

Il principio di conservazione della dinamica e le tecniche di riconciliazione di serie storiche nella stima dell'occupazione trimestrale per settore istituzionale*

Marianna Ascione, Giancarlo Lutero

Sommario

In questo lavoro è esposto il problema della riconciliazione di serie storiche nei conti economici nazionali, in riferimento alle stime trimestrali dell'occupazione per settore istituzionale. Dopo una panoramica sulla struttura dei conti istituzionali trimestrali, sono illustrate le fonti delle discrepanze che si possono originare quando nel processo di stima degli aggregati è necessario rispettare dei vincoli esogeni predeterminati. Una breve introduzione alle tecniche di bilanciamento di dati longitudinali più note è trattata unitamente alla metodologia di Denton, con particolare attenzione alla sua proprietà di conservazione della dinamica (movement preservation principle). Sono infine esaminate le tecniche di riconciliazione di serie storiche con discrepanze temporali e longitudinali, soffermandosi su metodologie a due passi le quali superano alcune difficoltà computazionali. Il lavoro è corredato da un esercizio empirico sulle Unità di lavoro equivalenti a tempo pieno, comparando due diversi approcci di riconciliazione a due passi.

Parole chiave: occupazione trimestrale per settore istituzionale, riconciliazione serie storiche, vincoli temporali e longitudinali, principio di conservazione della dinamica, bilanciamento a due stadi, unità di lavoro equivalenti.

Abstract

In this paper the methods of time series reconciliation in the presence of both contemporaneous and temporal aggregation constraints are discussed, relating to a set of Quarterly Sector Accounts (QSA) estimates. After a brief overview of Quarterly Sector Accounts, the sources of data discrepancies are stressed. A short introduction to the balancing and benchmarking methodologies adopted in operating context are treated. In addition, Denton's methodology is illustrated, remarking the movement preservation principle property. In order to solve computational impasses of a simultaneous solution, a two-steps reconciliation approach is introduced. An empirical exercise on full-time equivalent units is applied to compare several two-steps strategies.

Keywords: quarterly employment by institutional sector, time series reconciliation, temporal and contemporaneous constraints, movement preservation principle, two-steps balancing, full-time equivalent units.

* I risultati di questo lavoro sono stati presentati al workshop su Riconciliazione dei sistemi di serie storiche: sviluppi teorici e applicazioni in Contabilità Nazionale, svoltosi nell'Aula magna dell'ISTAT, il 23 giugno 2009, Roma. Questo contributo è stato realizzato grazie al lavoro congiunto di Marianna Ascione e Giancarlo Lutero. Le sezioni 1, 4, 5 sono state curate da Giancarlo Lutero, le sezioni 2, 3, 6 da Marianna Ascione.

1. Introduzione

In Italia le stime relative ai conti trimestrali nazionali seguono un approccio *top-down*, questo implica che i conti annuali nazionali sono il riferimento vincolante da cui si ricavano, con diverse tecniche ed approcci metodologici, tutte le altre stime prodotte dalla Contabilità Nazionale relativamente ai diversi sottosistemi (oltre ai conti trimestrali, i conti territoriali, i conti istituzionali, i conti ambientali). In altri contesti, la strategia nella produzione degli aggregati di Contabilità Nazionale può procedere in senso inverso, dal particolare al generale (approccio *bottom-up*): si stimano direttamente le variabili a un livello di analisi più dettagliato e, in seguito ad un processo più o meno complesso di aggregazione, si ottenengono le stime dei conti nazionali (pil, consumi, investimenti, risparmio, accumulazione, attività finanziarie, flussi del commercio estero, ecc.). I conti trimestrali per settore istituzionale costituiscono una rappresentazione integrata dei comportamenti dei diversi operatori del circuito economico nazionale, e sono un prototipo di dati soggetti ad una doppia restrizione contabile. Ciascun settore istituzionale è l'insieme delle singole unità istituzionali che possiedono delle caratteristiche economiche e dei comportamenti simili: i settori sono le famiglie, le imprese e la pubblica amministrazione. I conti per settore istituzionale consentono di evidenziare numerosi aggregati che non sono direttamente osservati in altri contesti come ad esempio il reddito disponibile lordo delle famiglie oppure la quota di profitti delle società non finanziarie. La stima degli aggregati annuali è condotta su una base informativa statistica assai diversificata; a livello trimestrale, invece, per integrare l'informazione diretta mancante, oppure troppo onerosa da stimare, si ricorre a dei metodi di calcolo indiretti basati su procedimenti di interpolazione dei dati (tecniche di disaggregazione temporale).¹

Nel processo di stima dei conti nazionali trimestrali, compresi quindi quelli per settore istituzionale, operando a livello di branca di attività economica, si originano delle discrepanze rispetto a delle grandezze esogene che assumono il ruolo di vincoli stringenti; di conseguenza per ogni trimestre è necessario tener conto di tali vincoli operanti come ad esempio le stime annuali per settore istituzionale o i conti trimestrali per il totale economia, ed ancora i conti trimestrali relativi alla pubblica amministrazione. Le restrizioni contabili possono essere di due tipi:

- *vincoli contemporanei*, in cui in ogni istante temporale, una grandezza deve essere la somma o la media di altre grandezze;
- *vincoli di aggregazione² temporale*, dove vi deve essere una consistenza fra una serie storica ad alta frequenza e la medesima formulata a frequenza temporale più bassa (che rappresenta il vincolo).

Tali disuguaglianze possono essere causate da diversi fattori: una differente metodologia di raccolta dei dati, l'uso di diverse tecniche di campionamento, oppure, in mancanza di rilevazione diretta o informazioni supplementari, da una cattiva specificazione delle stime ad un determinato livello di disaggregazione, o da trasformazioni funzionali come la pratica della destagionalizzazione, e in generale l'uso di modelli stocastici o previsivi nell'ambito delle tecniche di disaggregazione temporale.

¹ Per una introduzione si veda Bee Dagum e Cholette (2006).

² Il problema dell'aggregazione non concerne soltanto le serie temporali, ma altresì i dati longitudinali o spaziali, la costruzione degli indici compositi o il *combining forecasting*. È un fenomeno noto sin dagli albori delle scienze statistiche ed economiche applicate, ma è spesso sottovalutato nella pratica. Per una panoramica si veda Lutero (2010).

La necessità di rispettare un principio di coerenza formale fra le grandezze macroeconomiche prodotte nell'alveo dei conti nazionali, come l'eguaglianza fra risorse ed impieghi o la coerenza fra aggregati nazionali e settori istituzionali, conduce necessariamente all'adozione di metodi di bilanciamento dei dati. Con i termini *bilanciamento*, *benchmarking* e *riconciliazione* si indicano nella letteratura, in contesti diversi, le tecniche di ripristino della consistenza formale interna di un insieme di dati ordinati in un array a n dimensioni e, di fatto, i rispettivi totali marginali. Nello specifico per bilanciamento (*balancing*) si intende, in un dato istante temporale, il ripristino della consistenza fra un insieme di dati ed i suoi vincoli contemporanei. Invece le tecniche di *benchmarking* riguardano il ripristino della coerenza temporale fra la media o la somma di serie espresse ad alta frequenza (tipicamente trimestrale) e vincolo temporale a bassa frequenza (dati annuali). Infine i metodi di *riconciliazione* considerano entrambi i tipi di vincoli summenzionati e quindi concernono un sistema di dati più complesso che può causare, superata una certa soglia dimensionale, delle grosse difficoltà operative e computazionali. Le tecniche di riconciliazione a due passi possono ovviare a quest'ultimo inconveniente, semplificando il problema numerico senza alterare o preservando al meglio le caratteristiche dinamiche delle serie.

Lo scritto si articola come segue: la sezione 2 propone una panoramica introduttiva sui conti istituzionali così come si configurano nell'ambito delle stime relative ai conti nazionali; la sezione 3 chiarisce i motivi strutturali da cui si origina il problema della riconciliazione nei conti trimestrali istituzionali; la sezione 4 introduce il bilanciamento dei dati ed illustra sinteticamente alcune delle più note tecniche di riconciliazione di serie storiche, soffermandosi sulle caratteristiche specifiche afferenti la nota procedura di *benchmarking* di Denton; vengono inoltre definiti gli indici di discrepanza adottati per dare una misura del fenomeno della riconciliazione; la sezione 5 presenta un'applicazione, relativa alla stima trimestrale dell'occupazione per settore istituzionale, e commenta i risultati; infine la sezione 6 sintetizza il lavoro e dà indicazioni su possibili sviluppi futuri.

2. I Conti trimestrali per settore istituzionale

I conti nazionali per settore istituzionale costituiscono una rappresentazione integrata dei comportamenti dei diversi operatori economici. Essi illustrano le scelte operate dalle famiglie, dalle imprese e dalle amministrazioni pubbliche nelle fasi essenziali del processo economico: produzione, formazione, distribuzione, redistribuzione e utilizzazione del reddito, accumulazione non finanziaria.

Gli operatori economici sono visti come centri di decisione nel campo economico e finanziario. Essi costituiscono le unità istituzionali che sono raggruppate nei settori sulla base dell'uniformità di comportamento nell'esercizio delle loro funzioni principali. Sulla base delle funzioni prevalentemente svolte e del tipo di risorse utilizzate per il loro finanziamento (ricavi delle vendite, redditi da lavoro o da capitale-impresa, interessi bancari, premi di assicurazione, imposte e contributi sociali, quote associative) le unità sono raggruppate nei seguenti settori istituzionali:

- le *famiglie* che comprendono le famiglie consumatrici, le famiglie produttrici (imprese individuali, società semplici fino a cinque addetti e liberi professionisti) e le istituzioni senza scopo di lucro al servizio delle famiglie;

- le *società non finanziarie* che comprendono tutte le società di persone e di capitale e le imprese individuali con oltre cinque addetti che svolgono la loro attività nei settori diversi da quelli finanziari;
- le *società finanziarie* che includono le Autorità bancarie centrali, le altre istituzioni finanziarie monetarie, gli altri intermediari finanziari, le imprese di assicurazione e fondi pensione e gli ausiliari finanziari;
- le *amministrazioni pubbliche* che producono prevalentemente servizi non destinabili alla vendita (non market), redistribuiscono il reddito e la ricchezza fra gli operatori del sistema, contribuiscono al consumo finale ed al processo di accumulazione. Esse includono le amministrazioni centrali e locali, le amministrazioni di Stati federati ed infine gli enti di previdenza ed assistenza sociale.

A questi settori va aggiunto il settore definito *resto del mondo* che comprende le unità non residenti che effettuano operazioni con le unità istituzionali residenti classificate nei settori sopra elencati.

L'analisi dal punto di vista dei settori costituisce uno strumento utile all'interpretazione della realtà economica del Paese ed è pertanto di fondamentale importanza per l'analisi economica. Nella pratica sono stimati numerosi e differenziati aggregati economici che consentono di calcolare alcuni importanti saldi contabili non direttamente osservabili come il reddito disponibile ed il risparmio. La notevole mole di variabili stimate per settore istituzionale consente la definizione dell'intera sequenza dei conti utile a descrivere i comportamenti degli operatori economici nella formazione del prodotto, nella remunerazione dei fattori produttivi utilizzati, nell'impiego delle risorse da destinare a consumi e investimenti, nella formazione del risparmio e delle attività finanziarie. In generale le procedure di stima adottate definiscono gli aggregati dalla produzione fino al risultato di gestione per branca di attività economica e le operazioni rientranti nei circuiti successivi direttamente a livello di settore.

I dati sono elaborati a cadenza annuale e trimestrale. Soltanto i conti annuali per settore istituzionale sono diffusi nella loro interezza. A livello trimestrale si diffondono alcuni indicatori³ sui risultati ottenuti dai soli settori delle società non finanziarie e delle famiglie. Gli indicatori diffusi sono corretti per la componente stagionale e le serie trimestrali che li definiscono sono diffuse sia grezze che destagionalizzate. A partire dal 2006 in conformità a quanto stabilito nel Regolamento Ue n° 1161/2005, l'intero set di dati trimestrali è trasmesso all'Eurostat a t+90 giorni dal trimestre di riferimento contribuendo alla definizione dei conti economici trimestrali per settore istituzionale dell'Unione Europea (EU27) e dell'Unione Monetaria (EA15), rilasciati a t+120 giorni dal trimestre di riferimento dall'Eurostat e dalla Banca Centrale Europea.⁴

Pur basandosi sui medesimi concetti e definizioni dei corrispondenti conti annuali, i conti trimestrali per settore istituzionale sono stimati ricorrendo a metodi statistici per la limitatezza delle informazioni disponibili a cadenza trimestrale.⁵ La stima degli aggregati

³ Le serie trimestrali relative agli indicatori pubblicati per le società non finanziarie e per le famiglie sono disponibili all'indirizzo: http://www.istat.it/salastampa/comunicati/in_calendario/redisp/20101008_00/

⁴ Le serie trimestrali relative agli indicatori pubblicati dall'Eurostat sono disponibili al seguente indirizzo: http://epp.eurostat.ec.europa.eu/portal/page/portal/sector_accounts/data/quarterly_data/

⁵ Per la metodologia seguita per la stima degli aggregati trimestrali si rimanda all'inventario: http://epp.eurostat.ec.europa.eu/portal/page/portal/sector_accounts/methodology/from_national_to_european_accounts

annuali, infatti, si basa su diverse fonti statistiche che si rendono disponibili generalmente con una frequenza almeno pari all'anno; a livello trimestrale, invece, per integrare l'informazione mancante si ricorre ad appropriate tecniche statistiche definite tecniche di disaggregazione temporale.

Il metodo cui si ricorre per la stima trimestrale di un aggregato che concorre alla definizione dei Conti trimestrali per settore istituzionale è definito approccio indiretto e rappresenta un metodo di aggiustamento a 2 stadi in cui il dato annuale è trimestralizzato attraverso un indicatore disponibile ad alta frequenza. Una procedura di benchmarking è quindi già utilizzata nel processo di produzione dei dati trimestrali.

Per la produzione delle serie destagionalizzate, infine, si esegue generalmente l'identificazione, la stima e la rimozione della componente stagionale direttamente sugli indicatori di riferimento. Tuttavia ci sono degli aggregati per i quali le tecniche di destagionalizzazione si applicano direttamente all'aggregato trimestrale grezzo ovvero a quello ottenuto mediante la disaggregazione dei dati annuali con l'indicatore grezzo. In ogni caso la destagionalizzazione non assicura la consistenza dei dati destagionalizzati con il vincolo annuo.

Il ricorso alle tecniche di riconciliazione di seguito descritte garantisce la coerenza interna al complesso sistema di serie fin qui descritto.

3. Il problema della riconciliazione nei conti trimestrali istituzionali

Le rappresentazioni contabili, generalmente, necessitano sia di procedure di bilanciamento sia di *benchmarking*. Le procedure di bilanciamento rimuovono la discrepanza tra le variabili, quelle di *benchmarking* si occupano dell'aggiustamento interno alla dimensione temporale.⁶ Nell'ambito dei QSA (Quarterly Sector Accounts) nessun settore istituzionale è derivato a residuo, pertanto essi vanno riconciliati per rimuovere ogni possibile discrepanza nei dati.

Una procedura di riconciliazione automatica si rende necessaria in quanto ad ogni trimestre occorre tener conto dei seguenti vincoli:

- ASA (Annual Sectors Accounts):⁷ sono i conti annuali per settore istituzionale rilasciati nel mese di luglio dell'anno t ;
- QNA (Quarterly National Accounts):⁸ sono i conti trimestrali per il totale economia rilasciati a $t+70$ giorni dal trimestre di riferimento;
- QGGA (Quarterly General Government Accounts):⁹ sono i conti trimestrali per la pubblica amministrazione rilasciati a $t+90$ giorni dal trimestre di riferimento. I dati di questo settore sono integrati nei QSA e, oltre ad essere trasmessi ad Eurostat, sono diffusi a livello nazionale.

⁶ Si veda Bee Dagum e Cholette (2006), Di Fonzo (2003).

⁷ Le serie annuali degli ASA sono disponibili all'indirizzo:
http://www.istat.it/salastampa/comunicati/non_calendario/20100722_00/

⁸ Le serie trimestrali dei QNA sono disponibili all'indirizzo:
http://www.istat.it/salastampa/comunicati/in_calendario/contitri/20100910_00/

⁹ Le serie trimestrali della PA sono disponibili all'indirizzo:
http://www.istat.it/salastampa/comunicati/in_calendario/contoPA/20110110_00/

Con riferimento ai tre trimestri infrannuali, per ciascuna variabile è necessario assicurare che la somma dei settori istituzionali interni sia uguale al totale economia; mentre con riferimento al quarto trimestre oltre ad assicurare la somma dei settori istituzionali interni uguale al totale economia, per ciascuna variabile e per ciascun settore istituzionale va garantita la somma delle stime relative ai quattro trimestri che deve coincidere con le stime annuali.

Operativamente è come se si partisse da una matrice di dati preliminari ad alta frequenza da riconciliare con i *vincoli temporali* ovvero le stime annuali delle variabili di interesse (ASA) e con la serie ad alta frequenza che costituisce il *vincolo contemporaneo* (QNA) che deve essere garantito in ciascun periodo osservato.

Volendo schematizzare il problema attraverso una rappresentazione tabellare (Tav. 1): per un generico aggregato economico Y e per una generica branca di attività economica vale quanto segue ovvero che per ciascun istante temporale (il trimestre) deve valere che la somma dei settori società non finanziarie (SNF), società finanziarie (SFIN), famiglie (FT) e pubblica amministrazione (PA) deve coincidere con quanto stimato per il totale economia (TE).

Il generico elemento Y_{S,T,q_i} della tabella rappresenta la stima del trimestre q_i dell'anno T per il settore istituzionale S di un generico aggregato economico e per ciascun trimestre la somma della stima trimestrale di tutti i settori deve coincidere con la stima trimestrale Z_{TE,T,q_i} relativa al totale economia; inoltre per ciascun settore in corrispondenza del rilascio del quarto trimestre va garantita la coerenza con la stima annuale del settore $Y_{S,T}$ fornita dagli ASA.

La tavola 1 evidenzia, inoltre, che per i trimestri infrannuali q_1, q_2, q_3 quando non è disponibile il vincolo annuale ASA e si procede in estrapolazione si tiene conto soltanto del vincolo contemporaneo costituito dal totale economia fornito dai QNA.

Tavola 1 - Schema del problema della riconciliazione nei QSA

ANNO	TRIMESTRE	SNF	SFIN	FT	PA	TE
1	Q1	$Y_{SNF,1,q_1}$	$Y_{SFIN,1,q_1}$	$Y_{FT,1,q_1}$	$Y_{PA,1,q_1}$	$Z_{TE,1,q_1}$
	Q2	$Y_{SNF,1,q_2}$	$Y_{SFIN,1,q_2}$	$Y_{FT,1,q_2}$	$Y_{PA,1,q_2}$	$Z_{TE,1,q_2}$
	Q3	$Y_{SNF,1,q_3}$	$Y_{SFIN,1,q_3}$	$Y_{FT,1,q_3}$	$Y_{PA,1,q_3}$	$Z_{TE,1,q_3}$
	Q4	$Y_{SNF,1,q_4}$	$Y_{SFIN,1,q_4}$	$Y_{FT,1,q_4}$	$Y_{PA,1,q_4}$	$Z_{TE,1,q_4}$
	TOT	$Y_{SNF,1}$	$Y_{SFIN,1}$	$Y_{FT,1}$	$Y_{PA,1}$	$Z_{TE,1}$
2	Q1	$Y_{SNF,2,q_1}$	$Y_{SFIN,2,q_1}$	$Y_{FT,2,q_1}$	$Y_{PA,2,q_1}$	$Z_{TE,2,q_1}$
	Q2	$Y_{SNF,2,q_2}$	$Y_{SFIN,2,q_2}$	$Y_{FT,2,q_2}$	$Y_{PA,2,q_2}$	$Z_{TE,2,q_2}$
	Q3	$Y_{SNF,2,q_3}$	$Y_{SFIN,2,q_3}$	$Y_{FT,2,q_3}$	$Y_{PA,2,q_3}$	$Z_{TE,2,q_3}$
	Q4	$Y_{SNF,2,q_4}$	$Y_{SFIN,2,q_4}$	$Y_{FT,2,q_4}$	$Y_{PA,2,q_4}$	$Z_{TE,2,q_4}$
	TOT	$Y_{SNF,2}$	$Y_{SFIN,2}$	$Y_{FT,2}$	$Y_{PA,2}$	$Z_{TE,2}$
...	
T-1	Q1	$Y_{SNF,T-1,q_1}$	$Y_{SFIN,T-1,q_1}$	$Y_{FT,T-1,q_1}$	$Y_{PA,T-1,q_1}$	$Z_{TE,T-1,q_1}$
	Q2	$Y_{SNF,T-1,q_2}$	$Y_{SFIN,T-1,q_2}$	$Y_{FT,T-1,q_2}$	$Y_{PA,T-1,q_2}$	$Z_{TE,T-1,q_2}$
	Q3	$Y_{SNF,T-1,q_3}$	$Y_{SFIN,T-1,q_3}$	$Y_{FT,T-1,q_3}$	$Y_{PA,T-1,q_3}$	$Z_{TE,T-1,q_3}$
	Q4	$Y_{SNF,T-1,q_4}$	$Y_{SFIN,T-1,q_4}$	$Y_{FT,T-1,q_4}$	$Y_{PA,T-1,q_4}$	$Z_{TE,T-1,q_4}$
	TOT	$Y_{SNF,T-1}$	$Y_{SFIN,T-1}$	$Y_{FT,T-1}$	$Y_{PA,T-1}$	$Z_{TE,T-1}$
T	Q1	Y_{SNF,T,q_1}	Y_{SFIN,T,q_1}	Y_{FT,T,q_1}	Y_{PA,T,q_1}	Z_{TE,T,q_1}
	Q2	Y_{SNF,T,q_2}	Y_{SFIN,T,q_2}	Y_{FT,T,q_2}	Y_{PA,T,q_2}	Z_{TE,T,q_2}
	Q3	Y_{SNF,T,q_3}	Y_{SFIN,T,q_3}	Y_{FT,T,q_3}	Y_{PA,T,q_3}	Z_{TE,T,q_3}
	Q4	Y_{SNF,T,q_4}	Y_{SFIN,T,q_4}	Y_{FT,T,q_4}	Y_{PA,T,q_4}	Z_{TE,T,q_4}
	TOT	$Y_{SNF,T}$	$Y_{SFIN,T}$	$Y_{FT,T}$	$Y_{PA,T}$	$Z_{TE,T}$
	Q1	$Y_{SNF,T+1,q_1}$	$Y_{SFIN,T+1,q_1}$	$Y_{FT,T+1,q_1}$	$Y_{PA,T+1,q_1}$	$Z_{TE,T+1,q_1}$
	Q2	$Y_{SNF,T+1,q_2}$	$Y_{SFIN,T+1,q_2}$	$Y_{FT,T+1,q_2}$	$Y_{PA,T+1,q_2}$	$Z_{TE,T+1,q_2}$
	Q3	$Y_{SNF,T+1,q_3}$	$Y_{SFIN,T+1,q_3}$	$Y_{FT,T+1,q_3}$	$Y_{PA,T+1,q_3}$	$Z_{TE,T+1,q_3}$

4. Le metodologie di riconciliazione

4.1 Introduzione: il bilanciamento di matrici di dati

Il bilanciamento dei conti nazionali è una pratica diffusa ed applicata negli uffici statistici nazionali da diversi anni, a partire dai lavori pionieristici di Stone.¹⁰ Le procedure di bilanciamento dei dati devono risolvere un problema che può essere espresso nel modo seguente: data una matrice di grandezze P rettangolare o quadrata occorre trovare un'altra matrice finale Y , delle medesime dimensioni, in cui siano rispettate delle restrizioni lineari sui valori delle matrici stesse. Tipicamente le restrizioni implicano che le sommatorie per riga e/o per colonna debbano necessariamente eguagliare dei valori marginali, così ad esempio per ogni riga i potrebbe valere la seguente restrizione:

$$\sum_j y_{ij} = z_i \quad \forall i=1,2,\dots,r \quad (1)$$

con il vincolo espresso da un totale marginale di riga z_i ; al contempo, il medesimo discorso è valido per dei possibili vincoli di colonna: pertanto la necessità di una procedura di bilanciamento si pone quando l'espressione (1) non è rispettata anche per una sola riga i .

Vi sono diverse metodologie per risolvere il problema in questione; esse possono essere classificate in maniera molto sintetica in due modi:

- tecniche di *scaling*, che adottano un bilanciamento biproporzionale costante sino ad esaurimento¹¹ delle discrepanze; uno dei metodi più noti è il RAS;
- algoritmi di ottimizzazione che si fondano sulla minimizzazione di funzioni obiettivo vincolate, sulla base di una metrica che misura la distanza fra la matrice di dati originaria e quella finale bilanciata: il criterio più utilizzato è quello dei minimi quadrati e l'approccio più noto è quello di Stone.

Il criterio RAS è una metodologia di riproporzionamento iterativo che sinteticamente può essere descritto nei seguenti passi:

- i dati di input sono la matrice P ($i \times j$) dei dati da riproporzionare ed i vincoli di riga e di colonna z_i e z_j ;
- passo zero: si pone $s = 0$, la matrice P è quindi uguale alla matrice da stimare al passo zero Y^0 ;
- passo 1: il riproporzionamento si effettua dapprima per riga, calcolando per ogni riga i un peso che misura la discrepanza fra ogni singolo elemento della matrice ed

il valore che rappresenta il vincolo di riga pari a $\lambda_i^s = \frac{z_i}{\sum_j y_{ij}^s}$, il quale si utilizza

per aggiornare i valori di ogni singola elemento dell'array ponendo

$$y_{ij}^s = \lambda_i^s y_{ij}^s \quad \forall i, j;$$

¹⁰ Si vedano Stone et al (1942) e Stone (1961).

¹¹ In realtà viene fissato un valore soglia, molto prossimo allo zero, sotto cui il processo iterativo si arresta.

- passo 2: si effettua il riproporzionamento per colonna, stimando il peso

$$\mu_j^s = \frac{z_j}{\sum_i y_{ij}^s}$$

sulla base del vincolo di colonna, con cui si definisce la matrice Y^{s+1} ,

in cui ogni cella è pari a $y_{ij}^{s+1} = \mu_j^s y_{ij}^s \quad \forall i, j$;

- si pone $s = s+1$ e si riparte dal passo 1, sino a quando le discrepanze (le somme per riga e per colonna che non coincidono con i vincoli) sono ridotte ad un valore prossimo allo zero.

La tecnica RAS è relativamente semplice da computare ed attivare in un processo operativo, ma possiede degli inconvenienti che ne limitano l'uso. Tali limiti in sintesi sono:

- la tecnica funziona anche con valori negativi, i quali spesso hanno scarso significato con grandezze economiche o statistiche;
- le restrizioni possono essere espresse come semplici sommatorie e non come una qualsiasi combinazione lineare, fattore che implica la stima di un sistema di pesi;
- ad ogni valore della matrice P è assegnato il medesimo grado di affidabilità prima del bilanciamento.

Nonostante tali limitazioni la metodologia è molto diffusa in quanto molto intuitiva e, come detto, relativamente semplice da introdurre in una procedura automatizzata.

La tecnica di Stone rientra fra le procedure di bilanciamento di dati che utilizzano i modelli di regressione. Il problema può essere formalizzato nel seguente modo: dati una matrice di dati preliminari vettorizzata p , una matrice di varianza-covarianza delle discrepanze V , ipotizzata diagonale:¹²

$$V = \begin{bmatrix} \sigma_{p1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{p2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pn}^2 \end{bmatrix}$$

un vettore incognita di dati bilanciati y , una matrice H di aggregazione,¹³ un vettore di vincoli z , e un vettore di moltiplicatori di Lagrange λ , la funzione lagrangiana da minimizzare è la seguente:

$$L = (y - p)'V(y - p) + 2\lambda(Hy - z) \quad (2)$$

¹² Quindi per costruzione è ipotizzata eteroschedasticità nei dati, pur mantenendo l'ipotesi di assenza di correlazione longitudinale.

¹³ La matrice H di aggregazione assume una conformazione diversa a seconda se i dati sono flussi o stocks.

Lo stimatore BLUE di Stone (ovvero quello non distorto a varianza minima fra quelli della classe degli stimatori lineari) assume la seguente forma:

$$y = p - VH'(H'VH')^{-1}(Hp - z) \quad (3)$$

da cui si può vedere che i dati finali bilanciati sono uguali alle serie originarie corrette tramite una combinazione lineare del prodotto delle discrepanze ($Hp - z$) per una matrice di pesi. Moltiplicando ambo i membri della (3) per la matrice di aggregazione H si otterrà un'identità a posteriori $Hy = z$, relativamente ai vincoli. Questo stimatore può essere calcolato direttamente quando le dimensioni della matrice inversa di $H'VH'$ sono contenute oppure quando essa è una matrice densa, altrimenti è necessario ricorrere a tecniche iterative, come ad esempio quella adottata per il bilanciamento dei conti annuali italiani, basata sul metodo del gradiente coniugato.¹⁴

La matrice di varianza-covarianza V è di fondamentale importanza per individuare la grandezza e la direzione degli aggiustamenti: essa misura l'accuratezza dei valori preliminari rispetto ai "veri" valori bilanciati; va ricordato però che gli elementi di questa matrice non sono noti, di conseguenza Stone ha ipotizzato che essi siano posti pari al rispettivo valore preliminare $\sigma_{pi}^2 = p_i$. Inoltre si può operare una valutazione soggettiva "a priori" sull'attendibilità, e quindi sul grado di fiducia, che si ripone nel singolo vettore dei dati preliminari, modificando il valore di p_i con un fattore di correzione θ_i , quindi si pone:

$$\sigma_{pi}^2 = \theta_i p_i \quad 0 < \theta_i \leq 1 \quad \forall i$$

a cui si può attribuire, convenzionalmente, un valore compreso fra zero ed uno in modo da dargli un'interpretazione probabilistica, ma ciò non è obbligatorio, ovvero si possono assegnare anche valori superiori ad uno senza inficiare la stima. Come si vedrà nella sezione 4.3, l'opportunità di attribuire un valore soggettivo al peso relativo dei dati preliminari si prospetta addirittura come necessaria nel caso dei conti trimestrali per settore istituzionale.

4.2 Il benchmarking: la metodologia di Denton

La procedura di di Denton è una tecnica di interpolazione dei dati che può essere utilizzata in diversi contesti. I più noti sono i seguenti:

- nel contesto delle metodologie di dissaggregazione temporale, che consentono la produzione di stime infrannuali (mensili o trimestrali), data una collezione di stime a frequenza temporale trimestrale o annuale; in questa sede supporremo che la relazione sia fra dati trimestrali ed annuali;
- nell'ambito di procedure di *benchmarking* o più in generale nelle tecniche di bilanciamento che contemplino la presenza di serie temporali.

Essa si basa sulla minimizzazione quadratica della differenza fra una serie preliminare ed una serie aggiustata.¹⁵ Tale differenza può essere additiva (sui livelli) o proporzionale

¹⁴ Si vedano Byron (1978) e Nicolardi (2000).

¹⁵ Per approfondimenti si vedano Denton (1971) e Bee Dagum e Cholette (2006).

(sugli aggiustamenti relativi): nella tavola 2 è riportata una tassonomia delle principali funzioni di perdita che si possono associare al metodo di Denton. Generalizzando il problema, occorre minimizzare la seguente funzione lagrangiana:

$$L = (y - p)'M(y - p) + 2\lambda(Hy - z)$$

con y serie finale, p serie preliminare, H è la consueta matrice di aggregazione, z il vettore dei vincoli stringenti, λ consueto vettore dei moltiplicatori di Lagrange, ed M è un operatore lineare simmetrico non singolare cui accenneremo in seguito.

Tavola 2 - Tipologia funzioni di perdita associate con il metodo di Denton

TIPOLOGIA FUNZIONE DI PERDITA	Formula
Naïve	$(y - p)$
Additive first difference (AFD)	$\Delta(y - p)$
Additive second difference (ASD)	$\Delta^2(y - p)$
Additive third difference (ATD)	$\Delta^3(y - p)$
Pro-rata	$\left(\frac{y - p}{p} \right)$
Proportional first difference (PFD)	$\Delta \left(\frac{y - p}{p} \right)$
Proportional second difference (PSD)	$\Delta^2 \left(\frac{y - p}{p} \right)$
Proportional third difference (PTD)	$\Delta^3 \left(\frac{y - p}{p} \right)$

Brevemente le condizioni del primo ordine sono:

$$\begin{cases} \frac{\partial L}{\partial y} = 2My - 2Mp + 2\lambda H = 0 \\ \frac{\partial L}{\partial \lambda} = Hy - z = 0 \end{cases}$$

e svolgendo gli opportuni passaggi algebrici si ottiene lo stimatore di Denton pari a:

$$y = p + MH'(HMH')^{-1}(z - Hp) \tag{4}$$

La matrice M può assumere diverse configurazioni a seconda della caratterizzazione della funzione di perdita definita nella funzione obiettivo e delle ipotesi assunte sul primo valore del periodo campionario.¹⁶ Ad esempio l'utilizzo di un *benchmarking* semplice, naïf, può comportare la presenza di un salto nelle serie, uno scalino fra il quarto trimestre di un anno ed il primo dell'anno successivo, a causa di una differenza sensibile fra valori annuali contigui: ciò accade quando la matrice M è definita come una matrice identità. L'introduzione di un operatore differenza a vari ordini interi, o conformazioni proporzionali della funzione di perdita, minimizzano o eliminano questo problema del salto di serie: ad esempio l'uso di una differenza prima $\Delta = 1 - L$, con L operatore ritardo (lag) tale che $\Delta y_t = y_t - y_{t-1}$, implica che la funzione di perdita additiva (AFD) sarà la seguente:

$$\kappa(p, y) = \sum_{t=1}^T (\Delta y_t - \Delta p_t)^2 = \sum_{t=1}^T [\Delta(y_t - p_t)]^2$$

In termini matriciali la trasformazione si esprime ponendo $M = D'D$, dove l'operatore differenza è così strutturato:

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Le procedure di *benchmarking* non possono non tener conto della struttura temporale dei dati in serie storica, intesi questi ultimi come realizzazione finita di un vettore aleatorio a componenti infinite. Un intervento combinatorio su un processo stocastico distorce, anzi inficia completamente le caratteristiche dinamiche intrinseche di tale processo. Il metodo di Denton, in special modo nella sua variante PFD, che agisce sui tassi di crescita, è caratterizzato dalla proprietà di preservare la dinamica, ovvero la serie finale y deve riprodurre, o al più salvaguardare nel miglior modo possibile, il movimento ed il segno della serie originale p o della sua variazione. Il *movement preservation principle* è una delle caratteristiche più importanti della procedura di Denton e proprietà numerica fondamentale nelle tecniche di riconciliazione di dati.

4.3 La riconciliazione di sistemi di serie storiche: soluzione simultanea ed approccio a due passi

La riconciliazione di serie storiche è un ambito di studi teorici che sintetizza entrambe le metodologie discusse nelle precedenti sezioni. L'esigenza di riportare ad una condizione di coerenza interna un insieme di dati, nasce dalla pratica degli uffici nazionali nella produzione delle statistiche ufficiali di un paese, che producono e rilasciano un sempre maggior numero di panel di serie storiche classificati in relazione ad

¹⁶ Si rimanda sempre a Bee Dagum e Cholette (2006), cap. 6.

un attributo merceologico, territoriale o un criterio funzionale, come ad esempio i settori istituzionali. Nel processo di stima dei dati intervengono delle condizioni dalle quali si originano delle discrepanze rispetto a delle stime predeterminate esogene, le quali assumono il rango di vincolo inviolabile, causate o dal dettaglio di disaggregazione longitudinale a cui i dati sono compilati, oppure determinate da trasformazioni funzionali quali ad esempio:

- una cattiva specificazione delle stime ad un definito livello di disaggregazione;¹⁷
- la pratica della destagionalizzazione, utilizzata per rimuovere la componente stagionale nei dati congiunturali;¹⁸
- l'applicazione diretta di una procedura di *benchmarking*, per bilanciare serie temporali;
- un processo di interpolazione o di disaggregazione temporale, prassi utilizzata nella stima dei conti nazionali trimestrali quando si adotta una procedura indiretta, in assenza di dati direttamente disponibili;
- l'uso di modelli previsivi in totale assenza di una base informativa relativa alla variabile oggetto di stima.

Come affermato nella sezione 3, la stima trimestrale dei conti istituzionali è un ottimo paradigma della situazione ivi illustrata; alcuni aggregati sono prodotti a 32 branche di attività economica per ogni settore istituzionale e tutti e cinque i punti summenzionati possono intervenire nella filiera produttiva delle stime di tali conti, i quali necessariamente devono essere ricondotti in ogni trimestre sia ai valori delle stime dei conti trimestrali (per grandezze come ad esempio l'occupazione o i redditi), sia alle stime annuali per settore istituzionale, relativamente alle stesse variabili. Di conseguenza sono attivi due vincoli mentre, per i trimestri stimati in estrapolazione relativamente all'anno in corso non disponibile, è attivo il solo vincolo trimestrale. Le specifiche difficoltà nascono dal fatto che i dati possiedono una dimensione temporale ed una spaziale-sezionale, le quali richiedono un trattamento *ad hoc* per preservare al meglio la struttura informativa degli stessi.

I fattori che influenzano l'uso di un approccio simultaneo nella risoluzione di un sistema di serie storiche sono il numero delle serie e delle variabili (branche o distretti territoriali o settori istituzionali ad esempio), l'ampiezza del periodo campionario ed infine il numero dei vincoli operanti. Formalizziamo la questione considerando:

- m serie storiche da riconciliare;
- k unità longitudinali;
- n serie ad alta frequenza;
- N serie a bassa frequenza;
- s ordine di aggregazione temporale.

¹⁷ Qui entra in gioco la questione dell'aggregation bias, introdotta dall'aggregazione di stime risultanti da micromodelli stocastici contrapposta all'aggregation gain (cfr. Barker e Pesaran, 1990).

¹⁸ Questa prassi è giustificabile, anche se dipende dagli obiettivi della stima e dai fruitori dei dati: i cosiddetti dati grezzi possono essere più informativi in alcuni contesti, in particolare da quando la teoria dei sistemi dinamici, il paradigma della complessità e del caos hanno aperto la disciplina statistica ed econometrica a nuovi sviluppi. Il trattamento e la stima di dati possono essere condotti con modelli parametrici non lineari la cui generalizzazione è rappresentata dallo sviluppo in serie di Volterra (cfr. Granger e Teräsvirta, 1993).

Per stimare in modalità simultanea un sistema di serie storiche occorre risolvere un problema multivariato di minimizzazione quadratica vincolata come il seguente:

$$\begin{aligned} & \arg \min (R-P)' \Omega (R-P) \\ & \text{sub } HR = Y_a \end{aligned} \quad (5)$$

dove con R denotiamo le serie finali riconciliate (vettore di dimensione $nm \times I$), con P ($nm \times I$) le serie preliminari da sottoporre a bilanciamento, H (dimensione $kn+Nm \times nm$) è una matrice di aggregazione partizionata che contiene tutti i vincoli ed infine Y_a (dimensione $kn+Nm \times I$) è la variabile che esprime i vincoli opportunamente vettorizzati.¹⁹ La matrice di varianza-covarianza Ω , è espressione di una differenza prima proporzionale fra dato riconciliato e valore preliminare (*Proportional First Differences*, PFD):

$$\sum_{j=1}^M \sum_{t=2}^n \left(\frac{R_{j,t} - P_{j,t}}{P_{j,t}} - \frac{R_{j,t-1} - P_{j,t-1}}{P_{j,t-1}} \right)^2 = \sum_{j=1}^M \sum_{t=2}^n \left(\frac{R_{j,t}}{P_{j,t}} - \frac{R_{j,t-1}}{P_{j,t-1}} \right)^2$$

che, come già accennato, assicura la conservazione della dinamica, ed assume la seguente formulazione:

$$\Omega = \hat{P}^{-1} (I_m \otimes \Delta'_n \Delta_n) \hat{P}^{-1}$$

la quale a sua volta risulta come prodotto tensoriale della matrice identità di dimensione I_m , per la matrice Δ_n (di dimensione $(n-1) \times n$) operatore differenza prima,²⁰ che possiede la seguente forma:

$$\Delta_n = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

e dell'inversa della matrice $\hat{P} = \text{diag}(P)$.

Le condizioni del primo ordine dell'intero sistema sono:

$$\begin{bmatrix} \Omega & H' \\ H & 0 \end{bmatrix} \begin{bmatrix} R \\ \lambda \end{bmatrix} = \begin{bmatrix} \Omega P \\ Y_a \end{bmatrix}$$

¹⁹ Per i dettagli si vedano Bee Dagum e Cholette (2006) e Di Fonzo e Marini (2009).

²⁰ Tale matrice può assumere altre configurazioni, per approfondimenti si veda sempre Bee Dagum e Cholette (2006), cap.6.

La soluzione simultanea conduce al consueto stimatore che è espresso dalla seguente formula:

$$R = P + \Omega^{-1}H'(H\Omega^{-1}H')^{-1} (Y_a - HP) \quad (6)$$

La matrice Ω , ponendo $I_m=1$, ed ipotizzando frequenza trimestrale dei dati ($s = 4$), assume la seguente forma:

$$\Omega = \begin{bmatrix} \frac{1}{p_1^2} & -\frac{1}{p_1 p_2} & 0 & 0 \\ -\frac{1}{p_2 p_1} & \frac{2}{p_2^2} & -\frac{1}{p_2 p_3} & 0 \\ 0 & -\frac{1}{p_3 p_2} & \frac{2}{p_3^2} & -\frac{1}{p_3 p_4} \\ 0 & 0 & -\frac{1}{p_4 p_3} & \frac{1}{p_4^2} \end{bmatrix}$$

Lo stimatore possiede la nota struttura di correzione del dato preliminare P attraverso la distribuzione di una combinazione lineare delle discrepanze, con $(H\Omega^{-1}H')^{-1} = \Lambda$ matrice inversa generalizzata di Moore-Penrose. Inoltre tale matrice è una matrice sparsa, quindi presenta un'alta densità di valori nulli, e singolare (a causa di osservazioni ridondanti) per costruzione. Come è usuale in tutte le procedure di bilanciamento *tout court*, la dimensione della matrice partizionata inversa può essere fonte di problemi computazionali di un certo rilievo. Oltre un certo ordine di grandezza delle variabili coinvolte le dimensioni della matrice Λ , e delle osservazioni da elaborare, crescono esponenzialmente. In una procedura automatizzata si possono utilizzare algoritmi di calcolo o metodi diretti di risoluzione di sistemi lineari con matrici sparse, che impiegano ad esempio metodi di fattorizzazione delle matrici come la decomposizione QR.

Qualora non si volesse adottare una soluzione basata sulla fattorizzazione della matrice Λ , Quenneville e Rancourt (2005) hanno proposto una soluzione a due passi per la risoluzione di sistemi vincolati di serie storiche che può interpretarsi in termini di minimi quadrati ponderati (*Weighted Least Squares*). Tale procedura di riconciliazione a due passi (d'ora in avanti 2SR, *2-steps reconciliation*) è stata in seguito generalizzata e si articola di regola come segue:

- 1° step: si effettua un *benchmarking* univariato, quindi su una serie storica alla volta, per ristabilire l'additività temporale annuale, adottando la procedura di Denton, la quale, come sottolineato, se considerata nella sua variante PFD, assicura la proprietà di conservazione della dinamica;
- 2° step: si può adottare una riconciliazione con il metodo di Stone (o pro-rata bidimensionale o aggiustamento con minimi quadrati), un anno (quattro trimestri) alla volta.

Il secondo passo suggerisce la soluzione del problema numerico: la gestione di un bilanciamento un anno, o pochi anni, alla volta abbatte notevolmente il numero di incognite, aggirando i problemi di singolarità della matrice Λ e velocizzando i tempi di esecuzione della stima.

Introducendo la procedura di bilanciamento di Stone abbiamo menzionato la possibilità di intervenire sui valori della matrice di varianza-covarianza Ω che entra nel calcolo dello stimatore. Anche in questo stimatore è possibile applicare dei coefficienti di affidabilità o di alterabilità alle serie in oggetto. Nella stima trimestrale per settore istituzionale essi vanno interpretati più come grado di libertà negli aggiustamenti piuttosto che come misura della fiducia riposta nei confronti delle stime preliminari: tale assegnazione risulta addirittura imprescindibile, in quanto le stime relative ad alcuni settori istituzionali e/o alcune branche non possono essere in alcuna misura alterate in quanto già stimate o rilasciate nel contesto di altri domini.²¹ Ciò permette di alterare la misura della distribuzione degli aggiustamenti in relazione alla dispersione interna (*within*) piuttosto che esterna (*between*). Convenzionalmente il campo di variazione di questi coefficienti è posto fra zero ed uno: facciamo notare che nella realtà operativa il valore dei coefficienti di affidabilità deve essere prossimo a zero ma non nullo, altrimenti c'è il rischio di incorrere nella singolarità dell'inversa di Ω . L'operatore Ω , per costruzione, presenta eteroschedasticità e correlazione dei dati, quindi è possibile attribuire dei coefficienti di alterabilità anche al di fuori della diagonale principale.

4.4 Misure di discrepanza e di aggiustamento

Al fine di poter individuare la grandezza del fenomeno e fare dei confronti fra le metodologie disponibili, definiamo una misura delle discrepanze pre-riconciliazione con il seguente indice τ che esprime la grandezza percentuale della differenza fra la stima del totale dei settori istituzionali rispetto al *benchmark* (il vincolo contemporaneo, che potrebbe essere il totale marginale di riga z_i):

$$\tau = \frac{\sum_k y_{ik} - z_i}{z_i}$$

Date le serie riconciliate R_j e preliminari P_j , ed i rispettivi tassi di crescita r_j e p_j , definiamo degli indici che danno una misura degli aggiustamenti sui valori assoluti e sui tassi di crescita, espressi in termini percentuali. Di conseguenza definiamo lo *Mean Squared Percentage Adjustment* (MSPA) sui livelli:

$$MSPA(R_j, P_j) = 100 \times \left[\frac{1}{n} \sum_{t=1}^n \left(\frac{R_{jt}}{P_{jt}} - 1 \right)^2 \right]^{1/2}$$

²¹ Ad esempio le stime relative alla branca 28 del settore istituzionale della Pubblica Amministrazione.

ed il suo corrispettivo sui tassi di crescita *Mean Squared Adjustment*:

$$MSA(r_j, p_j) = 100 \times \left[\frac{1}{n-1} \sum_{t=2}^n (r_{jt} - p_{jt})^2 \right]^{1/2}$$

dove r_{jt} e p_{jt} sono i tassi di crescita delle serie preliminari e riconciliate:

$$r_{jt} = \frac{R_{jt} - R_{jt-1}}{R_{jt-1}} \quad p_{jt} = \frac{P_{jt} - P_{jt-1}}{P_{jt-1}}$$

5. Un'applicazione: le unità di lavoro equivalenti a tempo pieno (ULA) per settore istituzionale

I conti trimestrali per settore istituzionale coinvolgono la stima di una notevole mole di variabili per sei settori istituzionali. Per alcune di queste variabili, come ad esempio l'occupazione, la stima è articolata a trentadue branche di attività economica, due tipologie di condizione professionale (dipendenti ed indipendenti), in modalità grezza e destagionalizzata. In questa sede vengono riportati i risultati di una procedura di riconciliazione a due passi in cui vengono confrontati gli esiti relativi a due procedure che si differenziano solo nel secondo passo in cui viene eseguito un bilanciamento. Quella che abbiamo denominato strategia 2SR si articola nei due seguenti passi:

- 1° passo: *benchmarking* univariato con procedura di Denton, sulle differenze prime proporzionali (PFD);
- 2° passo: selezione di coefficienti di affidabilità per le unità longitudinali e riconciliazione con vincoli annuali e contemporanei, attraverso una procedura dei minimi quadrati, con matrice di varianza-covarianza proporzionale al quadrato delle serie preliminari.

La strategia che definiamo RAS, invece adotta nel secondo passo un bilanciamento iterativo biproportionale, a seguito di un *benchmarking* con Denton nel primo step.

La variabile prescelta per la stima è la più rappresentativa delle grandezze che determinano l'occupazione: le Unità di lavoro equivalenti a tempo pieno (ula, da ora in avanti). Le ula sono una delle grandezze utilizzate come misura più adeguata del lavoro umano applicato alla produzione di valore nell'ambito della contabilità nazionale; esse nel Sistema Europeo dei Conti (SEC) sono definite come segue "l'occupazione equivalente a tempo pieno, che è pari al numero di posizioni lavorative equivalenti a tempo pieno, è definita come il quoziente tra il totale delle ore lavorate e la media annuale del numero di ore lavorate in posizioni lavorative a tempo pieno nel territorio economico²²". In riferimento sempre alle definizioni del SEC aggiungiamo che "sebbene il totale delle ore lavorate costituisca il miglior parametro per la misurazione degli input

²² Cfr. Eurostat (1996), p. 265.

di lavoro, l'equivalenza a tempo pieno presenta il vantaggio di essere più semplice da calcolare, ciò che facilita i confronti internazionali con i paesi che possono stimare soltanto l'occupazione equivalente a tempo pieno". La stima delle uls si trova a monte del processo di produzione delle statistiche trimestrali per settore istituzionale, essendo necessaria sia al computo di alcune grandezze pro-capite, ed al contempo, come coefficiente di espansione all'universo. Oggetto di stima sono le uls trimestrali lorde per settore istituzionale, dove l'aggettivo lorde indica che si considera l'occupazione a tempo pieno considerando nel computo la *Cassa Integrazione Guadagni* e le cosiddette posizioni lavorative *part-time*.²³

L'applicazione empirica che qui proponiamo riporta quindi le uls lorde a frequenza trimestrale destagionalizzate,²⁴ relative al lavoro dipendente, stimate a 32 branche, ma in seguito aggregate a sei macrobranche esclusivamente per semplicità espositiva. Il periodo campionario considerato è di ampiezza 1991:1-2008:4. Le stime relative alla disaggregazione temporale ed alla riconciliazione²⁵ sono state svolte nell'ambiente ModelEasy+, la procedura di destagionalizzazione è stata realizzata con TRAMO-SEATS (release 5.0) nella versione installata su ModelEasy+, mentre tutta la fase di elaborazione successiva alle stime suddette ed i grafici sono stati eseguiti con la nota piattaforma *open source* R (release 2.9.0).²⁶

Per dare una misura dell'ordine di grandezza delle poste nella tavola 3 sono riportati i valori delle uls lorde dipendenti destagionalizzate distinte per settore istituzionale e branca, prima e dopo la procedura di riconciliazione, e le relative differenze assolute, relativamente all'anno 2008: nonostante i dati facciano riferimento all'ultimo anno del campione, in cui le discrepanze sono state meno rilevanti, si ha un totale di circa 11000 uls redistribuite fra i settori.

²³ Per qualsiasi definizione statistica relativa al sistema dei conti nazionali si faccia riferimento al già citato SEC95 ed a ISTAT (2004).

²⁴ La procedura di destagionalizzazione al momento utilizzata nei paesi appartenenti all'UE è TRAMO-SEATS, per una panoramica su questa metodologia si veda Maravall (1995).

²⁵ La procedura di riconciliazione dei dati dell'occupazione per settore istituzionale utilizza un procedimento a due passi, con codice scritto da Marco Marini nel linguaggio SpeakEasy.

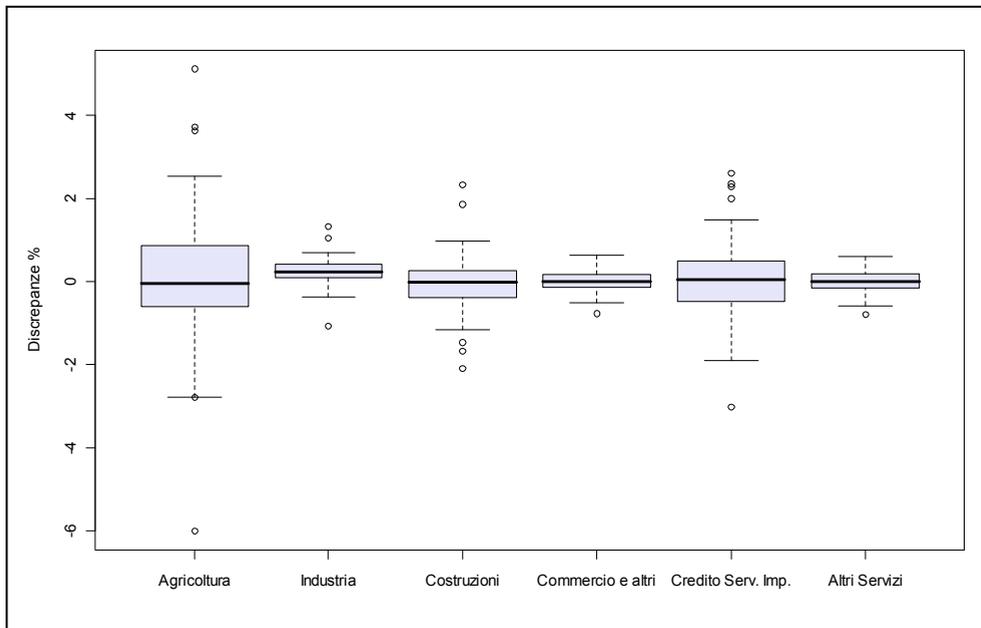
²⁶ URL <http://www.R-project.org>. Tale piattaforma, oltre a possedere un proprio linguaggio di programmazione di alto livello, presenta una vasta collezione di librerie matematiche e statistiche.

Tavola 3 - Valori assoluti delle ula lorde dipendenti destagionalizzate per settore istituzionale e branca, anno 2008: dati pre-riconciliazione, post-riconciliazione e differenze assolute

DATI SERIE PRELIMINARI							
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.	Totale Sett. Isti.
SNF	231994	4212161	1029788	3974463	1750535	688655	11887596
SF	0	588	0	0	518106	0	518694
PA	1554	6004	2650	15868	112992	3515018	3654086
FC	0	0	0	0	0	935974	935974
FP	245473	187666	264796	554095	193137	171829	1616996
ISP	7079	13747	1229	7906	10208	623020	663189
Totale branca	486100	4420166	1298463	4552332	2584978	5934496	19276535
DATI SERIE RICONCILIATE							
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.	Totale Sett. Isti.
SNF	232155	4203068	1029956	3976654	1750516	688209	11880558
SF	0	588	0	0	518106	0	518694
PA	1554	6005	2650	15868	111954	3515018	3653049
FC	0	0	0	0	0	936017	936017
FP	245527	187000	264959	553446	193101	171688	1615721
ISP	7079	13724	1229	7914	10207	621753	661906
Totale branca	486315	4410385	1298794	4553882	2583884	5932685	19265945
DIFFERENZE							
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.	Totale Sett. Isti.
SNF	161	-9093	168	2191	-19	-446	-7038
SF	0	0	0	0	0	0	0
PA	0	1	0	0	-1038	0	-1037
FC	0	0	0	0	0	43	43
FP	54	-666	163	-649	-36	-141	-1275
ISP	0	-23	0	8	-1	-1267	-1283
Totale branca	215	-9781	331	1550	-1094	-1811	-10590

Fonte: Elaborazione su dati Istat

Dal boxplot inserito nella figura 1 si può inferire immediatamente l'ordine di grandezza e la distribuzione delle discrepanze pre-riconciliazione: l'agricoltura, le costruzioni ed il credito sono quelle che presentano maggiore dispersione ed il più alto numero di valori anomali; l'ampiezza delle discrepanze dipende probabilmente dalla numerosità delle branche che entrano nell'aggregazione da 32 a 6 e dalla distribuzione delle discrepanze interna alle singole macrobranche.

Figura 1 - Boxplot della distribuzione delle discrepanze pre-riconciliazione per branca

Fonte: Elaborazione su dati Istat

Nelle tavole 4 e 5 sono riportati i valori delle misure di aggiustamento definite nella sezione 4.4, distinte per settore istituzionale e branca; da esse si può constatare come, sui dati disaggregati c'è una sostanziale parità fra i due metodi, mentre ragionando sui tassi di crescita c'è una netta prevalenza della strategia 2SR. Valutando l'impatto della riconciliazione sul dato aggregato per branche, è chiara una preminenza dell'approccio 2SR sui livelli e sui tassi di crescita (vedi Tav. 6).

Nelle tavole 7 e 8 viene comparato l'impatto che la modifica dei coefficienti di alterabilità ha sulle misure di aggiustamento; esso è stato condotto puramente a scopo didattico perché nella realtà operativa è stata già menzionata la necessità di lasciare inalterate determinate serie temporali relativamente ad alcuni settori istituzionali, scaricando tuttavia in tal modo sugli altri settori l'onere della riconciliazione. Ai settori istituzionali Società Non Finanziarie e Famiglie Produttrici è assegnato il valore massimo ($\theta = 1$) mentre agli altri settori è attribuito un valore di $\theta = 0.0001$: ad esempio, l'attribuzione di un coefficiente $\theta = 1$ anche alla PA e alle ISP, modifica la distribuzione degli aggiustamenti, il loro impatto sui settori istituzionali e sulle macrobranche, e di conseguenza, l'efficienza relativa degli indici.

Tavola 4 - Indice MSPA: valori per settore istituzionale e branca

MSPA 2SR						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Impr.	PA altri Serv.
SNF	1.20	0.42	0.82	0.32	1.32	0.85
SF	-	0.00	-	-	0.59	-
PA	0.00	0.03	0.16	0.00	0.91	0.00
FC	-	-	-	-	-	1.20
FP	1.90	0.48	0.42	0.13	0.24	0.23
ISP	0.02	0.29	0.00	0.27	0.03	0.41

MSPA RAS						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Impr.	PA altri Serv.
SNF	1.53	0.40	0.70	0.28	1.02	0.96
SF	-	0.34	-	-	1.02	-
PA	0.02	0.03	0.15	0.00	0.91	0.01
FC	-	-	-	-	-	0.90
FP	1.70	0.58	0.79	0.31	1.03	0.91
ISP	0.03	0.28	0.01	0.26	0.03	0.38

Fonte: Elaborazione su dati Istat

Tavola 5 - Indice MSA: valori per settore istituzionale e branca

MSA 2SR						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Impr.	PA altri Serv.
SNF	1.74	0.55	0.98	0.48	1.63	1.25
SF	-	0.00	-	-	0.72	-
PA	0.00	0.01	0.10	0.00	0.04	0.00
FC	-	-	-	-	-	1.79
FP	2.50	0.10	0.27	0.08	0.26	0.33
ISP	0.00	0.10	0.00	0.07	0.01	0.16

MSA RAS						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Impr.	PA altri Serv.
SNF	2.21	0.53	0.84	0.43	1.26	1.43
SF	-	0.53	-	-	1.26	-
PA	0.02	0.01	0.13	0.00	0.04	0.01
FC	-	-	-	-	-	1.35
FP	2.20	0.55	0.86	0.43	1.27	1.35
ISP	0.02	0.17	0.01	0.10	0.02	0.26

Fonte: Elaborazione su dati Istat

Tavola 6 - Confronto indici di aggiustamento per metodologie 2SR-RAS e per settore istituzionale

	MSPA						Tot
	SNF	SF	PA	FC	FP	ISP	
2SR	0.90	0.24	0.38	1.20	0.83	0.23	0.58
RAS	0.92	0.44	0.38	0.90	0.98	0.22	0.62

	MSA						Tot
	SNF	SF	PA	FC	FP	ISP	
2SR	1.21	0.29	0.05	1.79	1.04	0.08	0.73
RAS	1.27	0.56	0.06	1.35	1.26	0.13	0.80

Fonte: Elaborazione su dati Istat

Tavola 7 - Confronto indici MSPA per settore istituzionale e branca

	MSPA 2SR (Coeff. Alterabilità PA-ISP=0.0001)					
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.
SNF	1.20	0.42	0.82	0.32	1.32	0.85
SF	-	0.00	-	-	0.59	-
PA	0.00	0.03	0.16	0.00	0.91	0.00
FC	-	-	-	-	-	1.20
FP	1.90	0.48	0.42	0.13	0.24	0.23
ISP	0.02	0.29	0.00	0.27	0.03	0.41

	MSPA 2SR (Coeff. Alterabilità =1)					
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.
SNF	1.23	0.41	0.80	0.30	1.32	0.33
SF	-	0.00	-	-	0.60	-
PA	0.01	0.03	0.16	0.00	0.91	0.38
FC	-	-	-	-	-	0.09
FP	1.93	0.48	0.42	0.13	0.24	0.08
ISP	0.04	0.29	0.00	0.27	0.03	0.40

Fonte: Elaborazione su dati Istat

Tavola 8 - Confronto indici MSA per settore istituzionale e branca, per differenti valori dei coefficienti di alterabilità

MSA 2SR (Coeff. di Alterabilità PA-ISP=0.0001)						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.
SNF	1.74	0.55	0.98	0.48	1.63	1.25
SF	-	0.00	-	-	0.72	-
PA	0.00	0.01	0.10	0.00	0.04	0.00
FC	-	-	-	-	-	1.79
FP	2.50	0.10	0.27	0.08	0.26	0.33
ISP	0.00	0.10	0.00	0.07	0.01	0.16

MSA 2SR (Coeff. Alterabilità =1)						
	Agricoltura	Industria	Costruzioni	Comm. Alb.	Cred. Serv. Imp.	PA altri Serv.
SNF	1.79	0.55	0.94	0.45	1.62	0.16
SF	-	0.00	-	-	0.71	-
PA	0.01	0.01	0.10	0.00	0.11	0.57
FC	-	-	-	-	-	0.12
FP	2.55	0.10	0.26	0.07	0.26	0.02
ISP	0.04	0.10	0.00	0.07	0.02	0.18

Fonte: Elaborazione su dati Istat

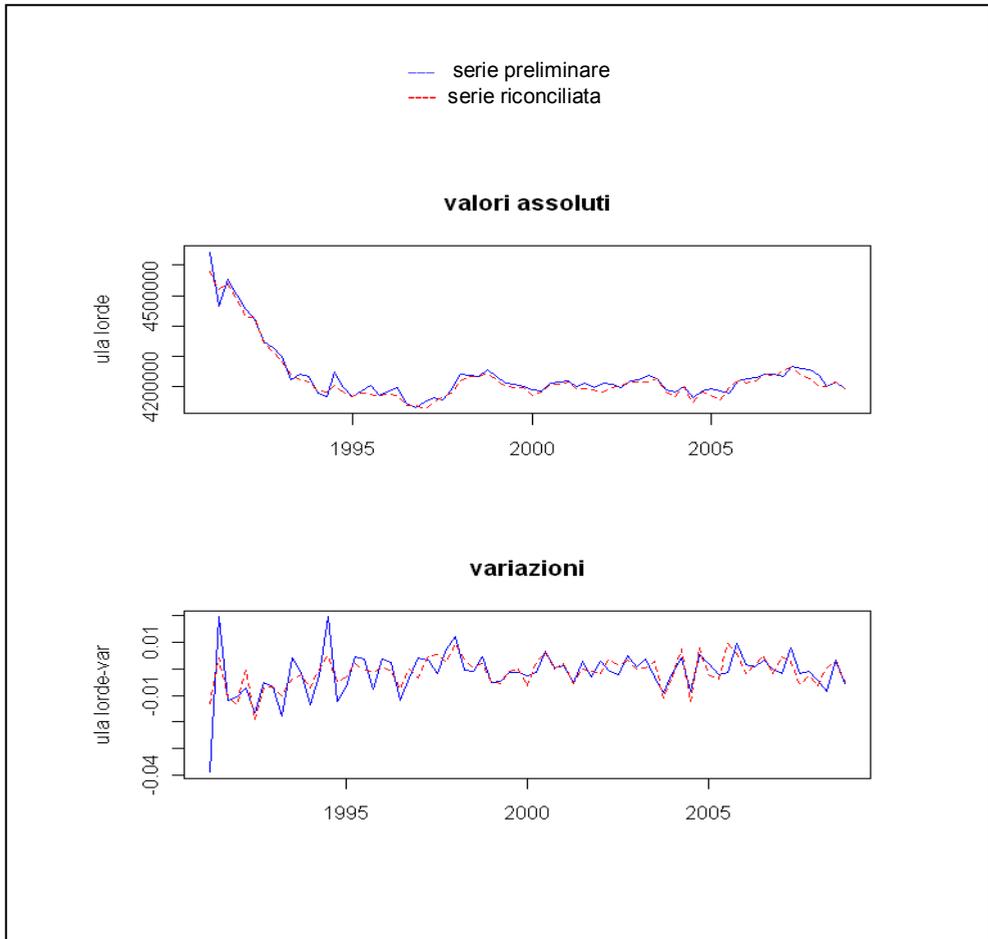
Nella figura 2 sono riportate le serie preliminari e riconciliate relative al comparto ed al settore istituzionale che possiede i valori e gli aggiustamenti più grandi (l'industria in senso stretto delle Società Non Finanziarie), dalla quale risulta immediatamente che la grandezza delle discrepanze pre-bilanciamento, in alcune finestre campionarie, non è affatto marginale e la procedura di riconciliazione può avere un impatto notevole nel ricondurre i dati ai vincoli trimestrale ed annuale.

Nelle figure 3-4 sono inoltre riprodotti i grafici delle serie dei tassi di crescita per le sei macrobranche relativamente ai due settori istituzionali (società non finanziarie e famiglie produttrici) a cui sono assegnati necessariamente, o meno, i maggiori aggiustamenti relativi nella riconciliazione: si può vedere chiaramente come nel settore istituzionale famiglie produttrici la metodologia 2SR presenta una minore volatilità dei tassi di variazione rispetto al RAS.

In conclusione, questa verifica empirica sembra dare delle indicazioni che sintetizziamo nei seguenti punti:

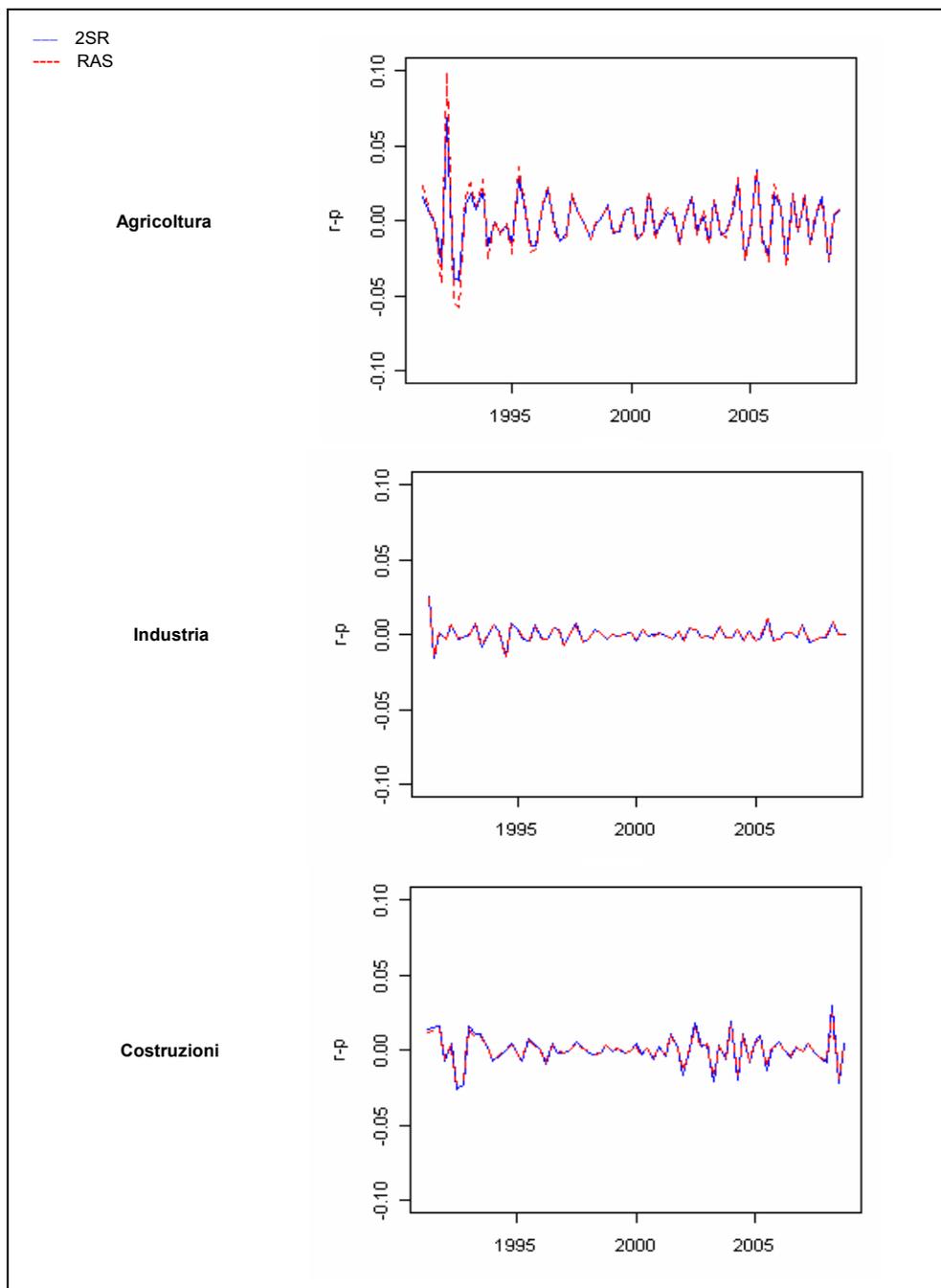
- la metodologia 2SR dà risultati soddisfacenti in termini assoluti e sui tassi di crescita rispetto alla metodologia alternativa RAS, sia per i singoli settori istituzionali che nella sua globalità;
- la calibrazione dei coefficienti di affidabilità, che ha un impatto sui valori degli indici di aggiustamento, va valutata anche in relazione alle revisioni delle serie, che nella stima dell'occupazione totale e per settore istituzionale riguarda gli ultimi 12 trimestri nel caso di dati grezzi, e l'intero periodo campionario nel caso di serie destagionalizzate;
- sviluppi futuri: è auspicabile l'introduzione nelle istituzioni che producono dati ufficiali della procedura in linguaggi e piattaforme più moderne e diffuse (R, Matlab, Octave ad esempio), sia in relazione all'efficienza operativa che in un'ottica di risparmio economico per la pubblica amministrazione.

Figura 2 - Serie temporale delle ula lorde dipendenti destagionalizzate: Società non finanziarie - branca Industria

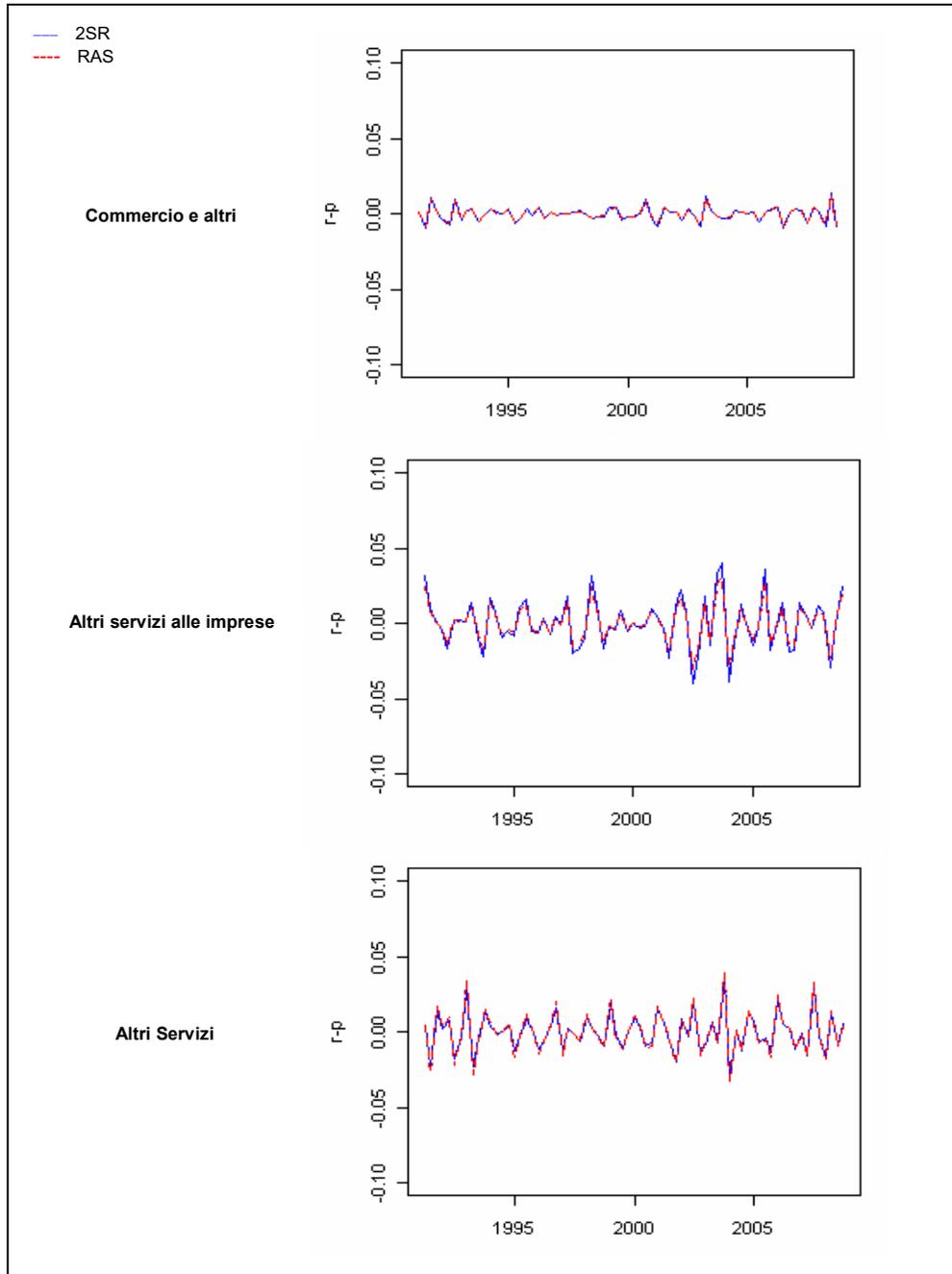


Fonte: Elaborazione su dati Istat

Figura 3 - Serie temporali degli aggiustamenti nei tassi di crescita delle ula lorde destagionalizzate: Società non finanziarie

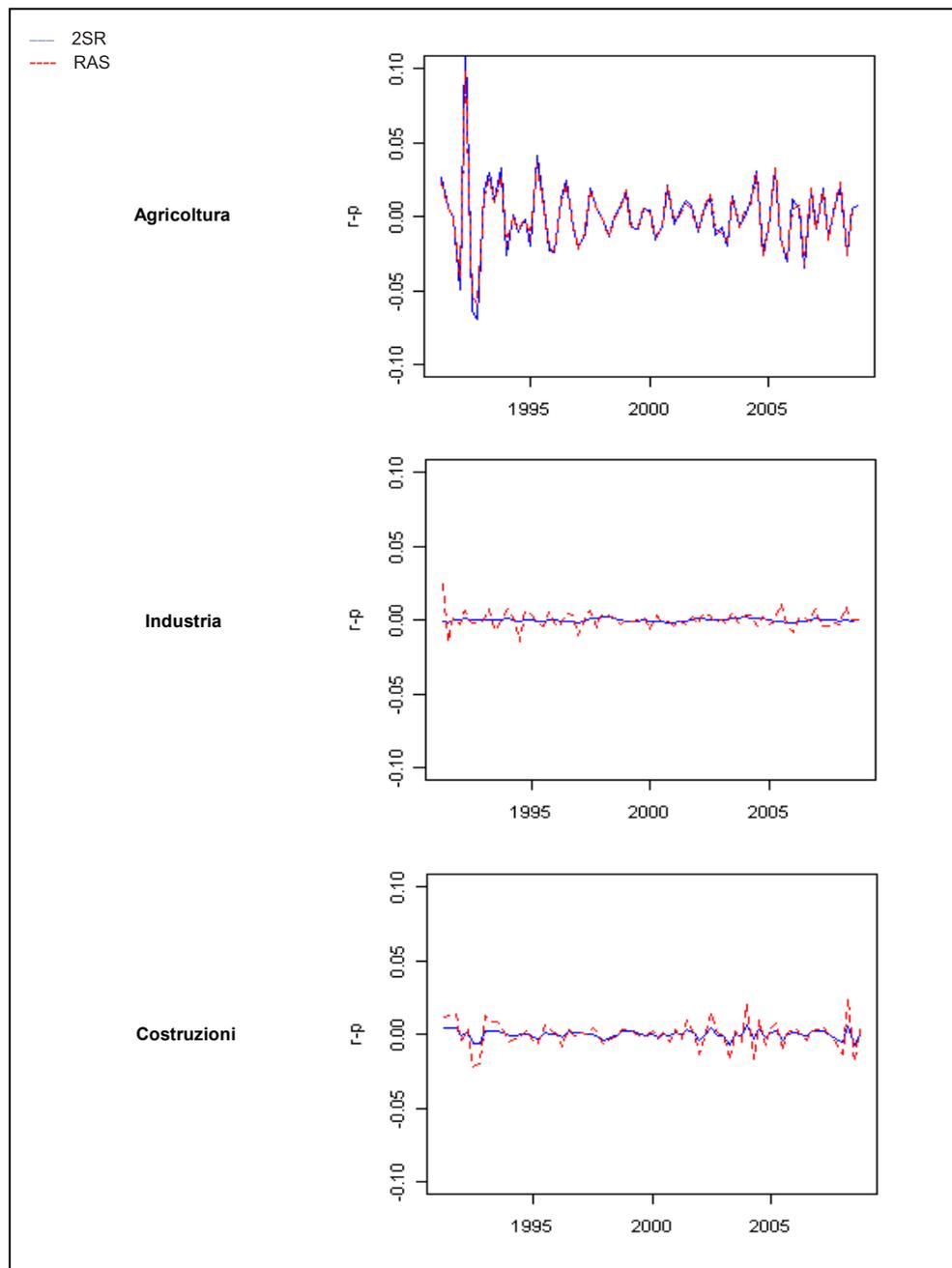


Fonte: Elaborazione su dati Istat

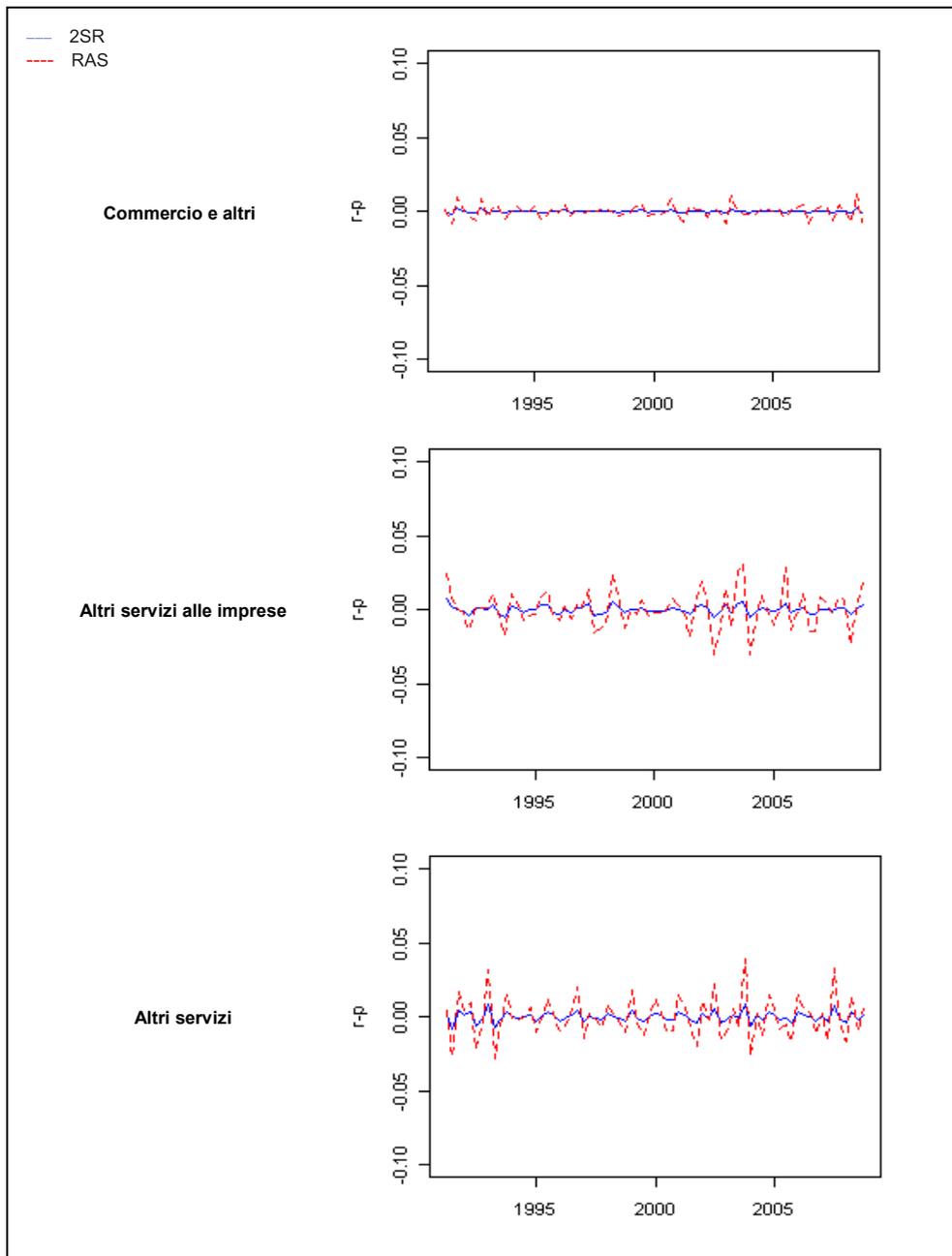
Figura 3 segue - Serie temporali degli aggiustamenti nei tassi di crescita delle ula lorde destagionalizzate: Società non finanziarie

Fonte: Elaborazione su dati Istat

Figura 4 - Serie temporali degli aggiustamenti dei tassi di crescita delle ula lorde destagionalizzate: Famiglie produttrici



Fonte: Elaborazione su dati Istat

Figura 4 segue - Serie temporali degli aggiustamenti dei tassi di crescita delle ula lorde destagionalizzate: Famiglie produttrici

Fonte: Elaborazione su dati Istat

6. Conclusioni

In questo lavoro è stato presentato il problema della riconciliazione di serie storiche applicato alle stime trimestrali per settore istituzionale. In tale ambito, infatti, coesistono vincoli contemporanei e temporali da rispettare. Dopo una breve introduzione alle tecniche di bilanciamento più note ed alla metodologia di Denton, sono stati messi a confronto due approcci in un'applicazione basata sui dati destagionalizzati delle Unità di lavoro equivalenti a tempo pieno. I dati analizzati, infatti, costituiscono un sistema di serie storiche soggetto a doppio vincolo contabile. Per poter riconciliare questo sistema di serie con i vincoli esogeni sono state applicate due differenti procedure di riconciliazione a 2 stadi che condividono la stessa tecnica di *benchmarking* (Denton PFD) al primo stadio - ampiamente discussa nella sezione 4.2 - ed entrambe basate sul metodo di aggiustamento che adotta i minimi quadrati al secondo stadio. Tali procedure al tempo stesso preservano la dinamica delle serie preliminari, requisito auspicabile quando nel processo di produzione di dati ad alta frequenza si ricorre all'approccio indiretto: nei QSA, infatti, è necessario conservare l'informazione trasmessa dagli indicatori utilizzati nel processo di trimestralizzazione.

Sulla base di specifiche misure calcolate sui livelli e sui tassi di variazione delle serie riconciliate sono state valutate le performances delle procedure messe a confronto. Tali indicatori sono stati, inoltre, calcolati sulle serie riconciliate - nella seconda parte dell'applicazione - simulando di non poter alterare alcune serie componenti il sistema considerato. L'analisi di questi indicatori è stata eseguita sia a livello di branca (aggregazione a sei branche di attività economica) sia a livello di settore istituzionale (massimo livello di aggregazione). Dall'analisi dei risultati è emerso che la natura della procedura statistica di riconciliazione impone la misura dell'aggiustamento nel senso che è direttamente collegato alla dimensione della variabile da riconciliare: maggiore (minore) è una variabile, maggiore (minore) sarà l'aggiustamento finale. Gli indicatori calcolati sui tassi di variazione evidenziano che l'approccio 2SR, alternativo a quello RAS, ha fornito risultati migliori complessivamente e a livello di settore istituzionale, mentre per quanto concerne l'indicatore sui valori assoluti, i due approcci si sono equivalsi. Il metodo di riconciliazione 2SR, infatti, in un sistema di serie storiche soggetto a vincoli temporali e contemporanei, ha restituito risultati migliori per quanto riguarda il *movement preservation principle*.

La dimensione del problema è un altro aspetto fondamentale da considerare quando un set di serie preliminari deve essere riconciliato tenendo conto di alcuni vincoli contabili: il numero di variabili coinvolte nel processo di riconciliazione, la lunghezza delle serie, il numero di vincoli. Quello trattato costituisce, pertanto, un problema di piccole dimensioni. Tuttavia altri aspetti connessi alla riconciliazione restano da investigare, come ad esempio la valutazione delle performances di tali procedure quando si procede in estrapolazione, ovvero quando non si dispone dell'ultimo dato annuale che costituisce il vincolo temporale e quindi del rispetto del *movement preservation principle* nei trimestri in corso d'anno.

Riferimenti bibliografici

- Barker T. e Pesaran M.H. (1990), *Disaggregation in econometric modelling*, Routledge, London and New York.
- Bee Dagum E. e Cholette P. A. (2006), *Benchmarking, temporal distribution and reconciliation methods for time series*, Springer.
- Bloem A., Dippelsman R. e Mæhle N. (2001), *Quarterly National Accounts manual concepts, data sources and compilation*, International Monetary Fund, Washington D.C.
- Byron R. (1978), *The estimation of large social account matrices*, Journal of the Royal Statistical Society, 141 (3), 359-367.
- Cholette P. A. (1984), *Adjusting sub-annual series to yearly benchmarks*, Survey Methodology 10, 35-49.
- Denton F. T. (1971), *Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization*. Journal of the American Statistical Association 66, 99-102.
- Di Fonzo T. e Marini M. (2003), *Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle*, Working Paper Series n.9, Università di Padova.
- Di Fonzo T. e Marini M. (2009), *Simultaneous and Two-step reconciliation of systems of time series*, Working Paper Series n.9, Università di Padova.
- Eurostat (1996), *Sistema europeo dei conti 1995*, Luxembourg.
- Eurostat (2008), *Quarterly National Accounts Inventory: sources and methods of Italian Quarterly National Accounts*, Luxembourg.
- Eurostat (2009), *European Inventory of Sources and Methods for Quarterly Sectors Accounts*, Luxembourg.
- Granger, C. W. J. e T. Teräsvirta (1993). *Modelling nonlinear economic relationship*. Oxford University Press.
- ISTAT (2004), *Metodologie di stima degli aggregati di contabilità nazionale a prezzi correnti*, Metodi e norme n. 21.
- ISTAT (2005), *I conti economici nazionali per settore istituzionale: le nuove stime secondo il Sec95*, Metodi e norme n. 23.
- Lutero G. (2010), *The aggregation problem in its historical perspective: a summary overview*, presentato alla Third Wye City Group Global Conference on Agricultural and Rural Household Statistics, Economic Research Service of the United States Department of Agriculture, Washington D.C.
- Maravall A. (1995), *Unobserved Components in Economic Time Series*, The Handbook of Applied Econometrics, Blackwell.
- Nicolardi V. (2000), *Balancing large accounting system: an application to the 1992 italian I-O table*, paper presentato alla 13th International conference on Input-Output techniques, Università di Macerata.

- Quenneville B. e Rancourt E. (2005), *Simple methods to restore the additivity of a system of time series*, paper presentato al workshop “Frontiers in benchmarking techniques and their application to official statistics”, Luxembourg April 2005.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Stone R., Champernowne D. e Meade J. (1942), *The precision of national income estimates*, *Review of economic studies*, 9, 111-125.
- Stone R. (1961), *Input-output and national accounts*, Oecd, Paris.