

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

*Corso di laurea in Scienze Statistiche, Economiche, Finanziarie ed Aziendali*

TESI DI LAUREA

RICONCILIAZIONE DI SERIE STORICHE: METODI ED  
UN'APPLICAZIONE

(Reconciliation of time series: methods and an application)

**Relatore**

Prof. Tommaso Di Fonzo

**Laureando**

Rocco Francesco Cerundolo

ANNO ACCADEMICO 2009/2010

# INDICE

|   |    |
|---|----|
| Capitolo 1: Introduzione.....   | 1  |
| 1.1: Cause del problema: stagionalità e dati ad alta e bassa frequenza.....                 | 2  |
| 1.2 Rassegna della letteratura .....  | 3  |
| Capitolo 2: Stagionalità e principali tecniche di destagionalizzazione.....                 | 5  |
| 2.1 Decomposizione di serie storiche e stagionalità.....                                    | 5  |
| 2.2 Principali metodi di destagionalizzazione.....  | 6  |
| 2.3 Destagionalizzazione diretta ed indiretta: metodi e test per il loro confronto.....     | 9  |
| 2.3.1 Metodologie.....  | 14 |
| Capitolo 3: il bilanciamento di matrici.....  | 18 |
| 3.1 Il problema: i vincoli contemporanei e temporali.....                                   | 19 |
| 3.2 Procedure per il bilanciamento: ottimizzazione RAS.....                                 | 21 |
| 3.3 Adattamento con i minimi quadrati di dati economici soggetti a restrizioni lineari..... | 21 |
| 3.4 Bilanciamento di un sistema di serie temporali con vincoli esogeni.....                 | 23 |
| Capitolo 4: Il benchmarking di serie storiche: descrizione dei vari modelli sviluppati..... | 26 |
| 4.1 Metodo di benchmarking di Cholette e Dagum.....   | 27 |
| 4.1.1 Una specificazione conservativa di trend deterministici.....                          | 28 |
| 4.1.2 Rappresentazione matriciale del modello.....  | 29 |
| 4.1.3 Operatori di somma temporale.....   | 30 |
| 4.1.4 Soluzione del modello.....  | 31 |
| 4.1.5 Modello additivo con benchmarking proporzionale.....                                  | 34 |
| 4.2 Metodo di benchmarking di Cholette e Dagum: il modello moltiplicativo.....              | 35 |
| 4.2.1 Il modello di benchmarking moltiplicativo.....  | 36 |
| 4.3 Benchmarking come un problema di minimi quadrati.....                                   | 38 |
| 4.4 Il metodo di Denton.....  | 39 |
| 4.4.1 il metodo di Denton e sue varianti: la versione a differenze prime additive.....      | 40 |
| Capitolo 5: Bilanciamento e riconciliazione di serie storiche.....                          | 43 |
| 5.1 Riconciliazione: un modello di regressione semplice.....                                | 47 |

|   |    |
|---|----|
| 5.2 Riconciliazione di sistemi di serie storiche ad una variabile.....  | 51 |
| 5.2.1 Il modello.....   | 52 |
| 5.3 Riconciliazione di sistemi di serie storiche con due variabili..... | 54 |
| 5.3.1 Il modello.....   | 55 |
| 5.4 il problema dei Vincoli ridondanti.....                             | 56 |
| Capitolo 6: Un'applicazione a dati reali.....                           | 58 |
| 6.1 <i>Il Monthly Retail Trade Survey</i> .....                         | 58 |
| 6.2 La raccolta dei dati.....   | 58 |
| 6.3 I dati.....   | 59 |
| 6.4 Primo passo: la destagionalizzazione dei dati.....                  | 60 |
| 6.5 La riconciliazione.....   | 73 |
| 6.5.1 La riconciliazione con la procedura a due stadi.....              | 76 |
| Capitolo 7: Conclusione.....  | 83 |

# CAPITOLO 1

## INTRODUZIONE

Uno degli obiettivi della ricerca statistica è lo studio dell'evoluzione nel tempo dei fenomeni dinamici, tra cui quelli di carattere socio-economico. Questo tipo di analisi ha come principale oggetto di interesse le serie storiche, sequenze di osservazioni dipendenti, misurate nel corso del tempo. Vista la loro importanza, non risulta difficile capire l'esigenza di processi di revisione delle serie, che, prima di essere pubblicate, vengono corrette per aumentare l'efficienza delle stime, ridurre la distorsione e replicarne i valori mancanti. Tra le principali tecniche di aggiustamento troviamo il benchmarking, l'interpolazione e la distribuzione temporale, la calendarizzazione e l'estrazione del segnale.

Come si capirà nel corso della tesi, le tecniche di benchmarking giocano un ruolo di primaria importanza, e per questo ce ne occuperemo più in dettaglio nei capitoli successivi. Ci limiteremo, invece, solo ad una breve descrizione delle altre tecniche, inserendo comunque in bibliografia i riferimenti più significativi.

Oltre alle tecniche di benchmarking, andremo ad analizzare il bilanciamento delle serie storiche, ovvero la procedura che permette di ricondurre le somme per riga o per colonna delle serie ai rispettivi totali marginali. Nel corso del capitolo dedicato al bilanciamento, si vedrà che il problema di soddisfare questi vincoli può derivare, ad esempio, dall'adozione di procedure di destagionalizzazione.

Infine si affronterà un terzo problema, quello della riconciliazione di sistemi di serie storiche, che comprende tanto il benchmarking quanto il bilanciamento.

L'ultimo capitolo della tesi sarà dedicato ad un'applicazione a dati reali canadesi, sottoposti preliminarmente alla destagionalizzazione, per poi considerare l'applicazione delle tecniche di bilanciamento di serie storiche.

Come detto, ora descriveremo per sommi capi alcune tecniche di riconciliazione, per poi approfondire le tecniche di benchmarking e bilanciamento.

### **1.1 Possibili cause del problema: stagionalità e dati ad alta e bassa frequenza**

Nell'operare con serie storiche economiche, convenzionalmente, i dati raccolti possono essere classificati, in base al loro timing temporale, in dati ad alta o bassa frequenza. Con riguardo a fenomeni macroeconomici, i primi hanno il problema di essere meno dettagliati, il che solitamente comporta l'uso di risorse per migliorare la loro qualità.

Nel caso delle serie storiche macro economiche, quelle a frequenza più elevata hanno, di solito, timing mensile o trimestrale, quelle a bassa hanno timing annuale. Per le serie finanziarie il timing cambia, e si parla di timing giornaliero o di sue frazioni ( minuti, secondi, "tick").

Avere la stessa serie, con timing differenti, spesso non permette che i vincoli previsti dalla teoria economica vengano rispettati dai dati, avendo così situazioni in cui la somma di dati ad alta frequenza (nel caso reale considerato mensili) risulta diversa dal valore aggregato a bassa frequenza. In questo caso si dice che non vengono rispettati vincoli temporali, e per risolvere il problema si deve ricorrere a tecniche di benchmarking. Inoltre, i dati, prima di essere pubblicati, subiscono spesso alcune trasformazioni utili per l'analisi economica. In particolare, fanno parte di queste analisi le procedure di destagionalizzazione, che permettono di eliminare i picchi periodici delle serie prese in considerazione. Le procedure di destagionalizzazione producono però una nuova forma di

discrepanza. Infatti, destagionalizzando le singole serie e, separatamente, la serie aggregata, ottenuta dalla somma delle serie componenti, si nota che la somma di tutte le serie destagionalizzate non coincide con la serie aggregata destagionalizzata. Si comprende quindi che questo problema origina una situazione in cui i vincoli di somma allo stesso istante temporale, e per questo chiamati contemporanei, non vengano rispettati, generando il problema del bilanciamento.

Dopo questa breve descrizione del problema dei vincoli temporali e contemporanei, risulta fondamentale dire che, nell'analisi di sistemi di serie storiche, si hanno di fronte entrambi i problemi, dunque sia quello del bilanciamento che quello del benchmarking. Le procedure per ovviare a questi due problemi prendono il nome di riconciliazione, alle quali sarà dedicato il terzo capitolo.

Passiamo ora a descrivere in modo più approfondito tutti gli elementi fin qui accennati, seguendo l'ordine tenuto nell'analisi dei dati ed iniziando quindi dalla destagionalizzazione e dal bilanciamento di serie storiche, per poi continuare con le tecniche di benchmarking e con la riconciliazione di sistemi di serie temporali.

## **1.2 Rassegna della letteratura**

Per facilitare la consultazione della tesi, si riportano i principali riferimenti bibliografici, seguendo l'ordine degli argomenti descritti. Questo paragrafo vuol essere un'indicazione di quelli che sono i principali lavori da cui si è preso spunto; ovviamente si rimanda alla bibliografia per una lista esaustiva di tutti gli articoli ed i libri esaminati per le nostre argomentazioni.

Il secondo capitolo descrive il problema della stagionalità, fornendone cause e metodi per risolverlo. Tra gli articoli studiati per poter affrontare l'argomento, ho trovato particolarmente interessante per semplicità e chiarezza, il lavoro di Astolfi, Ladiray e Mazzi, (2001) che descrive le principali tecniche di destagionalizzazione e le principali batterie di

test per valutare la bontà delle procedure utilizzate. Base di questo lavoro è la pubblicazione di Kuiper (1975), che permette una migliore comprensione delle statistiche utilizzate per il confronto tra le varie tecniche di destagionalizzazione.

Nel terzo capitolo si parla di bilanciamento di matrici; Il teorico principale dell'argomento è R. Stone (1961). Il problema del bilanciamento in anni più recenti è stato curato in Italia da Tommaso Di Fonzo e Marco Marini, in particolare in due lavori del 2005 e del 2009.

Il quarto capitolo parla del benchmarking di serie storiche, con i suoi due modelli principali, quello additivo e quello moltiplicativo. Per lo studio e l'esposizione di questi argomenti ho fatto riferimento al libro di due tra i principali esperti della questione, Cholette e Dagum (2006). Nell'ultimo paragrafo di tale lavoro viene descritto il metodo di Denton (1971).

Il quinto capitolo illustra il bilanciamento e la riconciliazione delle serie storiche; come anticipato nel terzo capitolo, principali teorici dell'argomento sono Di Fonzo e Marini (2005, 2009), affiancati da Quenneville e Rancourt (2005). Anche in questo caso queste due indicazioni non costituiscono una esaustiva panoramica sul problema; risulta quindi necessario rimandare anche in questo caso alla bibliografia.

# CAPITOLO 2

## STAGIONALITÀ E PRINCIPALI TECNICHE DI DESTAGIONALIZZAZIONE

In questo capitolo descriviamo in modo teorico ciò che poi è stato utilizzato nell'applicazione del sesto capitolo. Come si vedrà anche in seguito, in fase di raccolta dati il valore del totale nazionale risulta uguale alla somma di tutte le serie che compongono il sistema. Il problema nasce quando si procede alla destagionalizzazione delle serie componenti e di quella aggregata. Dopo aver destagionalizzato la serie componenti e quella aggregata, ci si accorge che la somma delle prime, fatta mese per mese, non combacia con il corrispondente mensile della seconda. Questo crea un problema di discrepanze e di mancato rispetto dei vincoli. Questa premessa permette di comprendere meglio perché, spesso, le tecniche di riconciliazione si applicano a dati grezzi e non destagionalizzati (ESS Guidelines on Seasonal Adjustment, 2009).

Secondo le indicazioni di Eurostat, la destagionalizzazione delle serie può essere riassunta in tre fasi:

- Scelta del metodo di aggiustamento stagionale
- Consistenza tra le serie grezze e le serie destagionalizzate
- Scelta tra un metodo diretto ed uno indiretto

Vista la problematica descritta, ci occuperemo della stagionalità, descrivendone le cause e i metodi che nel tempo sono stati sviluppati per trattarla.



## **2.1 Decomposizione di serie storiche e stagionalità**

In generale, la parte sistematica di una serie storica è considerata come somma, o prodotto, di tre componenti, che sono “direttamente non osservabili, ma chiaramente definibili sul piano concettuale” (Di Fonzo e Lisi, 2004), ovvero trend, ciclo e stagionalità. Dato il problema della tesi, non considereremo le prime due, ma ci soffermeremo sulla stagionalità. Questa è costituita da movimenti della serie nel corso dell’anno che, per effetto di alcuni fattori come i cicli climatici o fenomeni sociali, si pensi ad esempio a Pasqua o a Natale, “tendono a ripetersi in maniera pressoché analoga nel medesimo periodo”.

Nel corso di questa discussione citiamo solo brevemente i metodi minori di aggiustamento stagionale, e descriviamo invece i due principali approcci per l’eliminazione della componente stagionale.

## **2.2 Principali metodi di destagionalizzazione**

Per la descrizione delle procedure di destagionalizzazione facciamo riferimento alle linee guida dell’Eurostat, nella versione più aggiornata del 2009. Dopo il trattamento iniziale delle serie si passa infatti alla destagionalizzazione. I due metodi principali e maggiormente utilizzati per questi fini sono TRAMO-SEATS ed X-12-ARIMA. Il primo si basa su un approccio parametrico, il secondo no e questo, vedremo, si rivelerà più utile ai fini dell’applicazione dell’ultimo capitolo. Ulteriore possibilità è l’utilizzo di un modello strutturale per serie storiche, provvisto di procedure che permettano di trattare problemi di calendario e considerare anche valori anomali e che includa anche un adeguato set di diagnostiche. C’è da sottolineare che sia X-12-ARIMA che TRAMO-SEATS hanno già un’interfaccia sviluppata ed un buon set di diagnostiche, che permettono un’analisi completa dei risultati ottenuti. Bisogna però stabilire quale delle due procedure sia migliore. In realtà, vedremo, non esiste una risposta univoca che permetta di scegliere o uno o l’altro metodo; la scelta può essere dettata

dall'esperienza, da motivazioni personali e in base alle caratteristiche delle serie storiche. Secondo i suggerimenti di Eurostat, l'utilizzo di uno strumento tra TRAMO-SEATS ed ARIMA-X-12 è preferibile a quello di un modello strutturale.

Pensare che la stagionalità non produca effetti su tutta la serie è implausibile, soprattutto in modelli moltiplicativi applicati a serie in cui sono presenti valori anomali o effetti di calendario. Si può pensare di imporre che la somma dei dati destagionalizzati sia pari al valore grezzo, malgrado questo non abbia una vera giustificazione e possa produrre dati distorti, specialmente se i valori anomali o gli effetti di calendario risultano particolarmente rilevanti, può portare a dati non soddisfacenti e richiede l'utilizzo di ulteriori processi che garantiscano l'uguaglianza tra i valori. Il vantaggio di imporre vincoli ai dati si ha nella consistenza tra le serie di dati destagionalizzate e quelle grezze.

Da questa breve introduzione si comprende come, davanti a questo problema, l'analista debba optare fra tre possibilità; può infatti scegliere di non imporre vincoli e di lavorare su dati grezzi; scegliere di utilizzare vincoli tra dati destagionalizzati e dati grezzi, vincoli che necessitano l'utilizzo di tecniche di benchmarking; utilizzare tecniche di benchmarking che mantengano la stagionalità residua.

Tra queste tre possibilità, la seconda è consigliata ed è quella che verrà utilizzata nell'ultimo capitolo. C'è però da sottolineare che, su suggerimento delle linee guida che abbiamo preso in considerazione, la scelta migliore resta quella di non imporre vincoli ai dati e considerare i dati grezzi. Questo permette di comprendere meglio le applicazioni che non destagionalizzano le serie, ma considerano procedure di riconciliazione direttamente sui dati grezzi.

Ci sono anche altri metodi di destagionalizzazione minori, e per questo li accenniamo brevemente.

Il primo metodo di aggiustamento stagionale è il Metodo Burman; applicabile sia al modello moltiplicativo che a quello additivo; questa procedura

sottrae una media mobile a tredici termini dalle serie, eliminando in questo modo il trend e ottenendo una serie che fornisce i primi fattori stagionali. Le comunità europee utilizzano l'EEC; conosciuto come SEABIRD, è stato implementato da Bongard e Mesnage presso l'ufficio statistico europeo. I creatori del metodo avevano come problema il fatto che presso l'agenzia statistica della Comunità arrivavano dati da tutti i paesi membri dell'Unione; questo ha posto la necessità di avere un metodo unico, che si adattasse al trattamento della stagionalità per tutti i membri, capace di trattare le diversità esistenti tra i vari paesi, malgrado la grande varietà di caratteristiche delle serie. Il metodo si basa su una procedura a due stadi; al primo si eliminano i valori estremi ed al secondo si applica una media mobile a diciannove termini, creata da Bongard. Infine c'è il *Berlin Method*, che assume una decomposizione additiva tra le componenti delle serie storiche; vengono eliminati prima i valori considerati *outlier* e poi si stimano, mediante filtri, le componenti di trend e stagionalità.

Descriviamo ora il processo che ha portato all'implementazione dei due metodi principali, X-12-ARIMA e TRAMO-SEATS e vediamo in breve le loro principali caratteristiche.

X-11, sviluppato in seguito fino al più recente X-12-ARIMA, è stato introdotto nel 1965 dal *Census Bureau*; è il risultato di un processo di sviluppo condotto attraverso dodici metodi, partendo dal Metodo-primo presentato nel 1954. Alternativa al metodo del *Census-Bureau* è TRAMO-SEATS, implementato da Agustin Maravall e Gianluca Caporello. La principale differenza tra i due metodi è nel fatto che TRAMO SEATS si basa su un modello ARIMA specificato, mentre la famiglia del metodo X-11 applica alle serie un approccio non parametrico. Il dibattito tra i due metodi è, come detto, acceso. Come si vedrà nel prossimo paragrafo, non si può stabilire quale sia il metodo migliore. Per questo motivo si sta cercando di creare un programma sintesi dei due approcci, che ha come prototipo X-12-ARIMA/SEATS, prototipo del futuro X-13.

Vediamo ora più in dettaglio il problema del trattamento della stagionalità, per rendere più chiari i procedimenti delle analisi dei dati sviluppati nei prossimi capitoli.

### **2.3 Destagionalizzazione diretta ed indiretta: metodi e test per il loro confronto**

In molti casi, le serie storiche sono costruite come la somma di molte serie componenti; ci sono casi in cui la somma può avvenire in modo “orizzontale”, ad esempio come somma di provincie, o come aggregazione “verticale”, ovvero come somma di settori o fattori di produzione. È importante dire che bisogna analizzare il grado di aggregazione delle serie da destagionalizzare, perché se queste fossero troppo disaggregate, i movimenti irregolari della serie, ovvero quella parte della serie che viene considerata rumore, potrebbero essere piuttosto grandi rispetto ai fattori stagionali. Un processo di destagionalizzazione accettabile deve rispondere ad alcuni requisiti: prima di tutto la serie non deve più avere stagionalità residua, inoltre la serie deve essere stabile, non deve esserci bisogno di troppe modifiche qualora il modello di aggiustamento stagionale fosse riapplicato a nuove osservazioni aggiunte in un secondo momento.

Per destagionalizzare una serie si può operare in tre modi: in modo diretto, in cui la serie di interesse è prima costruita attraverso l’aggregazione per righe e viene destagionalizzata solo successivamente; in modo indiretto, in cui prima si destagionalizzano le serie storiche componenti (ad esempio, come si diceva prima, le provincie), tutte con lo stesso metodo e poi si ottiene l’indicatore cercato mediante l’aggregazione di tutte le serie già destagionalizzate; c’è poi un terzo modo, definito dagli studiosi come approccio spurio, in cui ogni serie viene

destagionalizzata, utilizzando strategie e metodi diversi e successivamente si ottiene l'indicatore aggregando le varie serie. Dalla descrizione di quest'ultimo, si capisce che anche questo metodo può essere considerato come destagionalizzazione indiretta, ma per la sua costruzione, in cui come si diceva prima si utilizzano tecniche differenti per le varie serie, viene ritenuto un metodo poco preciso e quindi poco usato. Per questo motivo ci soffermeremo sul confronto tra i primi due metodi. Se si sceglie di destagionalizzare in modo diretto, l'aggregato che ne risulta potrebbe non coincidere con la somma delle sue componenti destagionalizzate singolarmente. La destagionalizzazione diretta si comporta meglio quando tutte le componenti della serie mostrano pattern stagionali simili. Quando non è così è preferibile utilizzare la destagionalizzazione indiretta.

Come si può capire da questa breve premessa, la scelta su quale metodo utilizzare è stata al centro di molte discussioni, caratterizzate da pareri discordanti e non si è ancora giunti ad una scelta definitiva riguardo quale sia il metodo migliore. C'è anche da sottolineare che la scelta tra i due metodi non può basarsi solo su considerazioni statistiche. Esistono, ad esempio, casi in cui gli uffici statistici lavorano con serie di dati in cui sono presenti valori mancanti. L'approccio indiretto implica una stima dei valori mancanti e questo comporta una serie di modelli da prendere in considerazione, da controllare e da sottoporre a verifica, per garantire che la stima di questi dati avvenga in modo corretto. Il metodo di destagionalizzazione diretta ha un'applicazione facile e quindi molto comune, salvo casi in cui non ci sia una forte evidenza a favore del metodo indiretto.

C'è infine da segnalare che l'additività nei data set aggiustati può essere mantenuta con entrambi i metodi. La destagionalizzazione indiretta la mantiene per costruzione, la diretta la conserva, imponendo come vincolo di somma il valore totale precedentemente destagionalizzato.

Considerando la serie di un aggregato economico, questa può quindi essere ottenuta con tre differenti strategie:

- Con l' **Approccio diretto**, in cui si aggiusta direttamente l'aggregato; quest'ultimo, dopo la destagionalizzazione, lo indicheremo con  $A_T^D$ .

- Con l' **Approccio Indiretto**, in cui ogni componente della serie aggregata viene prima destagionalizzata, ricordiamo, con lo stesso metodo e lo stesso software, e la serie aggregata aggiustata viene quindi costruita come aggregazione delle singole serie. Questo equivale a dire che la serie può essere calcolata come:

$$A_T^I = \sum_{n=1}^N A_{n,j}$$

o come

$$A_T^I = \sum_{n=1}^N \omega_n A_{n,j}$$

con  $A_T^I = \sum_{n=1}^N A_{n,j}$  che evidenzia come la serie ottenuta per via indiretta è data dalla somma fatta sugli N istanti temporali della serie j; nella seconda formula si sono aggiunti i pesi  $\omega_n$ , attribuiti alle singole serie.

- Con un approccio sempre indiretto considerato **spurio**, in cui ogni serie viene destagionalizzata con un proprio metodo ed un proprio software.

Dobbiamo infine citare anche un approccio multivariato, che permette di ottenere contemporaneamente sia le serie destagionalizzate per l'aggregato che quelle per le sue componenti.

L'approccio spurio indiretto è una strategia molto popolare, ma a causa della difficoltà del processo di aggiustamento stagionale che ne sta alla base, non è particolarmente studiata. Le politiche di aggiustamento stagionale, seguendo la distinzione di Astolfi, Ladiray e Mazzi (2001), possono quindi essere riassunte nei seguenti punti:

- I metodi sono spesso diversi. Alcuni stati utilizzano un approccio di analisi basato sui modelli  $ARIMA^1$ , utilizzando software come *TRAMO-SEATS* o

---

<sup>1</sup>Si vedano i siti <http://www.bde.es/servicio/software/softwaree.htm> e <http://ftp.census.gov/pub/ts/x12a/final/pc/>.

*STAMP*<sup>2</sup>; altri, come quelli europei, utilizzano approcci non parametrici, facendo ricorso ai software da X-11 in poi, secondo il seguente ordine:

- Gli stati membri decidono da soli se destagionalizzare in modo diretto o indiretto
- La strategia per correggere gli effetti di calendario solitamente non è la stessa.

Per questi motivi è difficile paragonare l'approccio spurio di destagionalizzazione con gli altri due metodi. Quindi ci concentreremo sulla descrizione dei primi due metodi, tralasciando la tecnica di destagionalizzazione spuria.

A questo punto ci poniamo una domanda: esistono casi in cui l'approccio diretto e indiretto coincidono? In generale la risposta è no, specialmente nel caso di aggregati ottenuti come combinazione lineare delle varie componenti e in cui l'aggiustamento stagionale è un processo non lineare. La risposta potrebbe essere positiva solo sotto restrizioni molto forti; ad esempio se l'aggregato fosse una somma algebrica, se il modello venisse scomposto in modo additivo, se non ci fossero *outlier* nella serie e se il filtro usato nella destagionalizzazione fosse lo stesso per tutte le serie, allora i due metodi potrebbero essere considerati uguali. Queste condizioni risultano però poco consone alla realtà; infatti già il considerare il modello come additivo non rispecchia completamente ciò che accade nella maggior parte dei casi. Di solito infatti si tende a considerare modelli moltiplicativi. A riguardo bisogna dire che il modello moltiplicativo potrebbe diventare additivo utilizzando una trasformazione logaritmica; questa però non assicura l'equivalenza tra le due forme del modello. È poco vicino alla realtà anche il pensare che non ci siano *outliers*; questi infatti sono una componente molto frequente delle serie storiche, generati da modifiche strutturali, condizioni anomale e cambiamenti dovuti a fattori esterni. Infine, neanche pensare all'utilizzo di uno stesso filtro nella

---

<sup>2</sup> Si veda il sito <http://stamp-software.com/>

destagionalizzazione rispecchia la realtà. Spesso, soprattutto nei metodi di destagionalizzazione *model-based*, si associano a diverse serie filtri diversi, rendendo molto difficile l'analogia tra metodo diretto ed indiretto.

Soltanto in casi in cui alcuni blocchi di serie non abbiano troppo rumore al loro interno, dimostrino di avere trend simili, pattern stagionali simili e che siano affetti dagli stessi shock esterni, allora i due metodi di destagionalizzazione diretta ed indiretta possono essere considerati simili.

In generale, quindi, i metodi di destagionalizzazione diretta ed indiretta non possono essere considerati simili. Questo genera quindi il problema di riuscire a capire quale dei due metodi vada meglio per trattare i vari casi di serie storiche. Questa scelta viene fatta sia in base ad alcune considerazioni statistiche, sia in base a valutazioni non statistiche ed ad alcune proprietà a priori che risultano auspicabili per le serie prese in considerazione. Per questo motivo è utile capire, sempre seguendo il lavoro di Astolfi, Ladiray e Mazzi(2001), quando sia possibile scegliere fin dall'inizio quale metodo utilizzare e quando invece si renda necessaria un'analisi statistica per capire quale approccio sia più conveniente.

- I vincoli additivi e l'approccio indiretto

Per alcune serie, l'aver vincoli additivi risulta di fondamentale importanza. Come detto in precedenza, questa è una delle condizioni in cui si potrebbe pensare di applicare un approccio indiretto, approccio che risulta utile anche per il suo essere consistente riguardo ai tassi di crescita. Comunque anche il metodo diretto potrebbe andare bene, a patto che assicuri il vincolo di additività, magari attraverso l'utilizzo di appropriati metodi statistici.

- Alcune considerazioni statistiche



La scelta del metodo di destagionalizzazione può avvenire anche sulla base di alcune considerazioni statistiche. Ad esempio, se le componenti di una serie hanno caratteristiche diverse, o se l'importanza, in termini di pesi, delle singole serie cambia molto rapidamente nel tempo, allora è preferibile utilizzare l'approccio indiretto. Se le serie risultano simili ed hanno timing piuttosto simili, allora risulta più appropriato l'approccio indiretto.

Proprio per questa doppia possibilità, è stato studiato anche un approccio di tipo "misto", in cui prima si applica il metodo diretto a gruppi di serie che si dimostrano simili, secondo le considerazioni dette prima e poi si applica quello indiretto per la stima dell'aggregato finale destagionalizzato. Si può quindi capire che l'approccio diretto è più adatto ad aggregazione di tipo "orizzontale", quello indiretto per aggregazione di tipo "verticale".

### **2.3.1 Metodologie**

A questo punto dobbiamo però fornire una breve descrizione dei metodi attualmente disponibili per giudicare in modo empirico quale tra i due principali approcci sia quello migliore. Per poter utilizzare i metodi di analisi, bisogna prima risolvere alcuni problemi. Innanzitutto l'effetto della strategia di aggregazione deve essere isolato da altre modifiche apportate alla serie; inoltre ci sono dei piccoli problemi anche nella definizione stessa degli indicatori che vengono utilizzati per giudicare la bontà della destagionalizzazione; infine c'è anche da segnalare un problema puramente computazionale, visto che nessuno dei due principali metodi, *X-12-ARIMA* e *TRAMO-SEATS*<sup>3</sup> fornisce un set comune di statistiche che possano risolvere il problema. Per questo vediamo di elencare alcuni problemi e alcune tecniche per risolverli.

---

<sup>3</sup>Si vedano i siti <http://www.bde.es/servicio/software/softwaree.htm> e <http://ftp.census.gov/pub/ts/x12a/final/pc/>.

- **Vari casi di revisione**

Le serie destagionalizzate sono spesso soggette ad operazioni di revisione; questo accade perché frequentemente le serie mancano di alcuni dati, oppure per il cambiamento nelle politiche nazionali o comunitarie, o ancora per la modifica dei parametri di aggiustamento, delle metodologie per il trattamento degli *outliers*, degli effetti di calendario etc.

Per tali motivi si capisce perché, dovendo studiare la differenza tra i due metodi, risulti fondamentale isolare i cambiamenti nelle serie, cambiamenti dovuti solo ai problemi descritti, per poi lavorare con serie che risultino stabili, con assunzioni forti; bisogna prima eliminare dalla serie problemi come gli effetti di calendario o presenza di *outlier* e poi iniziare con la destagionalizzazione. Bisogna comunque dire che, più che l'applicare le operazioni descritte, il difficile è misurare l'importanza relativa di ogni fonte di revisione.

- **Misure di qualità (Astolfi, Ladiray e Mazzi, 2001)**

In letteratura ci sono molti test per valutare la qualità di una destagionalizzazione, visto che non esiste un metodo che riesca a spiegare più degli altri la qualità di un aggiustamento stagionale. C'è comunque da dire che alcuni aspetti della destagionalizzazione possono essere individuati, e per valutarne la bontà sono stati indicati alcuni criteri.

Per questi motivi cerchiamo di vederne alcuni utilizzati per valutare quale tra i due metodi descritti meglio si adatti alla destagionalizzazione.

**1) I metodi sono davvero così diversi?**

Per valutare se effettivamente i due metodi, diretto ed indiretto, siano diversi, si utilizza la media ed il massimo della serie delle deviazioni percentuali, calcolate come:  $[1 - \left(\frac{A_t^U}{A_t^I}\right)]$ , dove si utilizza la notazione vista in precedenza. Questa può essere calcolata per ogni possibile coppia di approcci (considerando quindi

anche un approccio spurio), per la serie destagionalizzate, per le stime di trend e ciclo, e per le componenti stagionali.

## 2) Inconsistenza

Per misurare il grado di consistenza del tasso di crescita, vengono calcolate due statistiche: come sottolineano Feldpausch, Hood, e Wills (2003), la statistica C1 calcola la percentuale totale della concordanza tra le serie aggiustate in modo diretto e quelle aggiustate in modo indiretto; la statistica C2 misura la percentuale di concordanza tra le serie aggiustate stagionalmente e le serie nazionali aggiustate.

## 3) Qualità dell'aggiustamento stagionale

X-12-ARIMA<sup>4</sup> propone un set di 11 statistiche M ed una statistica Q (Ladiray, Quenneville, 2001), calcolata come media ponderata delle 11 M, per verificare la qualità dell'aggiustamento stagionale. Queste statistiche, in alcuni casi, possono essere utilizzate anche se si utilizza TRAMO-SEATS.

## 4) Bontà dell'aggiustamento stagionale

Dagum (1979) propone due formule per verificare quanto siano accurati i risultati ottenuti dalle procedure di destagionalizzazione. Il primo è calcolato come:

$$R_1 = \sum_{t=2}^T (A_t - A_{t-1})^2 = \sum_{t=2}^T (\nabla A_t)^2.$$

Il secondo è invece basato sulla media mobile a 13 termini di Henderson, ed è calcolato come:

$$R_2 = \sum_{t=1}^T (A_t - H_{13}A_t)^2 = \sum_{t=1}^T (I - H_{13})A_t]^2.$$

## 5) Stabilità della destagionalizzazione

---

È importante sottolineare che in un processo di destagionalizzazione non dovrebbe rimanere stagionalità residua e tutti gli effetti di calendario dovrebbero essere eliminati dalla serie aggiustata. Un processo che non elimina tutti gli effetti di stagionalità è da considerarsi poco soddisfacente. Anche se tutti gli effetti di stagionalità e di calendario dovessero essere eliminati, si potrebbe comunque avere una destagionalizzazione poco accettabile. Infatti si può considerare accettabile il processo di aggiustamento della serie, quando la serie non cambia troppo nel momento in cui vengono inseriti valori che prima non erano disponibili e che quindi venivano stimati. Le revisioni frequenti possono causare a colui che studia serie storiche, una perdita di fiducia nell'utilità dei dati destagionalizzati. L'instabilità inoltre, potrebbe essere dovuta alla presenza di un'elevata variabilità stagionale o di forti movimenti nel trend della serie da destagionalizzare. Per questo è fondamentale riuscire a misurare la stabilità di una serie. *X-12-ARIMA* ha due diagnostiche per la stabilità: *sliding spans* e revisioni storiche (Findley, Monsell, Bell, Otto, Chen, 1998). Bisogna infine dire che le componenti irregolari non dovrebbero mostrare alcun segno di stagionalità o di particolare struttura residua. Sia in *TRAMO-SEATS* che in *X-12-ARIMA*<sup>5</sup> ci sono delle procedure automatiche per il controllo della componente irregolare, per permettere all'utente di analizzare questa componente e verificare la bontà del processo utilizzato per destagionalizzare.

---

<sup>5</sup> Si vedano i siti <http://www.bde.es/servicio/software/softwaree.htm>, <http://ftp.census.gov/pub/ts/x12a/final/pc/>.

## CAPITOLO 3

# IL BILANCIAMENTO DI MATRICI

L'errore presente nei modelli economici e le relazioni che i dati devono rispettare creano alcuni problemi in fase di studio di dati reali

In particolare, le relazioni, o vincoli, spesso non sono rispettate, poiché i dati subiscono vari passaggi durante le rilevazioni. Questo causa stime differenti delle stesse variabili, o, più in generale, la caduta di restrizioni lineari che in realtà dovrebbero essere soddisfatte.

Per i motivi descritti sopra, si rendono necessarie procedure di riconciliazione per fare in modo che i vincoli vengano rispettati. All'inizio la principale importanza della riconciliazione dei dati era stata data al problema dell'interpolazione delle tabelle input-output, per il quale il metodo RAS (Stone, 1961), che tra un po' descriveremo, è più utile da un punto di vista computazionale rispetto alla correzione con i minimi quadrati. Con l'evoluzione dei metodi però, ci si è accorti che per la riconciliazione meglio si adattano i minimi quadrati, spostando l'attenzione su nuove tecniche. Lo sviluppo di queste ultime si deve soprattutto al fatto che il metodo RAS, nella sua formulazione più classica, non considera alcuna misura di accuratezza preliminare delle stime.

Senza andare troppo nei dettagli, è sufficiente dire che per molto tempo le procedure si basavano sull'assunzione che la matrice di varianza e covarianza (o un altro indicatore di accuratezza delle stime) dei dati da riconciliare fosse nota. In questo caso, i dati erano aggiustati alla luce delle loro relative varianze così da soddisfare i vincoli lineari. Successivamente però ci si è posti un'altra domanda: se non dovessero esserci misure utilizzabili né provenienti da indagini né di natura soggettiva? In questo caso si deve ricorrere ad una soluzione a minimi quadrati, basandosi sostanzialmente su due approcci ( Di Fonzo, Marini, 2005):

1) Soluzioni matematiche: il data set è bilanciato minimizzando un criterio che comporta una matrice di covarianza (che è un semplice artefatto statistico)

2) Soluzioni data-based: la variabilità dei dati da riconciliare è stimata usando le osservazioni utilizzabili

Dopo questa introduzione, occupiamoci ora di definire meglio il problema ed i metodi utilizzati per risolverlo.

### 3.1 Il problema: i vincoli contemporanei e temporali

Il problema del bilanciamento delle matrici, data la sua importanza, ha attirato l'attenzione in vari ambiti, come l'economia, la statistica, la demografia, ma anche la pianificazione urbana e la modellazione stocastica (Schneider, Zenios, 1990).

Sintetizzando il problema, data una matrice Y con m righe e n colonne, come descritto in precedenza, si rende necessario il rispetto di due tipi di vincoli, contemporanei e temporali. Per capire meglio il problema passiamo ad una descrizione in termini matriciali.

Come detto si parte da una matrice Y, di dimensioni  $(n \times m)$ ,

$$\begin{bmatrix} y_{11} & \dots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \dots & y_{nm} \end{bmatrix}$$

Questa matrice deve rispettare dei vincoli, quelli che prima sono stati definiti contemporanei e temporali. Per capire meglio il significato di questi due vincoli, consideriamo prima un vettore, u, tale che  $u = [u_1, \dots, u_m]$ ; Consideriamo ora un vettore v, tale che  $v = [v_1, \dots, v_n]$ . Quello che si vuole ottenere è una nuova

matrice, chiamiamola  $Y^*$ , di dimensione  $(n \times m)$ , costruita in modo che  $\sum_{j=1}^m y_{t,j}^* = u_t$  per  $t=1, \dots, n$  e che  $\sum_{j=1}^m y_{i,j}^* = v_i$  per  $j=1, \dots, n$ .

C'è da dire che il problema potrebbe assumere caratteristiche diverse. Considerando infatti il caso  $n = m$ , l'obiettivo diventa trovare una matrice  $Y^*$ , di dimensione  $n \times n$ , tale che  $\sum_{j=1}^n y_{t,j}^* = \sum_{j=1}^n y_{i,j}^* = \sum_{i=1}^n y_{i,j}^*$

Sicuramente, per entrambi i problemi devono essere poste alcune restrizioni nei confronti degli aggiustamenti che possono essere fatti riguardo alla matrice  $Y$ , così che la richiesta di una matrice simile a quella di partenza sia ben definita. Nell'analisi input-output, per esempio, se una matrice di coefficienti diretti deve essere aggiornata/progettata, le matrici  $Y$  e  $Y^*$  e i vettori  $u$  e  $v$  devono essere non negativi ed è solitamente richiesta una condizione di "salvaguardia del segno" ( $y_{i,j}^* > 0$  solo se  $y_{i,j} > 0$ ). Questo non è il caso, come sempre, delle tabelle di contabilità nazionale, dove possono capitare entrate negative.

Diversi approcci algoritmici seguono naturalmente da diversi tipi di restrizioni. Inoltre, dovrebbe essere notato che, astraendosi da una specifica applicazione, un problema di riconciliazione dei dati può essere definito come un problema di programmazione matematica.

Le procedure per il bilanciamento di matrici possono essere divise in due classi:

- Gli algoritmi proporzionali come il RAS
- Gli algoritmi non proporzionali, che hanno come funzione il minimizzare la distanza tra la matrice  $Y$  e tutti gli elementi di una matrice candidata ad essere quella bilanciata. Detto in altri termini, questi algoritmi minimizzano una funzione di perdita, dove le condizioni di bilanciamento sono vincoli nel modello di ottimizzazione; questo fa in modo che la soluzione ottimale sia la matrice bilanciata più vicina a quella di partenza, rispettando il metro introdotto dalla funzione di perdita.

### 3.2 Procedure per il bilanciamento: ottimizzazione RAS (Stone, Toh, 1981)

La procedura di ottimizzazione RAS, per bilanciare le matrici, moltiplica in modo iterativo, per costanti positive, le righe e le colonne della matrice di partenza. L'obiettivo è quello di procedere con questi prodotti fino a trovare la matrice bilanciata. Con una descrizione un po' più astratta (Uribe, de Leeuw, Theil, 1970), si può dire che la procedura RAS genera una soluzione equivalente per minimizzare l'incremento dell'informazione di Kullback (Theil, 1967), data da:

$$\sum_{i=1}^m \sum_{j=1}^n y_{ij} \log \frac{y_{ij}^*}{y_{ij}}$$

Quest' algoritmo può essere utilizzato per risolvere il primo dei due problemi descritti. C'è però da dire che la procedura RAS mostra anche parecchi svantaggi (Sturaro, 2009); tra questi si devono ricordare la sua impossibilità di bilanciare qualcosa di diverso dai totali di riga e di colonna, il fatto che utilizzi solo somme di variabili e non loro combinazioni lineari, il non considerare alcuni valori più affidabili degli altri e il non poter essere applicata a valori negativi.

Per ovviare a questi problemi, si sono evolute nel tempo nuove tecniche, che riescono a fornire soluzioni più generali. Come detto, queste si basano su funzioni di perdita da minimizzare, garantendo che la matrice bilanciata sia più vicina possibile a quella di partenza.

### 3.3 Adattamento con i minimi quadrati di dati economici soggetti a restrizioni lineari

Sia  $\mathbf{y}$  un vettore di dimensioni  $(m \times 1)$  di dati non noti che dovrebbero soddisfare il sistema di restrizioni di contabilità indipendenti

$$A\mathbf{y} = \mathbf{a} \quad (1)$$



dove  $\mathbf{A}$  è una matrice  $k \times mn$ , con  $\text{rk}(\mathbf{A}) = k < mn$ , e  $\mathbf{a}$  è un vettore  $(k \times 1)$  di costanti note. Sia  $\mathbf{y}$  un vettore  $(mn \times 1)$  di dati osservati, che non soddisfa la restrizione lineare precedente, collegato ad  $\hat{\mathbf{y}}$  attraverso il modello lineare

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e} \quad (2)$$

dove  $\mathbf{e}$  è un vettore di disturbo di dimensioni  $(mn \times 1)$ , con media zero e matrice di varianze e covarianze nota  $\mathbf{V}$ .

è stato dimostrato che uno stimatore efficiente  $\mathbf{y}^*$  di  $\hat{\mathbf{y}}$  che minimizzi una funzione criterio del tipo:

$$(\mathbf{y}^* - \mathbf{y})' \mathbf{Q}^{-1} (\mathbf{y}^* - \mathbf{y}) \quad (3)$$

per  $\mathbf{Q} = \mathbf{V}$  e soddisfi il vincolo (2), è dato da:

$$\mathbf{y}^* = \mathbf{y} + \mathbf{V} \mathbf{A}' (\mathbf{A} \mathbf{V} \mathbf{A}')^{-1} (\mathbf{a} - \mathbf{A} \mathbf{y}) \quad (4)$$

$$\text{con } E(\mathbf{y}^* - \hat{\mathbf{y}}) = \mathbf{V} - \mathbf{V} \mathbf{A}' (\mathbf{A} \mathbf{V} \mathbf{A}')^{-1} \mathbf{V}$$

A differenza del metodo RAS, la procedura trovata da Stone *et al.* nel 1942 (Stone, Champerone, Meade, 1942) bilancia automaticamente la matrice quando le righe e le colonne sono formate da variabili endogene.

Un altro grande vantaggio di questo algoritmo è il suo essere basato sul principio dei minimi quadrati e pertanto si appoggia su una lunga e solida tradizione statistica. È sicuramente pregevole il fatto che i dati siano corretti alla luce delle loro relative varianze così da soddisfare le restrizioni lineari. Inoltre, sotto l'assunto di normalità, le stime riconciliate sono di massima verosimiglianza (Weale, 1985, 1988) e quindi possono essere applicate le procedure standard di inferenza statistica.

È inoltre piuttosto usuale semplificare l'assunzione sul termine d'errore e considerarlo indipendente, esattamente come fa Stone (1990). In questo caso la matrice di varianza e covarianza  $\mathbf{V}$  risulta diagonale. Considerando la definizione di

probabilità come un “grado di ragionevole convinzione”, Stone propone di determinare la varianza di ogni singola  $y_t$  come  $v_{it} = (\vartheta y_t)$ , per  $t = 1, \dots, mn$ , dove  $\vartheta$  è una valutazione determinata in modo soggettivo, che esprime il rapporto percentuale dello standard error con il precedente valore medio di  $y_t$ . Il principale inconveniente da un punto di vista concettuale è che non garantisce il mantenimento del segno delle variabili, che a volte può risultare problematico. Un'altra possibile debolezza della procedura dipende dall'assunzione che la matrice di varianza e covarianza delle varie stime sia nota.

### 3.4 Bilanciamento di un sistema di serie temporali con vincoli esogeni

Il problema del bilanciamento di serie temporali riguarda la stima diretta di un sistema di variabili anziché una singola. Più precisamente, si vogliono stimare  $M$  vettori ignoti con  $n$  righe ed una colonna di dati ad alta frequenza, anche riguardanti una serie temporale  $M$  di dati disaggregati, che devono soddisfare contemporaneamente i vincoli di aggregazione sia temporali che contemporanei.

L'informazione fondamentale è data da  $M + 1$  vettori aggregati:

1.  $Z$ , vettore di dati aggregati contemporanei di dimensione  $(n \times 1)$
2.  $y_{0j}$ ,  $j = 1, \dots, M$ , vettore di dati aggregati temporalmente (ad esempio dati annuali), di dimensione  $(N \times 1)$ .

Consideriamo il caso in cui la serie preliminare ad alta frequenza  $M$  (ad esempio una serie trimestrale),  $y_j$ ,  $j = 1, \dots, M$ , sia disponibile, dove  $\sum_{j=1}^M y_j \neq z$  e/o  $y_j$  che non si attiene a  $y_{0j}$ .

Chiamando  $y_j^*$ ,  $j = 1, \dots, M$  il vettore  $(n \times 1)$  di dati riconciliati da stimare, devono essere validi i seguenti vincoli di contabilità:

$$\sum_{j=1}^M y_j^* = z \quad (6)$$

$$y_j^* = y_{0j}, j = 1, \dots, M \quad (7)$$

dove  $J$  è la matrice di aggregazione dimensione  $(N \times n)$  che trasforma i dati ad alta frequenza in dati a bassa frequenza. Ogni elemento di  $y_{0j}$  può essere visto come una combinazione lineare di  $y_j^*$ , con coefficienti dati dal vettore  $j$  di dimensione  $(s \times 1)$ , dove  $s$  è l'ordine di aggregazione temporale.

Quindi possiamo definire matrice  $J$  pari a  $[I_N \otimes J' \mid \mathbf{0}]$ , dove  $\mathbf{0}$  è una matrice nulla  $(N \times (n - sN)M)$ .

Il vincolo (6) può essere riscritto come  $(\mathbf{1}'_M \times In)y^* = z$ , dove  $\mathbf{1}'_M$  è un vettore unitario di dimensioni  $(M \times 1)$  e  $y^* = [y_1^* \dots y_j^* \dots y_M^*]$ . Così come i vincoli di aggregazione temporale (7) sono costruiti, si ha

$$(I_N \times J)y^* = y_0 \text{ Con } y_0 = [y_{01} \dots y_{0j} \dots y_{0M}]'$$

Sia  $H$  la matrice di dimensione  $((n + NM) \times nM)$  composta da:

$$H = \begin{bmatrix} \mathbf{1}'_M \otimes I_n \\ I_M \otimes J \end{bmatrix}$$

e sia  $y_a$  il vettore definito come  $[z' y_0']$  di dimensione  $((n + NM) \times nM)$ , che contiene sia gli aggregati contemporanei sia quelli temporali. Il set completo di restrizioni tra i valori riconciliati e l'informazione aggregata utilizzabile può essere espressa come una matrice nella forma:

$$Hy^* = y_a.$$

Da notare che, l'aggregazione contemporanea di serie aggregate temporali implica che:

$$\sum_{j=1}^M y_{0jT} = \sum_{h=1}^s z_{s(T-1)+h} = z_{0,T}$$

Cioè, in forma matriciale

$$\mathbf{1}'_M \otimes \mathbf{I}_N = \mathbf{Jz},$$

Dove  $\mathbf{J}$  è la matrice di aggregazione temporale con  $\mathbf{j} = \mathbf{1}_s$ . Dalla relazione precedente segue che, solo  $r$  osservazioni aggregate, con  $r = n + N \times (M - 1)$ , sono “libere”, mentre  $N$  osservazioni sono ridondanti, quindi la matrice  $\mathbf{H}$  ha rango  $r$ .

Senza perdita di generalità, si può considerare il vettore  $(r \times 1) \mathbf{w}$ , cioè semplicemente il vettore aggregato  $\mathbf{y}_a$ , privato delle sue ultime  $M$  righe. Con facili operazioni, si può dimostrare come la restrizione che opera sul vettore  $\mathbf{y}^*$  in termini di matrice  $\mathbf{H}$  – che non ha rango pieno – possa ora essere espressa come un sistema di restrizioni lineari indipendenti come in (2) con  $\mathbf{A} \equiv \mathbf{H}_w$ , invece di  $\mathbf{y}$  e  $\alpha \equiv \mathbf{w}$ , cioè

$$\mathbf{H}\mathbf{y}^* = \mathbf{w}.$$

## CAPITOLO 4

# IL BENCHMARKING DI SERIE STORICHE: DESCRIZIONE DEI PRINCIPALI MODELLI

Dopo aver visto i principali metodi per il bilanciamento di serie storiche, passiamo adesso ad occuparci del problema del benchmarking di serie storiche. Ricordando quanto detto in precedenza, si ha un problema di benchmarking quando una serie storica di dati, per le stesse variabili obiettivo, è misurata ad intervalli differenti; questo naturalmente comporta diversi livelli di accuratezza. Infatti, descrivendo in altri termini il problema, spesso si hanno serie mensili da una fonte e annuali da un'altra. Le serie con maggiore frequenza sono meno utilizzabili delle altre. Per questo quelle a minor frequenza sono di solito considerate come serie di riferimento. Le discrepanze sono di solito osservate tra i benchmark annuali e la somma annuale delle serie più frequenti. Il benchmark combina l'efficacia delle due fonti di dati. utilizzando una notazione convenzionale, le serie sub-annuali possono essere scritte come  $s_t$  e le annuali come  $a_m$ . C'è da sottolineare che, oltre a serie sub-annuali, quindi come detto di durata inferiore all'anno, e annuali, ci sono anche casi in cui le prime potrebbero essere annuali e le altre quinquennali o decennali.

Nel considerare le tecniche di benchmarking, bisogna ricordare che la serie di riferimento annuale è utilizzabile soltanto parecchi mesi dopo la fine dell'anno

stesso. Questo vuol dire che non ci sono valori di riferimento per l'anno corrente, implicando la presenza di un dato mancante nelle serie.

I metodi più utilizzati tengono conto di questo problema, prevedendo tecniche di aggiustamento per l'anno corrente senza considerare un valore come riferimento, assicurando continuità tra le stime dell'anno precedente e quelle dell'anno corrente.

Dopo questa breve introduzione, passiamo ora a vedere i principali metodi di benchmarking, partendo da quello di Cholette e Dagum. Questo infatti può essere considerato un caso generale degli altri metodi sviluppati in seguito. I metodi vengono detti basati sulla regressione perché il modello è quello di una regressione nella sua forma più classica.

#### 4.1 METODO DI BENCHMARKING DI CHOLETTE E DAGUM: IL MODELLO ADDITIVO (Dagum e, Cholette, 2006)

Il modello di Cholette e dagum consiste nelle seguenti due equazioni:

$$s_t = \sum_{k=1}^H r_{tk} \beta_k + \theta_t + \varepsilon_t = \sum_{k=1}^H r_{tk} \beta_k + \theta_t + \sigma_t^\lambda \varepsilon_t^\dagger, \quad t = 1, \dots, T \quad (8)$$

$$\alpha_m = \sum_{t=i_{1m}}^{t_{1m}} j_{mt} \theta_t + \varepsilon_m \quad m = 1, \dots, M \quad (9)$$

Dove

$$E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_{t-l}) = \sigma_t^\lambda \sigma_{t-l}^\lambda \omega_l$$

e

$$E(\varepsilon_m) = 0, E(\varepsilon_m^2) = \sigma_{\varepsilon_m}^2, E(\varepsilon_m, \varepsilon_t) = 0$$

L'errore nella (9) ha un ruolo critico, poiché delinea il modo con il quale le discrepanze sono distribuite sulle osservazioni sub-annuali.

Quando  $\lambda = 0$ , l'errore è omoschedastico, altrimenti è eteroschedastico, con deviazione standard  $\sigma_t^2 - \lambda$  che può assumere valori 0,  $\frac{1}{2}$  o 1. Se  $\sigma_t^2$  è costante, il processo è omoschedastico, indipendentemente dal valore di  $\lambda$ .

L'errore è proporzionale a  $\sigma_t$  se  $\lambda = 1/2$  e a  $\sigma_t^2$  se  $\lambda = 1$ . Le autocorrelazioni dell'errore standardizzato sono quelle di un ARMA o del primo o del secondo ordine.

L'equazione (9) mostra che i benchmark  $a_m$  sono uguali alle somme temporali dei veri valori sub-annuali  $\sigma_t$ , sul periodo di riferimento, più un termine d'errore non correlato con varianza eteroschedastica.  $e_t$  ed  $e_m$  sono assunti essere mutualmente incorrelati, essendo i dati sub-annuali e annuali provengono da due fonti differenti. Il benchmark  $a_m$  è pienamente affidabile e vincolante se non è soggetto ad errore ( $\sigma_{e_m}^2 = 0$ ). Se questa uguaglianza non viene rispettata, allora la serie di riferimento non è più attendibile. In questo caso non si hanno misurazioni di riferimento, ma semplicemente misurazioni meno frequenti di una variabile obiettivo. Quando il benchmark non è vincolante, combina in modo ottimale le osservazioni annuali e sub-annuali. Con benchmark vincolanti, corregge le osservazioni sub-annuali al riferimento annuale.

#### 4.1.1 Una specificazione conservativa di trend deterministici

Nel caso del benchmarking, il regressore deterministico è tipicamente una costante che cattura il livello medio di differenza tra i dati annuali e i sub-annuali. Questa è sostanzialmente una media ponderata delle discrepanze. In un contesto survey, il parametro di regressione stimato  $\hat{\beta}_1$  è spesso chiamato "bias", distorsione; questo accade quando i dati sub-annuali sono ritenuti essere distorti rispetto al benchmark annuale dovuto alla sotto copertura del survey.

In alcuni casi c'è bisogno di un secondo regressore per catturare un trend deterministico nelle discrepanze. Chiaramente questo regressore deve essere conservativo del trend. Questa procedura ha però un grande svantaggio; inserire una funzione del tempo di grado maggiore di uno comporta un comportamento esplosivo alla fine delle serie temporali e subiscono pesanti revisioni quando si aggiungono nuovi dati. Da notare inoltre che, se si moltiplica l'andamento di un regressore per un coefficiente positivo, non si fa altro che ampliare o ridurre lo stesso movimento, in base al valore del coefficiente. In particolare, se il regressore considerato si stabilizza al tempo  $t$ , anche il prodotto del regressore per il coefficiente si stabilizza al tempo  $t$ .

#### 4.1.2 Rappresentazione matriciale del modello

Le equazioni (8) e (9) possono essere scritte in forma matriciale come:

$$s = R\beta + \theta + \varepsilon,$$

Con  $E(\varepsilon) = \mathbf{0}$ ,  $E(\varepsilon\varepsilon') = V_\varepsilon$

$$a = J\theta + \varepsilon,$$

con  $E(\varepsilon) = \mathbf{0}$ ,  $E(\varepsilon\varepsilon') = V_\varepsilon$ ,  $E(\theta\theta') = \mathbf{0}$ ,

dove  $s = [s_1 \dots s_T]'$ ,  $\theta = [\theta_1, \dots, \theta_T]'$ ,

$a = [a_1, \dots, a_m]'$  e così per altri vettori o matrici.

La matrice  $R$ , di dimensione  $(T \times H)$  contiene gli  $H$  regressori della matrice (8) e  $\beta = [\beta_1, \dots, \beta_H]$  gli  $H$  parametri di regressione. La matrice di varianza e covarianza  $V_\varepsilon$  delle serie sub-annuali considerate, dovrebbe derivare da un modello statistico applicato del benchmark, per esempio estrazione del segnale. Se una matrice di varianze e covarianze di questo tipo fosse utilizzabile, dovrebbe



essere usata nelle tecniche di benchmarking per preservare al massimo la struttura di dipendenza delle serie sub-annuali.

In molti casi la matrice di varianza e covarianza deve essere costruita da una funzione di autocorrelazione, e da una deviazione standard assunta costante o basata su un costante ( ovvero  $\theta_t = cxs_t$ ).

La matrice di varianza e covarianza ha quindi la struttura:

$$V_\varepsilon = \Sigma \Lambda \Sigma^T,$$

dove  $\Sigma$  è una matrice diagonale di deviazioni standard,  $\sigma_1, \dots, \sigma_T$  e  $\Lambda$  contiene le autocorrelazioni che solitamente corrispondono ad un modello ARMA con errore. Per un modello AR(1), con parametro  $|\varphi| < 1$ , gli elementi della matrice  $\Lambda$  sono dati da  $\omega_{ij} = \varphi^{|i-j|}$ . Per un modello AR(2), gli elementi  $\omega_{ij}$  sono dati in termini di ritardi  $l = |i - j|$ :  $\omega(l) = 1$ , per  $l = 0$ ,  $\omega(l) = \varphi_1 / (1 - \varphi_2)$ , per  $l = 1$ ,  $\omega(l) = \varphi_1 \omega_{l-1} + \varphi_2 \omega_{l-2}$ , per  $l = 2, \dots, T-1$ , dove  $|\varphi_1| < 1$ ,  $|\varphi_2| < 1$  e  $|\varphi_1 + \varphi_2| < 1$ , per garantire la stazionarietà. Il parametro

può essere uguale a  $0, 1/2, 1$ . Nel primo caso gli errori sono omoschedastici, negli altri due sono eteroschedastici.

#### 4.1.3 Operatori di somma temporale

La matrice temporale  $J$  è un operatore di somma temporale di dimensione  $M \times T$ , che contiene la frazione di copertura temporale:

$$J = \begin{bmatrix} j_{11} & j_{12} & \dots & j_{1T} \\ \vdots & \vdots & \ddots & \vdots \\ j_{M1} & j_{M2} & \dots & j_{MT} \end{bmatrix},$$

dove  $J_{mt} = \mathbf{0}$  per  $t < t_{1m}$  o  $t > t_{Lm}$  per  $t_{1m} \leq t \leq t_{Lm}$ . Ogni frazione di copertura  $J_{mt}$  è posta nella riga  $j$  e colonna  $t$  di  $J$ .

La matrice  $J$  di dimensione  $M \times T$  può essere utilizzata per riferimenti senza intervalli regolari e con differente durata.

#### 4.1.4 Soluzione del modello

Il modello (8) può essere scritto nella forma:

$$\begin{bmatrix} s \\ a \end{bmatrix} = \begin{bmatrix} R & I_T \\ \mathbf{0} & J \end{bmatrix} \begin{bmatrix} \beta \\ \theta \end{bmatrix} + \begin{bmatrix} e \\ \varepsilon \end{bmatrix};$$

O semplicemente come  $y = Xa + u$ ,  $E(u) = \mathbf{0}$ ,  $E(uu') = V_u = \text{block}(V_e V_\varepsilon)$ ,

dove si ha che:

$$y = \begin{bmatrix} s \\ a \end{bmatrix}, \quad X = \begin{bmatrix} R & I_T \\ \mathbf{0} & J \end{bmatrix}, \quad a = \begin{bmatrix} \beta \\ \theta \end{bmatrix}, \quad u = \begin{bmatrix} e \\ \varepsilon \end{bmatrix}, \quad V_u = \begin{bmatrix} V_e & \mathbf{0} \\ \mathbf{0} & V_\varepsilon \end{bmatrix}$$

e  $I_T$  è una matrice identità di dimensioni  $(T \times T)$ . Il modello che abbiamo appena presentato ha la forma di una regressione standard, con soluzione ai minimi quadrati generalizzati, data quindi da:

$$\hat{a} = (X'V_u^{-1}X)^{-1}X'V_u^{-1}Y \equiv \begin{bmatrix} \hat{\beta} \\ \hat{\theta} \end{bmatrix}, \quad e$$

$$\text{var}[\hat{a}] = (X'V_u^{-1}X)^{-1} \equiv \begin{bmatrix} \text{var}[\hat{\beta}] & \text{cov}[\hat{\beta} \ \hat{\theta}] \\ \text{cov}[\hat{\theta} \ \hat{\beta}] & \text{var}[\hat{\theta}] \end{bmatrix}.$$

I primi  $H$  elementi del vettore  $\hat{a} = [\hat{\beta}' \ \hat{\theta}']'$  contengono le stime dei parametri  $\hat{\beta}_1, \dots, \hat{\beta}_H$ ; gli ultimi  $T$  elementi della matrice sono invece le stime di  $\hat{\theta}_1, \dots, \hat{\theta}_T$ . In modo analogo le prime  $H$  righe e colonne della matrice  $\text{var}[\hat{a}]$  contiene la matrice di varianza e covarianza del vettore  $\hat{\beta}$ ; e le ultime  $H$  righe e colonne, la

matrice di varianza e covarianza del vettore  $\hat{\theta}$ . Da notare che la matrice  $\text{var}[\hat{a}]$  può essere ottenuta solo in caso in cui la matrice  $V_u$  invertibile, condizione che richiede che siano invertibili sia  $V_\varepsilon$  sia  $V_\xi$ . Quest'ultima condizione implica che tutti gli elementi diagonali della matrice  $V_\varepsilon$  devono essere maggiori di zero, ovvero il benchmark deve essere non vincolante.

In alcuni casi le matrici di varianza e covarianza di  $\varepsilon$  e di  $\xi$  sono note, quindi non c'è bisogno di stimare la varianza dei residui. Se le matrici non fossero note, bisognerebbe stimarle, ottenendo prima i residui, calcolandone la varianza e ristimando il modello, per ottenere intervalli di confidenza per le stime validi. Di sicuro, questo non è necessario se il metodo viene usato in modo numerico. In questo caso, in cui ogni varianza corrisponde alla varianza di un gruppo, si considerala media delle varianze originali e la media dei residui al quadrato, rispettivamente

$$\bar{\sigma}^2 (0) = \sum_{t=1}^T \sigma_t^2 / T, \quad \bar{\sigma}^2 (1) = \sum_{t=1}^T \hat{e}_t^2 / T.$$

Le deviazioni standard originali sono quindi moltiplicate per il rapporto  $\bar{\theta}^{(1)} / \bar{\theta}^{(0)}$  e il modello viene poi ristimato sulla base di  $\hat{V}_\varepsilon = \hat{\Sigma} \hat{\Omega} \hat{\Sigma}'$ .

Non vengono ristimate la varianze del benchmark, che sono di solito nulle o più piccole delle varianze sub-annuali. Con benchmark vincolanti, il ristimare il benchmark non cambia né le serie  $\hat{\theta}$  né i parametri  $\hat{\beta}$ , ma non cambia le loro varianze. È quindi sufficiente moltiplicare le matrici di varianza e covarianza  $\text{var}[\hat{\beta}]$  e  $\text{var}[\hat{\theta}]$  da  $(\bar{\theta}^{(1)} / \bar{\theta}^{(0)})^2$ .

La varianza di ogni combinazione lineare delle serie aggiustate e la sua varianza, sono date da:  $\hat{\theta}_L = L\hat{\theta}$ , e  $\text{var}[\hat{\theta}_L] = L\text{var}[\hat{\theta}]L'$ .

In particolare, la matrice  $L$  può essere un operatore di somma temporale, che considera le somme in base ad un anno o a trimestri basati sul calendario.

Guerrero, nel 1990 prima e nel 2000 poi, con Pena, ha fornito un test per verificare che il benchmark e la serie sub-annuale siano effettivamente compatibili. Continuando con la nostra notazione, il test consiste nel verificare se la statistica

$$G = (\mathbf{a} - \mathbf{J}\mathbf{s}^\dagger)' (\mathbf{J}\hat{\mathbf{V}}_e \mathbf{J}')^{-1} (\mathbf{a} - \mathbf{J}\mathbf{s}^\dagger)$$

Si distribuisca come un chi-quadro con  $M-H-L$  gradi di libertà, dove si ha che

$$\mathbf{s}^\dagger = \mathbf{s} - \mathbf{R}\hat{\boldsymbol{\beta}}$$

e  $L$  è il numero di parametri nei modelli ARMA. Se il valore di  $G$  è più piccolo del valore  $\chi^2_{M-H-L}(\alpha)$ , le serie sub-annuali e il benchmarking sono compatibili ad un livello di significatività  $\alpha$ . Il test assume quindi il benchmark, cioè  $\mathbf{V}_e = \mathbf{0}$  e la normale distribuzione degli errori.

Il modello di benchmarking additivo proposto da Cholette e Dagum, contiene pochi gradi di libertà: devono infatti essere stimati  $T+H$  parametri da  $T+M$  osservazioni. Questo vuol dire che i valori riferiti saranno molto vicini alle serie distorte-corrette. Questa "vicinanza" deve verificarsi per preservare l'andamento sub-annuale delle serie sub-annuali.

Inoltre, il fatto che  $\text{var}[\hat{\boldsymbol{\theta}}] = \mathbf{V}_e - \mathbf{V}_e \mathbf{J}' \mathbf{V}_e^{-1} \mathbf{J} \mathbf{V}_e$  mostra che, in assenza di un regressore deterministico, il valore della matrice delle serie riferite è più piccolo della matrice di varianza e covarianza originale, visto il sottrarre a queste ultime un valore positivo.

Il benchmarking, sia quello vincolante che il non vincolante, può essere interpretato soprattutto come un lisciamiento, in modo da ridurre la varianza delle serie aggiustate. D'altra parte l'equazione  $\text{var}[\hat{\boldsymbol{\theta}}] = \mathbf{V}_e - \mathbf{V}_e \mathbf{J}' \mathbf{V}_e^{-1} \mathbf{J} \mathbf{V}_e + \mathbf{W} \text{var}[\hat{\boldsymbol{\beta}}] \mathbf{W}'$ , mostra che in presenza di un regressore di tempo (cioè il caso in cui  $H > 0$ ), la  $\text{var}[\hat{\boldsymbol{\theta}}]$  può essere maggiore, uguale o minore di  $\mathbf{V}_e$ .

#### 4.1.5 Modello additivo con benchmarking proporzionale

Molte serie temporali hanno pattern stagionali con ampiezze tra loro differenti. Il benchmark additivo assume che tutte le osservazioni abbiano lo stesso peso nel calcolo delle discrepanze annuali, senza considerare i loro valori relativi. Questa assunzione non è sempre verificata; in alcuni casi il mancato rispetto di questa caratteristica potrebbe produrre benchmark negativi per variabili che invece sono definite positive. Il benchmarking proporzionale assume che, da un punto di vista stagionale, valori grandi siano più rilevanti per le discrepanze annuali rispetto a quelli più piccoli e raramente producono valori di riferimento negativi. Secondo le regole di benchmark proporzionale, la serie sub-annuale considerata viene moltiplicata per un fattore di scala  $b$ , dato da:

$$b = (\sum_{m=1}^M a_m) / (\sum_{m=1}^M \sum_{t=t_{1,m}}^{t_{L,m}} j_{mt} s_t)$$

La serie riscalata  $s_t^\dagger = b \times s_t$  porta la serie sub-annuale al livello medio dei benchmark e preserva esattamente i movimenti proporzionali e i tassi di crescita delle serie originali:  $s_t^\dagger / s_t - s_{t-1}^\dagger / s_{t-1} = 0$ ,

$$t = 1, \dots, T. \text{ Quindi, } s_t^\dagger / s_t = s_{t-1}^\dagger / s_{t-1} \text{ implica che } s_t^\dagger / s_t - 1 = s_{t-1}^\dagger / s_{t-1} - 1. \text{ È da}$$

notare che, in base al suo valore, il fattore di scala restringe o amplifica il pattern stagionale senza produrre valori negativi per  $s_t^\dagger$  malgrado  $s_t$  contenga valori negativi. La serie  $s_t^\dagger$  è quindi riferita usando il modello di benchmark additivo senza matrice di regressori  $\mathbf{R}$ . La matrice di covarianza usata nel modello è  $V_\varepsilon = \mathbf{E} \lambda \mathbf{\Omega} \mathbf{E}^\lambda$ , dove  $\lambda$  è posto uguale ad 1. La matrice  $\mathbf{E}$  è definita  $\text{diag}(\mathbf{c} \times \mathbf{s})$ , dove  $\mathbf{c}$  è un costante coefficiente di variazione uguale a 0.01. la matrice di autocorrelazione  $\mathbf{\Omega}$  è quella di un regolare processo auto regressivo del prim'ordine. Se il parametro auto regressivo  $\varphi$  è vicino ad 1, la serie riferita approssima bene il metodo di Denton (1971), che vedremo più avanti. Se invece

non è vicino ad uno, ad esempio 0.8, ci sarà comunque una situazione con benchmark proporzionale.

Infine con un regressore costante ( $H = \mathbf{1}$  e  $R = -\mathbf{1}$ ), la serie sarà riferita con un modello misto: si riscalerà la serie dovrebbe essere additiva, e l'aggiustamento dei residui sarà proporzionale.

#### **4.2 Metodo di benchmarking di Cholette e Dagum: il modello moltiplicativo**

Il modello additivo visto nel precedente paragrafo risulta appropriato quando le varianze dell'errore delle serie sub-annuali sono indipendenti dal livello della serie. Nel caso in cui le varianze siano proporzionali, non indipendenti, al livello della serie, risulta preferibile un modello di tipo moltiplicativo, per evitare che tecniche additive producano valori negativi per matrici definite positive. In questi casi, tutte le componenti del modello di benchmarking e le stime che ne derivano tendono ad avere coefficienti di variazione costanti. Il modello moltiplicativo preserva bene i tassi di crescita, solitamente utilizzati per confrontare il comportamento di vari indicatori socio-economici.

Ciò che distingue i due modelli è il fatto che il modello moltiplicativo può includere nella regressione un termine deterministico moltiplicativo, mentre il modello proporzionale ne può contenere solo uno additivo. Nel modello moltiplicativo inoltre si ha il vantaggio che le discrepanze proporzionali sono standardizzate, facilitando quindi la possibilità di confrontare le varie serie.

Vediamo ora in dettaglio questa variante di Cholette e Dagum.

#### 4.2.1 Il modello di benchmarking moltiplicativo

Anche questo modello, così come quello additivo, consiste di due equazioni; la prima in questo caso è di tipo moltiplicativo, ovvero:

$$\begin{aligned} s_t &= b_t \times \theta_t \times e_t = b_t \times \theta_t \times (1 + e_t^*) \\ &= b_t \times \theta_t \times (1 + c_t e_t^\dagger), \end{aligned}$$

Dove  $b_t = \prod_{h=1}^H \beta_h^{r_{th}}$  è la parte del regressore deterministico. L'errore  $e_t$  è centrato ad uno ed è pari a  $(1 + e_t^*) - (1 + c_t e_t^\dagger)$  dove  $e_t^*$  è una deviazione da 1 molto vicina a zero. L'errore standardizzato  $e_t^\dagger$  ha media zero e varianza unitaria e  $c_t$  è la deviazione standard di  $e_t^*$ , con  $E(e_t^*) = 0, E(e_t^*, e_{t-1}^*) = c_t c_{t-1} \omega_1$ , dove  $\omega_1$  è l'autocorrelazione dell'errore standardizzato. Applicando la trasformazione logaritmica si ottiene:

$$\ln s_t = \ln b_t + \ln \theta_t + \ln(1 + e_t^*) = \ln b_t + \ln \theta_t + \ln(1 + c_t e_t^\dagger).$$

E prendendo la trasformazione logaritmica di  $b_t = \prod_{h=1}^H \beta_h^{r_{th}}$  si ottiene  $\ln b_t = \sum_{h=1}^H r_{th} \ln \beta_h$ . Sostituendo si ricava che:

$$\begin{aligned} s_t^* &= \sum_{h=1}^H r_{th} \beta_h^* + \theta_t^* + c_t e_t^\dagger, \\ s_t^* &= \sum_{h=1}^H r_{th} \beta_h^* + \theta_t^* + e_t^*, \quad t=1, \dots, T, \end{aligned}$$

La seconda equazione resta pari a quella del modello additivo, ovvero:

$$a_m = \sum_{t=t_{1m}}^{t_{Lm}} j_{mt} \theta_t + \varepsilon_m, \quad m=1, \dots, M,$$

Dove  $E(\varepsilon_m) = 0$  e  $E(\varepsilon_m^2) = \sigma_{\varepsilon_m}^2$

Il modello moltiplicativo richiede che sia le osservazioni sub-annali che quelle annuali siano positive. Inoltre la serie è contaminata dall'errore e dalla presenza di un effetto deterministico.  $\beta_h^*$  denota il vero valore che soddisfa i vincoli

annuali. L'errore può essere espresso come il prodotto di un coefficiente di variazione per l'errore standardizzato. Quindi il coefficiente di variazione del modello moltiplicativo gioca il ruolo di deviazione standard nel modello additivo.

Il regressore deterministico è tipicamente costante, e cattura il livello di differenza proporzionale media tra i dati annuali e sub-annuali. Questa quantità è una media ponderata delle discrepanze proporzionali annuali,

$$d_m^{(p)} = a_m / (\sum_{t=t_{1m}}^{t_{Lm}} j_{mt} s_t), \quad m=1, \dots, M.$$

Quando i dati sono survey, il parametro di regressione stimato  $\tilde{\beta}_1 = \exp(\hat{\beta}_1)$  è interpretato come una "distorsione" moltiplicativa.

In alcuni casi viene utilizzato un secondo regressore per catturare un trend deterministico nelle differenze proporzionali,  $d_m^{(p)}$ , o il regressore  $R$  potrebbe essere assente. L'errore  $e_t^*$  determina la distribuzione delle discrepanze annuali sulle osservazioni sub-annuali. L'errore è omoschedastico se  $\lambda = 0$ , eteroschedastico altrimenti. Il parametro  $\lambda$  assume valori 0,  $\frac{1}{2}$  o 1. Le autocorrelazioni dell'errore standardizzato sono quelle di un processo AR(1) o AR(2).

#### 4.3 Benchmarking come un problema di minimi quadrati



In questo paragrafo descriviamo il problema del benchmarking, introducendo l'approccio che si basa sulla minimizzazione quadratica, proposto da Denton nel 1971. Prima di entrare nel dettaglio del metodo, vediamo però in modo più completo il problema.

Assumiamo che i dati utilizzabili,  $y_j$ , siano distribuiti senza distorsione intorno alla vera serie  $\hat{y}_j$ , in accordo con il modello

$$y_j = \hat{y}_j + e_j,$$

$$j = 1, \dots, M$$

dove  $e_j$  è un vettore d'errore casuale con media zero di dimensione  $(n \times 1)$ , con matrice di varianza e covarianza nota e denotata con  $V_{ij}$ . Mettendo insieme le M relazioni abbiamo un modello completo

$$y_j = \hat{y}_j + e_j$$

con valore atteso pari a zero e matrice di varianza e covarianza  $V$ . Le serie riferite simultaneamente sono la soluzione del problema a minimi quadrati

$$\min (y - y^*)' V^{-1} (y - y^*)$$

con il vincolo  $Hy = y_a$

Come sempre un'estensione dei risultati classici è dovuta al rango delle matrici coinvolte nella procedura.

Si può mostrare che le stime aggiustate possono essere espresse come:

$$y^* = y + VH'(HVH')^{-1} (y_a - Hy)$$

dove  $(HVH')$  denota l'inversa generalizzata di  $V_a = HVH'$ . Una soluzione equivalente e che non coinvolge matrici singolari da invertire può essere espressa in termini di r vincoli "liberi". Infatti la matrice singolare  $(HVH')$  può essere

riscritta come  $\mathbf{RH}_w\mathbf{VH}'_w\mathbf{R}' = \mathbf{RV}_w\mathbf{R}'$  dove  $\mathbf{V}_w = \mathbf{H}_w\mathbf{VH}'_w$  è una matrice a rango pieno ( $r \times r$ ). Inoltre, si può facilmente controllare che  $\mathbf{V}_w^{-1}$  è univocamente data da:

$$\mathbf{V}_w^{-1} = (\mathbf{H}\mathbf{V}\mathbf{H}')^{-1} = \mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{V}_w^{-1}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'$$

Con alcuni facili passaggi e considerando che  $\mathbf{R}'\mathbf{H} = \mathbf{R}'\mathbf{R}\mathbf{H}_w$ , troviamo la formula di aggiustamento più utilizzabile

$$\mathbf{y}^* = \mathbf{y} + \mathbf{VH}'_w\mathbf{V}_w^{-1}(\mathbf{w} - \mathbf{H}_w\mathbf{y})$$

Le stime aggiustate sono quindi ottenute distribuendo una combinazione lineare delle discrepanze che si riferiscono a  $r$  osservazioni non vincolate del vettore  $\mathbf{y}_w$  sui dati originali non aggiustati. Si dovrebbe notare che l'espressione (13) comporta l'inversione solo di matrici di pieno rango e rispetta i vincoli sia temporali che contemporanei.

#### 4.4 Il metodo di Denton

In questo paragrafo vogliamo descrivere il metodo di Denton, nelle sue due varianti, quella additiva e quella proporzionale, sia considerando il caso di differenze prime che considerando quello con differenze seconde. Il lavoro di Denton (1971) e la successiva estensione delle procedure di aggiustamento, vista come una semplice generalizzazione per  $M$  serie temporali, con  $M$  maggiore di uno, inizia dal "*movement preservation principle*" definito nel 1971; in questo lavoro si definisce la teoria che le dinamiche temporali delle serie riconciliate dovrebbero essere quanto più possibile vicino ai dati preliminari.

Secondo questo principio, il valore riferito  $\theta_t$  dovrebbe essere il più vicino possibile a  $s_t$ . Come anticipato, Denton propose sia una versione additiva che una proporzionale per la conservazione dei movimenti delle serie. Inoltre considerò sia le differenze prima della serie sia quelle seconde. Per capire meglio questo concetto vediamo più in dettaglio il suo lavoro.

#### 4.4.1 il metodo di Denton e sue varianti: la versione a differenze prime additive

L' estensione delle procedure di aggiustamento di Denton può essere visto come una semplice generalizzazione per M serie temporali, con M maggiore di uno, del "movement preservation principle " definito da Denton nel 1971, in accordo al quale le dinamiche temporali delle serie riconciliate dovrebbero essere quanto più possibile vicino ai dati preliminari. Come noto (Bloem et al.,2001) questo principio si applica semplicemente concentrando la propria attenzione sui cambi "da periodo a periodo":

$$(\hat{y}_{j,t}^* - \hat{y}_{j,t-1}^*) - (y_{j,t} - y_{j,t-1}) \equiv (\hat{y}_{j,t}^* - y_{j,t}) - (\hat{y}_{j,t-1}^* - y_{j,t-1}), \quad j = 1, \dots, M,$$

o i cambi proporzionali periodo per periodo,

$$\frac{\hat{y}_{j,t}^* - y_{j,t}}{y_{j,t}} - \frac{\hat{y}_{j,t-1}^* - y_{j,t-1}}{y_{j,t-1}} \equiv \frac{\hat{y}_{j,t}^*}{y_{j,t}} - \frac{\hat{y}_{j,t-1}^*}{y_{j,t-1}}, \quad j = 1, \dots, M$$

Le funzioni obiettivo da minimizzare sono quindi date da:

$$\sum_{j=1}^M \sum_{t=2}^n [(\hat{y}_{j,t}^* - y_{j,t}) - (\hat{y}_{j,t-1}^* - y_{j,t-1})]^2$$

Chiamata differenze prime additive (dall'inglese Additive First Differences, AFD)

e

$$\sum_{j=1}^M \left( \frac{\hat{y}_{j,t}^*}{y_{j,t}} - \frac{\hat{y}_{j,t-1}^*}{y_{j,t-1}} \right)^2$$

Chiamata differenze prime proporzionali (dall'inglese Proportional First Differences, PFD).

Usando la notazione matriciale, consideriamo la matrice  $\Omega_{AFD}$  di dimensioni  $(Mn \times Mn)$  pari a

$I_M \otimes (\Delta' \Delta)$  e  $\Omega_{PFD} = \hat{y}^{-1} [I_M \otimes (\Delta' \Delta)] \hat{y}^{-1} = \hat{y}^{-1} \Omega_{AFD} \hat{y}^{-1}$ , dove  $\hat{y} = \text{diag}(y)$  e  $\Delta$  è la matrice di dimensione

$((n-1) \times n)$  che crea le differenze prime:

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

Le serie aggiustate simultaneamente possono essere ottenute risolvendo il problema di minimizzazione

$$\text{Min} (y^* - y)' \Omega^{-1} (y^* - y) \quad \text{sotto}$$

il vincolo  $H_w y^* = w$

Dove  $\Omega = \Omega_{AFD}$  per varianti additive, o  $\Omega = \Omega_{PFD}$  per la formulazione proporzionale.

In entrambi i casi la matrice  $\Omega$  non ha rango pieno e l'analogia con la soluzione (13)-con  $V = \Omega^{-1}$ - non può essere stabilita immediatamente. Per convenienza computazionale, in linea con la proposta di Denton (1971), possiamo considerare una matrice di differenze prime "approssimata" di dimensione  $(n \times n)$ ,  $D$ , data da

In questo caso entrambe le matrici  $\tilde{\Omega}_{AFD} = I_M \otimes (D'D)$  e  $\tilde{\Omega}_{PFD} = \hat{y}^{-1} [I_M \otimes (D'D)] \hat{y}^{-1}$  hanno rango pieno, e la riconciliazione ai minimi quadrati può essere calcolata assumendo che  $V = \Omega^{-1}$ .

La formula di aggiustamento multivariata "approssimata" di Denton quindi può essere espressa come:

$$y^* = y + \tilde{\Gamma}^{-1} H_w (w - H_w y)$$

Anche usando questa approssimazione, le dimensioni della matrice coinvolta nel calcolo può essere considerevole in situazioni pratiche, possibilmente riducendo i carichi computazionali. Come sempre, una preziosa riduzione di tempo computazionale o di area di memoria può essere ottenuto utilizzando la partizione della matrice (Di Fonzo, 2003).

## **CAPITOLO 5**

# **BILANCIAMENTO E RICONCILIAZIONE DI SERIE STORICHE**

In molti casi le serie di sistemi devono soddisfare vincoli di aggregazione *cross-section* (Di Fonzo, 1990), soprattutto devono sommare al valore del loro benchmark annuale, rispettando quelli che in precedenza abbiamo chiamato vincoli di aggregazione temporale.

Si possono avere sistemi di serie classificati in base ad una variabile, ad esempio la Provincia, o sistemi classificati in base a due variabili, ad esempio Industrie e Provincie, e considerare dati marginali riguardo a due variabili, che includono i totali per Provincia, per industria e quello complessivo delle due classificazioni.

La tabella seguente indica un sistema di serie classificato in base a due variabili, industrie e provincie, con i totali marginali delle due variabili e i totali complessivi, tutto ai margini della tabella.

| Prov.<br>Ind.   | Prov. 1           | Prov. 2           | ... | Prov. P-1           | Ind.<br>Totals    |
|-----------------|-------------------|-------------------|-----|---------------------|-------------------|
| Ind. 1          | $\{s_{11,t}\}$    | $\{s_{12,t}\}$    | ... | $\{s_{1,P-1,t}\}$   | $\{s_{1P,t}\}$    |
| Ind. 2          | $\{s_{21,t}\}$    | $\{s_{22,t}\}$    | ... | $\{s_{2,P-1,t}\}$   | $\{s_{2P,t}\}$    |
| ⋮               | ⋮                 | ⋮                 | ⋮   | ⋮                   | ⋮                 |
| Ind. N-1        | $\{s_{N-1,1,t}\}$ | $\{s_{N-1,2,t}\}$ | ... | $\{s_{N-1,P-1,t}\}$ | $\{s_{N-1,P,t}\}$ |
| Prov.<br>Totals | $\{s_{N1,t}\}$    | $\{s_{N2,t}\}$    | ... | $\{s_{N,P-1,t}\}$   | $\{s_{NP,t}\}$    |

All'interno delle celle ci sono invece le serie elementari e ogni cella contiene il valore della serie sub-annuale al tempo t, riferita all'industria n e alla provincia p. I vincoli di aggregazione temporale richiedono che i valori delle serie sommino ai totali marginali per ogni periodo di tempo t, ovvero, il sistema deve soddisfare N set di vincoli di aggregazione sulle provincie per ogni industria e P set di vincoli di aggregazione sull'industria per ogni provincia.

Ricordando quanto detto in precedenza, i vincoli possono essere formalizzati come:

$$s_{n,p,t} = \sum_{p=1}^{P-1} s_{np,t} \quad t = t_{1,n,p}, \dots, t_{L,n,p} \quad n = 1, \dots, N \quad (10)$$

$$s_{Np,t} = \sum_{n=1}^{N-1} s_{np,t} \quad t = t_{1,nP}, \dots, t_{L,nP} \quad p = 1, \dots, P \quad (11)$$

Dove  $t_{1,nP}$  e  $t_{L,nP}$  sono rispettivamente il primo e l'ultimo valore della serie  $s_{np,t}$ . Il numero di osservazioni per la generica serie  $s_{np,t}$  è quindi  $T_{np} = t_{L,nP} - t_{1,nP} + 1$ . Il numero  $T_{np}$  varia da cella a cella. Infatti potrebbe essere zero, se una tale industria non esistesse in una determinata provincia.

Va sottolineato inoltre che il numero di osservazioni per totale complessivo,  $T_{NP}$ , può essere più grande, uguale o più piccolo di  $T_{np}$  per ogni n e p, visto che i periodi sub-annuali t coperti dal totale complessivo sono l'unione di tutti i periodi sub-annuali soddisfatti da tutte le altre serie nel sistema. In modo analogo si comporta la serie  $T_{nP}$ , ovvero quella che determina il numero di osservazioni per i totali industriali. Questa può essere più grande, uguale o più piccola di  $T_{np}$  per ogni data industria n, e il numero di osservazioni per i totali provinciali, i.e.  $T_{Np}$ , può essere più grande, uguale o più piccolo di  $T_{np}$  per ogni provincia. In altre parole ci sono  $T_{nP}$  vincoli individuali nell'n-simo set di vincoli industriali e  $T_{Np}$  nel p-esimo set di vincoli provinciali.

Per i sistemi di serie classificati per un solo attributo, ad esempio le provincie, viene omesso l'apice n.

In molte situazioni, ad esempio i casi di contabilità nazionale, le serie di un sistema non soddisfano i vincoli di aggregazione. Ciò che bisogna fare è rimuovere le discrepanze *cross-section* e temporali e gli errori residui, dovuti al fatto che le osservazioni possono essere ottenute da entrate, uscite o livello di produzione.

La riconciliazione emerge anche in quei casi in cui le serie vengono riclassificate in base ad i nuovi criteri statistici, che potrebbero nascere in base a nuove esigenze dei vari stati.

Consideriamo ad esempio un sistema classificato in base a due variabili, contenente N set di discrepanze industriali per le industrie e P set di discrepanze per le provincie. Le discrepanze possono essere scritte come:

$$d_{n,t} = s_{np,t} - \sum_{p=1}^{P-1} s_{np,t} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad n = 1, \dots, N \quad (12)$$

$$d_{p,t} = s_{Np,t} - \sum_{n=1}^{N-1} s_{np,t} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad p = 1, \dots, P \quad (13)$$

Queste discrepanze di aggregazione sono spesso espresse in termini di proporzione, per ridurre le dimensioni, ottenendo quindi:

$$d_{nt}^{(p)} = s_{np,t} / \sum_{p=1}^{P-1} s_{np,t} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad n = 1, \dots, N \quad (14)$$

$$d_{nt}^{(p)} = s_{np,t} / \sum_{p=1}^{P-1} s_{np,t} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad n = 1, \dots, N \quad (15)$$

dove l'apice (p) indica la provincia e non l'industria.

I vincoli di additività *cross-section* sono spesso implementati attraverso l'approccio di adattamento proporzionale iterativo (*Iterative Proportional Fitting*). Questo si svolge in 5 passi:

- 1) Le serie elementari delle righe  $n=1, \dots, N$  sono moltiplicate per le corrispondenti discrepanze provinciali; in questo modo queste vengono eliminate.
- 2) Le discrepanze relative all'industria sono ricalcolate visto che l'applicare il primo passaggio le altera.
- 3) Le serie delle colonne  $p=1, \dots, P-1$  sono moltiplicate per le corrispondenti discrepanze proporzionali industriali. Questo elimina le discrepanze industriali.
- 4) Le discrepanze proporzionali provinciale vengono ricalcolate
- 5) Vengono ripetuti i passaggi



Presentiamo un metodo generale per la riconciliazione dei sistemi di serie storiche. Questo ha le seguenti proprietà:

- 1) Include l'IPS come metodo particolare
- 2) Alcune serie devono essere viste come esogene
- 3) I totali marginali devono essere visti come endogeni, cioè stimati dalle serie elementari nel processo di riconciliazione
- 4) Le serie potrebbero avere valori negativi (per es. il profitto)
- 5) Il metodo mantiene i vincoli di aggregazione temporale, per permettere alle serie di soddisfare il loro benchmark
- 6) In base alla matrice di varianza e covarianza utilizzata, i diversi criteri sono ottimizzati
- 7) Avere matrici di covarianza valide fa sì che il metodo modifichi le serie originarie in accordo con la loro covarianza e produca valide matrici di covarianza per le serie riconciliate
- 8) dopo può essere usato per imporre vincoli

### 5.1 Riconciliazione: un modello di regressione semplice

Spesso la riconciliazione è ottenuta da alcune semplici stime; mostriamo come queste possano essere ottenute attraverso una regressione ponderata (Quenneville, Rancourt, 2005).

Supponiamo che ci siano 3 osservazioni  $y_0, y_1, y_2$ , tali che la somma di  $y_1$  e  $y_2$  sia uguale a  $y_0$ . Un modo per riconciliare i valori osservati  $y_1$  e  $y_2$  con  $y_0$  è dato semplicemente dalla stima di  $y_1$  e  $y_2$ , dove il valore corretto di  $y_1$  è preso uguale a  $b_1 = \frac{y_1 y_0}{(y_1 + y_2)}$ , quello di  $y_2$  è preso uguale a  $b_2 = \frac{y_2 y_0}{(y_1 + y_2)}$ , e così  $b_1 + b_2 = y_0$ .

È facile costruire un modello di regressione per svolgere la procedura computazionale numerica per la stima. Sia  $y_1 = b_1 + e_1$ ,  $y_2 = b_2 + e_2$  e  $y_0 = b_1 + b_2$ ,

$\epsilon_1 \sim (0, \gamma_1)$ ,  $\epsilon_2 \sim (0, \gamma_2)$ , dove  $\epsilon_i \sim (0, \gamma_i)$  indica che l'errore  $\epsilon_i$  ha media 0 e varianza  $\gamma_i$ . Il parametro  $b_2$  può essere eliminato per ottenere il modello semplificato  $y_1 = b_1 + \epsilon_1$ ,  $y_0 - y_2 = b_1 + \epsilon_2$ . Assumendo che  $\epsilon_1$  e  $\epsilon_2$  siano incorrelati, il miglior stimatore lineare non distorto (*BLUE, Best Unlinear Unbiased Estimator*) di  $b_1$  è una media ponderata di  $y_1$  e  $y_0 - y_2$  dove i pesi sono inversamente proporzionali alle varianze:

$$b_1 = \left( \frac{1}{\gamma_1} + \frac{1}{\gamma_2} \right)^{-1} \left( \frac{y_1}{\gamma_1} + \frac{y_0 - y_2}{\gamma_2} \right) = y_1 \frac{\gamma_2}{\gamma_1 + \gamma_2}$$

Questo modello di regressione è quindi un mezzo meccanico per implementare la computazione numerica per la stima. Vediamone ora una sua estensione adatta a trattare casi più complessi.

Il modello generale di Dagum e Cholette (2006) usa i coefficienti di alterabilità per modellare la varianza dell'errore associato ad un'osservazione, o per modificare una restrizione obbligatoria. In questo caso sia  $\epsilon_1 \sim (0, a_1 \gamma_1)$ ,  $\epsilon_2 \sim (0, a_2 \gamma_2)$ ,  $y_3 = b_1 + b_2 + \epsilon_0$ ,  $\epsilon_0 \sim (0, a_0 \gamma_0)$ , dove  $a = (a_0, a_1, a_2)$  è un vettore noto di coefficienti di alterabilità. In questo caso è facile mostrare che

$$b_1 = y_1 + \frac{a_1 \gamma_1}{a_0 \gamma_0 + a_1 \gamma_1 + a_2 \gamma_2} [y_0 - (y_1 + y_2)],$$

$$b_2 = y_2 + \frac{a_2 \gamma_2}{a_0 \gamma_0 + a_1 \gamma_1 + a_2 \gamma_2} [y_0 - (y_1 + y_2)],$$

ma adesso

$$b_1 + b_2 = y_0 - \frac{a_0 \gamma_0}{a_0 \gamma_0 + a_1 \gamma_1 + a_2 \gamma_2} [y_0 - (y_1 + y_2)].$$

L'uso di un coefficiente di alterabilità per  $y_3$  non consente di soddisfare in modo completo il vincolo. Porre  $a_0$  pari ad un numero piccolo è un trucco per renderlo il più vicino possibile al vincolo. Inoltre, il modello di regressione può

essere decisamente semplificato considerando  $y_0$  come un'osservazione con un errore molto piccolo invece di restrizioni sui parametri.

Quando si ha una classificazione ad un'entrata, potrebbe essere più facile applicare le stime senza avere il problema di scrivere un modello di regressione per implementare l'aritmetica.

Con tabelle di serie storiche di dimensioni più grandi, dovrebbe essere più facile costruire un modello di regressione usando l'idea utilizzata in precedenza, mediante un insieme di regressioni per implementare la computazione.

Mostriamo come un approccio di questo tipo semplifichi la sequenza computazionale, quando si ha a che fare con una tabella di valori di dimensione  $2 \times 2$ .

Supponiamo che i valori siano forniti con una classificazione che utilizza una tabella a doppia entrata come nell'esempio seguente:

|           |           |           |
|-----------|-----------|-----------|
| $y_{1,1}$ | $y_{1,2}$ | $y_{1,.$  |
| $y_{2,1}$ | $y_{2,2}$ | $y_{2,.$  |
| $y_{.,1}$ | $y_{.,2}$ | $y_{.,.}$ |

Dove  $y_{i,j}$   $i,j = 1,2$  sono i valori osservati per la cross-classificazione  $i,j = 1,2$ ,  $y_{i,}$   $i = 1,2$  sono le colonne totali, e  $y_{.,}$  è il totale complessivo della tabella.

Il bilanciamento di questa tabella di valori è ottenuto con il seguente modello di regressione:

$$y_{1,1} = b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1}y_{1,1})$$

$$y_{1,2} = b_{1,2} + e_{1,2}, e_{1,2} \sim (0, a_{1,2}y_{1,2})$$

$$y_{2,1} = b_{2,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1}y_{2,1})$$

$$y_{2,2} = b_{2,2} + e_{2,2}, e_{2,2} \sim (0, a_{2,2}y_{2,2})$$

$$y_{1.} = b_{1,1} + b_{1,2} + e_{1.}, e_{1.} \sim (0, a_{1.} y_{1.})$$

$$y_{2.} = b_{2,1} + b_{2,2} + e_{2.}, e_{2.} \sim (0, a_{2.} y_{2.})$$

$$y_{.1} = b_{1,1} + b_{2,1} + e_{.1}, e_{.1} \sim (0, a_{.1} y_{.1})$$

$$y_{.2} = b_{1,2} + b_{2,2} + e_{.2}, e_{.2} \sim (0, a_{.2} y_{.2})$$

$$y_{..} = b_{1,1} + b_{1,2} + b_{2,1} + b_{2,2} + e_{..}, e_{..} \sim (0, a_{..} y_{..})$$

Ci sono 9 osservazioni, 9 coefficienti di alterabilità noti  $(a_{ij}, i, j = ., 1, 2)$  e 4 parametri  $(b_{ij}, i, j = ., 1, 2)$  da stimare. Così come  $a_{ij} > 0, i, j = ., 1, 2$ , può essere implementata una regressione non ristretta e i valori predetti sono le stime bilanciate o riconciliate.

Se uno dei coefficienti di alterabilità è pari a zero, allora l'equazione corrispondente dovrebbe essere scritta come un vincolo.

Per illustrare questo consideriamo il bilanciamento di una tabella  $2 \times 2$  soggetta a totali marginali fissati dalle medie di una regressione ponderata.

L'equazione (1) diventa:

$$y_{1,1} = b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1} y_{1,1})$$

$$y_{1,2} = b_{1,2} + e_{1,2}, e_{1,2} \sim (0, a_{1,2} y_{1,2})$$

$$y_{2,1} = b_{2,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1} y_{2,1})$$

$$y_{2,2} = b_{2,2} + e_{2,2}, e_{2,2} \sim (0, a_{2,2} y_{2,2})$$

$$y_{1.} = b_{1,1} + b_{1,2}$$

$$y_{2.} = b_{2,1} + b_{2,2}$$

$$y_{.1} = b_{1,1} + b_{2,1}$$

$$y_{1,2} = b_{1,2} + b_{2,2}$$

Da notare che non c'è bisogno dell'equazione  $y_{1,}$  perché si assume che  $y_{1,} = y_{1,1} + y_{1,2} = y_{1,1} + y_{1,2}$ .

Usando le relazioni tra i parametri e i totali di controllo, si può riscrivere l'equazione (2) come:

$$y_{1,1} = b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1}y_{1,1})$$

$$y_{1,} - y_{1,2} = b_{1,2} + e_{1,2}, e_{1,2} \sim (0, a_{1,2}y_{1,2})$$

$$y_{1,1} - y_{2,1} = b_{1,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1}y_{2,1})$$

$$y_{2,2} + y_{1,} - y_{1,2} = b_{1,1} + e_{2,2}, e_{2,2} \sim (0, a_{2,2}y_{2,2})$$

Lo stimatore BLUE di  $b_{1,1}$ , è quindi la media ponderata di  $y_{1,1}$ ,  $y_{1,} - y_{1,2}$ ,  $y_{1,1} - y_{2,1}$ ,  $y_{2,2} + y_{1,} - y_{1,2}$ , dove i pesi sono inversamente proporzionali alle loro varianze:

$$b_{1,1} = \frac{\left( \frac{y_{1,1}}{a_{1,1}y_{1,1}} + \frac{y_{1,} - y_{1,2}}{a_{1,2}y_{1,2}} + \frac{y_{1,1} - y_{2,1}}{a_{2,1}y_{2,1}} + \frac{y_{2,2} + y_{1,} - y_{1,2}}{a_{2,2}y_{2,2}} \right)}{\left( \frac{1}{a_{1,1}y_{1,1}} + \frac{1}{a_{1,2}y_{1,2}} + \frac{1}{a_{2,1}y_{2,1}} + \frac{1}{a_{2,2}y_{2,2}} \right)}$$

Le stime per i rimanenti parametri risultano dai vincoli dell'equazione 2.

Per una tabella  $r \times c$ , sarà facile creare il modello di regressione descritto dalle equazioni (1) o (2) e usare un pacchetto di regressioni per implementare le computazioni. L'equazione (1) e la (2) differiscono solo per la costruzione di vincoli sui parametri. L'equazione (2) assume vincoli obbligatori mentre l'equazione (1) assume vincoli non obbligatori; ovviamente, dovrebbe esserci un insieme delle due equazioni dove, per esempio, i totali in riga potrebbero essere non obbligatori e i totali in colonna si.

## 5.2 Riconciliazione di sistemi di serie storiche ad una variabile.

In questo paragrafo ci dedicheremo alla riconciliazione di sistemi di serie storiche, classificati sulla base di una variabile. Senza perdere di generalità, assumiamo un sistema classificato per provincie (Dagum, Cholette, 2006). Questo deve soddisfare un set di vincoli di aggregazione *cross-section*:

$$d_{n,t} = s_{P,t} - \sum_{p=1}^{P-1} s_{p,t} \quad t = t_{1,P}, \dots, t_{L,P} \quad (16)$$

Dove P denota il totale,  $t_{1,p}$  e  $t_{L,p}$  rappresentano la prima e l'ultima osservazione utilizzabile per la serie  $s_{p,t}$ . Il numero di osservazioni per una data serie  $s_{p,t}$  è quindi  $T_p = t_{L,p} - t_{1,p} + 1$ . Questo numero varia da provincia a provincia e il numero dei vincoli è uguale al numero delle osservazioni  $T_p$  per le serie totali.

Il sistema contiene un set di discrepanze di aggregazione *cross-section* e un set di discrepanze di aggregazione proporzionali *cross-section*:

$$d_t = s_{P,t} - \sum_{p=1}^{P-1} s_{p,t} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad n = 1, \dots, N \quad (17)$$

$$d_t^{(n)} = \frac{s_{P,t}}{\sum_{p=1}^{P-1} s_{p,t}} \quad t = t_{1,MP}, \dots, t_{L,MP} \quad n = 1, \dots, N \quad (18)$$

Bisogna sottolineare che l'apice sta per proporzione e non per provincia.

I vincoli di additività *cross-section* sono implementati attraverso l'*iterative Proportional Fitting* descritto in precedenza. Nel caso di classificazione ad una variabile questo si applica semplicemente moltiplicando la serie elementare per la discrepanza corrispondente. Questo metodo implica che il totale delle serie sia pre-determinato, quindi esogeno.

### 5.2 Il modello

A questo punto del discorso si cerca di dare una soluzione analitica al modello di riconciliazione di un sistema di serie storiche classificate con una variabile. Per un sistema contenente  $K=P$  provincie, il modello consiste di due equazioni:

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_p \end{bmatrix} = \begin{bmatrix} I_{T_1} & 0 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & \\ 0 & & \dots & I_{T_P} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad (19)$$

Dove  $I_{T_p}$  è una matrice identità di dimensione  $T_p \times T_p$ ,  $s_p$  è la serie totale originale e  $E(\varepsilon_p) = 0$   $E(\varepsilon_p \varepsilon_p') = V_{\varepsilon_p}$  e covarianze tra gli errori pari a zero. Inoltre:

$$g = [G_1 \quad G_2 \quad \dots \quad -G_p] \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} + \varepsilon$$

Dove il  $E(\varepsilon) = 0$   $E(\varepsilon \varepsilon) = 0$  e  $E(\varepsilon \varepsilon') = V_\varepsilon$  risulta essere di solito pari a zero. Le matrici  $G_p$  sono solitamente matrici di selezione, ovvero matrici identità con numero di righe al più uguale al numero delle colonne.

Nei periodi in cui le serie non esistono, si rimuovono le colonne. In questo modo si possono far partire prima e/o dopo le serie che non hanno tutti i valori.

Adesso resta da ottenere una soluzione per l'equazione (19). per fare questo rendiamo espliciti alcuni prodotti:

$$V_\varepsilon G' = \begin{bmatrix} V_{\varepsilon 1} & 0 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & \\ 0 & & \dots & V_{\varepsilon p} \end{bmatrix} \begin{bmatrix} G_1' \\ G_2' \\ \vdots \\ -G_p' \end{bmatrix} = \begin{bmatrix} V_{\varepsilon 1} G_1' \\ V_{\varepsilon 2} G_2' \\ \vdots \\ -V_{\varepsilon p} G_p' \end{bmatrix}$$

$$G(V_\varepsilon G') = \left[ \sum_{p=1}^P G_p V_{\varepsilon p} G_p' \right]$$

$$V_d = \left[ \sum_{p=1}^P G_p V_{sp} G_p' + V_\varepsilon \right]$$

Con semplici passaggi algebrici (Dagum, Cholette, 2006), si ottiene che:

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_P \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_P \end{bmatrix} + \begin{bmatrix} L'_1 \\ L'_2 \\ \vdots \\ -L'_P \end{bmatrix} [V_d^{-1} d]$$

Con  $L_p = G_p V_{sp}$

Si ottiene quindi:

$$\hat{\theta}_p = s_p \delta_{pP} V_{sp} G_p' [V_d^{-1} d]$$

dove  $\delta_{pP}$  È uguale a -1 se  $p=P$ , a 1 altrimenti.

Il vettore delle discrepanze cross-section è dato da:

$$d = g - \sum_{p=1}^P \delta_{pP} G_p s_p$$

Con  $g = 0$ .

### 5.3 Riconciliazione di sistemi di serie storiche con due variabili

I sistemi di serie storiche classificati con due variabili, sono il risultato di indagini in cui vengono considerati due attributi. I totali marginali sono pubblicati qualche settimana prima delle serie elementari, in modo da delimitare le serie più importanti all'interno di una fascia più corta, e quelle meno importanti in "deadline" più lunghe.

Volendo esemplificare la questione, consideriamo un modello di riconciliazione marginale a due entrate, che si ha per tabelle del tipo:



| Prov.<br>Ind. | Prov. 1        | Prov. 2        | Ind. Totals    |
|---------------|----------------|----------------|----------------|
| Ind. 1        | $\{s_{11,t}\}$ | $\{s_{12,t}\}$ | $\{s_{13,t}\}$ |
| Ind. 2        | $\{s_{21,t}\}$ | $\{s_{22,t}\}$ | $\{s_{23,t}\}$ |
| Ind. 3        | $\{s_{31,t}\}$ | $\{s_{32,t}\}$ | $\{s_{33,t}\}$ |
| Prov. Totals  | $\{s_{41,t}\}$ | $\{s_{42,t}\}$ | $\{s_{43,t}\}$ |

Dove si ha un sistema di serie storiche valutato sulla base di tre industrie e due province; nell'ultima colonna ci sono i totali delle industrie, nell'ultima riga i totali delle province, e all'incrocio di questi ci sono i totali complessivi marginali.

In questo caso si hanno due set di vincoli di aggregazione cross-section, uno sulle province e uno sulle industrie. Questi possono essere scritti come:

$$s_{NP,t} = \sum_{p=1}^{P-1} s_{Np,t} \quad t = 1, \dots, t_{NP}$$

$$s_{np,t} = \sum_{p=1}^{P-1} s_{np,t} \quad t = 1, \dots, t_{NP}$$

$T_{NP}$  Rappresenta il numero di osservazioni per il totale marginale. Analogamente al modello ad una variabile, se definiscono le discrepanze, per le province e per le industrie:

$$d_{N,t}^{(p)} = \frac{s_{NP,t}}{\sum_{p=1}^{P-1} s_{Np,t}}, \quad t = t_{1,NP}, \dots, t_{L,NP}$$

$$d_{P,t}^{(p)} = s_{NP,t} / \sum_{n=1}^{N-1} s_{nP,t} \quad t = t_{1,NP}, \dots, t_{L,NP}$$

Dove l'esponente (p) sta per proporzione non per province.

È possibile anche in questo caso imporre i totali su province e industrie in base all' *iterative proportional fitting*, seguendo i seguenti passi:

- 1) I totali industriali sono moltiplicati per le discrepanze provinciali per eliminare queste ultime discrepanze
- 2) Visto che il primo passo le modifica, le discrepanze industriali vengono ricalcolate.
- 3) I totali provinciali vengono moltiplicati per le discrepanze industriali, per eliminare quest'ultime
- 4) Si ricalcolano le discrepanze proporzionali
- 5) Si ripetono i primi passi finchè le discrepanze provinciali sono vicine a uno.

### 5.3.1 Il modello

Seguendo il modello di regressione visto in precedenza, possiamo rappresentare questo nuovo modello di riconciliazione nel seguente modo

$$\begin{bmatrix} s_{41} \\ s_{42} \\ s_{13} \\ s_{23} \\ s_{33} \\ s_{43} \end{bmatrix} = \begin{bmatrix} I_{T41} & 0 & 0 & 0 \\ 0 & & & \\ 0 & & & \\ 0 & & & I_{T43} \end{bmatrix} \begin{bmatrix} \theta_{41} \\ \theta_{42} \\ \theta_{13} \\ \theta_{23} \\ \theta_{33} \\ \theta_{43} \end{bmatrix} + \begin{bmatrix} e_{41} \\ e_{42} \\ e_{13} \\ e_{23} \\ e_{33} \\ e_{43} \end{bmatrix}$$

dove  $E(e_{np}) = 0$ ,  $E(e_{np}e_{np}') = V_{e_{np}}$ ,  $E(e_{ij}e_{np}') = 0$  per  $i$  diverso da  $n$  o  $j$  diverso da  $p$ .

$$\begin{bmatrix} g_A \\ g_B \end{bmatrix} = \begin{bmatrix} G_{41} & G_{42} & 0 & 0 & 0 & -G_{43} \\ 0 & 0 & G_{13} & G_{23} & G_{33} & G_{43} \end{bmatrix} \begin{bmatrix} \theta_{41} \\ \theta_{42} \\ \theta_{13} \\ \theta_{23} \\ \theta_{33} \\ \theta_{43} \end{bmatrix} + \begin{bmatrix} \varepsilon_4 \\ \varepsilon_3 \end{bmatrix}$$

Dove  $E(\varepsilon_4) = 0$ ,  $E(\varepsilon_3) = 0$ ,  $E(\varepsilon_4\varepsilon_4') = V_{\varepsilon_4}$ ,  $E(\varepsilon_3\varepsilon_3') = V_{\varepsilon_3}$ ,  $E(\varepsilon_4\varepsilon_3') = 0$ , con  $V_{\varepsilon_4}$  e  $V_{\varepsilon_3}$  sono matrici diagonali. Solitamente queste matrici sono pari a zero. Questo non accade se si vuole imporre restrizioni non vincolanti. La matrice  $G_{np}$  è anche in questo caso matrice di selezione, corrispondente a  $T_{43}$ , ma con alcune colonne

escluse. L'utilizzo di questa matrice è uguale a quello visto per la matrice G nel caso univariato.

Come nel caso univariato, forniamo ora una soluzione analitica al problema della riconciliazione (Dagum, Cholette, 2006).

$$\hat{\theta} = s + V_G G' [V_G^{-1} d]$$

#### 5.4 il problema dei Vincoli ridondanti

Come detto all'inizio, questo modello ha vincoli *cross-section* sia per le provincie che per le industrie, rispettivamente:

$$\theta_{NP,t} - \sum_{p=1}^{P-1} \theta_{NP,t} = 0 \quad t = 1, \dots, t_{NP}$$

$$\theta_{NP,t} - \sum_{n=1}^{N-1} \theta_{NP,t} = 0 \quad t = 1, \dots, t_{NP}$$

Anche in questo caso bisogna dividere in due casi il problema; infatti ci si può trovare nella situazione in cui non ci siano vincoli temporali, e quella in cui questi siano presenti. In questo caso si considerano serie della stessa lunghezza, con benchmark annuali e con vincoli temporali restrittivi impliciti nella matrice di covarianza delle serie che necessitano riconciliazione.

Se non ci sono vincoli temporali allora non ci sono vincoli ridondanti; nel caso di presenza di vincoli temporali, allora ci sono vincoli ridondanti in entrambe le serie. Per questo motivo il problema può essere trattato in molti modi:

- 1) Si può ad esempio omettere un vincolo per anno in ognuno dei due set di vincoli
- 2) Rendere un vincolo non restrittivo in ognuno dei due set di vincoli
- 3) Omettendo tutti i vincoli temporali
- 4) Rendendo tutti i vincoli non restrittivi

Nell'applicare questi quattro punti, i vincoli omessi o quelli resi non restrittivi, sono inclusi già nei rimanenti. Se questo non dovesse accadere, allora vuol dire che si è in presenza di dati contraddittori; ad esempio si può essere nella situazione di benchmark che non soddisfano i vincoli di aggregazione cross-section.

# CAPITOLO 6

## UN'APPLICAZIONE

Dopo aver visto in modo teorico i principali argomenti, vediamo ora un'applicazione pratica applicando i concetti visti a dati reali. In questo capitolo conclusivo descriveremo l'applicazione delle tecniche di riconciliazione ai dati mensili sul commercio al dettaglio in Canada, considerando i dati dal gennaio 1991 al dicembre 2003. Vista la natura dei dati, si inizia con le tecniche di destagionalizzazione, per poi passare alla riconciliazione.

### **6.1 Il *Monthly Retail Trade Survey***

Come detto nella premessa del capitolo, i dati sono presi dall'indagine mensile sul commercio al dettaglio in Canada. Quest'indagine è chiamata *MRTS*, acronimo di *Monthly Retail Trade Survey*, e fornisce il livello di produzione del commercio canadese<sup>6</sup>. Le stime di vendita fornite dai venditori costituiscono un indicatore fondamentale per comprendere i modelli di acquisti dei consumatori canadesi. Le vendite al dettaglio, oltretutto, rappresentano un importante componente del prodotto interno lordo nazionale.

### **6.2 La raccolta dei dati**

Seguendo ancora la descrizione di Sturaro<sup>7</sup>, evidenziamo che la raccolta dei dati viene effettuata dagli uffici regionali statistici canadesi. I questionari vengono inviati per posta o somministrati per via telefonica. Le unità contattate forniscono il valore delle loro vendite e delle merci presenti in magazzino. In particolare, i dati

---

<sup>6</sup> Per approfondimenti sul sistema di campionamento e raccolta delle risposte si rimanda a Sturaro, *op. cit.*, 2009

<sup>7</sup> Vedi *supra*.

utilizzati per questo esempio vanno dal gennaio 1991 all' Agosto 2004. Si vedrà però che per motivi puramente computazionali, si utilizzeranno effettivamente i dati fino al dicembre 2003. Questo perché, come spiegato nelle tecniche di benchmarking, vengono imposti vincoli temporali visti come somme annuali delle serie grezze. Si comprende quindi che nei vincoli non sarebbe corretto inserire la somma dei primo otto mesi del 2004. Per questo motivo quindi si dovrebbe imporre al programma utilizzato di arrotondare la serie di riferimento al dicembre 2003. Pertanto necessario non considerare nell'analisi tutti i dati, ma solo quelli fino al Dicembre 2003.

Come già spiegato in precedenza, l'analisi si basa su due procedure fondamentali: la prima è quella della destagionalizzazione dei dati e la seconda è la riconciliazione delle serie destagionalizzate. In particolare si applicheranno le due procedure ai dati riguardanti i settori produttivi canadesi.

### **6.3 I dati**

I dati forniti dall'agenzia statistica canadese sono pubblicati in base a due classificazioni: per provincia (PR) o per settore produttivo (TG). I dati considerati, forniti, come detto, da *Statistic Canada*, sono quelli grezzi (dall'inglese *RAW*). Nella prima parte dell'analisi si è provveduti a destagionalizzarli, nella seconda abbiamo applicato le procedure di riconciliazione.

Come accennato nel paragrafo precedente le serie coprono 164 mesi, anche se poi effettivamente ne considereremo solo 156, dal gennaio 1991 al dicembre 2003.

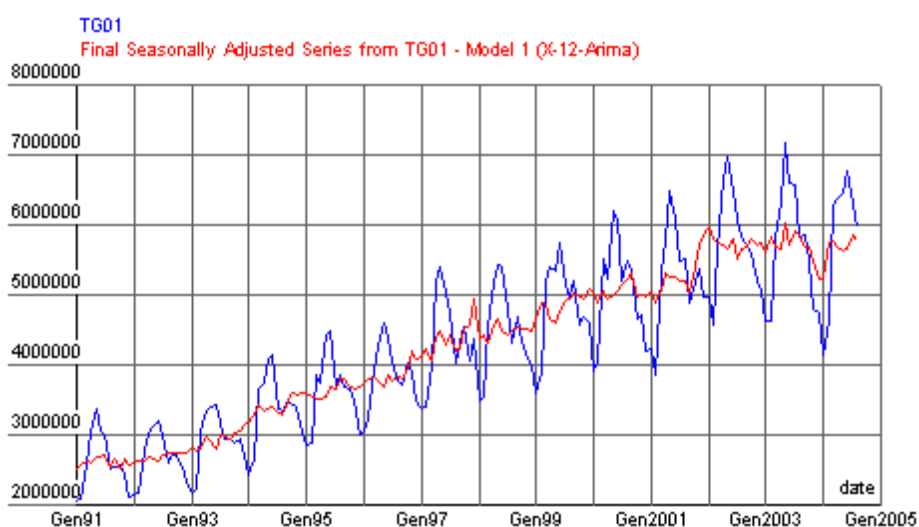
## 6.4 La destagionalizzazione dei dati

Come già accennato, i dati sono composti da 13 serie per le provincie e 19 serie per le industrie. Ai territori dello Yukon, quelli del nord-ovest e del Nunavut si uniscono le provincie del Terranova e Labrador, l'Isola del Principe Edoardo, la Nuova Scozia, il Nuovo Brunswick, il Quebec, l'Ontario, la Manitoba, il Saskatchewan, l'Alberta e la Columbia Britannica. Da sottolineare che il territorio del Nunavut è distaccato dai territori del nord-ovest dal 1999.

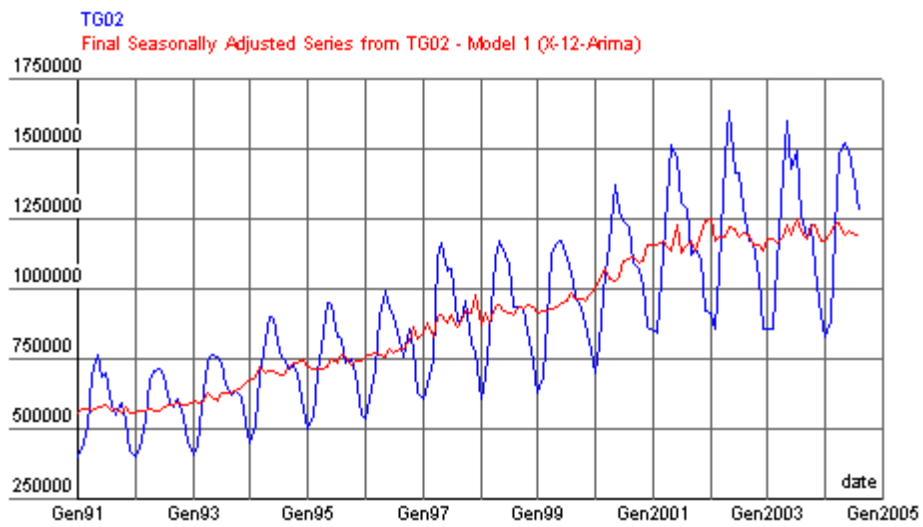
Le 19 Provincie sono invece catalogate in venditori di auto nuove, di auto usate e pezzi di ricambio, negozi di mobili, di articoli per la casa, di computer e software, di elettrodomestici, ferramenta, negozi di materiale per l'edilizia ed il giardinaggio, supermercati, gastronomie, negozi di birra vino e liquori, farmacie e negozi per la cura personale, stazioni di servizio, negozi di abbigliamento, negozi di scarpe accessori e gioiellerie, negozi di articoli sportivi, hobbistica, libri e musica, grandi magazzini, altri venditori al dettaglio ed altri negozi generalisti.

I grafici dei dati grezzi, in blu e destagionalizzati, in rosso, dei settori di produzione sono riportati di seguito. La destagionalizzazione è stata realizzata con il software Demetra<sup>®</sup> 2.2

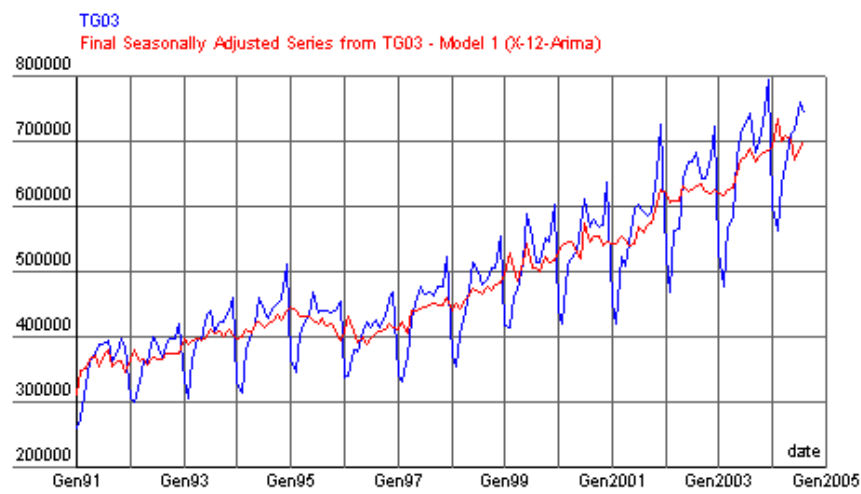
I grafici dei settori di produzione, nei quali sono riportate le serie grezze in blu e quelle destagionalizzate in rosso, sono riportati di seguito:



**Figura 1: venditori di auto nuove**



**Figura 2: venditori di auto usate**



**Figura 3: negozi di mobili**



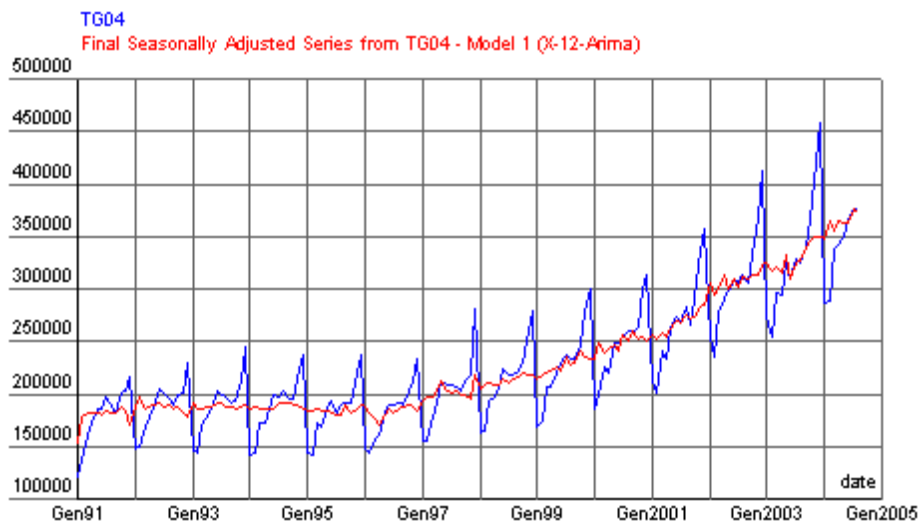


Figura 4: negozi di articoli per la casa

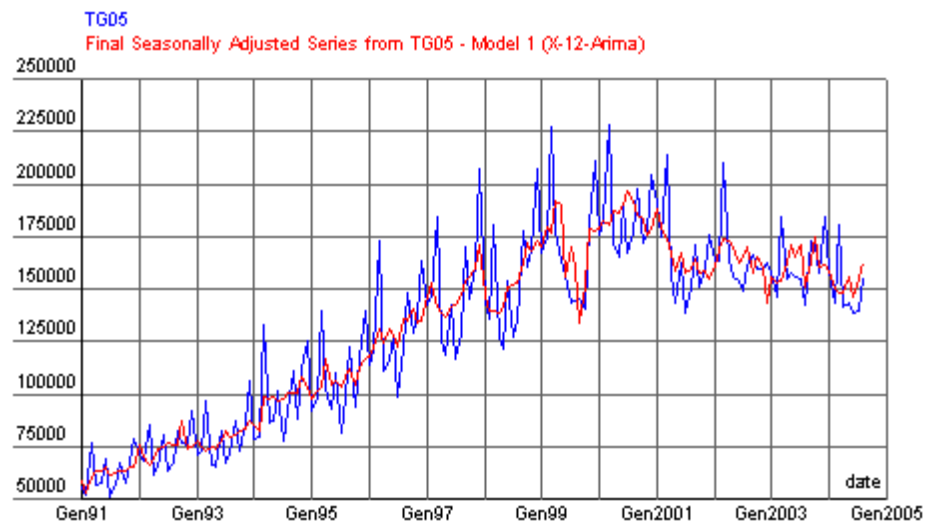
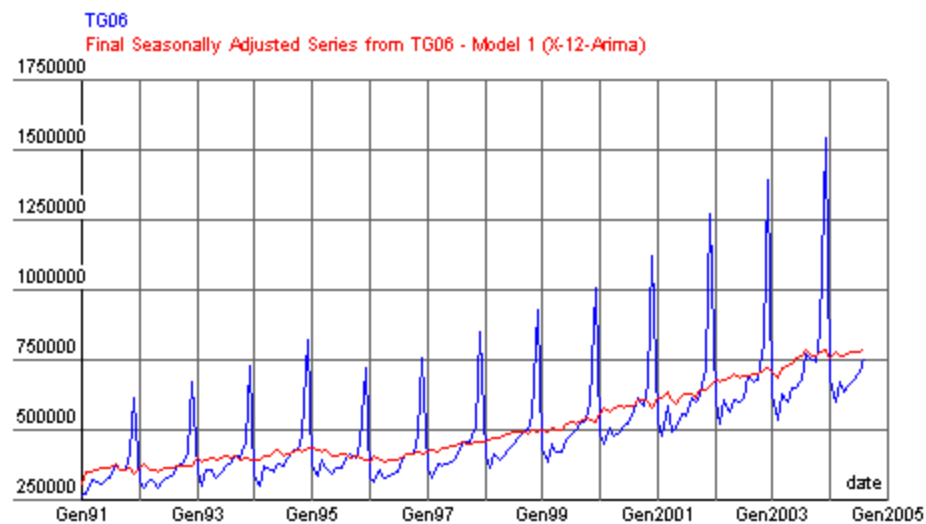
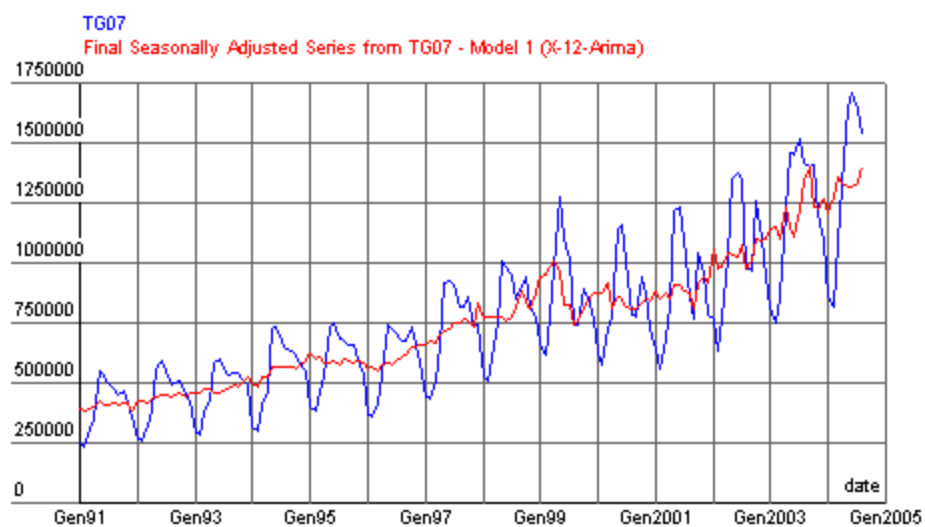


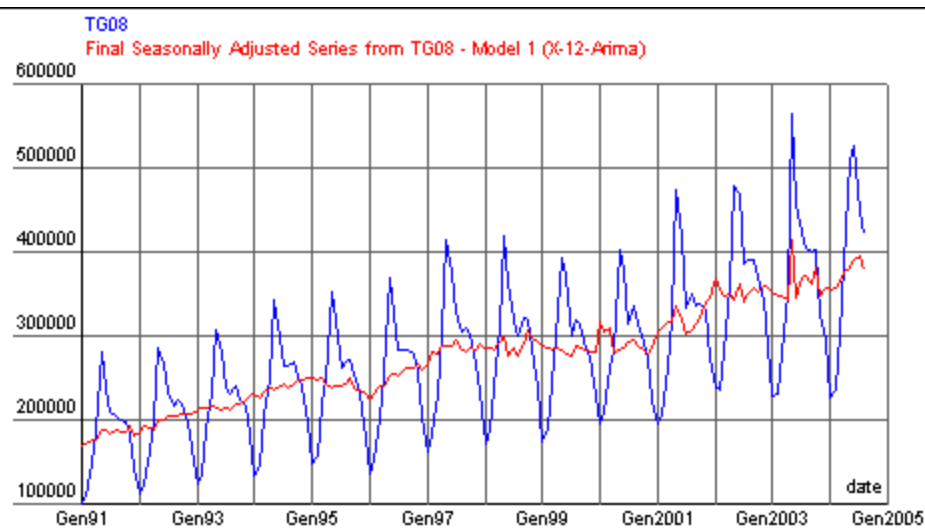
Figura 5: negozi di computer e software



**Figura 6: negozi di elettrodomestici ed elettronica**



**Figura 7: ferramenta**



**Figura 8: negozi di materiale per l'edilizia e ed il giardinaggio**

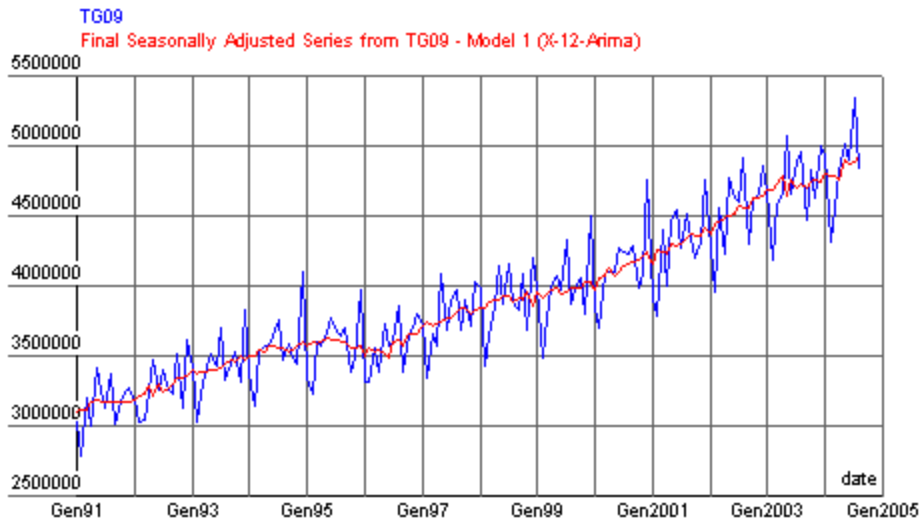


Figura 9: supermercati

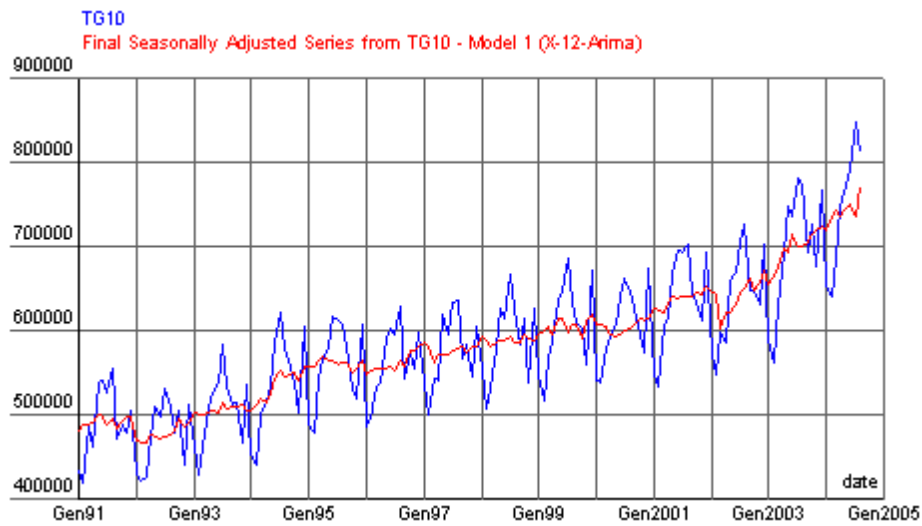


Figura 10: gastronomie

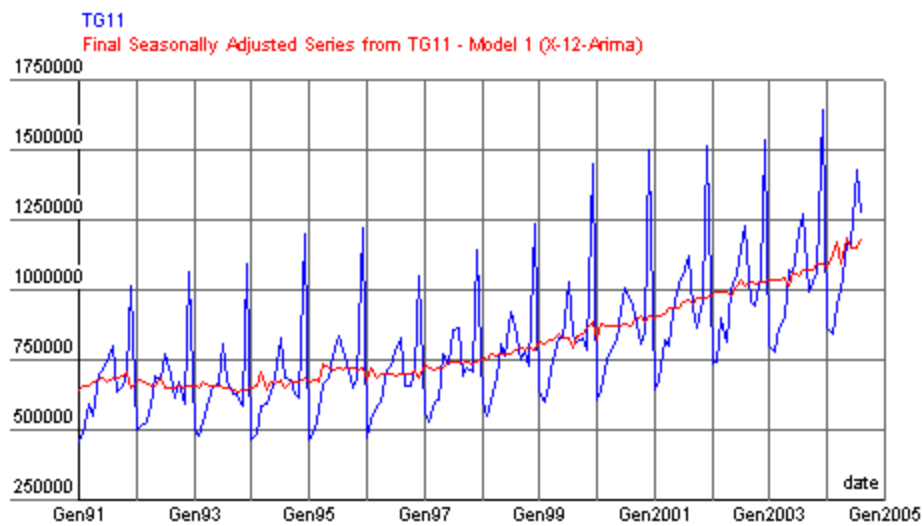


Figura 11: negozi di birra

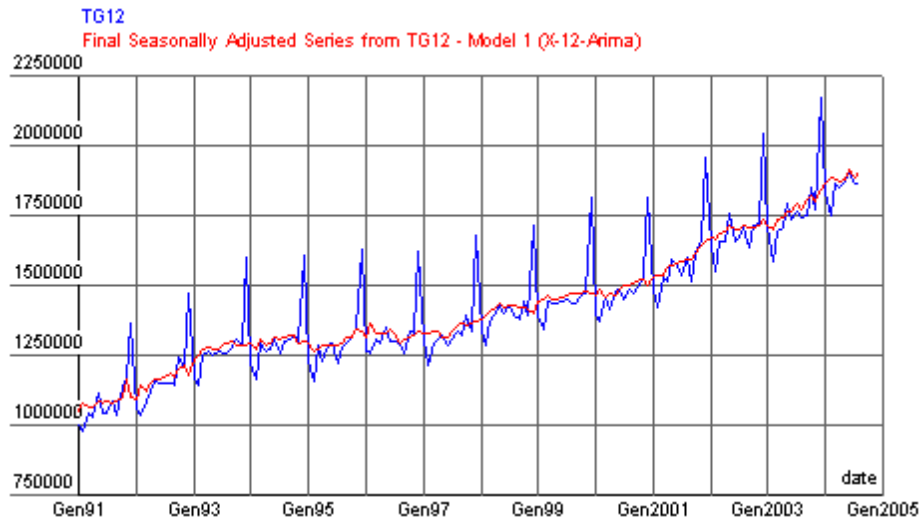


Figura 12: farmacie

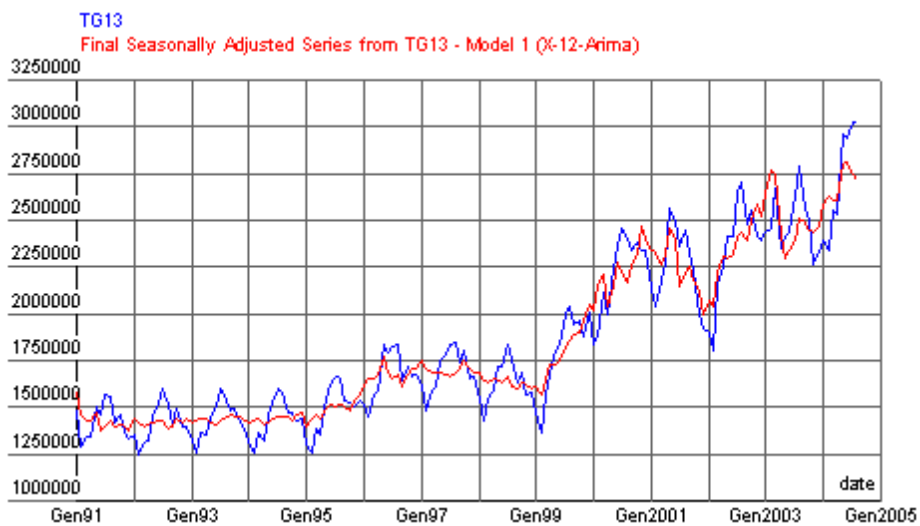


Figura 13: stazioni di servizio

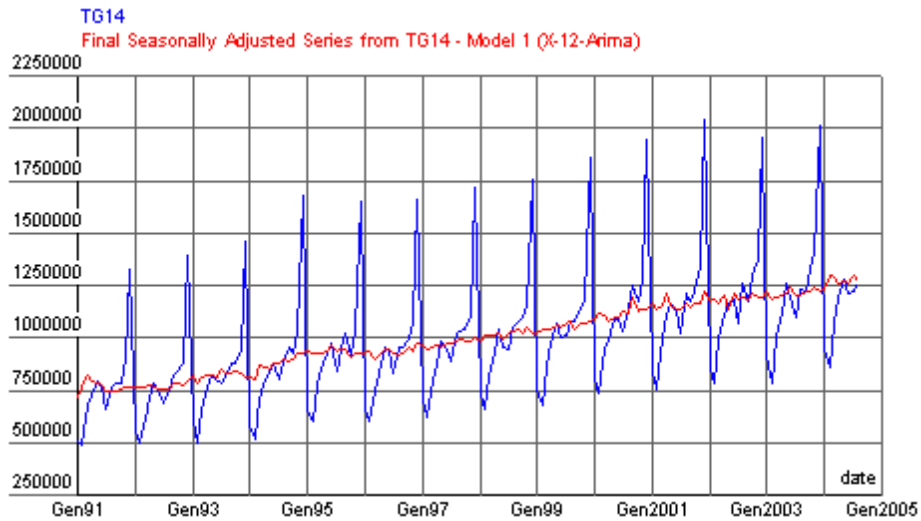


Figura 14: abbigliamento

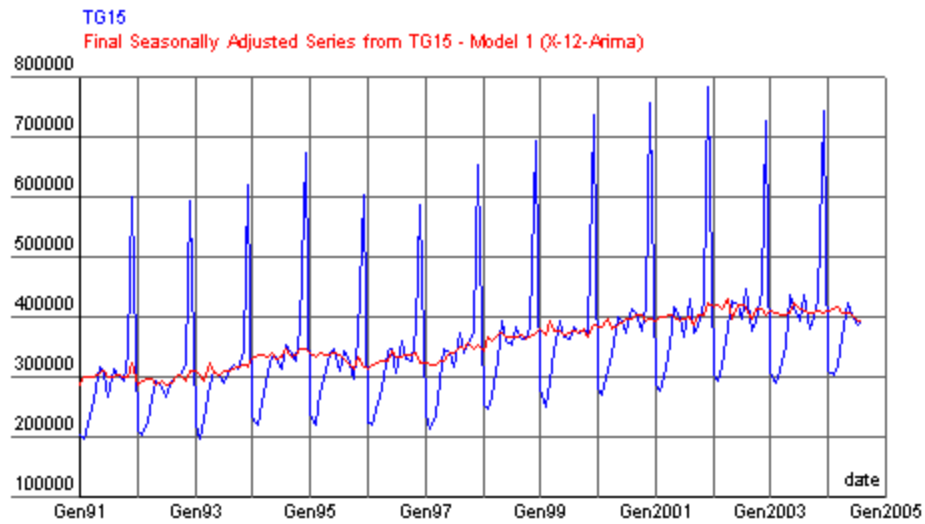
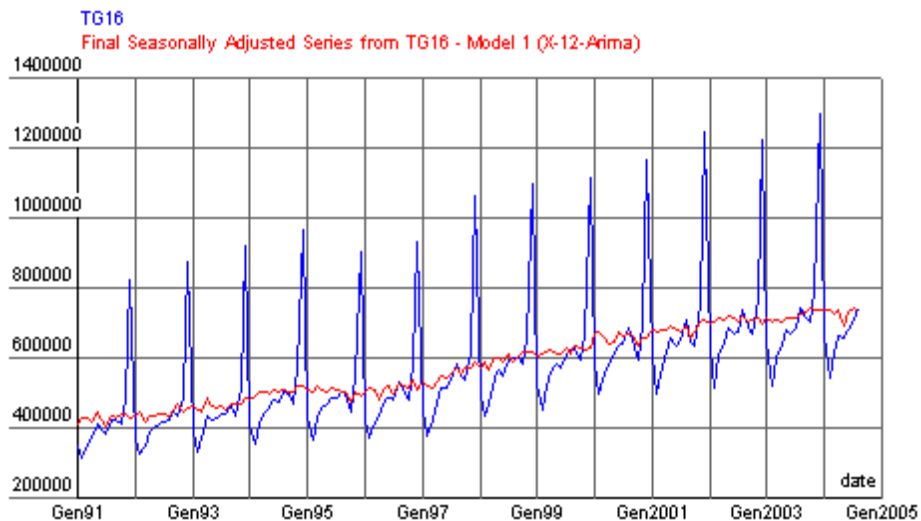
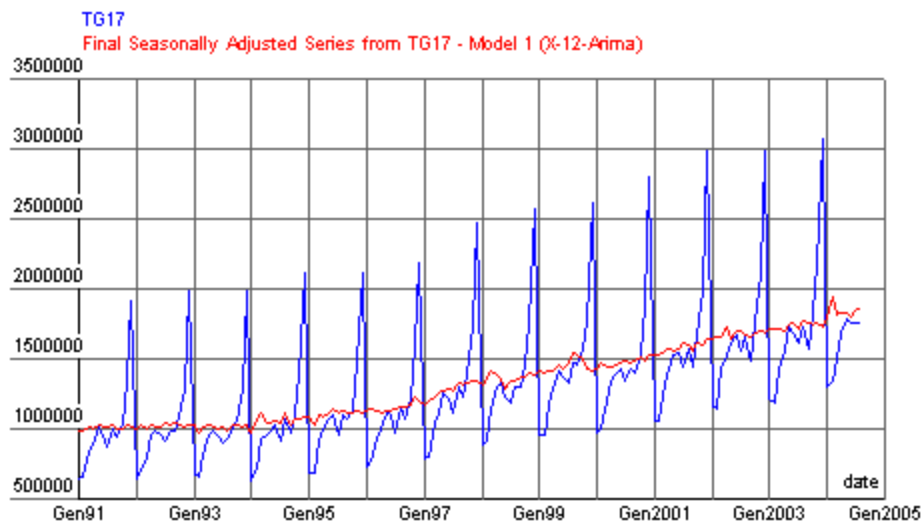


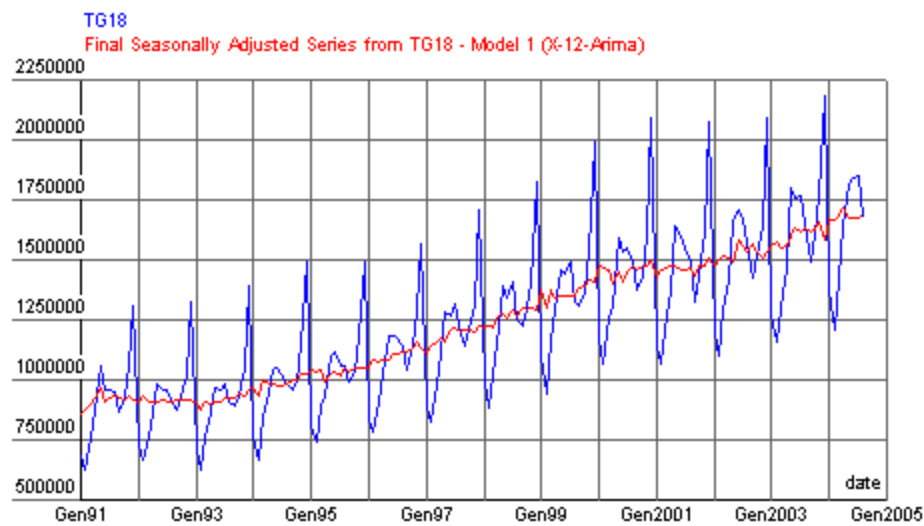
Figura 15: negozi di scarpe



**Figura 16: negozi di articoli sportivi**



**Figura 17: grandi magazzini**



**Figura 18: altri negozi**

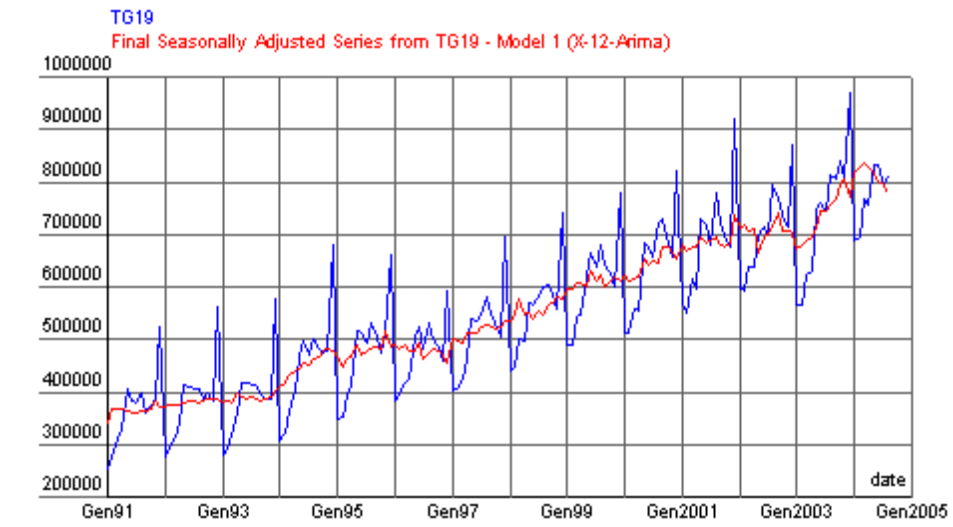


Figura 19: Altri venditori al dettaglio

I dati nazionali sono riportati nelle ultime due figure:

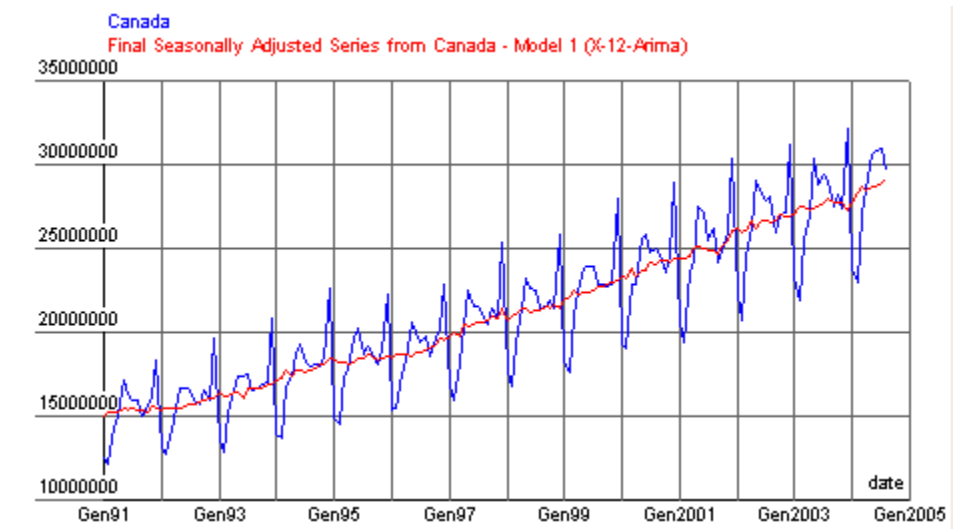
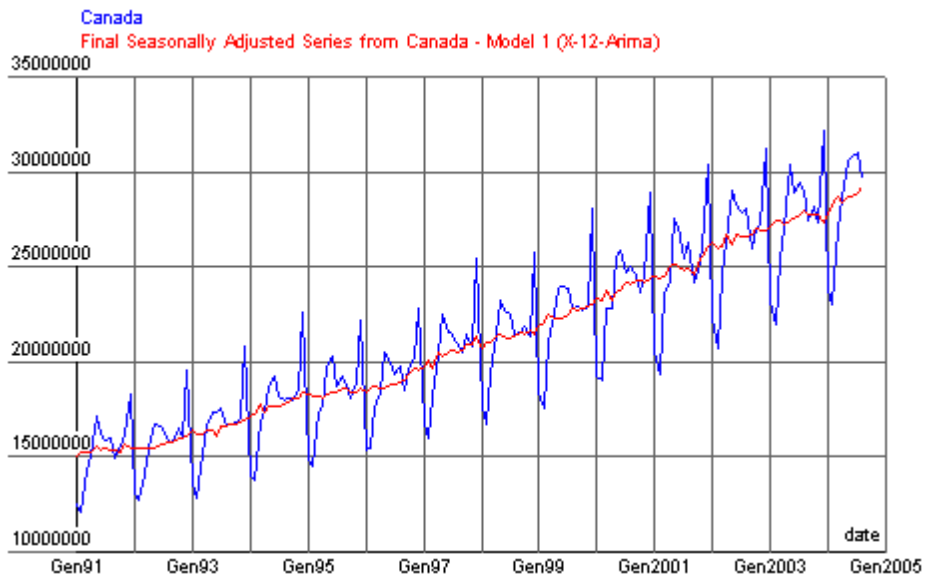


Figura 20: totale Canada per settori di produzioni



**Figura 21: Totale Canada per provincie**

Come detto per la destagionalizzazione, abbiamo utilizzato il programma Demetra, applicando sia X-12 che TRAMO-SEATS; questo per poter valutare quale dei due metodi meglio si adatti ai dati utilizzati. Come si vedrà dalla batteria di test proposta in seguito, non si può concludere univocamente a favore di uno o dell'altro metodo. Poiché però il metodo X-12 meglio si adatta alla teoria discussa nei capitoli precedenti, si è scelto di utilizzare questo come metodo principale.

Demetra fornisce in automatico degli indicatori che permettono di valutare l'effettiva riuscita della destagionalizzazione. In particolare, in automatico, viene fornita un indice che permette di valutare la qualità della destagionalizzazione. Quest'ultima viene considerata accettabile se il valore dell'indice resta all'interno dell'intervallo (0, 10). Tutti i valori della statistica per le serie considerate sono accettati.

Anche le statistiche di Ljung-box, condotte sia sulle serie sui residui, che sui loro quadrati, permettono di confermare che la procedura di destagionalizzazione può ritenersi soddisfacente.



Nel secondo capitolo abbiamo parlato dei due metodi principali per il trattamento della stagionalità, ovvero le procedure derivate dall'X-11-ARIMA e TRAMO SEATS. Esistono dei test per stabilire quali dei due metodi produca una destagionalizzazione migliore, test descritti nel secondo capitolo di questa tesi.

Per la nostra analisi abbiamo utilizzato sia la procedura X-12-ARIMA che quella prevista da TRAMO-SEATS. Resta quindi da stabilire quali di questi due metodi meglio si adatti ai dati e se sia meglio applicare una destagionalizzazione di tipo diretto o indiretto. Fatte queste scelte si potrà procedere con le procedure di riconciliazione. È facilmente intuibile come la scelta tra un approccio parametrico o non o tra un metodo di aggiustamento stagionale diretto e quello indiretto non sia facile e i test siano spesso in contraddizione fra di loro.

Questo problema è emerso anche nei dati analizzati in questa tesi. Nel seguito sono riportati i valori di alcuni dei test descritti nel secondo capitolo. Questi permettono di confrontare le varie procedure utilizzate:

|             | TG_x12      | TG_TS       | PR_x12     | PR_TS       |
|-------------|-------------|-------------|------------|-------------|
| Mean_APD_SA | 2.95E-05    | 0.000186141 | 2.03E-05   | 0.00043     |
| Mean_APD_TC | 0.000319013 | 3.04E-05    | 0.00026    | 0.000191158 |
| max_APD_SA  | 0.014049836 | 0.014819149 | 0.013368   | 0.008038    |
| max_APD_TC  | 0.007227221 | 0.013648017 | 0.00430102 | 0.008952341 |
| R1_dir      | 8.61E+14    | 8.75E+14    | 8.61E+14   | 8.75E+14    |
| R1_ind      | 8.50E+14    | 8.54E+14    | 8.64E+14   | 8.60E+14    |
| R3_dir      | 3.87E+17    | 3.76E+12    | 3.87E+12   | 3.76E+12    |
| R3_indir    | 4.32E+16    | 4.32E+16    | 7.36E+16   | 7.30E+16    |
| Mar1_dir    | 8.48E+14    | 8.55E+14    | 8.48E+14   | 8.55E+14    |
| Mar1_ind    | 8.43E+14    | 8.43E+14    | 8.44E+14   | 8.50E+14    |
| Mar2_dir    | 7.17E+29    | 7.26E+29    | 7.17E+29   | 7.26E+29    |
| Mar2_ind    | 7.08E+29    | 7.07E+29    | 7.10E+29   | 7.19E+29    |

**Figura 22: batteria di test per la valutazione dei metodi di destagionalizzazione**

Per comprendere meglio i valori riportati, vediamo per sommi capi gli indicatori in figura 4, descrivendo successivamente le conclusioni a cui questi permettono di giungere.

Con APD abbiamo indicato gli indicatori di deviazione percentuale assoluta (SA per le serie destagionalizzate e TC per le serie trend), Mean rappresenta la

deviazione percentuale media e Max la massima. Successivamente abbiamo riportato le statistiche R1 ed R3, così come sono state descritte nel capitolo 2. Infine con gli indicatori Mar1 e Mar2 si vuole mostrare se i dati siano dettagliati o grossolani. Dai valori tutti prossimi a zero si comprende che i dati non mostrano problemi da un punto di vista di aggiustamento stagionale e dettaglio.

Gli indicatori visti servono anche a comprendere quale metodo e quale procedura sia meglio utilizzare per la destagionalizzazione. Per questo motivo abbiamo evidenziato i valori che indicano quale tra x-12 e TRAMO-SEATS meglio si adatta ai nostri dati e se sia meglio un approccio diretto o indiretto. Come si vede c'è una leggera tendenza a preferire il metodo X-12 sia per la destagionalizzazione dei dati relativi ai settori di produzione che per quella dei dati relativi alle industrie, ma non si può dire che un metodo sia nettamente migliore dell'altro. Lo stesso discorso vale per la decisione tra la destagionalizzazione diretta e quella indiretta. Anche in questo caso i test non conducono ad una decisione univoca. Alcuni sembrano consigliare un metodo di destagionalizzazione diretta, altri indiretta.

Da questa breve analisi si vede in modo empirico quel che anticipavamo dal punto di vista teorico; non esiste infatti, almeno per ora, un metodo che riesca a condurre in modo univoco ad una scelta.

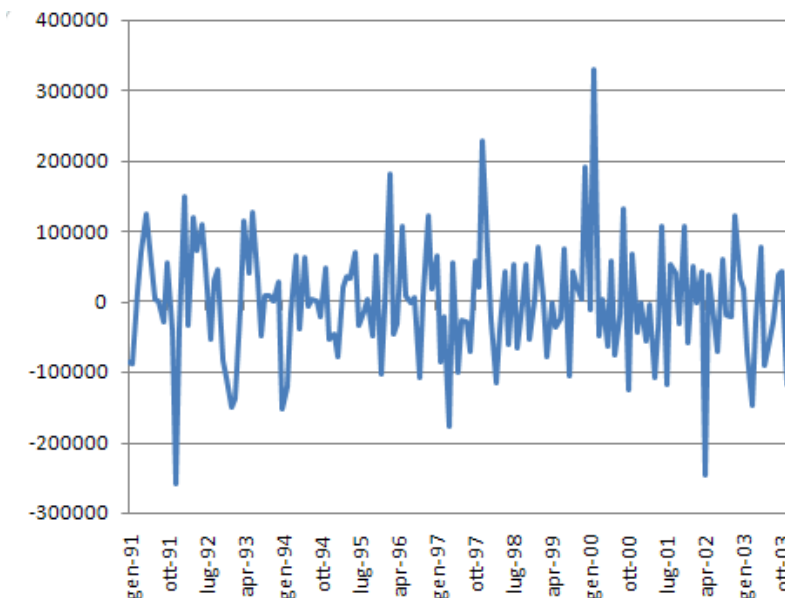
La tabella seguente mostra le differenze, in milioni di euro, tra le somme annuali delle serie dei settori di produzione destagionalizzate e i totali annui delle corrispondenti serie grezze:

|      | TG_01 | TG_02 | TG_03 | TG_04 | TG_05 | TG_06 | TG_07 | TG_08 | TG_09 | TG_10 | TG_11 | TG_12 | TG_13 | TG_14 | TG_15 | TG_16 | TG_17 | TG_18 | TG_19 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1991 | 0     | 1     | -1    | 0     | 1     | 0     | -1    | 0     | 1     | 0     | -1    | -1    | 2     | 0     | 1     | 1     | -1    | -1    | 0     |
| 1992 | -2    | -1    | 0     | 1     | -1    | 1     | 1     | 0     | -2    | 0     | -2    | -1    | -1    | 0     | -1    | 0     | -2    | 0     | -1    |
| 1993 | -3    | 1     | 2     | 0     | 0     | 0     | 1     | -2    | -4    | -1    | -1    | -3    | 0     | -1    | 0     | 1     | -1    | -2    | 0     |
| 1994 | -2    | 0     | 1     | 1     | 0     | 0     | -1    | 1     | -3    | -1    | -1    | -2    | -1    | -1    | -1    | -1    | 0     | 1     | 1     |
| 1995 | -1    | -1    | -1    | 0     | 1     | 0     | 1     | 0     | -2    | -1    | -1    | 0     | 0     | 0     | 1     | 0     | -2    | -1    | 0     |
| 1996 | -1    | 0     | 0     | 1     | -1    | -1    | 0     | 0     | -2    | 1     | 0     | 1     | 0     | -1    | -1    | -1    | 0     | -1    | -1    |
| 1997 | -3    | 1     | 0     | 1     | 0     | 0     | 0     | -1    | -2    | -1    | -1    | -2    | -1    | 0     | 0     | 0     | -2    | -1    | 0     |
| 1998 | -3    | -1    | 1     | 0     | 0     | 1     | -1    | -1    | -2    | -1    | 0     | -1    | -1    | -1    | 2     | 1     | 0     | -1    | 0     |
| 1999 | -3    | -1    | -1    | 1     | 0     | 0     | 0     | 0     | -1    | 0     | -1    | 0     | -1    | 1     | 0     | 0     | 0     | -1    | -1    |
| 2000 | -5    | 0     | -1    | 1     | 0     | 0     | 0     | 0     | -3    | 0     | 0     | -1    | -1    | 1     | 0     | 0     | -1    | -1    | 0     |
| 2001 | -2    | -1    | -1    | -1    | 0     | 2     | 1     | 1     | -3    | -1    | -2    | -1    | -1    | 0     | 0     | -1    | -1    | 0     | -1    |
| 2002 | -3    | -1    | -1    | 1     | -1    | -1    | 0     | -1    | -3    | 0     | -1    | -1    | -1    | -1    | 1     | 2     | -1    | -1    | -2    |
| 2003 | -2    | -1    | -1    | -2    | -1    | 1     | 0     | -1    | -2    | 0     | -1    | -1    | -3    | 2     | 0     | 0     | -2    | 0     | -1    |

**Figura 23: Discrepanze tra le somme annuali delle serie dei settori di produzione destagionalizzate e i totali annui delle corrispondenti serie grezze**

Come si può notare le differenze sono sempre piccole, e raggiungono i quattro milioni di dollari, nel 1993, nel settore 09, ovvero quello dei supermercati e i cinque milioni di dollari nel 2000 nel settore delle auto nuove; le altre differenze sono tutte entro i tre milioni di dollari.

Riportiamo anche le differenze mensili, sempre in milioni di dollari, tra la somma delle serie dei settori di produzione e il totale nazionale destagionalizzato.



## 6.5 La riconciliazione

Dopo aver destagionalizzato i dati, li abbiamo riconciliati.

In particolare è stata fatta la riconciliazione del sistema MRTS per settore industriale, considerando un vincolo esogeno. Per questo tipo di analisi abbiamo usato il software Matlab<sup>®</sup> 7, R 2007 a.

Come è già stato descritto nei capitoli dedicati alla teoria, le statistiche economiche sono spesso caratterizzate da relazioni contabili, definite da teorie economiche o vincoli di additività in base al livello di disaggregazione delle serie.

Solitamente, però, i vincoli non sono soddisfatti dalle variabili osservate, a causa di osservazioni più frequenti o di dati raccolti con metodi diversi. Il mancato rispetto dei vincoli causa problemi di discrepanze che devono essere corrette per evitare problemi in fase di pubblicazione dei dati.

Questa premessa sottolinea, ancora una volta, l'esigenza di correggere i data-set in modo tale da rendere soddisfatte le restrizioni di contabilità.

Come già descritto in precedenza ed anticipato in questo capitolo, i vincoli possono essere di due tipi: contemporanei e temporali. I primi hanno la forma di combinazioni lineari delle variabili che devono essere soddisfatti in ogni periodo osservato; i secondi richiedono che le serie ad alta frequenza riconciliate siano uguali ai valori degli aggregati rilevati a bassa frequenza.

Le procedure statistiche per ristabilire la consistenza tra le variabili ed all'interno delle variabili sono molto simili, ma i due problemi sono trattati, a livello pratico, con le stesse procedure, malgrado, in questa tesi e, in generale, in letteratura vengono trattati in modo separato. Per questo motivo ricordiamo che il processo di aggiustamento per la dimensione temporale è chiamato benchmarking, quello di riconciliazione come bilanciamento.

Le maggiori difficoltà che si possono incontrare nella soluzione di questi problemi, sono causate dalla grandezza delle matrici che entrano in gioco. Maggiore, infatti, è il numero delle variabili, più le procedure vengono complicate a livello computazionale.

Negli ultimi anni è stato implementato un metodo per la risoluzione del problema che prevede l' applicazione di una procedura a due stadi (Quenneville e Rancourt, 2005); questa procedura divide la riconciliazione in due passi: nel primo, le serie preliminari vengono sottoposte ad una procedura di benchmarking, considerando come serie di riferimento gli aggregati temporali noti, in modo che il livello temporale venga alterato il meno possibile venga garantito il rispetto dei vincoli contemporanei per ogni periodo a bassa frequenza, anno per anno ad esempio.

Questa procedura a due stadi ha due vantaggi: infatti utilizzando questo metodo il problema viene ridotto, poiché si considerano due momenti di applicazione; inoltre l'applicazione del processo a due stadi non ha bisogno di conservare l'andamento delle caratteristiche della serie (ad esempio quello del tasso di crescita mensile) al secondo stadio, perché questo è già garantito al primo passo.

Per chiarire meglio la situazione descritta, descriviamo la situazione da un punto di vista teorico, per poi verificare ciò che si verificherà sui dati dell' MRTS.

Come già spiegato, un sistema di serie storiche a frequenza inferiore ad un anno, deve risultare coerente con i vincoli di aggregazione noti. Quello che stiamo considerando è quindi un set di vettori  $\mathbf{M}$ , con  $n$  righe e una colonna, di dati ad alta frequenza,  $\mathbf{Y}_j$ ,  $j=1,\dots,M$ , che devono soddisfare sia il vincolo temporale che quello contemporaneo. Quello che si conosce è un vettore di dati ad alta frequenza  $M$ ,  $M$  aggregati temporali e alcune serie di aggregazione contemporanea. Sia ora  $\mathbf{G}$  la matrice, con  $k$  righe ed  $M$  colonne, di costanti note che definisce le relazioni tra  $\mathbf{Y}$ , con  $k$  il numero di relazioni da soddisfare. Sia inoltre  $\mathbf{Z}_h$ , con  $h=1,\dots,k$ , i vettori ( $n \times 1$ ) ad alta frequenza di quantità note, associati ai  $k$  vincoli di contabilità in  $\mathbf{G}$ . considerando il vettore  $\mathbf{Y} = [\mathbf{Y}'_1 \mathbf{Y}'_2 \dots \mathbf{Y}'_M]'$ , formato da tutte le serie componenti non note, i vincoli di aggregazione contemporanea possono essere scritti come:

$$(\mathbf{G} \otimes \mathbf{I}_n) \mathbf{Y} = \mathbf{Z}$$

Assumiamo inoltre che i vettori a bassa frequenza  $\mathbf{Y}_{0j}$  sono disponibili; ogni vettore può essere descritto come una combinazione non lineare dei vettori  $\mathbf{Y}_j$ , con coefficienti dati dal vettore  $\mathbf{c}$  con  $s$  righe ed una colonna, con  $s$  l'ordine di aggregazione temporanea (12 se mensile, 3 se trimestrale etc.). Definiamo

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_N \otimes \mathbf{c}' \\ \mathbf{0} \end{bmatrix}$$

la matrice di aggregazione con  $N$  righe ed  $n$  colonne che converte le serie ad alta frequenza in quelle a bassa frequenza, con  $\mathbf{0}$  una matrice nulla con  $N$  righe ed  $(n - Ns)$  colonne che permette di rendere compatibili, ad esempio le serie preliminari ad alta frequenza con i periodi a bassa frequenza più recenti non noti.

I vincoli temporali che collegano le serie componenti ad alta frequenza con le corrispondenti serie aggregate possono essere espressi come  $\mathbf{C}\mathbf{Y}_j = \mathbf{Y}_{0j}$ ,  $j=1, \dots, M$ , ovvero come:

$$(\mathbf{I}_M \otimes \mathbf{C}) \mathbf{Y} = \mathbf{Y}_0.$$

Consideriamo adesso la matrice di aggregazione  $\mathbf{H}$ , rappresentata a blocchi come segue:

$$\mathbf{H} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{I}_n \\ \mathbf{I}_M \otimes \mathbf{C} \end{bmatrix}$$

e  $\mathbf{Y}_a = [\mathbf{Z}'\mathbf{Y}'_0]$  sia il vettore contenente sia i vincoli temporali che contemporanei. Il set completo di vincoli tra le serie componenti ad alta frequenza non note e le informazioni disponibili, possono essere scritti come:

$$\mathbf{H}\mathbf{Y} = \mathbf{Y}_a$$

I vincoli di aggregazione delle serie aggregate temporalmente implicano che:

$$(\mathbf{G}_h \otimes \mathbf{I}_N) \mathbf{Y}_0 = \mathbf{C}\mathbf{Z}_h,$$

Con  $h=1, \dots, k$ , dove  $\mathbf{G}_h$  è l' $h$ -esimo vettore riga della matrice  $\mathbf{G}$ . Si può quindi considerare l'intero set di vincoli

$$(\mathbf{G} \otimes \mathbf{I}_N) \mathbf{Y}_0 = (\mathbf{I}_k \otimes \mathbf{C}) \mathbf{Z}$$

Che mostra come le informazioni esogene debbano essere consistenti con i vincoli del sistema. Questo vuol dire che le serie componenti a bassa frequenza, quando vengono aggregate con l'ausilio della matrice  $\mathbf{G}$ , devono essere uguali alle serie ottenute dall'aggregazione temporale delle serie ad alta frequenza in  $\mathbf{Z}$ . detto in altro modo,  $\mathbf{Y}_0$  e  $\mathbf{Z}$  rappresentano tutti i vincoli contemporanei e temporali.

### 6.5.1 La riconciliazione con la procedura a due stadi

Come già accennato, il problema maggiore delle procedure di riconciliazione è dato dalla dimensione della matrice di dati che si sta trattando. Nel caso di matrici con un numero elevato di righe e colonne, applicare una riconciliazione con una procedura simultanea può creare problemi computazionali; per questo motivo Quenneville e Rancourt hanno proposto una soluzione a due stadi per garantire l'additività di un sistema di serie storiche destagionalizzato; la procedura consiste in due passi: il primo prevede una procedura di benchmarking (nel nostro caso considerando il modello di Denton AFD), per garantire l'additività temporale di ogni serie destagionalizzata; nel secondo caso le componenti delle serie SA vengono riconciliate anno per anno usando una procedura di bilanciamento basata sui minimi quadrati.

Per chiarire meglio il primo stadio analizziamo il problema da un punto di vista matriciale. Come detto, il sistema di partenza è dato dalle 19 serie dei settori industriali del sistema MRTS. Il sistema di partenza è costituito dalla matrice:

$$\begin{matrix} y_{1,1} & \cdots & y_{19,1} \\ \vdots & \ddots & \vdots \\ y_{1,n} & \cdots & y_{19,n} \end{matrix}$$

Con  $t=1,\dots,n$  che rappresenta i mesi,  $j=1,\dots, M$  rappresenta i settori industriali e  $T=1,\dots,N$  che rappresenta gli anni. I valori annuali delle serie si ottengono come somma dei dati grezzi. Questi saranno quindi rappresentati con una matrice come segue:

$$\begin{matrix} y_{0,1,1} & y_{0,2,1} & \cdots & y_{0,19,1} \\ y_{0,1,2} & y_{0,2,2} & \cdots & y_{0,19,2} \\ \vdots & \vdots & \cdots & \vdots \\ y_{0,1,N} & y_{0,2,N} & \cdots & y_{0,19,N} \end{matrix}$$

Le serie ottenute dopo la destagionalizzazione, che indicheremo con  $sa_{ij}$ , costituiscono la nuova matrice:

$$\begin{matrix} sa_{1,1} & \cdots & sa_{19,1} \\ \vdots & \ddots & \vdots \\ sa_{1,n} & \cdots & sa_{13,n} \end{matrix}$$

Il problema nasce dal fatto che, dopo la procedura di destagionalizzazione, si ha che l'uguaglianza

$$\sum_{t \in T} sa_{jt} = y_{0JT}$$

non è più verificata.

Per questo motivo c'è la necessità di applicare, come primo passo, una procedura di benchmark univariato, che permetta di ottenere dei valori

$$\sim sa_{jt}$$

Tali per cui valga



$$\sum_{t \in T} \tilde{sa}_{jt} = y_{0JT}$$

Sempre considerando la notazione di Di Fonzo e Marini, sia  $\mathbf{B}$  il vettore con  $Mn$  righe ed una colonna che contiene tutte le serie sottoposte alla procedura di benchmarking temporale; il secondo passo è un problema di minimizzazione vincolata considerando non tutti i vincoli ma solo un vincolo annuale per volta. Come tutti i problemi di minimizzazione, la funzione obiettivo da minimizzare è data da:

$$F_T = \sum_{j=1}^M \sum_{t=(T-1)s+1}^{T_s} \frac{(R_{j,t} - B_{j,t})^2}{B_{j,t}}$$

Dove  $T$  denota che la procedura viene fatta considerando i periodi separatamente. Resta comunque ovvio che, questa procedura a due stadi sia valida anche nel caso in cui si considerino tutti i vincoli insieme ed in questo caso coincide con la riconciliazione con procedura simultanea.

Un aspetto interessante è che la procedura di riconciliazione a due passi produce correzioni che cambiano in base alla dimensione della variabile riconciliata; infatti la natura dell'aggiustamento è sempre proporzionale; quindi, assumendo una relazione diretta tra il livello e la varianza della variabile considerata, risulta ovvio che più grande è la variabile, maggiore sarà il suo aggiustamento e viceversa.

In particolare, (Kim et al. , 1994) se un elemento è  $k$  volte più grande di un altro, il suo livello di aggiustamento sarà  $k^2$  volte più grande degli altri e il suo aggiustamento proporzionale sarà  $k$  volte più grande.

Se lo stesso elemento è  $k$  volte più grande e ed è  $m$  volte meno accurato, il livello di aggiustamento sarà  $k^2 m^2$  volte più grande e quello proporzionale sarà  $km^2$ .

Chiaramente, anche considerando modelli a più dimensioni, il problema resta uguale, poiché le dimensioni maggiori non aumentano il livello di conoscenza circa l'aggiustamento.

Come abbiamo fatto per la destagionalizzazione, anche il risultato della riconciliazione va valutato, per avere più chiaro il risultato ottenuto. Stuckey et al. (2004), ad esempio, considerano che:

- 1) Le stime delle serie riconciliate devono essere vicine al livello delle serie preliminari.
- 2) Le serie riconciliate non abbiano grandi differenze nel passaggio da un periodo a quello successivo
- 3) Serie con grande volatilità siano corrette di più di quelle con volatilità minore.

Seguendo sempre il lavoro di Di Fonzo e Marini, abbiamo considerato due indicatori, per valutare la riconciliazione delle serie precedentemente destagionalizzate sia per le serie riconciliate sia per i tassi di cambio percentuali. Le due statistiche sono la radice quadrata dell'aggiustamento percentuale (acronimo inglese MSPA) e la radice dell'aggiustamento medio (acronimo MSA).

Le due misure sono espresse, rispettivamente, dalle formule:

$$MSPA(R_j, P_j) = 100x \sqrt{\frac{1}{n} \sum_{t=1}^n \left( \frac{R_{j,t}}{P_{j,t}} - 1 \right)^2}$$

$$MSA(r_j, p_j) = 100x \sqrt{\frac{1}{n-1} \sum_{t=2}^n (r_{j,t} - p_{j,t})^2}$$

dove

$$r_{j,t} = \frac{R_{j,t}}{R_{j,t-1}} - 1$$

e

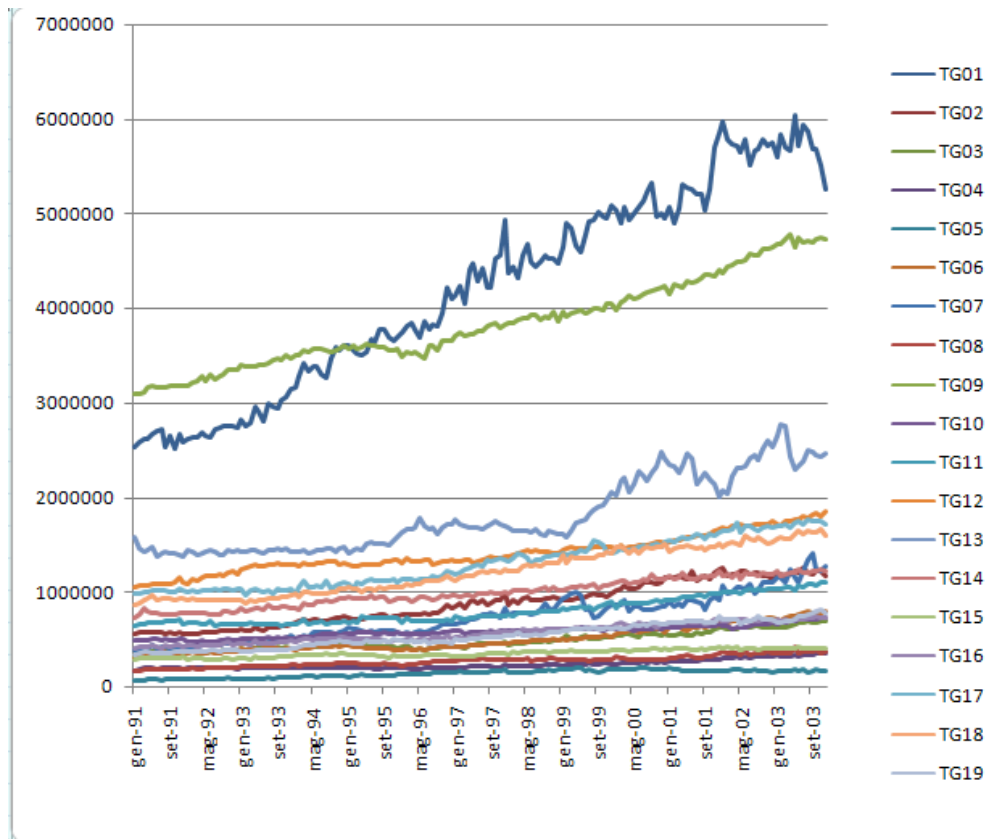
$$p_{j,t} = \frac{P_{j,t}}{P_{j,t-1}} - 1$$

sono i tassi di crescita delle serie riconciliate e di quelle preliminari.

Considerando come esempio pratico il data set del MRTS descritto all'inizio, abbiamo utilizzato dati dal gennaio 1991 al dicembre 2003. Dopo la destagionalizzazione descritta, abbiamo applicato la procedura di riconciliazione a due stadi. Al primo passo abbiamo applicato la procedura di benchmarking alle 13 serie destagionalizzate delle provincie secondo la procedura di Denton a differenze prime additive, in modo da eguagliare i totali annuali a quelli della somma delle serie componenti. In questo modo si impone che il totale nazionale sia coerente con i dati del sistema delle 19 industrie; questa procedura serve a rendere più evidenti valori anomali o rotture strutturali nel modello. I vincoli totali annuali sono soddisfatti automaticamente.

Riassumendo, si ha un numero di serie (M) con 19 righe, 156 mesi e 13 anni. Inoltre c'è un vincolo contemporaneo. Il numero delle serie preliminari da riconciliare sono quindi 2964 (ottenuto dal prodotto di 19 per 156).

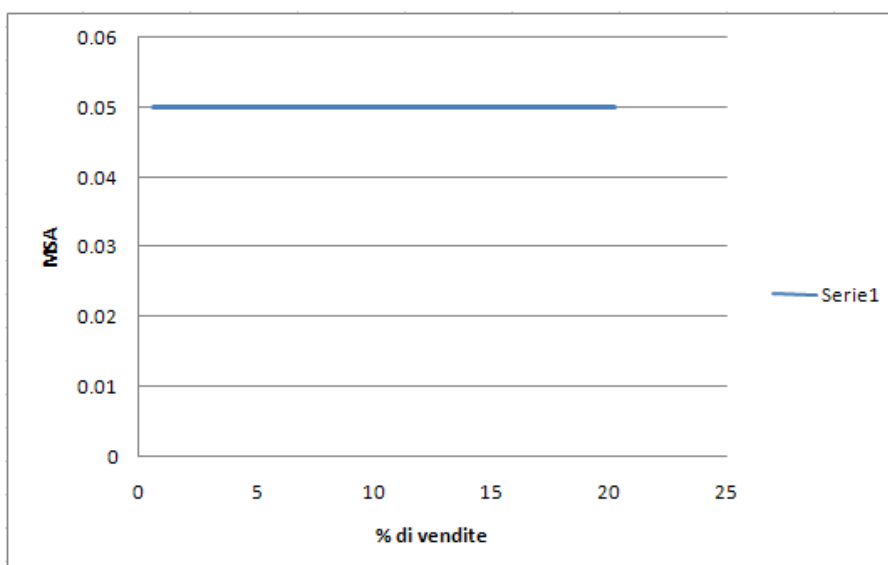
I grafici delle serie ottenute sono riportate nel seguente grafico:



**Figura 3: Grafico delle serie riconciliate**

Dal grafico si può notare il livello più alto delle serie riconciliate dei venditori di auto nuove e dei supermercati. Da sottolineare come, ovviamente, anche in queste serie, la stagionalità è assente dalle 19 serie dei settori produttivi.

Riportiamo infine il grafico dell' MSA delle serie riconciliate



Come ci si aspettava, il grafico dell' MSA è costante e vale circa 0.05.

## **CAPITOLO 7**

### **CONCLUSIONE**

La tesi, che giunge in questo capitolo alla sua conclusione, considera un problema di primaria importanza all'interno della produzione di statistiche ufficiali, quello della riconciliazione dei conti economici.

Spesso, infatti, le tavole di contabilità nazionale, ottenute dai produttori statistici, non soddisfano i vincoli di aggregazione, sia temporali che contemporanei. Questo ha creato negli anni vari filoni di studio per sviluppare le

tecniche di benchmarking e di riconciliazione delle serie storiche ed ovviare ai problemi sopracitati.

I dati considerati hanno però un'ulteriore problema. Infatti, come la teoria sulle serie storiche spiega, hanno al loro interno, oltre alle componenti di ciclo e trend, una componente stagionale. Per questo motivo, prima di procedere alle procedure di riconciliazione è opportuno procedere alla destagionalizzazione delle serie considerate.

Questa operazione può essere fatta con due procedure e tre modi diversi. Le procedure principali sono quelle derivate dal metodo *ARIMA-X-11*, sviluppato dal *Census Bureau* e *TRAMO-SEATS*, creato da Maravallo e Caporello; la componente stagionale può inoltre essere eliminata con tre metodi, così come descrivono Astolfi, Ladiray e Mazzi:

- Con un metodo diretto, in cui vengono destagionalizzate tutte le serie considerate, quindi sia quelle delle provincie sia quelle dei settori produttivi, e anche la corrispondente serie grezza del totale nazionale.
- Con un metodo indiretto, in cui, dopo aver destagionalizzato le serie componenti, si considera come serie totale destagionalizzata quella ottenuta dalla somma delle singole serie destagionalizzate
- Con un metodo spurio, formato da un misto delle due opzioni precedenti. C'è da sottolineare che questo metodo è quello che per ora riscontra meno consenso e quindi risulta anche meno usato.

La molteplicità di possibilità per risolvere la questione, pone il problema di trovare strumenti adatti a giudicare quale metodo sia migliore nel trattare i dati in esame; in realtà, come visto nel secondo capitolo, le batterie di test che cercano di dare risposta a questo quesito, non danno risposta univoca; questo implica che in ogni situazione la scelta sulla procedura e sul metodo è del tutto soggettiva.

Nel terzo capitolo abbiamo descritto le procedure di bilanciamento di serie storiche. All' inizio abbiamo esposto il problema dei vincoli contemporanei e temporali, seguendo la descrizione fatta da Schneider e Zenios prima e da Di Fonzo e Marini poi; abbiamo suddiviso le tecniche di bilanciamento sulla base degli algoritmi utilizzati; questi sono di due categorie: quelli proporzionali, come il RAS, e quelli non proporzionali. Questa distinzione ci ha suggerito una rapida panoramica su questi due approcci.

Il quarto capitolo entra nel vivo del problema trattato nella tesi. Si inizia a parlare di benchmarking di serie storiche, seguendo la descrizione di Cholette e Dagum, principali teorici della questione. Vengono descritti i due modelli principali, quello additivo, della quale se ne vede anche una variante con benchmarking proporzionale ed uno moltiplicativo. Inoltre viene descritto il metodo di Denton, con la sua variante a differenze prime, procedura fondamentale per l'analisi pratica condotta nel sesto capitolo.

Nel quinto capitolo si parla di riconciliazione di serie storiche. Gli autori di riferimento per questo argomento sono Quenneville e Rancourt, che, insieme a Dagum e Cholette hanno implementato un metodo di regressione per la soluzione del problema, al quale hanno poi affiancato le tecniche di riconciliazione di sistemi di serie storiche ad una e a due variabili. Queste tecniche sono di fondamentale importanza per comprendere ciò che è stato fatto nel capitolo successivo, dove si è applicata la procedura di riconciliazione ad una variabile ai dati sui settori produttivi canadesi.

Il sesto ed ultimo capitolo presenta un' applicazione di ciò che abbiamo descritto da un punto di vista teorico ai dati del *MRTS* del Canada. L'ordine seguito nell'analisi è quello dei capitoli teorici. Dunque si è partiti destagionalizzando le serie riguardanti il settore industriale dei dati canadesi; secondo passo è stato applicare le procedure di benchmark univariato, secondo il metodo di Denton. Infine abbiamo applicato le procedure di riconciliazione perché fossero soddisfatti i vincoli temporali e di aggregazione.

## BIBLIOGRAFIA

AA.VV. (2008). Matlab 7, getting started guide, mathworks

Astolfi, R., Ladiray, D., Mazzi, G.L. (2004). *Seasonal adjustment European Aggregates: Direct versus Indirect Approach*, Lussemburgo

Aston, J.A.D, Findley, D.F., McElroy, T.S., Wils, K.C. e Martin, D.E.K. (2007). New ARIMA models for seasonal time series and their applications to seasonal adjustment and forecasting, research report series, *Statistics #2007-14*

Bacharach, M. (1970). *Biproportional matrices and input-output change*, University Press, Cambridge

Bloem, A.M. e Dippelsman, R.J. e Mæhle O. N. (2001). Quarterly national accounts manual, concepts, data sources, and compilation, 2001 Washington, DC: International Monetary Fund. Sito web: [www.imf.org/external/pubs](http://www.imf.org/external/pubs)

Box, G.E.P., Hillmer, S.C. e Tiao, G.C. (1975) Analysis and modeling of seasonal time series, Working paper

Burnett, M. Seasonal adjustment of UK monetary aggregates: direct versus indirect approach

Ciaburro, G. (2005). Manuale Matlab, Dispensa on line, indirizzo: <http://www.ciaburro.it>

Dagum, E.B. (1979). "On the Seasonal Adjustment of economic Time Series Aggregates: A Case Study of the Unemployment Rate", *Counting of the Labour Force, National Commission on Employment and Unemployment Statistics, Appendix 2*, 317-344, Washington

Dagum, E. B. e Cholette, P. A. (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*, Springer Science, New York

Denton, F.T. (1990). 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. (1990). The estimation of M disaggregate time series when contemporaneous and temporal aggregates are known, *The Review of Economics and Statistics, Notes*

Di Fonzo, T. (2002), Temporal disaggregation of a system of time series when the aggregate is known: Optimal VS. adjustment methods, *Workshop on Quarterly National Accounts, Eurostat, Theme 2 Economy and Finance: 63-77*

Di Fonzo, T. (2003) Temporal disaggregation of economic time series: towards a dynamic extension, Dipartimento di Scienze Statistiche, Università di Padova

Di Fonzo, T. (2003). *Temporal disaggregation using related series: log-transformation and dynamic extensions*, RISEC, Volume 50, No 3, 371-400

Di Fonzo, T e Lisi, F. (2004). *Introduzione alle serie storiche*, Carocci, Padova

Di Fonzo, T e Marini, M. (2005). Benchmarking Systems of Seasonally Adjusted Time Series, *Journal of Business cycle Measurement and Analysis-Vol. 2*, Pagine 89-123

Di Fonzo, T e Marini, M. (2006). Benchmarking a system of time series: Denton's movement preservation principle vs a data based procedure, Working Paper 2006, Eurostat, Luxemburg

Di Fonzo, T e Marini, M. (2008), Least-squares reconciliation of large systems of sub-annual time series: alternative solutions, XLIV Riunione Scientifica Società Italiana di Statistica, 27 Giugno 2008

Di Fonzo, T e Marini, M. (2009), Simultaneous and Two Step Reconciliation of Systems of Time series, Department of Statistical Sciences, University of Padua

Feldpausch, R.M., Hood, K.M. e Wills, K.C. Diagnostics for Model-Based seasonal adjustment, U.S. Census Bureau, Washington, D.C. 20233-6200, 2003

Findley, D. F. (2005). *Some recent developments and directions in seasonal adjustment*, *Journal of Official Statistics*, vol. 21, No. 2, pp 343-365

Findley, D. F. e Monsell, B. C. e Bell, W. R. e Otto, M. C. e Chen, B. (1998) New capabilities and methods of the X-12- ARIMA seasonal Adjustment Program, *Journal of Business and Economic Statistics*, 16, 127-177; U. S. Census Bureau, *X-12-ARIMA Reference Manual, reference 0.2.8*, Washington DC, U.S. Bureau of the Census}

Findley, D. F. e Hood, K.M., (2002), X-12-ARIMA and its application to some Italian indicator series, U.S. Census Bureau

Findley, D. F. e Monsell, B. C. (2006) Modeling stock trading day effects under flow day of week effects constraints, *Research Report Series , Statistics #2006-10*

Guerrero, V.M. (1990). Temporal disaggregation of time series: an ARIMA-based Approach, *International Statistical Review* 58:29-46

Hood, K.M e Findley, D.F. (2002) Comparing direct and indirect seasonal adjustment of aggregate series, U.S. Census Bureau, Washington, DC 20233

Granger, W.J.C. (1978). Seasonality: causation, interpretation and implications, *University of California*, La Jolia

Guerrero, V.M. e Pena, V. (2000). Linear combination of restrictions and forecasts in time series analysis, *Journal of Forecasting* 19:103-122

Hindrayanto, I. (2004). Seasonal adjustment: direct or indirect or multivariate method?, U.S. Census Bureau

Hood, C. C e Feldpausch, R.M. (2003). Some properties of the seasonal adjustment diagnostics in X-12-ARIMA, proceedings of statistics Canada symposium

Karlsen, T., Pham, D. Q., Skjerpen, T. Seasonal adjustment and smoothing of manufacturing investments series from the quarterly Norwegian national accounts

Kuiper, J. (1975). A survey and comparative analysis of various methods of seasonal adjustment, University of Ottawa

Ladiray, D. e Quenneville, B. (2001) *Seasonal Adjustment with X-11 Method*, Lecture Notes in Statistics, Springer-Verlag, New York

Monsell, B.C. e Aston J.A.D. e Koopman, J.S. (2005) Toward X-13?, Working Paper, Washington DC

Otranto, E. e Triacca, U. (2002), Measures to evaluate the discrepancy between direct and indirect model-based seasonal adjustment, *Journal of Official Statistics*, Vol. 18, No. 4, pagine 511-530

Otranto, E. e Triacca, U. (2003). A distance-based method for the choice of the direct or indirect seasonal adjustment, Istituto nazionale di statistica

Quenneville, B. e Rancourt, E. (2005). Simple method to Restore the Additivity of Time series, Workshop on Frontiers in Benchmarking Techniques and their application to official statistics, 2005, Statistic Canada

Quenneville, B. e Fortier, S. Balancing seasonally adjusted series as a complement to the direct and indirect approaches to seasonal adjustment, Statistics Canada, Ottawa, Ontario, K1A 0T6, Canada

Schneider, M. e Zenios, S. (1990). A comparative study of algorithms for matrix Balancing, *Operations Research*, 38, 439–455

Stone, R. (1961). *Input-Output and National Accounts*, OECD, Paris

Stone, R. e Champernowne, R. e Meade, J. (1942). “*The precision of national income estimates*”, *Review of Economic Studies*, 9, 111–125

Toh, M.H. The RAS Approach to Updating Input-Output Matrices: an Instrumental Variable Interpretation and Analysis of Structural Changes, *Economic System Research*, 10, 63-68

Uribe, P. C. e de Leeuw e Theil, P. C. (1970). The information theoretic approach to the prediction of interregional trade flows. *Rotterdam, Econometric Institute of the Netherlands School of Economics, Report 6507*