}^N = \sum_{i \in S} C_i\,\pi_{i,t}^N$$

$$C_k = V_h\left(I_{k,t}^X\right) + V_h\left(I_{k\,,t}^{\pi_D}\right) + \frac{V_h\left(I_{k,t}^X\right)V_h\left(I_{k\,,t}^{\pi_D}\right)}{100}$$

Queste considerazioni, basate sull'approccio analitico presentato, sono alla base dell'implementazione di una procedura di *editing selettivo* e imputazione di mancate risposte influenti sulla variazione dell'indice generale, utilizzata nell'Indagine Oros quale fase finale di controllo delle stime prodotte[5].

5 Per una spiegazione approfondita delle caratteristiche di questo metodo cfr. Istat 2019, paragrafo 6.6.

4. Conclusioni

Quale strumento di analisi delle serie storiche di indicatori congiunturali di variabili rapporto, lo studio è stato finalizzato a ricavare alcune regole di scomposizione dell'indice semplice calcolato su tale tipologia di variabili, consentendo di farne emergere alcune criticità e potenzialità. D'altra parte, è importante tenere conto e monitorare i singoli elementi di questa tipologia di indici (e relativi tassi di variazione) e, inoltre, la scomposizione è un passaggio fondamentale per poter interpretare correttamente il fenomeno congiunturale oggetto di studio. In generale, si può concludere che:

- l'utilizzo di un indice semplice per variabili rapporto consente di tenere conto anche della dinamica della struttura occupazionale (o in generale della variabile denominatore), il che non sarebbe possibile utilizzando un indice di valore;
- non è possibile definire un sistema di aggregazione che garantisca coerenza interna tra gli indici e le variazioni settoriali rispetto a un aggregato generale. La conseguenza principale è la possibilità di osservare *fuori range* settoriali negli indici e nei tassi di variazione. Per avere garantita questa proprietà va utilizzata una differente metodologia di calcolo dell'indice. La scomposizione ricavata ha fatto emergere le cause dei *fuori range* e offerto uno strumento per monitorare le serie storiche;
- la scomposizione, sia nel caso degli indici sia dei tassi di variazione, permette di calcolare il contributo di ogni singola componente dell'aggregato, strumento utile per l'analisi delle serie storiche e per il monitoraggio della qualità degli indicatori.

Appendice A - Definizioni utili

L'indice semplice, a base fissa, calcolato sulle serie storiche di variabili rapporto può essere utilizzato per valutare la dinamica di una quantità misurata in termini relativi, ad esempio un costo unitario, per il cui calcolo si abbiano a disposizione soltanto le variabili relative a "valore" e "quantità" espresse come totali. Ne sono un esempio le retribuzioni per Ula calcolate nel contesto dell'Indagine Oros.

Sia X_t una variabile osservata in serie storica a intervalli regolari (mensili, trimestrali, eccetera) dove t indica l'istante di tempo relativo all'osservazione. Definiremo il suo indice semplice a base fissa come:

$$I_t^X = \frac{X_t}{X_b} \quad \text{(A1)}$$

dove b rappresenta un preciso riferimento temporale, detto base, rispetto al quale vengono rapportati i valori osservati a ogni istante t[6]. L'indice solitamente viene calcolato in termini percentuali.

Sia ora la variabile X_t, ottenuta dal rapporto di due variabili, definite rispettivamente N_t e D_t.

L' indice semplice della variabile rapporto X_t, sarà definito come segue:

$$I_t^X = \frac{X_t}{X_b} = \frac{N_t/D_t}{N_b/D_b} \quad \text{(A2)}$$

Definiamo inoltre il tasso di variazione di passo h, indicato con il simbolo $V_h(\cdot)$, la differenza relativa tra l'indice della variabile X al tempo $t + h$, e l'indice rispetto al tempo t in valore percentuale (ad esempio $h = 1$ per le variazioni congiunturali, $h = 4$ per le variazioni tendenziali di osservazioni trimestrali e $h = 12$ per le variazioni tendenziali di osservazioni mensili) definita dalla relazione seguente:

$$V_h(I_t^X) = \left(\frac{I_{t+h}^X}{I_t^X} - 1\right)100 \quad \text{(A3)}$$

Si osserva che la relazione che esprime il tasso di variazione della variabile X calcolata sull'indice restituisce lo stesso risultato se calcolata sulla variabile stessa, per via della semplificazione dei termini riferiti all'anno base[7].

$$V_h(I_t^X) = \left(\frac{I_{t+h}^X}{I_t^X} - 1\right)100 = \left(\frac{\frac{X_{t+h}}{X_b}}{\frac{X_t}{X_b}} - 1\right)100 = \left(\frac{X_{t+h}}{X_t} - 1\right)100 = V_h(X_t) \quad \text{(A4)}$$

6 Di norma nel calcolo di indici a base fissa, quale valore osservato in un istante temporale b, si considera la media delle osservazioni di X in un determinato anno, per cui b si definisce anno base.

7 Occorre precisare che l'uguaglianza descritta può non essere verificata in presenza di arrotondamenti intermedi effettuati sull'indice prima del calcolo della variazione. Traendo spunto dalla teoria sui metodi numerici, si verifica che l'errore originario sugli indici, causato dall'operazione di arrotondamento, si propaga tramite il calcolo della variazione tendenziale in misura inversamente proporzionale alla differenza degli indici (più la differenza è piccola più l'errore si amplifica). Nei risultati ottenuti nel calcolo delle variazioni degli indici sui dati dell'Indagine Oros, questo errore rimane spesso contenuto e inferiore o uguale a 0,2.

Di norma gli indicatori di carattere infra-annuale vengono prodotti e rilasciati secondo un piano di diffusione che contempla indici aggregati e, insieme a essi, indici relativi a sotto insiemi di classificazione. Una delle classificazioni oggetto di pubblicazione maggiormente utilizzata è la classificazione delle attività economiche, attualmente Ateco 2007.

In tale ambito, è interessante ricavare una relazione che leghi gli indici (o i tassi di variazione) della variabile riferita a unità statistiche classificate in insiemi disgiunti (ad esempio le divisioni o sezioni dell'Ateco) con l'indice (o il tasso di variazione) relativo all'insieme unione delle unità che costituiscono l'aggregato.

Supponiamo di avere una partizione delle unità statistiche in n sottoinsiemi disgiunti di S, indicati con S_1, …, S_n. Pensiamo ad esempio a S come il macro settore dell'industria (aggregato B-F) e gli n sottoinsiemi come le sezioni Ateco che lo compongono (sezioni B, C, D, E, F). Un altro esempio di aggregazione del macro settore industria potrebbe essere tra il macro settore dell'industria in senso stretto (B-E) e il settore delle costruzioni (F).

Definiamo l'indice relativo a un sotto insieme S_i di S, come l'indice calcolato sulla variabile osservata sul settore S_i all'istante t rispetto all'anno base. Quest'ultima variabile si definisce come l'ammontare della variabile Y misurato sulle unità j appartenenti al sotto-insieme i-esimo[8]. Indicheremo con $m_{i,t}$ la numerosità dell'insieme S_i[9].

$$I_{S_i,t}^{Y} = \frac{Y_{S_i,t}}{Y_{S_i,b}} = \frac{\sum_{j=1}^{m_{i,t}} y_{j,t}}{\sum_{j=1}^{m_{i,t}} y_{j,b}}$$

Si dimostra facilmente che la scomposizione degli indici e dei tassi di variazione nelle componenti dell'aggregato per variabili espresse come totali è esprimibile come una media ponderata di indici elementari.

$$I_{S,t}^{Y} = \frac{Y_{S,t}}{Y_{S,b}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_{i,t}} y_{j,t}}{\sum_{i=1}^{n}\sum_{j=1}^{m_{i,t}} y_{j,b}} = \frac{\sum_i Y_{S_i,t}}{\sum_i Y_{S_i,b}} = \sum_i \frac{Y_{S_i,t}}{Y_{S_i,b}} \frac{Y_{S_i,b}}{\sum_i Y_{S_i,b}} =$$

$$= \sum_i I_{S_i,t}^{Y} \pi_{S_i,b}^{Y} \quad \text{(A5)}$$

Lo stesso vale per i tassi di variazione e la dimostrazione è fornita come esempio dell'applicazione della formula generale nell'Appendice B (B12).

Sia X la variabile rapporto definita in precedenza si ha:

$$I_{S_i,t}^{X} = \frac{X_{S_i,t}}{X_{S_i,b}} = \left. \frac{\sum_{j=1}^{m_{i,t}} N_{j,t}}{\sum_{j=1}^{m_{i,t}} D_{j,t}} \middle/ \frac{\sum_{j=1}^{m_{i,t}} N_{j,b}}{\sum_{j=1}^{m_{i,t}} D_{j,b}} \right.$$

---

8 La formulazione si adatta al caso di indagini censuarie. Nel caso di indagini campionarie il valore $Y_{j,t}$ deve essere sostituito dalla stima campionaria.

9 La numerosità $m_{i,t}$, per come è definita, sta a indicare che le unità appartenenti all'insieme S_i al tempo t non sono necessariamente le stesse unità appartenenti all' insieme in un diverso istante temporale: in questa formulazione rientrano entrambi i casi di indagine *panel* e *non panel* in cui si considera la demografia di impresa.

mentre l'indice dell'aggregato $S = \bigcup_{i=1}^{n} S_i$ sarà uguale a:

$$I_{S,t}^{X} = \frac{X_{S,t}}{X_{S,b}} = \left. \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_{i,t}} N_{j,t}}{\sum_{i=1}^{n} \sum_{j=1}^{m_{i,t}} D_{j,t}} \middle/ \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_{i,t}} N_{j,b}}{\sum_{i=1}^{n} \sum_{j=1}^{m_{i,t}} D_{j,b}} \right.$$

Per semplicità scriveremo $N_{si,t}$ per indicare $\sum_{j=1}^{m_{i,t}} N_{j,t}$. La stessa notazione è usata in maniera speculare per la variabile D.

La dimostrazione delle formule di scomposizione per indici semplici di variabili rapporto sono inserite nel paragrafo 2.3.

Appendice B - Calcolo della scomposizione dei tassi di variazione di un indice in funzione delle componenti: un approccio analitico

In questa sezione viene affrontato il problema di scomposizione delle variazioni di un indice generale nelle sue componenti utilizzando un approccio analitico. In sintesi qualunque sia la funzione di scomposizione dell'indice generale negli indici componenti, utilizzando una precisa formula è possibile calcolare la relazione di scomposizione per i tassi di variazione. L'approccio fa uso della stima con espansione in serie di *Taylor* applicati a una funzione in più variabili.

Supponiamo di avere un elenco di variabili Z_i per $i = 1, .., k$ e di osservarne le realizzazioni a intervalli regolari, e supponiamo di poter interpolare le osservazioni con delle funzioni del tempo $Z_i(t)$ che sono per ipotesi almeno due volte derivabili in t. Supponiamo inoltre di osservare le realizzazioni di una variabile Y la quale è in relazione con le Z_i tramite una funzione f per ipotesi differenziabile due volte. Definiamo con $Y(t)$ la relativa interpolazione delle osservazioni, per ipotesi due volte differenziabile come funzione delle $Z_i(t)$ e pari alla f:

$$Y(t) = f\left(Z_1(t), Z_2(t), \dots, Z_k(t)\right) \quad \text{(B1)}$$

Avremo allora che $Y(t)$ è derivabile rispetto a t come composizione di una funzione differenziabile applicata a un vettore di funzioni derivabili. In seguito scriveremo per semplicità di notazione Z^i_t per indicare $Z_i(t)$.

Utilizzando la notazione $V_h(Z^i_t)$ per indicare il tasso di variazione della variabile Z^i_t di passo h, come definita in (A3), e la notazione $\dot{Z}^i_t$ per indicare $\frac{dZ_i(t)}{dt}$, la derivata della variabile $Z_i(t)$ rispetto al tempo t, si noti come le due quantità sono in relazione:

$$\lim_{h\to 0}\frac{V_h(Z^i_t)}{h} = \frac{1}{Z^i_t}\frac{dZ_i(t)}{dt} = \frac{d\ln(Z_i(t))}{dt} \quad \text{(B2)}$$

Data la f saremo interessati a ricavare una funzione g tale che:

$$V_h(Y_t) = V_h\left(f\left(Z^1_t, Z^2_t, \dots, Z^k_t\right)\right) = g\left(V_h(Z^1_t), V_h(Z^2_t), \dots, V_h(Z^k_t)\right) \quad \text{(B3)}$$

Sappiamo che per la derivabilità di Z^i_t vale la seguente stima al primo ordine:

$$Z^i_{t+h} - Z^i_t = h\,\dot{Z}^i_t + o_i(h) \quad \text{(B4)}$$

Da cui possiamo ricavare una stima del tasso di variazione di passo h in funzione della derivata:

$$V_h(Z^i_t) = \frac{h\,\dot{Z}^i_t + o_i(h)}{Z^i_t}\,100 \quad \text{(B5)}$$

Saremo interessati alla relazione inversa ossia:

$$\dot{Z}^i_t = \frac{V_h(Z^i_t)\,Z^i_t}{100\,h} - \frac{o_i(h)}{h} \quad \text{(B6)}$$

Nell'ipotesi in cui la variabile Z^i_t sia due volte derivabile, utilizzeremo anche la stima al second'ordine:

$$Z^i(t+h) - Z^i(t) = h\,\dot{Z}^i_t + \frac{1}{2}h^2\ddot{Z}^i_t + o_i\,(h^2) \quad \text{(B7)}$$

Ora calcoliamo una stima al secondo ordine per la funzione *f* come funzione di *t*, sfruttandone la differenziabilità. La funzione, di una variabile, è il risultato della composizione di una funzione differenziabile in più variabili applicata a *k* funzioni di *t* derivabili due volte e di *k* funzioni di una variabile in *t*. Indicando con il termine <·,·> il prodotto scalare tra due vettori, con il termine ∇ il gradiente di una funzione in più variabili ovvero il vettore *k-dimensionale* avente per componenti le derivate parziali rispetto alle singole componenti, con il termine ∇^2 l'Hessiano della funzione *f* come funzione delle Z^i_t e con il termine $\underline{Z}_t$ il vettore degli Z^i_t per $i = 1, .., k$, avremo:

$$f(\underline{Z}_{t+h}) - f(\underline{Z}_t) = h < \nabla f, \underline{\dot{Z}}_t > + \frac{1}{2}h^2 < \underline{\dot{Z}}_t, \nabla^2 f \cdot \underline{\dot{Z}}_t > + \frac{1}{2}h^2 < \nabla f, \underline{\ddot{Z}}_t > + o(h^2)$$

Mettendo a fattore il gradiente e semplificando otteniamo:

$$f(\underline{Z}_{t+h}) - f(\underline{Z}_t) = < \nabla f, h\underline{\dot{Z}}_t + \frac{1}{2}h^2\underline{\ddot{Z}}_t > + \frac{1}{2}h^2 < \underline{\dot{Z}}_t, \nabla^2 f \cdot \underline{\dot{Z}}_t > + o(h^2) \quad \text{(B8)}$$

Inoltre, utilizzando la (B5) e la (B6) possiamo inserire nella formula il tasso di variazione. Infatti:

$$\frac{V_h(Z^i_t)Z^i_t}{100} = Z^i_{t+h} - Z^i_t = h\,\dot{Z}^i_t + \frac{1}{2}h^2\ddot{Z}^i_t + o_i(h^2)$$

$$\frac{V_h(Z^i_t)Z^i_t}{100} = Z^i_{t+h} - Z^i_t = h\,\dot{Z}^i_t + o_i(h) \Rightarrow \dot{Z}^i_t = \frac{V_h(Z^i_t)Z^i_t}{100\,h} - \frac{o_i(h)}{h}$$

Sostituendo nella (B8) otteniamo:

$$\frac{V_h(Y_t)Y_t}{100} = Y_{t+h} - Y_t =$$

$$= f(\underline{Z}_{t+h}) - f(\underline{Z}_t) = < \nabla f, h\underline{\dot{Z}}_t + \frac{1}{2}h^2\underline{\ddot{Z}}_t > + \frac{1}{2}h^2 < \underline{\dot{Z}}_t, \nabla^2 f \cdot \underline{\dot{Z}}_t > + o(h^2) =$$

$$= \sum_i \frac{\partial f}{\partial Z^i_t}\left(\frac{V_h(Z^i_t)Z^i_t}{100} - o_i\,(h^2)\right) + \frac{1}{2}h^2\sum_{i,j}\frac{\partial f}{\partial Z^i_t}\frac{\partial f}{\partial Z^j_t}\left(\frac{V_h(Z^i_t)Z^i_t}{100\,h} - \frac{o_i(h)}{h}\right)\left(\frac{V_h(Z^j_t)Z^j_t}{100\,h} - \frac{o_i(h)}{h}\right) + o(h^2) \quad \text{(B9)}$$

Divideremo l'equazione in due parti, la prima parte che rappresenta l'equazione (B3) e la seconda costituita da tutti i residui, che vengono accorpati in un'unica funzione *R*. Per semplicità utilizzeremo la notazione *f* per indicare Y_t:

$$V_h(Y_t) = \frac{Y^i_{t+h} - Y^i_t}{Y^i_t}100 = \text{per la (B9)} =$$

$$= \frac{1}{f}\sum_i \frac{\partial f}{\partial Z^i_t}\; V_h(Z^i_t)Z^i_t + \frac{1}{200\,f}\sum_{i,j}\frac{\partial f}{\partial Z^i_t}\frac{\partial f}{\partial Z^j_t}V_h(Z^i_t)\,V_h(Z^j_t)\,Z^i_t Z^j_t + R(\nabla f, \nabla^2 f, \underline{Z}_t, h) =$$

$$= g\big(V_h(Z^1_t), V_h(Z^2_t), \ldots, V_h(Z^k_t)\big) + R(\nabla f, \nabla^2 f, \underline{Z}_t, h) \quad \text{(B10)}$$

Si può notare come la (B10) ha la forma dell'equazione (B3) a meno di un residuo $R(\cdot)$ da stimare che, come sarà dimostrato successivamente, è un o-piccolo di h^2. Prima di effettuare questa stima, verrà mostrata qualche applicazione della formula appena ricavata.

Somma di variabili (i totali rientrano in questa casistica):

$$f = Z_t^1 + Z_t^2 + \cdots + Z_t^k$$

$$\nabla f = \left(\frac{\partial f}{\partial Z_t^1}, \frac{\partial f}{\partial Z_t^2}, \ldots, \frac{\partial f}{\partial Z_t^k}\right) = (1,1,\ldots,1) \text{ vettore unitario}$$

$$\nabla^2 f = \begin{matrix} \frac{\partial f}{\partial^2 X_t^1} & \cdots & \frac{\partial f}{\partial X_t^1 \partial X_t^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial X_t^k \partial X_t^1} & \cdots & \frac{\partial f}{\partial^2 X_t^k} \end{matrix} = \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{matrix} \text{ matrice nulla}$$

$$V_h(Y_t) = \frac{1}{\sum_{i=1}^k Z_t^i} \sum_{i=1}^k V_h\left(Z_t^i\right) Z_t^i + \frac{1}{200\, f} < V_h\left(Z_t^i\right) Z_t^i, \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{matrix} \cdot V_h\left(Z_t^i\right) Z_t^i >$$

$$= \frac{1}{\sum_{i=1}^k Z_t^i} \sum_{i=1}^k V_h\left(Z_t^i\right) Z_t^i$$

Possiamo indicare:

$$\pi_t^i = \frac{Z_t^i}{\sum_{i=1}^k Z_t^i}$$

Considerando che stiamo calcolando la variazione al tempo $t + h$ rispetto al tempo t, il termine rappresenta il peso al tempo passato su cui si rapporta il valore presente, in questo modo otteniamo la relazione:

$$V_h(Y_t) = \sum_{i=1}^k V_h\left(Z_t^i\right) \pi_t^i$$

La variazione della somma di variabili (o di indici) è pari alla media ponderata delle variazioni degli addendi, la stessa formula di scomposizione per gli indici di variabili espresse come totali ottenuta con approccio *top-down* in (A5).

Facciamo un altro esempio, il prodotto di n indici (il prodotto tra due indici è ricavabile anche con approccio *top down*, con più di due indici risulta più difficile).

Prodotto di variabili (o dei numeri indici):

$$f = \prod_{i=1}^k I_t^l = I_t^1 I_t^2 \cdots I_t^k$$

$$\nabla f_i = \begin{cases} \frac{I_t^1 I_t^2 \cdots I_t^k}{I_t^i} & , se\ I_t^i \neq 0 \\ 0 & , se\ I_t^i = 0 \end{cases}$$

$$\nabla^2 f_{i,j} = \begin{cases} 0 \quad , & se\ i = j\ oppure\ I_t^i = 0\ oppure\ I_t^j = 0 \\ \frac{I_t^1 I_t^2 \cdots I_t^k}{I_t^i I_t^j} \quad , & i \neq j\ e\ I_t^i \neq 0\ e\ I_t^j \neq 0) \end{cases}$$

$$V_h(Y_t) = \frac{1}{\prod_{i=1}^k I_t^i} \sum_{i=1}^k \frac{I_t^1 I_t^2 \cdots I_t^k}{I_t^i} V_h(I_t^i) I_t^i + \frac{1}{200 \prod_{i=1}^k I_t^i} \sum_{\substack{i,j=1 \\ i \neq j}}^k \frac{I_t^1 I_t^2 \cdots I_t^k}{I_t^i I_t^j} V_h(I_t^i) I_t^i V_h(I_t^j) I_t^j$$

Che con le dovute semplificazioni diventa pari a:

$$V_h(Y_t) = \sum_{i=1}^k V_h(I_t^i) + \frac{1}{200} \sum_{\substack{i,j=1 \\ i \neq j}}^k V_h(I_t^i) V_h(I_t^j)$$

Ricaviamo la relazione nel caso in cui l'indice generale sia una media ponderata di sotto indici (ad esempio l'Indice di *Laspeyres*).

Media ponderata di indici:

f $= \sum_{i=1}^k I_t^i \pi_p^i$ dove π_p^i è un qualsiasi sistema di pesi $i = 1, \ldots k$ riferito al tempo p, che sommano a 1.

Se p è uguale all'anno base b l'indice di tipo *Laspeyres* rientra in questa casistica.

$$\nabla f = \left(\frac{\partial f}{\partial I_t^1}, \frac{\partial f}{\partial I_t^2}, \ldots, \frac{\partial f}{\partial I_t^k}\right) = (\pi_p^1, \pi_p^2, \ldots, \pi_p^k)$$

$$\nabla^2 f = \begin{matrix} \frac{\partial f}{\partial^2 I_t^1} & \cdots & \frac{\partial f}{\partial I_t^1 \partial I_t^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial I_t^k \partial I_t^1} & \cdots & \frac{\partial f}{\partial^2 I_t^k} \end{matrix} = \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix} \text{ matrice nulla}$$

$$V_h(Y_t) = \frac{1}{\sum_{i=1}^k I_t^i \pi_p^i} \sum_{i=1}^k V_h(I_t^i) I_t^i \pi_p^i + \frac{1}{200\, f} < V_h(I_t^i) I_t^i, \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \cdot V_h(I_t^i) I_t^i > =$$

$$= \sum_{i=1}^k V_h(I_t^i) \frac{I_t^i \pi_p^i}{\sum_{i=1}^k I_t^i \pi_p^i}$$

La relazione tra le variazioni è ancora una volta una media ponderata con pesi pari a:

$$\pi'_{i,p} = \frac{I_t^i \pi_p^i}{\sum_{i=1}^k I_t^i \pi_p^i} \quad \text{(B12)}$$

Il risultato di scomposizione ottenuto coincide con quello presente nel lavoro di Brunetti (2010).

Abbiamo visto quali sono le relazioni tra le variazioni una volta note le relazioni tra gli indici, sia per l'operazione di somma, sia per l'operazione di prodotto e infine per la media ponderata di indici.

Vogliamo utilizzare le formule precedenti per ricavare l'equazione (11), combinando opportunamente i risultati ottenuti. Va specificato che, la formula di scomposizione degli indici ricavata nella (10) non rappresenta una media ponderata di indici elementari bensì una media ponderata di prodotti di indici.

In questo caso le componenti della funzione *f*, sono gli *n* indici della variabile rapporto relativa ai sotto-settori e altrettanti *n* indici della variabile composizione relativa ai sotto-settori, pertanto $k = 2n$.

$$\underline{Z}_t = (I^X_{1,t}, I^X_{2,t}, \dots, I^X_{n,t}, I^{\pi_D}_{1,t}, I^{\pi_D}_{2,t}, \dots, I^{\pi_D}_{n,t}).$$

Abbiamo scritto per semplicità *i* al posto di s_{ie} in seguito indicheremo π^i_b al posto di $\pi^N_{i,b}$.

Pertanto avremo:

$$Y_t = I^X_{S,t} = f(\underline{Z}_t) = \sum_{i=1}^n Z^i_t\, Z^{n+i}_t\, \pi^i_b \;=\; \sum_{i=1}^n I^X_{i,t}\, I^{\pi_D}_{i,t}\, \pi^i_b$$

È possibile, combinandoli opportunamente, utilizzare in sequenza i risultati di scomposizione ottenuti per la media ponderata di indici e per il prodotto di indici.

Applicando il risultato ottenuto per la media ponderata di indici si ha:

$$V_h(Y_t) = V_h\left(\sum_{i=1}^n I^X_{i,t}\, I^{\pi_D}_{i,t}\, \pi^i_b\right) = \sum_{i=1}^n V_h\left(I^X_{i,t}\, I^{\pi_D}_{i,t}\right) \frac{I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_b}{\sum_{i=1}^k I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_b}$$

Semplificando il peso di ponderazione si ha:

$$\frac{I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_p}{\sum_{i=1}^k I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_p} = \frac{\dfrac{\frac{\frac{N_{i,t}}{D_{i,t}} \frac{D_{i,t}}{\sum_i D_{i,t}}}{\frac{N_{i,p}}{D_{i,p}} \frac{D_{i,p}}{\sum_i D_{i,p}}} \dfrac{N_{i,p}}{\sum_i N_{i,p}}}{}}{\sum_{i=1}^k I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_p} = \frac{\sum_i D_{i,p}}{\sum_i D_{i,t} \cdot \sum_i N_{i,p}} \frac{N_{i,t}}{\sum_{i=1}^k I^X_{i,t}\, I^{\pi_D}_{i,t} \pi^i_p} =$$

$$= \frac{\sum_i D_{i,p}}{\sum_i D_{i,t} \cdot \sum_i N_{i,p}} \frac{N_{i,t}}{\frac{\sum_i D_{i,p}}{\sum_i D_{i,t} \cdot \sum_i N_{i,p}} \sum_{i=1}^k N_{i,t}} = \frac{N_{i,t}}{\sum_{i=1}^k N_{i,t}} = \pi^i_t$$

La formula semplificata diventa:

$$V_h(Y_t) = \sum_{i=1}^n V_h\left(I^X_{i,t}\, I^{\pi_D}_{i.t}\right) \pi^i_t$$

Ora, applicando la formula di scomposizione nel caso di prodotto di numeri indice si ottiene la formula (11) già ottenuta con approccio *top down* nel paragrafo 2.3:

$$V_h(Y_t) = \sum_{i=1}^n V_h\left(I^X_{i,t}\, I^{\pi_D}_{i,t}\right) \pi^i_t \;= \sum_{i=1}^n \left(V_h\left(I^X_{i,t}\right) + V_h\left(I^{\pi_D}_{i,t}\right) + \frac{1}{200} 2 V_h\left(I^X_{i,t}\, I^{\pi_D}_{i,t}\right)\right) \pi^i_t \;=$$

$$= \sum_{i=1}^n V_h\left(I^X_{i,t}\right) \pi^i_t \; + \sum_{i=1}^n V_h\left(I^{\pi_D}_{i,t}\right) \pi^i_t \; + \frac{1}{100} \sum_{i=1}^n V_h\left(I^X_{i,t}\right) V_h\left(I^{\pi_D}_{i,t}\right) \pi^i_t$$

Per terminare calcoliamo una stima in termini di *o*-piccoli del residuo. Prendiamo la formula (B10) e prendiamo i soli termini del residuo.

Avremo :

$$\boldsymbol{R\left(\nabla f, \nabla^2 f, \underline{Z}_t, h\right) = -\frac{100}{f} \sum_i \frac{\partial f}{\partial Z_t^i}\, o_i\left(h^2\right) + \frac{100}{2f} h^2 \sum_{\substack{i,j \\ i \neq j}} \frac{\partial f}{\partial Z_t^i} \frac{\partial f}{\partial Z_t^j} \left(\frac{o_i(h)}{h}\right)\left(\frac{o_j(h)}{h}\right) + o\left(h^2\right)}$$

$$\boldsymbol{= -\frac{100}{f} \sum_i \frac{\partial f}{\partial Z_t^i}\, o_i\left(h^2\right) + \frac{100}{2f} \sum_{\substack{i,j \\ i \neq j}} \frac{\partial f}{\partial Z_t^i} \frac{\partial f}{\partial Z_t^j} o_i(h) o_j(h) + o\left(h^2\right)}$$

Ora come noto dalla definizione di *o*-piccolo, una trasformazione lineare applicata a un *o*-piccolo restituisce ancora un *o*-piccolo dello stesso ordine. Inoltre il prodotto tra due *o*-piccoli di ordine h restituisce un *o*-piccolo di ordine h^2. Per cui avremo:

$$\boldsymbol{R\left(\nabla f, \nabla^2 f, \underline{Z}_t, h\right) = -\frac{100}{f} \sum_i \frac{\partial f}{\partial Z_t^i}\, o_i\left(h^2\right) + \frac{100}{2f} \sum_{\substack{i,j \\ i \neq j}} \frac{\partial f}{\partial Z_t^i} \frac{\partial f}{\partial Z_t^j} o_i(h) o_j(h) + o\left(h^2\right) =}$$

$$\boldsymbol{= o\left(h^2\right) + o\left(h^2\right) + o\left(h^2\right) = o(h^2)}$$

Con tutte le proprietà che ne derivano.

Riferimenti bibliografici

Biffignandi, S. 1993. *Aspetti metodologici e interpretativi della tecnica shift-share*. Padova, Italia: CEDAM.

Brunetti, A. 2010. "The decomposition of the chained price index rate of change: generalization and interpretative effectiveness". *Rivista di statistica ufficiale/Review of official statistics*, N. 1/2010: 17-34. https://www.istat.it/it/files/2011/05/1_2010_02.pdf.

Eurostat. 2006. "Methodology of short-term business statistics. Interpretation and guidelines". *Methods and Nomenclatures*. Luxembourg: Office for Official Publications of the European Communities. https://ec.europa.eu/eurostat/documents/3859598/5895945/KS-BG-06-001-EN.PDF/609c47c3-220a-48b4-874f-d9c589bf302c?version=1.0.

Istituto Nazionale di Statistica - Istat. 2019. "La rilevazione trimestrale Oros su occupazione e costo del lavoro: indicatori e metodologie". *Letture Statistiche - Metodi*. Roma, Italia: Istat. https://www.istat.it/it/files/2019/03/La-rilevazione-trimestrale-oros.pdf.

Istituto Nazionale di Statistica - Istat. 2010. "Retribuzione pro capite nelle grandi imprese: effetti di composizione dell'occupazione. Anni 2001-2009". *Approfondimenti*. Roma, Italia: Istat. https://www.istat.it/wp-content/uploads/2011/01/testointegrale20100330.pdf.

Istituto Nazionale di Statistica - Istat. 2009. "Classificazione delle attività economiche Ateco 2007, derivata dalla Nace Rev. 2". *Metodi e Norme*, N. 40. Roma, Italia: Istat. https://www.istat.it/wp-content/uploads/2022/03/volume_integrale_ATECO2007.pdf.

Keller, R. 2009. "How shifting occupational composition has affected the real average wage". *Monthly Labor Review*, Volume 132, N. 6: 26-38. https://www.bls.gov/opub/mlr/2009/06/art2full.pdf.

Predetti, A. 2006. *I numeri indici. Teoria e pratica dei confronti temporali e spaziali*. Milano, Italia: Giuffrè editore.

Informazioni per le autrici e per gli autori

La collana è aperta alle autrici e agli autori dell'Istat e del Sistema statistico nazionale e ad altri studiosi che abbiano partecipato ad attività promosse dall'Istat, dal Sistan, da altri Enti di ricerca e dalle Università (convegni, seminari, gruppi di lavoro, etc.).

Coloro che desiderano pubblicare su questa collana devono sottoporre il proprio contributo al Comitato di redazione degli Istat working papers, inviandolo per posta elettronica all'indirizzo: iwp@istat.it.

Il saggio deve essere redatto seguendo gli standard editoriali previsti (disponibili sul sito dell'Istat), corredato di un sommario in Italiano e in Inglese e accompagnato da una dichiarazione di paternità dell'opera.

Per le autrici e gli autori dell'Istat, la sottomissione dei lavori deve essere accompagnata da un'e-mail della/del propria/o referente (Direttora/e, Responsabile di Servizio, etc.), che ne assicura la presa visione.

Per le autrici e gli autori degli altri Enti del Sistan la trasmissione avviene attraverso la/il responsabile dell'Ufficio di statistica, che ne prende visione. Per tutte le altre autrici e gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione.

Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del Chicago Manual of Style.

Attraverso il Comitato di redazione, tutti i lavori saranno sottoposti a un processo di valutazione doppio e anonimo che determinerà la significatività del lavoro per il progresso dell'attività statistica istituzionale.

La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line gratuitamente.

Gli articoli pubblicati impegnano esclusivamente le autrici e gli autori e le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.