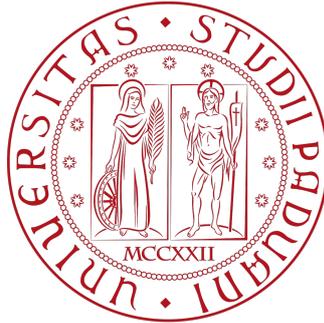


UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per le Tecnologie e per le Scienze



RELAZIONE FINALE  
**Riconciliazione ottimale cross-sezionale delle previsioni per serie  
storiche gerarchiche e raggruppate**

*Relatore: Prof. Tommaso Di Fonzo  
Dipartimento di Scienze Statistiche*

*Laureando: Andrea Frasson  
Matricola N.: 1190901*

**ANNO ACCADEMICO 2021/2022**



## *Sommario*

La chiave per ottenere il successo in molti campi economici, dove le serie storiche gerarchiche rappresentano la quasi totalità delle strutture di dati presenti, è la tempestiva definizione delle politiche future e la programmazione dettagliata, che devono essere basate sulle informazioni raccolte. La formulazione di strategie di lungo periodo necessita di previsioni con un orizzonte più “lontano”, che devono essere aggregate attraverso una gerarchia secondo un principio fondamentale, la coerenza delle previsioni. Per poter prendere decisioni che siano allineate per qualunque livello della gerarchia, infatti, le previsioni devono sommarsi in maniera appropriata non solo secondo la gerarchia, ma anche attraverso diversi istanti di tempo. In questo studio si presenta un nuovo approccio per la previsione di serie storiche gerarchiche in grado di produrre stime ottimali che siano coerenti per le serie gerarchiche, migliori di quelle ottenute con metodi più tradizionali, come i metodi top-down o bottom-up. Questo nuovo approccio consiste nel generare in maniera indipendente le previsioni per tutti i livelli della gerarchia e, successivamente, riconciliarle incorporando le informazioni della matrice di varianza-covarianza dell’errore di previsione. Sotto semplici assunzioni, le previsioni finali ottenute dopo la riconciliazione, sono non distorte e con errore quadratico medio minimo tra tutte le possibili previsioni. Successivamente, verranno generate le previsioni per i dati sul flusso turistico australiano, disaggregati per regione geografica, e, una volta riconciliate con il metodo proposto, verranno confrontate con quelle ottenute con i metodi tradizionali per valutarne comparativamente le *performance*.



# Indice

<b>Sommario</b>	<b>iii</b>
<b>Indice</b>	<b>v</b>
<b>Elenco delle figure</b>	<b>vii</b>
<b>Elenco delle tabelle</b>	<b>ix</b>
<b>1 Introduzione</b>	<b>1</b>
<b>2 Serie storiche gerarchiche e raggruppate</b>	<b>5</b>
2.1 Serie storiche gerarchiche . . . . .	5
2.2 Serie storiche raggruppate . . . . .	7
<b>3 Metodi tradizionali di previsione per serie gerarchiche</b>	<b>9</b>
3.1 Approccio bottom-up . . . . .	9
3.2 Approccio top-down . . . . .	12
3.3 Approccio middle-out . . . . .	15
<b>4 Riconciliazione ottimale per serie storiche gerarchiche e raggruppate</b>	<b>17</b>
4.1 Approccio a combinazione ottimale . . . . .	19
4.2 Riconciliazione ottimale a traccia minima . . . . .	23
4.3 Riconciliazione ottimale non negativa . . . . .	29
<b>5 Previsioni sul flusso turistico australiano</b>	<b>33</b>
5.1 Configurazione dell'esperimento . . . . .	35
5.2 Risultati . . . . .	37
5.3 Conclusioni . . . . .	40
<b>6 Appendice</b>	<b>43</b>
6.1 Teorema A.1 . . . . .	43
6.2 Algoritmo rolling-origin per la valutazione dei metodi di riconciliazione . . . . .	45
6.3 Ripartizione geografica dell'Australia . . . . .	53
<b>Bibliografia</b>	<b>55</b>



## Elenco delle figure

2.1	Diagramma di gerarchia a due livelli . . . . .	5
2.2	Rappresentazione alternativa di una struttura raggruppata a due livelli . . . . .	7
5.1	Flusso turistico australiano misurato in milioni, per alcune delle zone geografiche selezionate. . . . .	34
5.2	Accuratezza dei metodi per la riconciliazione delle previsioni per metodo utilizzato e per livello di aggregazione. . . . .	39



## Elenco delle tabelle

5.1	Numero di serie per livello nel dataset del turismo australiano. . . . .	33
5.2	Performance dei vari metodi di riconciliazione delle previsioni per il dataset "TourismData v4" dopo aver applicato un approccio di valutazione a finestra scorrevole (media di 36 valutazioni di previsioni con orizzonte di previsione uguale a 12). Le valutazioni vengono misurate in termini di MASE e AMSE, stimati sia per livello della gerarchia sia media complessiva.	37
6.1	Wickamasurya et al. (2019) [5]. Le righe non numerate sono escluse dal dataset. . . . .	53



# 1. Introduzione

Nei campi economici e finanziari capita spesso di incontrare serie storiche che possono essere naturalmente disaggregate secondo una gerarchia, usando alcune tra le caratteristiche interne alle serie stesse, ad esempio in base alla tipologia di prodotto venduto o in base alla ripartizione geografica. Esemplicativo di questa struttura di aggregazione dei dati è il flusso turistico australiano, una risorsa di vitale importanza per la nazione. Considerando solo questo settore e il contributo che ha portato nel 2016, con \$55.3 miliardi di dollari, è pari al 3.2% sul PIL totale (Kourentzes e Athanasopoulos, 2019 [3]). Ovviamente sostenere un turismo dinamico ma rispettoso dei diversi paesaggi e biomi risulta essere un esercizio molto costoso. Per esempio, nel periodo 2016-2017, ci furono 204 progetti ideati dal governo australiano specificamente per il settore turistico, dal valore stimato di \$37.8 miliardi. Per poter prendere decisioni che possano portare a maggiori benefici economici servono precisione e accuratezza nei campi decisionali, come per la suddivisione del denaro a disposizione. Per ottenere i risultati sperati ad inizio programmazione, è superfluo dire che le decisioni devono essere prese in base a delle previsioni calcolate sui dati in possesso del governo australiano. Così come per il resto del mondo, il traffico turistico australiano può essere disaggregato geograficamente, rendendo di fatto le serie storiche relative ai dati turistici delle serie storiche gerarchiche. Il livello più aggregato, riferito come livello 0, rappresenta il totale di tutto il flusso turistico del Paese. Il livello 1 comprende le osservazioni disaggregate in 7 stati, che sono successivamente suddivise in 21 zone al livello 2 e, al livello 3, in 76 regioni. In totale, quindi, i dati sulla domanda turistica in Australia, disposti in maniera gerarchica e disaggregati geograficamente, sono formati da 105 serie storiche, ognuna con caratteristiche uniche. Con strutture di questo tipo, in cui ogni serie storica rappresenta un ambito geografico differente, ogni componente del totale è diversa da un'altra in termini di natura, di scala e di caratteristiche proprie della singola serie storica. Generare delle previsioni accurate per ogni componente della gerarchia e fare in modo che queste rispettino i vincoli imposti, ad esempio facendo in modo che le previsioni siano tutte maggiori di zero, è la chiave di una buona programmazione su tutti i livelli di aggregazione.

Quando si modellano queste strutture di dati, una proprietà molto importante è quella che definisce la somma delle serie storiche di livello più basso uguale alla somma delle serie storiche di livello più alto, che viene definita come coerenza, tale caratteristica è sempre vera quando si valutano le serie storiche gerarchiche dei dati osservati. Tuttavia, lo stesso non si può dire quando vengono prese in considerazione le previsioni generate utilizzando le strutture di dati gerarchiche. Storicamente i principali metodi per generare previsioni coerenti per questa tipologia di serie storiche possono essere riassunti in tre diverse categorie distinte: *bottom-up*, *top-down* e *middle-out*. Nel metodo *bottom-up* vengono generate

le previsioni per le serie al livello inferiore della gerarchia, cioè per le serie più disaggregate, e successivamente vengono aggregate per poter ottenere le previsioni per i livelli più alti della struttura. Nel metodo *top-down* il procedimento è inverso, le previsioni vengono generate per la serie “totale”, quella con il livello massimo di aggregazione, e vengono poi disaggregate scendendo lungo la gerarchia. In questo caso ci sono diversi modi per disaggregare le serie, il più comune è quello di calcolare delle proporzioni di disaggregazione, dei pesi, basandosi sui dati. Se, per esempio la variabile A è mediamente composta dal 65% del Totale, allora si moltiplicherà la previsione generata per il Totale per 0.65. Il problema è che le proporzioni calcolate non tengono conto di come queste potrebbero variare nel tempo, per cui i metodi basati su un approccio del tipo top-down tendono a produrre risultati meno accurati per le serie ai livelli più bassi. Per risolvere questo problema, sono stati descritti dei metodi per poter generare delle previsioni per le proporzioni (Athanasopoulos, Ahmed e Hyndman, 2009 [12]). Il terzo approccio tradizionale è l’approccio *middle-out*. Il primo passo consiste nello scegliere un livello intermedio della gerarchia, e generare le previsioni per le serie che compongono quel livello in maniera indipendente. Una volta fatto ciò si aggregano le previsioni per trovare le previsioni per i livelli superiori, mentre per trovare le previsioni scendendo nella gerarchia si utilizzano gli stessi principi dei metodi top-down. Nel corso degli anni si sono svolti diversi studi per poter valutare quale di questi metodi fosse il migliore per poter generare previsioni per questo tipo di serie storiche, tuttavia, anche se gli studi hanno portato a dire che i metodi bottom-up sono migliori solo in alcune circostanze (Shlifer e Wolff, 1979 [16]) o se si ha a che fare con dati mancati e non affidabili (Schwarzkopf, Teresine e Morris, 1988 [15]), queste condizioni non sono sufficienti per poter concludere che un metodo sia migliore di un altro. Questi metodi, infatti, soffrono di diversi tipi di inconvenienti: il primo è che non tengono conto della struttura di correlazione intrinseca nelle serie che compongono la gerarchia, il secondo è che non è facile calcolare intervalli di confidenza per le previsioni e il terzo è che anche se generassimo in maniera indipendente le previsioni per ogni livello della gerarchia, queste non sarebbero coerenti tra di loro, in particolare se aggregassimo le previsioni partendo dai livelli più bassi, queste non sarebbero uguali alle previsioni generate per i livelli più alti. L’ultimo difetto, in particolare, è il punto critico dei metodi, perché non avere una coerenza nelle previsioni si traduce nell’impossibilità di poter prendere delle decisioni in maniera corretta quando si parla di previsioni, indipendentemente dall’orizzonte temporale. Per risolvere tutte queste problematiche è stato ideato un metodo per generare previsioni per le serie storiche gerarchiche che: (1) fornisca previsioni puntuali da poter riconciliare in modo coerente attraverso le gerarchie, (2) permetta di poter specificare e valutare la correlazione e l’interazione tra le serie ad ogni livello della gerarchia, (3) fornisca le stime dell’incertezza delle previsioni riconciliate attraverso i livelli della gerarchia e (4) che sia sufficientemente flessibile da permettere aggiustamenti ad hoc, in base ad informazioni che si posseggono sulle singole serie che formano la gerarchia.

L'idea, chiamata riconciliazione ottimale delle previsioni (Hyndman, Ahmed Athanopoulos e Shang, 2011 [11]), funziona nel modo seguente. Per prima cosa si generano, in maniera indipendente, le previsioni per ogni componente della gerarchia. Queste previsioni verranno successivamente definite con il termine 'di base'. A questo punto le previsioni di base vengono riconciliate in modo tale da risultare coerenti ai vincoli e tengano conto delle informazioni fornite dalla matrice di varianza e covarianza dell'errore di previsione.

Generando in maniera indipendente delle previsioni per ogni serie della gerarchia si può tener conto delle diverse caratteristiche di ogni componente di ciascun livello, permettendo di acquisire informazioni sulle correlazioni ed effettuare aggiustamenti ad hoc sulle previsioni, (Athanopoulos, Hyndman, Kourentzes e Petropoulos, 2017 [14]). Inizialmente il metodo ideato da Hyndman et al. (2011) [11] utilizzava il metodo dei minimi quadrati generalizzati per calcolare uno stimatore a varianza minima, dimostrando come il metodo da loro proposto era migliore quando confrontato con i metodi più tradizionali. Nel corso degli anni vari approcci sono stati implementati, ad esempio Hyndman, Lee e Wang (2016) [1] suggerirono di usare i minimi quadrati pesati, tenendo conto degli elementi sulla diagonale della matrice della varianza covarianza, tralasciando gli elementi al di fuori di essa, tuttavia non fornirono una giustificazione teorica della scelta fatta.

Wickramasuriya, Athanopoulos e Hyndman (2019) [5], proposero un metodo chiamato a traccia minima, che non solo fornì una giustificazione teorica al precedente lavoro di Hyndman et al (2016) [1], ma anche un approccio per il calcolo di un insieme di previsioni non distorte e a varianza minima, per tutti gli orizzonti di previsione e lungo tutta la gerarchia. In questo lavoro, oltre a presentare gli aspetti teorici del metodo proposto per la riconciliazione delle previsioni per serie storiche gerarchiche, si applica la teoria sui dati reali sulla domanda turistica australiana, fornendo previsioni della serie totale secondo la ripartizione geografica, rispettando la proprietà di coerenza.

Il lavoro è organizzato nel modo seguente.

Nel capitolo 1 si introdurranno i concetti di serie storica gerarchica e raggruppata, le strutture su cui è possibile applicare gran parte dei metodi di previsione che verranno discussi in seguito, e verrà scelta la notazione usata poi in tutto il lavoro. Nel capitolo 2 si discuteranno i metodi tradizionali di previsione, *bottom-up*, *top-down* e *middle-out* cercando di mettere in luce le idee che hanno portato alla loro origine erano ampiamente utilizzati, prima del lavoro di Hyndman et al. (2011) [11]. Nel capitolo 4 verranno riassunti i concetti chiave che hanno portato l'idea della riconciliazione ottimale delle previsioni ad essere il metodo più competitivo nel campo dei problemi di previsione per serie storiche. Infine, nel capitolo 5, utilizzando i dati sul flusso turistico australiano si confronteranno i vari metodi di previsione con l'obiettivo di valutarne le prestazioni.



## 2. Serie storiche gerarchiche e raggruppate

In questo capitolo si introduce il concetto di serie storica gerarchica e raggruppata, insieme alle loro caratteristiche principali. Questo tipo di strutture sono ormai quasi onnipresenti nei campi economici e finanziari, diventa quindi una necessità per chiunque voglia avvicinarsi a questo mondo conoscere le loro caratteristiche e come sfruttarle a proprio vantaggio.

Inoltre, la notazione definita in questa prima parte verrà mantenuta con lo stesso significato per tutto il resto del lavoro.

### 2.1 Serie storiche gerarchiche

Moltissime serie storiche possono essere disaggregate naturalmente secondo diversi attributi di interesse, se prendiamo ad esempio il totale delle vendite di un negozio di biciclette, possiamo decidere di disaggregare la struttura per tipologia di biciclette vendute. In questo modo i dati sulle vendite totali saranno distribuiti in bici da corsa, mountain bike, bici da bambini e ibride. Ognuna di queste categorie potrebbe essere a sua volta divisa in sottocategorie più specifiche, ad esempio le bici ibride potrebbero essere divise in city bike e bici da cicloturismo. Utilizzando questo tipo di ragionamento per disaggregare i dati in sottocategorie, viene creata una tipologia specifica di serie storiche, le serie storiche gerarchiche. La Figura 2.1 mostra un esempio di gerarchia a  $K = 2$  livelli. Il livello più alto della struttura (chiamato livello 0 o *top-level*) è il "Totale", il livello dei dati più aggregato, rappresentato con  $y_t$ . Scendendo possiamo notare come il Totale si divida in due serie al livello 1, e successivamente ognuna di queste due serie si divide a sua volta in, rispettivamente, tre e due serie nel livello più basso della gerarchia, chiamato *bottom-level*. Sotto al livello 0, utilizziamo  $y_{j,t}$  per denotare la  $t$ -esima osservazione delle serie corrispondenti al nodo  $j$ .  $y_{A,t}$  denota perciò la  $t$ -esima osservazione della serie storica corrispondente al nodo A al livello 1, e così via.

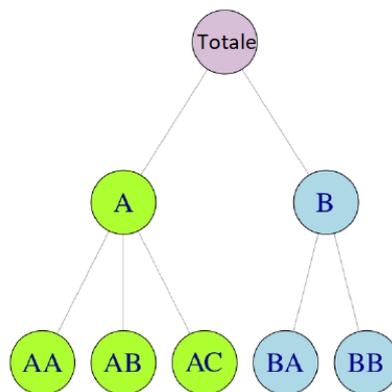


FIGURA 2.1: Diagramma di gerarchia a due livelli

In questo piccolo esempio, il numero totale di serie che compongono la struttura è  $n = 1 + 2 + 5 = 8$ , mentre il numero di serie al livello più basso è  $m = 5$ . Per ogni istante di tempo  $t$ , le osservazioni ai livelli superiori della gerarchia sono date dalla somma delle osservazioni delle serie di livello più basso. In questo caso vale che

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t} \quad (2.1)$$

Inoltre, possiamo anche scrivere

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \quad \text{e} \quad y_{B,t} = y_{BA,t} + y_{BB,t} \quad (2.2)$$

Sostituendo (2.2) dentro (2.1), otteniamo inoltre che  $y_t = y_{A,t} + y_{B,t}$ . Queste espressioni possono essere viste come dei vincoli di aggregazione, e possono essere scritte in maniera più efficiente usando la rappresentazione matriciale. Considerando una matrice  $n \times m$ , chiamata "summing matrix", o matrice di somma,  $\mathbf{S}$ , che descrive il modo in cui i livelli inferiori sono aggregati per formare i livelli superiori. Per la struttura in Figura 2.1, possiamo scrivere che

$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix}$$

Che, in forma compatta, diventa

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t \quad (2.3)$$

dove  $\mathbf{y}_t$  è un vettore  $n$ -dimensionale contenente tutte le osservazioni nella struttura gerarchica al tempo  $t$ ,  $\mathbf{S}$  è la matrice di somma, e  $\mathbf{b}_t$  è un vettore  $m$ -dimensionale con tutte le osservazioni delle serie al livello più basso, o *bottom-level*, della gerarchia. Notiamo come la prima riga della matrice di somma rappresenta l'equazione (2.1) mentre la seconda e la terza riga rappresentano l'equazione (2.2). Le ultime  $m$  righe della summing matrix possono essere viste come una matrice identità  $m$ -dimensionale,  $\mathbf{I}_m$ , per fare in modo che ogni osservazione del livello più basso sia uguale solo a se stessa.

## 2.2 Serie storiche raggruppate

Prendendo lo stesso esempio del negozio di biciclette, se quest'ultimo avesse delle filiali sparse nel mondo, si potrebbe pensare di disaggregare la struttura dei dati per Paese, successivamente per regione ed infine per città. In questo caso il negozio di bici potrebbe decidere di disaggregare le proprie attività sia per tipologia di bici sia per locazione. A questo punto non solo la struttura non avrebbe più un modo univoco per essere disaggregata, ma questa risulterebbe più complicata, perché i fattori per cui si applica la disaggregazione sono sia nidificati che incrociati tra di loro. In questo caso si ha di fronte un nuovo tipo di struttura, chiamata serie storiche gerarchiche raggruppate, o *grouped time series*. La Figura 2.2 mostra una struttura raggruppata a  $K = 2$  livelli. Come per le serie gerarchiche, al livello 0 troviamo il Totale, il livello di aggregazione più alto dei dati. In questo caso il Totale può essere disaggregato usando gli attributi (A, B) formando le serie  $y_{A,t}$  e  $y_{B,t}$ , oppure usando gli attributi (X, Y) formando così le serie  $y_{X,t}$  e  $y_{Y,t}$ . Al livello successivo, i dati sono disaggregati per entrambi gli attributi. Le serie storiche raggruppate possono essere pensate come serie storiche gerarchiche che non impongono una struttura univoca, nel senso che l'ordine con il quale ogni serie può essere disaggregata (o aggregata) non è univoco.

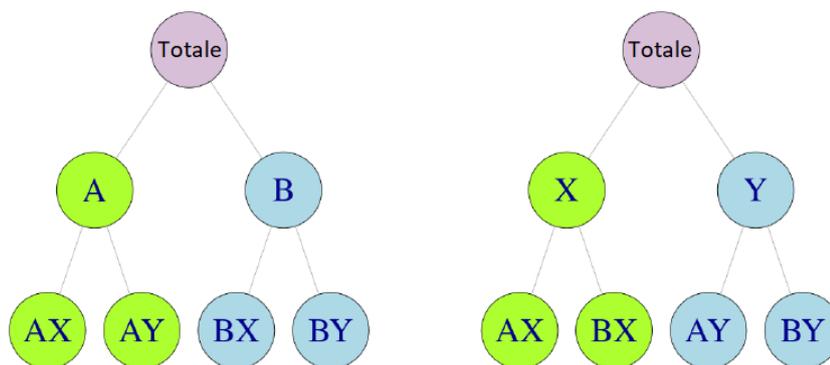


FIGURA 2.2: Rappresentazione alternativa di una struttura raggruppata a due livelli

Questo esempio mostra che ci sono diversi modi di aggregare questa tipologia di strutture di dati. Anche in questo caso vale che

$$y_t = y_{AX,t} + y_{AY,t} + y_{BX,t} + y_{BY,t}$$

Tuttavia, quando andiamo a valutare il primo livello della struttura otteniamo che

$$y_{A,t} = y_{AX,t} + y_{AY,t} \quad \text{e} \quad y_{B,t} = y_{BX,t} + y_{BY,t} \quad (2.4)$$

Ma vale anche

$$y_{X,t} = y_{AX,t} + y_{BX,t} \quad \text{e} \quad y_{Y,t} = y_{AY,t} + y_{BY,t} \quad (2.5)$$

Queste uguaglianze possono essere rappresentate ancora una volta dalla matrice di somma **S**. In questo caso il numero totale di serie è  $n = 9$ , con  $m = 4$  serie al livello più basso. Per rappresentare la struttura in Figura 2.2 usando le matrici si può scrivere

$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}$$

cioè

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

Le equazioni (2.4) e (2.5) sono rappresentate rispettivamente dalla seconda e terza riga della matrice **S**, e dalla quarta e quinta riga.

### 3. Metodi tradizionali di previsione per serie gerarchiche

Tutte le serie storiche, sia gerarchiche che raggruppate, presentano delle proprietà che le caratterizzano e le differenziano le une dalle altre. Una di queste proprietà, in particolare, è la coerenza. La coerenza dei dati si ottiene quando la serie al livello 0 è uguale alla somma delle serie al livello più basso. Questa proprietà nasce dalle strutture utilizzate e dai vincoli di aggregazione, e non esistono serie storiche, gerarchiche o raggruppate, dove questa proprietà non sia rispettata. Tuttavia, se noi generassimo per ogni serie storica delle previsioni per un istante di tempo di cui non conosciamo il vero valore  $t + h$ , allora la somma delle previsioni al livello più basso non sarebbe uguale alla serie al livello più alto. Questo è un punto critico, perché avere dei risultati differenti a seconda dei livelli della gerarchia potrebbe portare a scelte inconsistenti. Fortunatamente esistono diversi metodi per generare delle previsioni che rispettino la naturale proprietà di coerenza, che caratterizza i dati nelle serie storiche gerarchiche. I metodi tradizionali, che si dividono in tre categorie a seconda dell'approccio utilizzato: *top-down*, *bottom-up* e *middle-out*. Questi metodi vengono detti a combinazione nel senso che le previsioni vengono generate indipendentemente solo per un livello della gerarchia, che può essere il livello 0, il livello più basso o un livello intermedio, e successivamente vengono aggregate o disaggregate in modo da ottenere le previsioni per gli altri livelli della gerarchia. Per tutte le formule e gli esempi applicativi verrà considerata la serie storica gerarchica a due livelli proposto in Figura 2.1.

#### 3.1 Approccio bottom-up

Il primo metodo che viene illustrato per generare in maniera semplice delle previsioni, che rispettino la proprietà di coerenza è l'approccio *bottom-up*. Questo approccio consiste nel generare le previsioni per ogni serie al livello più basso della struttura e, successivamente, sommarle seguendo la struttura gerarchica in modo da ottenere le previsioni per le altre serie che compongono la gerarchia. Se prendiamo la struttura in Figura 2.1, la prima cosa da fare è generare le previsioni per l'orizzonte di previsione  $h$  per ognuna delle serie al *bottom level* della struttura. Chiameremo tale valore  $\hat{y}_{j,h}$ , utilizzando tale notazione per le previsioni generate in maniera indipendente per il livello  $j$ , al relativo tempo  $T + h$  e calcolato usando le informazioni disponibili fino al tempo  $T$ . L'insieme delle previsioni indipendenti per il *bottom-level* della gerarchia è

$$\hat{y}_{AA,h}, \hat{y}_{AB,h}, \hat{y}_{AC,h}, \hat{y}_{BA,h}, \hat{y}_{BB,h}$$

Partendo da queste previsioni, si generano le altre sommando lungo la struttura della gerarchia, in modo da ottenere previsioni coerenti:

$$\begin{aligned}\tilde{y}_h &= \hat{y}_{AB,h} + \hat{y}_{AC,h} + \hat{y}_{BA,h} + \hat{y}_{BB,h} \\ \tilde{y}_{A,h} &= \hat{y}_{AB,h} + \hat{y}_{AC,h} \\ \tilde{y}_{B,h} &= \hat{y}_{BA,h} + \hat{y}_{BB,h}\end{aligned}\tag{3.1}$$

si noti che abbiamo usato  $\tilde{y}_h$  per indicare una previsione che rispetta la proprietà di coerenza.

L'equazione (3.1) si può scrivere in forma matriciale, utilizzando la matrice di somma

$$\begin{pmatrix} \tilde{y}_t \\ y_{\tilde{A},t} \\ y_{\tilde{B},t} \\ y_{\tilde{AA},t} \\ y_{\tilde{AB},t} \\ y_{\tilde{AC},t} \\ y_{\tilde{BA},t} \\ y_{\tilde{BB},t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{\hat{AA},t} \\ y_{\hat{AB},t} \\ y_{\hat{AC},t} \\ y_{\hat{BA},t} \\ y_{\hat{BB},t} \end{pmatrix}$$

che, in forma compatta diventa

$$\tilde{\mathbf{y}}_h = \mathbf{S}\hat{\mathbf{b}}_h,$$

dove  $\tilde{\mathbf{y}}_h$  è un vettore  $n$ -dimensionale di previsioni coerenti ottenute per l'orizzonte di previsione  $T + h$ , mentre  $\hat{\mathbf{b}}_h$  è un vettore  $m$ -dimensionale (dove  $m$  è il numero di serie storiche al livello più basso della gerarchia) contenente i valori delle previsioni generate indipendentemente per le serie più disaggregate della gerarchia. Prendendo in considerazione lo stesso esempio, è utile specificare un'ulteriore notazione per questo tipo di approccio alla previsione di serie storiche gerarchiche. Supponiamo di aver calcolato le previsioni in maniera indipendente per ogni serie ad ogni livello della struttura, denoteremo il vettore che contiene l'insieme delle previsioni indipendenti con  $\hat{\mathbf{y}}_h$ . Il vettore delle previsioni, ottenute usando un approccio *bottom-up* può essere calcolato tramite

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_h\tag{3.2}$$

dove  $\mathbf{S}$  è la *summing matrix* che descrive la struttura della gerarchia,  $\hat{\mathbf{y}}_h$  è il vettore delle previsioni generate in maniera indipendente per ogni serie in ogni livello della gerarchia e  $\mathbf{P}$  è una matrice definita come

$$\mathbf{P} = [\mathbf{0}_{m \times (n-m)} | \mathbf{I}_{m \times m}]\tag{3.3}$$

dove  $\mathbf{0}_{m \times (n-m)}$  è una matrice nulla con  $m$  righe e  $n - m$  colonne e  $\mathbf{I}_{m \times m}$  è la matrice identità di ordine  $m$ . In questo caso  $\mathbf{P}$  estrae solamente le previsioni per il livello più basso della struttura e, grazie ad  $\mathbf{S}$ , queste vengono sommate in modo tale da avere previsioni coerenti. Un sicuro vantaggio di questo metodo è che si generano previsioni per il livello più basso della struttura, per cui nessuna informazione verrà persa per "colpa" dell'aggregazione. Studi empirici dimostrano infatti come questa metodologia abbia un'efficacia maggiore rispetto ad altri metodi più tradizionali. Dangerfield e Morris (1992) [8] costruirono una gerarchia artificiale a due livelli usando i dati della M-competition con due serie al livello inferiore e dimostrarono che le previsioni ottenute con i metodi *bottom-up* erano più accurate, soprattutto quando le serie sono molto correlate tra loro. Tuttavia questa tipologia di approccio soffre di un difetto fondamentale. Le serie storiche che compongono i livelli più disaggregati della gerarchia contengono sì molte informazioni su trend e stagionalità, ma sono tipicamente più variabili e influenzate da errori casuali, il che ne rende molto difficile l'interpretazione e la modellazione. Per questo motivo, prima che fosse sviluppato un metodo ottimale, anche se molti studi provavano a sottolineare la maggiore efficienza degli approcci *bottom-up* per il calcolo di previsioni per serie gerarchiche, non è mai stata riconosciuta in maniera univoca la sua superiorità sugli altri metodi tradizionali di previsione.

### 3.2 Approccio top-down

Un secondo insieme di metodi per la generazione di previsioni coerenti è l'approccio *top-down*. La particolarità di questo approccio è che può essere applicato sulle serie storiche gerarchiche, e non sulle serie raggruppate. L'idea su cui questa metodologia si applica è opposta rispetto a quella che ispira l'approccio *bottom-up*. In questo caso, infatti, il primo passo è quello di generare le previsioni in maniera indipendente solamente per la serie del Totale, la serie al livello più alto di aggregazione, e successivamente disaggregare i risultati scendendo nella gerarchia. Poniamo  $p_1, \dots, p_m$ , con  $m$  il numero totale di serie storiche presenti al *bottom level* della struttura, come un insieme di proporzioni di disaggregazione, cioè un insieme di proporzioni che indichi come le previsioni del Totale debbano essere distribuite in modo da ottenere le previsioni per i livelli inferiori. Tenendo l'esempio definito in Figura 2.1 otteniamo che

$$\tilde{y}_{AA,h} = p_1 \hat{y}_h, \quad \tilde{y}_{AB,h} = p_2 \hat{y}_h, \quad \tilde{y}_{AC,h} = p_3 \hat{y}_h, \quad \tilde{y}_{BA,h} = p_4 \hat{y}_h, \quad \tilde{y}_{BB,h} = p_5 \hat{y}_h$$

usando la notazione matriciale e inserendo le proporzioni in un vettore  $m$ -dimensionale  $\mathbf{p} = (p_1, \dots, p_m)'$ , possiamo scrivere

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{P}\hat{y}_h$$

con

$$\mathbf{P} = [\mathbf{p} | \mathbf{0}_{m \times (n-1)}], \quad (3.4)$$

In questo modo si ottengono le previsioni  $h$ -passi-avanti, che vengono quindi sommate con la matrice  $\mathbf{S}$  per generare il resto delle previsioni nella gerarchia. I metodi più comuni per la previsione in serie storiche gerarchiche che implementano un approccio *top-down* specificano le proporzioni di disaggregazione sulla base della proporzione storica dei dati, che garantiscono buone performance, come risulta da Gross, Sohl e Jeffrey (1990) [7]. Il primo modo per valutare le proporzioni di disaggregazione è quello di porle uguali alla media delle proporzioni storiche, in questo caso ogni peso  $p_j$  riflette la media della proporzione della serie al livello più basso della gerarchia  $y_{j,t}$  rispetto alla serie del Totale  $y_t$ , su tutto il periodo  $t = 1, \dots, T$ .

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t} \quad (3.5)$$

Il secondo modo per calcolare le proporzioni storiche è calcolare le proporzioni delle medie delle serie storiche al *bottom-level* e usarle come proporzioni storiche. In questo caso ogni proporzione  $p_j$  cattura il valore medio delle serie di livello inferiore  $y_{j,t}$  rispetto al valore medio dell'aggregato totale  $y_t$ .

$$p_j = \frac{\sum_{t=1}^T \frac{y_{j,t}}{T}}{\sum_{t=1}^T \frac{y_t}{T}} \quad (3.6)$$

Sicuramente l'aspetto più vantaggioso della metodologia *top-down* è la sua semplicità. Tutto quello di cui si ha bisogno per poter modellare i dati, seguendo questa logica, è la previsione per le serie più aggregate. Tuttavia, questi metodi presentano degli svantaggi. Il primo degli svantaggi è il fatto che si perde dell'informazione: visto che utilizziamo solamente i dati aggregati, non si è in grado di cogliere alcune dinamiche delle singole serie come eventi speciali. Questo fenomeno è reso possibile anche perché le proporzioni storiche usate per disaggregare le previsioni non tengono conto del fatto che queste potrebbero cambiare nel corso del tempo e questo produce previsioni meno accurate ai livelli inferiori rispetto alla tecnica *bottom-up*. Per ovviare alla natura statica delle proporzioni storiche usate per disaggregare il livello più alto della gerarchia, si introdusse un metodo in grado di produrre proporzioni per disaggregare le previsioni basate sulle proporzioni delle previsioni delle serie storiche al livello più disaggregato della gerarchia. Consideriamo sempre l'esempio in Figura 2.1, l'intuizione dietro al metodo delle proporzioni di previsione consiste nel generare prima di tutto le previsioni per ogni componente della gerarchia, dopodiché calcolare la proporzione storica come il rapporto tra la previsione indipendente  $h$  passi avanti al bottom-level e la somma di tutte le previsioni generate in maniera indipendente per quel livello. Indicando con  $\hat{y}_{j,h}^{(l)}$  la previsione base corrispondente al nodo  $l$  livelli sopra a  $j$  e, usando lo stesso principio, si definisce  $\hat{S}_{j,h}^{(l)}$  la somma delle previsioni generate base con orizzonte di previsione  $T + h$  dei nodi direttamente inferiori al nodo che si trova  $l$  livelli sopra a  $j$ . Con l'esempio usato abbiamo che

$$\begin{aligned} \hat{y}_{A,h}^{(1)} &= \hat{y}_{B,h}^{(1)} = \hat{y}_h \\ \hat{y}_{AA,h}^{(1)} &= \hat{y}_{AB,h}^{(1)} = \hat{y}_{AC,h}^{(1)} = \hat{y}_{A,h} \\ \hat{y}_{AA,h}^{(2)} &= \hat{y}_{AB,h}^{(2)} = \hat{y}_{AC,h}^{(2)} = \hat{y}_h \\ \hat{S}_{AA,h}^{(1)} &= \hat{S}_{AB,h}^{(1)} = \hat{S}_{AC,h}^{(1)} = \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h} \\ \hat{S}_{AA,h}^{(2)} &= \hat{S}_{AB,h}^{(2)} = \hat{S}_{AC,h}^{(2)} = \hat{S}_{A,h}^{(1)} = \hat{S}_{B,h}^{(1)} = \hat{y}_{A,h} + \hat{y}_{B,h} \end{aligned} \quad (3.7)$$

Possiamo quindi scrivere che le previsioni coerenti per le serie al *bottom-level* della gerarchia sono date da

$$\tilde{y}_{AA,h} = \left( \frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \tilde{y}_{A,h} = \left( \frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \tilde{y}_h$$

Conseguentemente si ha che la proporzione per il primo elemento del livello più disaggregato della struttura è

$$p_1 = \left( \frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \left( \frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \quad (3.8)$$

Nonostante il metodo delle proporzioni di previsione riesca a limitare la staticità dei metodi che utilizzano proporzioni storiche, lo svantaggio che si vuole sottolineare è il fatto che i metodi *top-down* forniscono sempre delle previsioni distorte e che di fatto con questa tecnica non è garantita la positività delle previsioni, caratteristica che in ambiti economici e finanziari non solo è ragionevole, è anche essenziale per dare senso ai valori ottenuti. Per dimostrare il primo dei due lati negativi di questa tecnica assumiamo di aver generato in maniera indipendente delle previsioni non distorte per ogni serie e che siano raccolte in  $\hat{\mathbf{y}}_h$  (cioè che  $E[\hat{\mathbf{y}}_h] = E[\mathbf{y}_h]$ ), e che è ragionevole volere delle previsioni coerenti che siano anche non distorte a loro volta, stiamo quindi richiedendo che  $E[\tilde{\mathbf{y}}_h] = E[\mathbf{y}_h] = \mathbf{S}E[\mathbf{b}_h]$ , con  $\mathbf{b}_h$  i veri valori delle serie a livello più disaggregato della disaggregazione. Supponiamo che  $\beta_h = E[\mathbf{b}_h]$  il valore atteso delle vere osservazioni all'istante temporale  $T + h$ . Allora  $E[\tilde{\mathbf{y}}_h] = \mathbf{S}PE[\hat{\mathbf{y}}_h] = \mathbf{SPS}\beta_h$ . La condizione di non distorsione sarà mantenuta fintanto che

$$\mathbf{SPS} = \mathbf{S} \quad (3.9)$$

Questa condizione è vera per i metodi *bottom-up*, con  $\mathbf{P}$  ottenuta come descritto con (3.3). Tuttavia, usando un approccio *top-down* con la matrice  $\mathbf{P}$  data da (3.4) possiamo dire che  $\mathbf{SPS} \neq \mathbf{S}$  per qualsiasi  $\mathbf{P}$  fissato. Per questo motivo, qualsiasi siano le stime indipendenti, i metodi che seguono un approccio *top-down* non riescono a fornire delle previsioni non distorte (Hyndman et al., 2011 [11]).

### 3.3 Approccio *middle-out*

L'approccio *middle-out* combina gli approcci *top-down* e *bottom-up* visti in precedenza. Il primo passo è scegliere un "livello intermedio" da cui generare le previsioni in maniera indipendente, per ogni serie a quel livello. Per le serie al di sopra del livello intermedio selezionato, si seguono le regole descritte per l'approccio *bottom-up* e si somma secondo la struttura gerarchica per ottenere delle previsioni coerenti.

Per quanto riguarda le previsioni per le serie ai livelli al di sotto del livello intermedio, le previsioni coerenti sono generate seguendo un approccio *top-down* disaggregando le previsioni generate dal livello intermedio verso il basso. Anche in questo caso, le previsioni coerenti che si ottengono possono essere scritte come  $\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{j,h}$ , dove  $\hat{\mathbf{y}}_{j,h}$  sono le previsioni base, generate in maniera indipendente per un livello intermedio  $j$ , la matrice invece  $\mathbf{P}$  è una combinazione delle matrici definite per gli approcci *top-down* e *bottom-up*. Usando un insieme di proporzioni definite a priori, le previsioni base verranno disaggregate per poter trovare le previsioni coerenti per i livelli al di sotto di quello intermedio, e poi, usando la matrice  $\mathbf{S}$ , verranno sommate man mano che si sale nella gerarchia.



## 4. Riconciliazione ottimale per serie storiche gerarchiche e raggruppate

Ognuno dei metodi tradizionali visti in precedenza presenta propri vantaggi e svantaggi, tuttavia tutti quanti i metodi falliscono nel tener conto di tutte le informazioni contenute nei dati nei vari livelli della struttura, utilizzando solo alcune di esse. I metodi *bottom-up*, ad esempio, utilizzano solamente le informazioni contenute nel *bottom-level*, che contiene serie storiche generalmente più variabili e influenzate da errori casuali, ma con importanti informazioni su trend e stagionalità. Gli approcci *top-down*, al contrario, utilizzano solamente le previsioni per il Totale, che sono generalmente più stabili, ma possiedono poche informazioni sulle serie ai livelli più bassi della gerarchia. In più, nessuno di questi metodi tiene conto delle correlazioni tra le serie in livelli diversi della struttura. Per anni si sono studiati i vari approcci per la previsione di serie storiche gerarchiche, e quale tra questi fosse il migliore, ma nessuno studio ha portato ad risultato significativo: l'efficienza di un metodo dipende molto dal tipo di struttura su cui si applica.

Nello studio di Hyndman et al. (2011) [11], viene presentato *"un nuovo approccio per la previsione per serie storiche gerarchiche, basato su previsioni indipendenti per tutte le serie su tutti i livelli e successivamente riconciliate"*.

Per poter presentare il metodo proposto inizialmente da Hyndam et al. nel 2011 e successivamente implementato ed aggiornato nel corso degli anni assumiamo di essere sempre nel caso riportato in Figura 2.1, una serie storica gerarchica a  $K = 2$  livelli, con  $n = 8$  numero complessivo di serie storiche nella struttura e  $m = 5$  serie storiche al livello più disaggregato, o bottom-level. Usando le notazioni definite fino ad adesso indichiamo con  $\mathbf{y}_t$  il vettore  $n$ -dimensionale contenente tutte le osservazioni al tempo  $t$ , mentre  $\mathbf{b}_t$  è il vettore  $m$ -dimensionale contenente le serie al bottom-level, per  $t = 1, \dots, T$ . Con l'esempio scelto abbiamo che  $\mathbf{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$ , mentre abbiamo che  $\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{AC,t}, y_{BA,t}, y_{BB,t}]'$ . Alternativamente possiamo scrivere, usando la notazione matriciale che

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t,$$

dove  $\mathbf{S}$  è la matrice di somma di ordine  $n \times m$  che indica come le serie al livello più basso della gerarchia vengono aggregate. In questo modo si vede in maniera molto più esplicita come la proprietà di coerenza sia sempre verificata quando si prendono in considerazione le osservazioni.

Sia  $\hat{\mathbf{y}}_h$  un vettore di previsioni  $h$  passi avanti generate in maniera indipendente per ogni serie nella struttura, i metodi tradizionali per ottenere previsioni coerenti possono essere scritti come

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_h$$

dove  $\mathbf{P}$  è opportunamente scelta. Il ruolo di  $\mathbf{P}$  è quello di mappare le previsioni indipendenti in modo da ottenere delle previsioni per il livello più disaggregato della struttura che rispettino la proprietà di coerenza, che possano essere sommate tramite  $\mathbf{S}$ .

## 4.1 Approccio a combinazione ottimale

Questo approccio per la generazione di previsioni coerenti è stato introdotto da Hyndman et al. (2011) [11] per poter ottenere delle previsioni che fossero migliori di quelle ottenute con i metodi tradizionali. Questo approccio combina in maniera ottimale le previsioni generate in maniera indipendente per produrre un nuovo insieme di previsioni riconciliate il più vicine possibile a quelle iniziali, che rispettino la proprietà di coerenza. A differenza degli altri approcci, questo (1) utilizza tutte le informazioni disponibili nella gerarchia; (2) permette correlazioni e interazioni tra le serie ad ogni livello della gerarchia; (3) lascia la possibilità di effettuare aggiustamenti ad hoc per le previsioni ad ogni livello e, se le previsioni generate in modo indipendente sono non distorte, (4) fornisce delle previsioni finali non distorte per ogni livello della gerarchia. Inoltre, questo approccio produce delle stime per l'incertezza delle previsioni che sono consistenti su tutti i livelli.

L'idea generale è quella di rappresentare le previsioni generate in maniera indipendente tramite un modello di regressione lineare, definiamo quindi

$$\hat{\mathbf{y}}_h = \mathbf{S}\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_h, \quad (4.1)$$

dove  $\boldsymbol{\beta}_h = E[\mathbf{b}_h | \mathbf{y}_1, \dots, \mathbf{y}_T]$ , il valore atteso ignoto delle osservazioni future delle serie al livello più disaggregato della gerarchia, il *bottom-level*.  $\boldsymbol{\varepsilon}_h$  è l'errore di aggregazione, cioè la differenza tra le previsioni originali e il loro valore atteso se le previsioni fossero riconciliate. Assumendo gli errori indipendenti dalle osservazioni, con media nulla e matrice di varianza-covarianza uguale a  $\text{var}(\boldsymbol{\varepsilon}_h | \mathbf{y}_1, \dots, \mathbf{y}_T) = \boldsymbol{\Sigma}_h$ . Se  $\boldsymbol{\Sigma}_h$  fosse nota, lo stimatore ottenuto per  $\boldsymbol{\beta}_h$  usando il metodo dei minimi quadrati generalizzati sarebbe non distorto e con varianza minima tra gli stimatori non distorti, ed in particolare la stima si otterrebbe con

$$\hat{\boldsymbol{\beta}}_h = (\mathbf{S}'\boldsymbol{\Sigma}_h^\dagger\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}_h^\dagger\hat{\mathbf{y}}_h \quad (4.2)$$

dove  $\boldsymbol{\Sigma}_h^\dagger$  è l'inversa generalizzata di Moore-Penrose di  $\boldsymbol{\Sigma}_h$ . È necessario usare l'inversa generalizzata perché  $\boldsymbol{\Sigma}_h$  è spesso (vicino) singolare, questo è dovuto all'aggregazione di  $\mathbf{y}_h$ . Questo risultato ci permetterebbe di avere delle stime riconciliate utilizzando

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}_h^\dagger\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}_h^\dagger\hat{\mathbf{y}}_h = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_h \quad (4.3)$$

La difficoltà di questo metodo è che richiede la conoscenza di  $\boldsymbol{\Sigma}_h$ , o almeno una buona stima, conoscenza che risulta molto onerosa da ottenere o addirittura impossibile al crescere della grandezza della gerarchia.

Hyndman et al. (2011) [11] cercarono un modo per identificare la matrice di varianza-covarianza, che allo stesso tempo limitasse la complessità computazionale.

Per fare ciò assunsero che gli errori di aggregazione, in (4.1), potessero essere definiti come  $\boldsymbol{\varepsilon} \approx \mathbf{S}\boldsymbol{\varepsilon}_{m,h}$ , dove  $\boldsymbol{\varepsilon}_{m,h}$  è il vettore  $m$ -dimensionale che contiene gli errori di previsione per le

serie al livello più basso della gerarchia. Così definiti, anche gli errori avrebbero soddisfatto gli stessi vincoli di aggregazione dei dati, un vincolo che secondo il lavoro pubblicato dovrebbe essere rispettato da ogni insieme ragionevole di previsioni.

Il problema con questo ragionamento è che, a causa dei vincoli di aggregazione, la matrice di varianza-covarianza  $\Sigma_h$  non è identificabile. Non esistendo un modo univoco per poter stimare la matrice utilizzando le osservazioni, l'unica strada che risulta percorribile per trovare una soluzione è quella dell'approssimazione, Wickramasuriya et al. (2019) [5]. Nel paragrafo successivo verrà descritto nel dettaglio l'errore commesso.

Nonostante il ragionamento proposto da Hyndman et al. (2011) [11] fu corretto definitivamente da Wickramasuriya et al. (2019) [5], nel lavoro proposto da Hyndman et al. (2016) [1] ci fu il primo tentativo di usare un'approssimazione per matrice  $\Sigma_h$  invece che una sua stima, che come abbiamo detto non è possibile calcolare in modo univoco.

Questo approccio alternativo prevede di utilizzare i minimi quadrati ordinari o i minimi quadrati pesati. Il vettore delle stime riconciliate è quindi ottenuto attraverso

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{y}}_h, \quad (4.4)$$

si ha, di conseguenza, che  $\mathbf{P} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ .

Hyndman et al. (2016) [11] mostrarono che, sotto l'assunto che  $\Sigma_h = \mathbf{S}\Omega_h\mathbf{S}'$  (cioè quando assumiamo che gli errori rispettano il vincolo di aggregazione dei dati), la soluzione ottenuta attraverso i minimi quadrati ordinari è uguale a quella ottenuta attraverso i minimi quadrati generalizzati, e che la combinazione ottimale delle previsioni di base è indipendente dai dati osservati. Questo è facilmente osservabile anche sulle gerarchie semplici, se si prende una gerarchia con un solo livello di disaggregazione, con  $K = 1$ , e con  $m_1$  nodi al livello 1, i pesi sono dati da

$$\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' = (m_1 + 1)^{-1} \begin{pmatrix} m_1 & 1 & 1 & \dots & \dots & 1 \\ 1 & m_1 & -1 & -1 & \dots & -1 \\ 1 & -1 & m_1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & -1 & \dots & -1 & m_1 & -1 \\ 1 & -1 & \dots & \dots & -1 & m_1 \end{pmatrix}$$

Allo stesso modo, si possono ottenere set di pesi per ogni particolare gerarchia, indipendentemente dai dati osservati. Tuttavia, sapendo che l'idea di avere errori che soddisfino il vincolo di aggregazione sarà poi dimostrata come errata, il risultato fondamentale che si ottenne attraverso questo modo alternativo di scrivere le stime riconciliate fu che venne evidenziato come la soluzione trovata attraverso i minimi quadrati ordinari era una soluzione consistente al problema di riconciliazione delle previsioni, Hyndman et al. (2016) [1].

Il metodo proposto, però, non mancava di aspetti critici e problemi dal punto di vista computazionale. La principale difficoltà che si incontra per poter implementare questo metodo

di previsione, dato (4.4), è che la matrice  $\mathbf{S}$  può essere molto grande e il calcolo dell'inversa di  $\mathbf{S}'\mathbf{S}$  potrebbe risultare proibitivo dal punto di vista computazionale. Hyndman et al. (2016) [1] propongono diverse soluzioni a questo problema.

La prima soluzione tiene conto del fatto che dei  $m \times n$  elementi che compongono la *summing matrix*  $\mathbf{S}$ , solamente  $m \times K$ , con  $K$  numero di livelli della gerarchia, sono non nulli. Siccome il numero di livelli della gerarchia è tendenzialmente molto minore del numero di serie di cui è composto il *bottom-level*, allora  $\mathbf{S}$  è quasi certamente una matrice sparsa. Si possono dunque implementare diversi algoritmi per poter memorizzare in maniera efficiente la matrice, in modo da salvare memoria, ridurre il tempo di computazione ed utilizzare dei metodi per risolvere l'equazione (4.7), ad esempio quello implementato da Koenker e Ng (2003) [10]. Questa soluzione si presta bene nel caso di matrici moderatamente grandi, ma nella pratica, quando si incontrano matrici ad elevata dimensionalità, si possono incontrare molte difficoltà.

La seconda soluzione introdotta fa riferimento al metodo di di Lanczos (1952) [19] originariamente sviluppato per risolvere un sistema lineare di equazioni e successivamente esteso a diversi scenari. Questo algoritmo iterativo viene adattato per risolvere un problema di minimi quadrati usando delle matrici sparse. La soluzione che fornisce è un'approssimazione della soluzione diretta, tuttavia nella maggior parte dei casi la differenza è trascurabile.

L'ultimo approccio proposto per risolvere il problema dell'invertibilità consiste nel riparametrizzare il modello di regressione definito in (4.1). Invece di usare un parametro per rappresentare ogni serie al livello più disaggregato della gerarchia si usa un parametro per ogni serie della gerarchia, imponendo dei vincoli di somma uguali a zero per evitare una sovrapparametrizzazione. Ogni parametro misurerà il contributo del livello considerato sulle serie al livello sottostante.

A questo punto il modello di regressione lineare diventa un modello ANOVA vero e proprio, i parametri possono essere stimati usando i minimi quadrati, evitando di dover invertire matrici. Questo approccio funziona per ogni tipo di gerarchia, ma gli stimatori che si calcolano risultano molto complicati se la gerarchia non è bilanciata.

Per questo motivo, nel lavoro proposto da a Hyndman et al. (2016) [1], viene consigliato di usare algoritmi per le matrici sparse se la gerarchia non è di dimensione troppo elevata, di usare il modello ANOVA se la gerarchia è bilanciata e di utilizzare l'approccio iterativo se la gerarchia considerata è non bilanciata e di elevata dimensionalità.

Non è difficile notare che il modello lineare formalizzato in (4.1) può essere utilizzato anche in situazioni differenti rispetto a serie storiche gerarchiche. Considerando la serie storica raggruppata in Figura 2.2, questi dati possono formare due diverse strutture gerarchiche differenti che dipendono da quale variabile si utilizza prima per disaggregare i dati. Ma anche in questa situazione si può scrivere lo stesso modello di regressione, modificando opportunamente le quantità all'interno, in questo caso particolare si otterrebbe

$$\mathbf{y}_t = \begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}, \mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ e } \mathbf{b}_t = \begin{pmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{pmatrix}$$

Ciascun gruppo aggregato d'interesse è incluso nel vettore  $\mathbf{y}_t$ , e la riga corrispondente di  $\mathbf{S}$  definisce come le serie nel *bottom-level* sono aggregate per formare quell'elemento (Hyndman et al. 2016 [1]). La struttura di raggruppamento può essere molto più complicata, più semplice, gerarchica o raggruppata, oppure un mix tra i due, in ogni caso si potrà rappresentare in maniera simile a quella proposta da Hyndman et al. (2011) [11], e quindi i risultati ottenuti saranno validi.

## 4.2 Riconciliazione ottimale a traccia minima

Come abbiamo visto, Hyndman et al. (2011) [11] proposero un approccio basato sul modello di regressione lineare, per generare previsioni riconciliate che fossero coerenti e sfruttassero una quantità maggiore di informazioni sulla gerarchia rispetto ai metodi tradizionali di previsione. Il loro metodo si basava su un assunto, che gli errori seguissero la struttura di aggregazione dei dati e prevedeva che le stime riconciliate potessero essere calcolate come

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\Sigma_h^\dagger\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^\dagger\hat{\mathbf{y}}_h \quad (4.5)$$

dove  $\Sigma_h^\dagger$  è l'inversa generalizzata di  $\Sigma_h$ , ottenendo  $\mathbf{P} = (\mathbf{S}'\Sigma_h^\dagger\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^\dagger$ .

Questo ragionamento venne corretto da Wickamasurya et al. (2019) [5], che dimostrarono l'impossibilità di stimare  $\Sigma_h$  in maniera univoca, sotto l'assunto di aggregazione degli errori.

Esplicitando in maniera estesa i residui del modello, cioè gli errori di coerenza, otteniamo infatti che

$$\tilde{\boldsymbol{\varepsilon}}_h = \hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h = (\mathbf{I}_n - \mathbf{S}\mathbf{P})\hat{\mathbf{y}}_h \quad (4.6)$$

dove  $\hat{\mathbf{y}}_h$  sono le stime base, ottenute in maniera indipendente per ogni nodo della gerarchia, e  $n$  è il numero totale di serie presenti nella gerarchia. Questo porta a definire la varianza degli errori come

$$\text{var}(\tilde{\boldsymbol{\varepsilon}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T) = (\mathbf{I}_n - \mathbf{S}\mathbf{P})\Sigma_h(\mathbf{I}_n - \mathbf{S}\mathbf{P})' \quad (4.7)$$

È proprio in questo passaggio che viene riscontrato un problema per poter identificare e invertire la matrice  $\Sigma_h$ . Per poter definire al meglio il perché l'identificazione di un'inversa della matrice di varianza-covarianza per le previsioni sia impossibile, serve prima effettuare un ragionamento sulle stime base utilizzate nel metodo a combinazione ottimale da Hyndman et al. (2011) [11].

Identificando con

$$\hat{\boldsymbol{\varepsilon}}_h = \mathbf{y}_h - \hat{\mathbf{y}}_h \quad (4.8)$$

per gli errori di previsione per le stime indipendenti, o *base forecast errors*, è di interesse assumere che il loro valore atteso sia pari a zero, esplicitando la non distorsione delle previsioni. Questo significa che  $E[\hat{\boldsymbol{\varepsilon}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = 0$ , e quindi che  $E[\hat{\mathbf{y}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = E[\mathbf{y}_h | \mathbf{y}_1, \dots, \mathbf{y}_T]$ . A questo risultato si può collegare il fatto che, se le previsioni per le serie al *bottom-level* della gerarchia sono  $\hat{\mathbf{b}}_h$ , queste hanno valore atteso pari a  $E[\hat{\mathbf{b}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = \boldsymbol{\beta}_h$ , vero valore atteso delle osservazioni. Unendo questo risultato alla formalizzazione della struttura gerarchica definita in 2.3, otteniamo che  $E[\hat{\mathbf{y}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = E[\mathbf{y}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = \mathbf{S}\boldsymbol{\beta}_h$ .

Seguendo la stessa logica, un insieme di previsioni riconciliate sarà non distorto solamente

se

$$E[\tilde{\mathbf{y}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = \mathbf{SPS}\boldsymbol{\beta}_h = \mathbf{S}\boldsymbol{\beta}_h \quad (4.9)$$

questa definizione impone quindi un vincolo sulle matrici  $\mathbf{P}$  e  $\mathbf{S}$ , in particolare il vincolo imposto è  $\mathbf{SPS} = \mathbf{S}$ , o equivalentemente che  $\mathbf{PS} = \mathbf{I}_m$ .

Tornando al problema di partenza, sappiamo che  $\Sigma_h$  è identificabile solamente se  $(\mathbf{I}_n - \mathbf{SP})$  è invertibile e perciò a rango pieno. Per trovare il rango della matrice risultante scriviamo che

$$\text{Rank}(\mathbf{I}_n - \mathbf{SP}) = \text{tr}(\mathbf{I}_n - \mathbf{SP}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_m) = n - m \quad (4.10)$$

Si vede chiaramente che  $\mathbf{I}_n - \mathbf{SP}$  ha rango non massimo, e conseguentemente la matrice  $\Sigma_h$  non può essere identificata, Wickramasuriya, et al. (2019) [5].

Nel lavoro proposto da Wickramasuriya, et al. (2019) [5], inoltre, viene segnalato un altro errore fondamentale commesso da Hyndman, et al. (2011) [11].

Definiamo gli errori di previsione, dato l'orizzonte di previsione  $h$ , delle previsioni riconciliate come

$$\tilde{\mathbf{e}}_h = \mathbf{y}_h - \tilde{\mathbf{y}}_h \quad (4.11)$$

Assumendo che  $\mathbf{y}_h$  è il vettore delle osservazioni per ogni serie della gerarchia al tempo  $T + h$ , mentre  $\tilde{\mathbf{y}}_h$  è il vettore delle previsioni riconciliate, ottenute usando le informazioni fino al tempo  $T$ . È opportuno specificare che gli errori definiti in (4.11) sono logicamente differenti dagli errori di coerenza definiti in (4.6). I primi sono essenzialmente gli errori di previsione delle stime riconciliate, mentre i secondi definiscono la differenza dovuta all'incoerenza delle stime generate in maniera indipendente, (Wickramasuriya et al. 2019 [5]). Hyndman et al. (2011) [11] confusero questi due errori. Sappiamo che, per ogni matrice  $\mathbf{P}$  tale per cui vale il vincolo che chiameremo di "non distorsione", cioè che  $\mathbf{SPS} = \mathbf{S}$ , la matrice di varianza-covarianza degli errori di stima riconciliati con orizzonte di previsione  $h$  è uguale a

$$\text{var}[\mathbf{y}_h - \tilde{\mathbf{y}}_h | \mathbf{y}_1, \dots, \mathbf{y}_T] = \mathbf{SPW}_h \mathbf{P}' \mathbf{S}' \quad (4.12)$$

*Dimostrazione.* Riscriviamo per comodità gli errori di previsioni con orizzonte di tempo  $h$  sia per le previsioni generate in maniera indipendente sia per le previsioni riconciliate, questi sono rispettivamente

$$\begin{aligned} \hat{\mathbf{e}}_h &= \mathbf{y}_h - \hat{\mathbf{y}}_h \\ \tilde{\mathbf{e}}_h &= \mathbf{y}_h - \tilde{\mathbf{y}}_h \end{aligned}$$

A questo punto possiamo scrivere che

$$\begin{aligned}
\tilde{\mathbf{e}}_h &= \mathbf{y}_h - \tilde{\mathbf{y}}_h \\
&= \mathbf{y}_h - \mathbf{SP}\hat{\mathbf{y}}_h + \hat{\mathbf{y}}_h - \hat{\mathbf{y}}_h \\
&= \hat{\mathbf{e}}_h + (\mathbf{I} - \mathbf{SP})\hat{\mathbf{y}}_h \\
&= \hat{\mathbf{e}}_h + (\mathbf{I} - \mathbf{SP})[\mathbf{y}_h - \hat{\mathbf{e}}_h] \\
&= \hat{\mathbf{e}}_h + \mathbf{y}_h - \mathbf{SP}\mathbf{y}_h - \hat{\mathbf{e}}_h + \mathbf{SP}\hat{\mathbf{e}}_h \\
&= \mathbf{SP}\hat{\mathbf{e}}_h + (\mathbf{I} - \mathbf{SP})\mathbf{y}_h
\end{aligned}$$

Dal momento che possiamo scrivere  $\mathbf{y}_h = \mathbf{S}\mathbf{b}_h$  dove  $\mathbf{b}_h$  sono le osservazioni all'istante temporale  $T + h$  delle serie al *bottom-level* della gerarchia, otteniamo che  $(\mathbf{I} - \mathbf{SP})\mathbf{S} = \mathbf{0}$ . Per cui possiamo dire che  $\tilde{\mathbf{e}}_h = \mathbf{SP}\hat{\mathbf{e}}_h$ , ottenendo quindi che la loro varianza, se poniamo  $\text{var}[\hat{\mathbf{e}}_h] = \mathbf{W}_h$ , è

$$\text{var}[\tilde{\mathbf{e}}_h] = \mathbf{SPW}_h\mathbf{P}'\mathbf{S}'$$

□

Grazie a questa formula si riesce quindi ad ottenere una stima della varianza degli errori di previsione per le serie riconciliate che, se unito ad uno stimatore ragionevole per  $\mathbf{W}_h$ , permette di costruire intervalli di confidenza indipendentemente dal tipo di approccio usato per generare previsioni, se si è assunta la Normalità degli errori. Inoltre, se le previsioni di base sono costruite in modo da essere non distorte, allora anche quelle riconciliate lo saranno.

A questo punto l'incognita più grossa è quel valore di  $\mathbf{P}$  che minimizza la traccia di  $\text{var}[\mathbf{y}_h - \tilde{\mathbf{y}}_h]$ , cioè la somma degli elementi sulla diagonale, che coincide con lo stimatore non distorto e a varianza minima delle previsioni riconciliate, (Wickramasuriya et al. 2019 [5]).

Seguendo quanto scritto da Wickramasuriya et al. (2019) [5], se adottiamo la notazione usata fino ad ora, la matrice  $\mathbf{P}$  che minimizza  $\text{tr}[\mathbf{SPW}_h\mathbf{P}'\mathbf{S}']$ , che coincide con la soluzione ottimale trovata da Hyndman et al. (2011) [11], è

$$\mathbf{P} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}, \quad (4.13)$$

che, dal punto di vista computazionale, può richiedere molte risorse e molto tempo perché bisogna invertire due matrici di ordine  $m \times m$  e di ordine  $n \times n$ , dove  $n$  e  $m$  sono rispettivamente il numero complessivo di serie presenti nella gerarchia e il numero di serie al *bottom-level*. Per cui viene proposta una rappresentazione alternativa della formula, una rappresentazione che sia significativamente meno onerosa in termini di costo computazionale.

Partizioniamo  $\mathbf{S} = \begin{bmatrix} \mathbf{C}_{(m-n) \times n} \\ \mathbf{I}_n \end{bmatrix}$ ,  $\mathbf{J} = \begin{bmatrix} \mathbf{0}_{n \times (m-n)} & \mathbf{I}_n \end{bmatrix}$  e  $\mathbf{U}' = \begin{bmatrix} \mathbf{I}_{m-n} & -\mathbf{C}_{(m-n) \times n} \end{bmatrix}$ , otteniamo quindi che

$$\mathbf{P} = \mathbf{J} - \mathbf{J}\mathbf{W}_h\mathbf{U}(\mathbf{U}'\mathbf{W}_h\mathbf{U})^{-1}\mathbf{U}' \quad (4.14)$$

A questo punto se si vuole calcolare le previsioni riconciliate si ottiene che

$$\tilde{\mathbf{y}}_h = \mathbf{S}[\mathbf{J} - \mathbf{J}\mathbf{W}_h\mathbf{U}(\mathbf{U}'\mathbf{W}_h\mathbf{U})^{-1}\mathbf{U}']\hat{\mathbf{y}}_h = \mathbf{S}\tilde{\mathbf{b}}_h \quad (4.15)$$

Per la dimostrare come l'equazione (4.13) possa essere riscritta in questo modo si prende tutto il procedimento direttamente dall'appendice di Wickramasuriya et al. (2019) [5], che definiscono questo metodo a "traccia minima", o MinT. Per la dimostrazione completa riferirsi al teorema riportato in appendice, sezione 6.1.

Come abbiamo detto in precedenza, nello studio di Hyndman et al. (2011) [11] si sono confusi gli errori di previsione con gli errori di coerenza. In questo modo al momento del calcolo della soluzione con i minimi quadrati generalizzati, si è utilizzata la matrice  $\mathbf{W}_h$ , cioè la matrice di varianza-covarianza degli errori di previsione delle stime generate in maniera indipendente per ciascuna serie della gerarchia, invece che la vera ed ignota matrice  $\Sigma_h$ . In questo modo la soluzione che è stata ottenuta è definita dall'equazione (4.13), giustificando l'uso dei minimi quadrati ordinari attraverso l'assunzione di additività degli errori, assumendo quindi che gli errori seguendo la struttura dei dati. Tuttavia, se assumiamo di avere le previsioni indipendenti per le serie al *bottom-level*, e le chiamiamo  $\hat{\mathbf{a}}_h$ , se assumiamo l'additività degli errori, allora possiamo scrivere  $\hat{\mathbf{e}}_h = \mathbf{S}\mathbf{V}_h\mathbf{S}'$ , dove  $\mathbf{S}$  è la solita notazione per la *summing matrix* e  $\mathbf{V}_h = \text{var}[\hat{\mathbf{a}}_h]$ . Quando andiamo a calcolare la matrice di varianza-covarianza per gli errori di previsione usando le previsioni riconciliate, però, otteniamo che

$$\text{var}[\mathbf{y}_h - \tilde{\mathbf{y}}_h] = \mathbf{S}\mathbf{P}\mathbf{S}'\mathbf{V}_h(\mathbf{S}\mathbf{P}\mathbf{S})', \quad (4.16)$$

per il vincolo di non distorsione otteniamo quindi  $\text{var}[\mathbf{y}_h - \tilde{\mathbf{y}}_h] = \mathbf{S}\mathbf{V}_h\mathbf{S}'$ , che è indipendente da  $\mathbf{P}$  e, da ciò, qualsiasi matrice che soddisfi il vincolo di non distorsione, cioè che  $\mathbf{S}\mathbf{P}\mathbf{S} = \mathbf{S}$ , risulta una soluzione a varianza minima.

Il lavoro presentato da Wickramasuriya et al. (2019) [5] è stato un punto di svolta nell'ambito dei problemi di previsione per serie storiche ed a gruppi perché fino a quel momento l'unico approccio ideato che sfruttasse, o che provasse a sfruttare, le informazioni sulle correlazioni intrinseche nelle gerarchie era l'approccio a combinazione ottimale, che è stato dimostrato impossibile da realizzare per via delle condizioni imposte da Hyndman et al. (2011) [11]. Il risultato fondamentale trovato da Wickramasuriya et al. (2019) [5] è che seppur utilizzando un approccio differente da quello ideato originariamente da Hyndman et al., la soluzione analitica è identica a quella che si otterrebbe con il metodo dei minimi

quadrati generalizzati. Inoltre, si è presentata una rappresentazione alternativa dello stimatore, una che riduce in maniera significativa il costo computazionale del procedimento, perché prevede l'inversione di una sola matrice, rendendo così la soluzione scalabile anche quando i dati aumentano di dimensione.

Nonostante questo, la stima della matrice di varianza-covarianza dei "base forecast errors", cioè gli errori di previsione delle stime generate in maniera indipendente, risulta difficile, specialmente quando l'orizzonte di previsione  $h$  diventa maggiore di 1, e non sempre è definita positiva. Per far fronte a questo problema, nello stesso articolo in cui hanno illustrato il nuovo metodo per il calcolo delle previsioni riconciliate Wickramasuriya et al., discutono diverse alternative per il calcolo della matrice  $\mathbf{W}_h$ . (Wickramasuriya et al. 2019 [5])

Il punto di partenza, e primo metodo, per trovare un metodo per identificare la matrice in maniera più efficiente è quello di porre  $\mathbf{W}_h = k_h \mathbf{I} \forall h$ , qui si impone che il peso  $k_h > 0$ . Questa soluzione, per quanto sia la più semplice, fa in modo che lo stimatore ottenuto attraverso il MinT coincida con quello ottenuto utilizzando i minimi quadrati ordinari di Hyndman et al. (2011). Tuttavia, questo è possibile solamente solo se alcune condizioni molto speciali vengono rispettate, ad esempio che gli errori di previsioni per le stime indipendenti siano incorrelate ed equivarianti. Si può dimostrare come queste condizioni siano impossibili da soddisfare nei casi reali di dati gerarchici o raggruppati.

Il secondo metodo presentato è denominato come  $\text{WLS}_v$ , e consiste nel calcolare  $\mathbf{W}_h = k_h \text{diag}(\hat{\mathbf{W}}_1) \forall h$ , dove  $k_h > 0$  e la matrice  $\hat{\mathbf{W}}_1$  è calcolata come

$$\hat{\mathbf{W}}_1 = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{e}}_t \hat{\mathbf{e}}_t', \quad (4.17)$$

cioè lo stimatore non distorto della varianza degli errori di previsione ad un passo. Hyndman et al. (2016) [1] idearono uno stimatore simile, ma non portarono una giustificazione teorica. In questo caso, invece, il MinT può essere visto come uno stimatore ottenuto attraverso i minimi quadrati pesati, applicando ad esempio una varianza scalata.

Il terzo metodo di stima discusso da Wickramasuriya, et al. [5] è stato proposto precedentemente da Athanasopoulos, et al. (2017) [14] per le gerarchie temporali, e assume che ogni errore di previsione nelle serie al *bottom-level* della gerarchia abbia varianza che dipende solo dall'istante di previsione  $k_h$  e che siano tra di loro incorrelate. In particolare, il metodo proposto dipende esclusivamente dal peso assegnato e dalla struttura, e viene quindi definito come uno stimatore che applica una trasformazione strutturale,  $\text{WLS}_S$ . Si assume, quindi, equivarianza solamente per gli errori al livello più disaggregato della gerarchia, al contrario di quanto succede con i minimi quadrati ordinari, che assume equivarianza su tutta la gerarchia. Poniamo, in pratica,  $\mathbf{W}_h = k_h \Lambda \forall h$ , assumendo che  $k_h > 0$  e  $\Lambda = \text{diag}(\mathbf{S1})$ , specificando che  $\mathbf{1}_m$  è un vettore colonna unitario di dimensione  $m$ .

Calcolando  $\mathbf{W}_h = k_h \hat{\mathbf{W}}_1 \forall h$ , sempre definendo  $k_h$  maggiore di 0 si ottiene lo stimatore della varianza campionaria non ristretta per  $h > 1$ , (Wickramasuriya et al. 2019 [5]). Chiamato con il nome di "MinT(Sample)" questo stimatore è molto semplice da realizzare, tuttavia

studi empirici hanno dimostrato che, se  $n > T$  o  $n$  è dello stesso ordine di grandezza di  $T$ , non risulta un buono stimatore.

L'ultimo metodo proposto utilizza un metodo di *shrinkage* per "contrarre" le entrate della matrice di varianza-covarianza per gli errori di previsione ad un passo  $\mathbf{W}_1$ . In particolare, se impostiamo  $\lambda_D$  come il parametro di intensità e  $\mathbf{W}_{1,D}$  la matrice diagonale che comprende solamente gli elementi sulla diagonale di  $\mathbf{W}_1$ , otteniamo che

$$\hat{\mathbf{W}}_{1,D}^* = \lambda_D \mathbf{W}_{1,D} + (1 - \lambda_D) \mathbf{W}_1 \quad (4.18)$$

è lo stimatore *shrunked*, che "contrae" verso lo zero tutti gli elementi fuori dalla diagonale di  $\mathbf{W}_1$ . A questo punto, se vogliamo calcolare la matrice di varianza-covarianza per gli errori di previsioni delle *base forecast* per un generico orizzonte di previsione  $h$ , possiamo scrivere  $\mathbf{W}_h = k_h \hat{\mathbf{W}}_{1,D}^*$ . Studi empirici hanno dimostrato come il metodo MinT(Shrink) sia quello che in generale riesce a performare in maniera migliore rispetto agli altri.

In termini generali, il metodo per la riconciliazione per serie storiche gerarchiche proposta da Wickramasuriya et al. (2019) [5] è stato validato in molte situazioni empiriche e ha dimostrato come quasi nella totalità dei casi sia riuscito ad ottenere risultati almeno "buoni" quanto metodi di previsione tradizionali, oltre al fatto che è stato dimostrato che se si parte da un'insieme di previsioni non distorte, fornisce sicuramente previsioni riconciliate non distorte. Per questo motivo l'utilizzo di previsioni riconciliate dovrebbe essere sempre preferito ai vari metodi *bottom-up*, *top-down* o *middle-out* perché queste tecniche oltre ad essere applicabili in moltissime situazioni e moltissimi settori diversi, riescono ad incorporare informazioni intrinseche nella struttura dei dati che tutti gli altri metodi non fanno, rendendo dal punto di vista concettuale molto più complete le stime delle previsioni derivanti dai metodi di riconciliazione.

Gli approcci che implementano la riconciliazione delle previsioni forniscono delle metodologie semplici per problemi di stima multivariati, perché suddividono il problema generale in una serie di problemi di stima univariati, a cui fa seguito la fase di riconciliazioni delle previsioni che, attraverso l'uso della matrice di varianza-covarianza degli errori di previsione, riesce a catturare le relazioni tra i nodi della gerarchia ed a sfruttare queste nuove informazioni per rendere le previsioni finali migliori e, spesso, metodi di previsione univariata riescono a performare meglio rispetto ai metodi multivariati perché nelle situazioni reali il numero di parametri da dover stimare può risultare proibitivo, specialmente per serie ad elevata dimensionalità. (Wickramasuriya et al. 2020 [13])

### 4.3 Riconciliazione ottimale non negativa

Nelle sezioni precedenti si è studiato come risolvere il problema della riconciliazione delle previsioni, che consiste nell'assicurare che le previsioni per i livelli più disaggregati della gerarchia che i dati formano si sommino in maniera tale da essere uguali alle previsioni per i livelli più alti di aggregazione. Questa proprietà, definita come coerenza, è di vitale importanza in molti ambiti economici, perché se è vero che nelle osservazioni risulta sempre verificata, nel momento in cui si genereranno delle previsioni, queste difficilmente rispetteranno questa proprietà. Abbiamo visto come i metodi più tradizionali a combinazione, come i metodi *bottom-up*, *top-down* e *middle-out*, riescano a garantire questa proprietà, ma risultino spesso incompleti perché utilizzano solamente le informazioni fornite ad un livello della gerarchia.

Una soluzione è stata individuata inizialmente da Hyndman et al. (2011) [11], che proposero di riconciliare le previsioni generate per ogni serie nella gerarchia, rendendole così coerenti e potendo sfruttare tutte le informazioni sulle correlazioni intrinseche nella struttura di aggregazione dei dati. La riconciliazione ai minimi quadrati proposta forniva delle previsioni coerenti che fossero il più simili possibile, dal punto di vista della differenza al quadrato, alle previsioni generate in maniera indipendente, le previsioni di base. Successivamente il lavoro proposto inizialmente fu esteso da Wickramasuriya et al. (2019) [5], che mostrarono come, attraverso un metodo da loro realizzato chiamato MinT, lo stimatore che ne risulta è a varianza minima tra gli stimatori non distorti.

Nella maggior parte delle applicazioni in cui questi metodi vengono applicati, si ha a che fare con dati strettamente non negativi, gli esempi più usati sono serie storiche sulle vendite di un negozio disaggregate per tipologia dei prodotti, flussi turistici in un Paese e così via, imponendo un vincolo logico sulle previsioni, cioè che anche queste siano strettamente non negative. Sfortunatamente, l'approccio MinT e le sue varianti non garantiscono che le previsioni riconciliate siano non negative, neppure quando le previsioni generate in maniera indipendente sono anch'esse non negative.

Il metodo proposto da Wickramasuriya et al. (2020) [13] per risolvere questo problema si basa sul fatto di esplicitare il vincolo di non negatività delle previsioni. È opportuno far notare che l'obiettivo in questo caso non è quello di migliorare la precisione delle stime, ma solamente quello di poter usare questi metodi nelle situazioni pratiche, come quelle viste prima.

Per poter assicurare che tutte le componenti di  $\tilde{\mathbf{y}}_h$ , vettore con le previsioni riconciliate al tempo  $T + h$ , siano non negative, è sufficiente garantire che tutte le entrate del vettore  $\tilde{\mathbf{b}}_h$  siano non negative, cioè che le previsioni al tempo  $T + h$  delle serie al *bottom-level* della gerarchia siano maggiori o uguali a zero.

Ricordando quanto trovato nell'equazione (4.18), possiamo scrivere che le previsioni riconciliate per le serie al *bottom-level* della gerarchia si calcolano come

$$\begin{aligned}\tilde{\mathbf{b}}_h &= [(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}]\hat{\mathbf{y}}_h \\ &= [\mathbf{J} - \mathbf{J}\mathbf{W}_h\mathbf{U}(\mathbf{U}'\mathbf{W}_h\mathbf{U})^{-1}\mathbf{U}']\hat{\mathbf{y}}_h\end{aligned}\quad (4.19)$$

dove  $\mathbf{W}_h$  è una matrice di pesi definita positiva, che nel caso del MinT risulta essere la matrice di varianza-covarianza degli errori di previsioni per le previsioni di base. Per quanto riguarda la seconda rappresentazione, quella usata da Wickamasuriya et al. per ridurre la complessità computazionale, sottolineiamo che  $\mathbf{S} = \begin{bmatrix} \mathbf{C}_{(m-n)\times n} \\ \mathbf{I}_n \end{bmatrix}$ ,  $\mathbf{J} = \begin{bmatrix} \mathbf{0}_{n\times(m-n)} & \mathbf{I}_n \end{bmatrix}$  e in ultimo  $\mathbf{U}' = \begin{bmatrix} \mathbf{I}_{m-n} & -\mathbf{C}_{(m-n)\times n} \end{bmatrix}$ . Nei casi più generali, ai pesi possono essere assegnati dei valori ad hoc da chi deve risolvere un problema di previsione, oppure possono essere calcolati.

Come detto poco fa, per poter essere sicuri che tutti i valori di  $\tilde{\mathbf{y}}_h$  siano positivi è sufficiente assicurarsi che i valori di  $\tilde{\mathbf{b}}_h$  siano maggiori o uguali a zero. Per cui possiamo scrivere che le stime  $\tilde{\mathbf{b}}_h$  possono essere viste come la soluzione di un problema ai minimi quadrati generalizzati della regressione definita come

$$\begin{aligned}\min_{\check{\mathbf{b}}} & \frac{1}{2}[\hat{\mathbf{y}}_h - \mathbf{S}\check{\mathbf{b}}_h]'\mathbf{W}_h^{-1}[\hat{\mathbf{y}}_h - \mathbf{S}\check{\mathbf{b}}_h] \\ &= \min_{\check{\mathbf{b}}} \frac{1}{2}\check{\mathbf{b}}_h'\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\check{\mathbf{b}}_h - \check{\mathbf{b}}_h'\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h + \frac{1}{2}\hat{\mathbf{y}}_h'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h\end{aligned}\quad (4.20)$$

che si traduce in un problema di programmazione quadratica

$$\min_{\check{\mathbf{b}}} q(\check{\mathbf{b}}) := \frac{1}{2}\check{\mathbf{b}}_h'\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\check{\mathbf{b}}_h - \check{\mathbf{b}}_h'\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h \quad \text{soggetto a } \check{\mathbf{b}} \geq 0 \quad (4.21)$$

Il problema così definito è detto ai minimi quadrati non negativi (NNLS).

In questo contesto, un vettore che soddisfa tutte le condizioni ed è soluzione del problema ai minimi quadrati è detto soluzione "possibile", e l'insieme di tutte le soluzioni così definite viene chiamata "regione possibile". Un problema di programmazione quadratica è detto quindi "possibile" se l'insieme delle soluzioni non è vuoto. Nel problema in questione, applicando i vincoli alla matrice dei pesi, si riesce ad assicurare che la regione possibile sia non vuota, e quindi che il problema sia sempre risolvibile, con un'unica soluzione di minimo globale.

Per derivare le condizioni per cui il vettore sia la soluzione per il problema NNLS bisogna considerare la funzione lagrangiana

$$\mathcal{L}(\check{\mathbf{b}}, \lambda) = q(\check{\mathbf{b}}) - \lambda'\check{\mathbf{b}} \quad (4.22)$$

e da questa ricavare le condizioni ottimali per cui una soluzione sia accettabile. Per una lettura più approfondita nell'ambito dei problemi di riconciliazione ottimale si riconduce a

Wickramasuriya et al. (2020) [13], in seguito verranno spiegati in maniera generale i principali risultati ed algoritmi ideati nella loro ricerca.

Nella loro pubblicazione, una volta esplicitati i vincoli che la soluzione deve rispettare per essere anche soluzione a NNLS, Wickramasuriya, Hyndman e Turla implementano tre algoritmi per raggiungere in maniera numerica questa soluzione. Il primo algoritmo implementato è l'algoritmo di perno sul blocco principale (*block principal pivoting*), la cui idea di base è quella di trasformare i problemi di disuguaglianza vincolata in una serie di problemi di uguaglianza vincolata (per ulteriori informazioni, consultare Lawson e Hanson (1974) [6]).

Il secondo è l'algoritmo di proiezione del gradiente (*conjugate gradient projection*), che si basa su due passi: il primo consiste nell'aggiornare la soluzione possibile il più velocemente possibile attraverso una discesa del gradiente, e il secondo passo consiste nel risolvere un sottoproblema agguistando i vincoli imposti sulla regione possibile.

L'ultimo algoritmo implementato è quello della proiezione scalata del gradiente (*Scaled gradient projection*). L'idea principale è quella di determinare una direzione in cui far decrescere la soluzione basandosi sul valore corrente, metodo ideato da Bonettini, Zanella e Zanni (2008) [17].

I risultati che ottennero Wickramasuriya, Hyndman e Turlac furono che l'algoritmo più efficiente sotto vari punti di vista come il tempo e il numero di istruzioni fu l'algoritmo *block principal pivoting*. Questo guadagno in termini di efficienza fu attribuito alla rappresentazione alternativa del MinT, e non direttamente all'algoritmo utilizzato per trovare una soluzione con valori espressamente non negativi.

Anche se non è stato implementato un algoritmo che soddisfi a pieno dal punto di vista dell'efficienza, in questo studio vengono espressi punti molto interessanti. Si è visto, infatti, che anche se viene scartato il vincolo di non distorsione, l'errore delle previsioni non è così impattante come si prevedeva. Infatti nei livelli più disaggregati della struttura si ha addirittura un guadagno, che viene gradualmente perso quando si risale la struttura di aggregazione, Wickramasuriya et al. (2020) [13]. Inoltre viene specificato che anche se questi guadagni e queste perdite non sono significative, un'insieme di previsioni riconciliate non negative è estremamente utile nel momento in cui si devono prendere decisioni importanti in campi in cui previsioni negative non sono sensate, campi economici per esempio, e quindi questo tipo di approccio deve essere preso in considerazione quando si ha da risolvere problemi di questa tipologia.



## 5. Previsioni sul flusso turistico australiano

Per valutare le *performance* delle tecniche che si basano sulla riconciliazione ottimale delle previsioni, consideriamo un *dataset* chiamato "*TourismData v4*", dove sono contenuti i dati sul flusso turistico australiano per un periodo che va dal Gennaio 1998 al Dicembre 2017. In particolare, si hanno 240 misurazioni mensili del totale di notti spese fuori dalla propria residenza, in milioni, che formano una gerarchia a quattro livelli dove il livello più disaggregato della gerarchia, il livello 3, è composto da 76 serie storiche, che rappresentano le regioni. Le regioni possono essere aggregate in maniera tale da formare 21 zone che formano il livello 2, che a loro volta possono essere aggregate in 7 stati, che corrisponde al livello 1 della gerarchie. Il livello 0 è la serie del totale, dove si hanno i dati a livello nazionale.

<b>Livello della Gerarchia</b>	<b>Numero di serie</b>
Livello 0	1
Livello 1	7
Livello 2	21
Livello 3	76
Totale	105

TABELLA 5.1: Numero di serie per livello nel dataset del turismo australiano.

La tabella 5.1 riassume le caratteristiche strutturali della serie storica gerarchica, mentre in Figura 5.1 vengono mostrate alcune delle caratteristiche delle serie interne. Come si può vedere nel pannello in alto a sinistra, il totale delle notti dei turisti, quindi al massimo livello di aggregazione, mostra in maniera marcata la sua forte componente stagionale, con i picchi in corrispondenza del mese di gennaio, che coincide con le vacanze estive in Australia. Si può notare, inoltre, un trend crescente che sembra partire dal 2010, e che dura fino alla fine del campione. Man mano che si scende nella gerarchia queste caratteristiche diventano meno marcate e sempre più difficili da modellare, perché il rapporto tra il segnale ed il rumore decresce. Per questo motivo, le serie dei livelli 2 e 3, illustrate nei riquadri in basso, mostrano una variazione più casuale e caratteristiche meno stabili, quando confrontate con le serie ai livelli superiori.

Lo scopo di questo esperimento è quello di valutare come diverse tecniche di combinazione e riconciliazione delle previsioni si comportano nei casi reali, e quali di queste riescano a performare in maniera più efficiente. In questo contesto usare un metodo di riconciliazione che riesca a catturare in maniera efficace le diverse caratteristiche delle singole serie storiche, e che fornisca previsioni coerenti, risulta fondamentale per la programmazione. La suddivisione completa dei dati è riportata nell'appendice, Tabella 6.1.

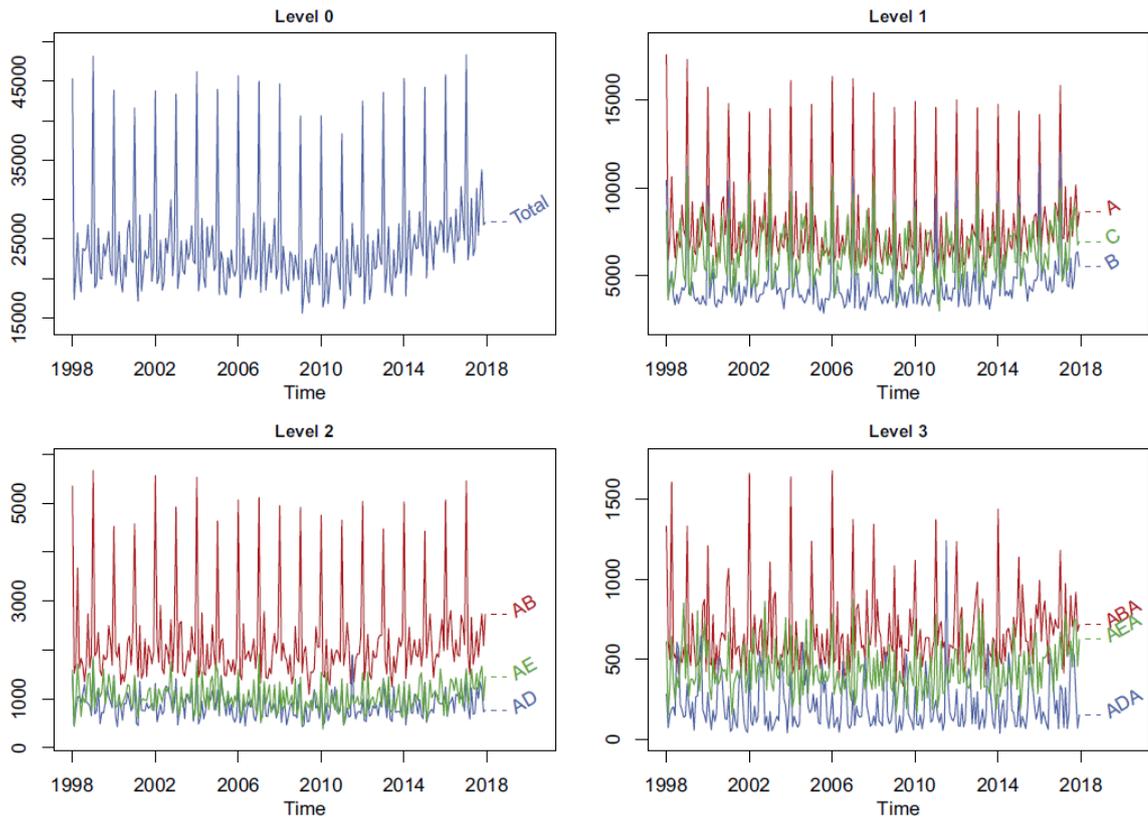


FIGURA 5.1: Flusso turistico australiano misurato in milioni, per alcune delle zone geografiche selezionate.

## 5.1 Configurazione dell'esperimento

Dato il dataset descritto precedentemente, per valutare l'accuratezza delle previsioni, si decide di utilizzare un test fuori campione a origine variabile. Di questa procedura è stato scritto, infatti, che definisce regole più precise per la divisione delle serie, permette di distinguere la distribuzione degli errori in base al tempo e desensibilizza gli errori dovuto ad un determinato punto di origine da cui generare le previsioni, Tashman (2000) [20].

Inoltre, usando questo approccio, si riescono ad ottenere dei risultati che rispecchino in maniera più simile possibile alla realtà e approssimare le *performance* dei vari metodi in situazioni a lungo termine. L'approccio *rolling-origin* per la valutazione delle *performance* dei metodi di previsioni viene implementato nella seguente maniera: viene diviso il *dataset* in due parti. La prima parte definita di *train*, che comprende le osservazioni dall'istante 1 all'istante  $T$ , viene usata per produrre in maniera automatica le previsioni per un'orizzonte di previsione  $h$ . La parte rimanente del *dataset* di partenza viene assegnata come insieme di *test*, in questo esempio comprende le osservazioni dal tempo  $T + 1$  fino al tempo  $T + h$  e viene usato per il confronto con le previsioni generate. Una volta confrontato il primo insieme di previsioni, si modifica l'insieme di *train*, aggiornando l'istante dal quale si generano le previsioni. In sostanza, in questa seconda iterazione il *dataset* di *train* è composto dalle osservazioni dal tempo 1 al tempo  $T + 1$ , mentre quello di *test* è formato dalle osservazioni dall'istante  $T + 2$  all'istante  $T + h + 1$ . Il processo è ripetuto un numero di volte  $K$  finché il dataset di *test* non è vuoto, ossia finché  $N + h + K - 1 \leq n$ , dove  $n$  è il numero di osservazioni totali nell'insieme delle osservazioni iniziale non diviso. Osservando la frequenza delle osservazioni e la lunghezza del *dataset* sul turismo australiano, viene considerata una versione personalizzata dei valori iniziali per la divisione del dataset in parte di *train* e parte di *test*. In particolare, anche se l'obiettivo è quello di valutare la precisione delle previsioni per un anno completo (12 osservazioni mensili), si sceglie di impostare 3 anni per la valutazione progressiva della precisione delle stime. Si imposta quindi  $T_1 = 204$  mesi e  $h_1 = 12$  mesi, il primo valore per avere un dataset di *train* abbastanza robusto per poter generare previsioni in maniera adeguata e un *dataset* di *test* sufficientemente grande per poter ottenere delle valutazioni più precise sui metodi presi in considerazione, mentre il secondo perché si è scelto un orizzonte di previsione di 12 mesi, per poter generare le stime per tutto l'anno successivo all'origine. Per riassumere, alla prima iterazione il dataset di *train* avrà 105 variabili da 204 osservazioni ciascuna, mentre il dataset di *test* comprenderà le 36 osservazioni rimanenti per le 105 variabili.

Per produrre le previsioni base vengono utilizzati dei modelli ARIMA, implementati per dal pacchetto *forecast* di R, Hyndman et al.(2022) [4]. Dal momento che lo scopo dell'esperimento è quello di confrontare i diversi metodi di riconciliazione e combinazione per l'ottenimento di previsioni coerenti, viene tralasciato il compito di proporre il modello da cui generare le previsioni, lasciando che sia il programma in sé a decidere quale modello sia il più adeguato per ogni serie della gerarchia, imponendo come unico vincolo che faccia

parte della classe ARIMA.

Una volta generate le previsioni base per ogni serie della gerarchia, si utilizzano cinque diversi metodi di riconciliazione per riconciliare le previsioni. I metodi utilizzati per il confronto sono il metodo *bottom-up* (BU), il metodo *top-down* (TD), calcolando le proporzioni come *average historical proportion*, e quattro varianti del MinT, cioè: lo stimatore che coincide con quello utilizzato da Hyndman et al. (2011) [11], cioè lo stimatore i minimi quadrati ordinari (denominato con *OLS*), lo stimatore che utilizza i minimi quadrati pesati applicando una varianza scalata (denominato con *WLS<sub>u</sub>*) implementato per la prima volta da Hyndman et al. (2016) [1], lo stimatore proposto per la prima volta da Athanasopoulos et al. (2017) [14] che utilizza la matrice di covarianza strutturale, che dipende solamente dalla struttura della gerarchia (denominato con *MinTSS*) e lo stimatore che contrae con un metodo di *shrinkage* i valori fuori dalla diagonale della matrice di varianza e covarianza (denominato con *MinTShrink*).

Per tutte le procedure di riconciliazione si utilizza il pacchetto R FoReco, che contiene al suo interno metodi per la riconciliazione cross-sezionale e cross-temporale, possibilmente con vincoli di non negatività (Di Fonzo e Girolimetto 2022 [18]).

Per valutare le *performance* dei metodi scelti per generare previsioni coerenti, invece, vengono implementati degli indicatori che consentano di basarsi sull'accuratezza e la deviazione assoluta delle previsioni dai veri valori.

In particolare, seguendo quanto descritto da da Spiliotis et al. (2021) [9], ci si affida al MASE, *Mean Absolute Scaled Error* ed al AMSE, *Absolute Mean Scaled Error*, per ottenere delle valutazioni oggettive sulle *performance* dei metodi selezionati.

$$\text{MASE} = \frac{n-s}{h} \frac{\sum_{t=n+1}^{n+h} |y_t - f_t|}{\sum_{t=s+1}^n |y_t - y_{t-s}|},$$

$$\text{AMSE} = \frac{n-s}{h} \frac{|\sum_{t=n+1}^{n+h} (y_t - f_t)|}{\sum_{t=s+1}^n |y_t - y_{t-s}|},$$

Dove  $y_t$  e  $f_t$  sono rispettivamente le osservazioni e le previsioni all'istante temporale  $t$ ,  $n$  è la dimensione dell'insieme di osservazioni usato per generare le previsioni,  $s$  è il periodo e  $h$  è l'orizzonte di previsione considerato.

Le varie tipologie di errore utilizzate sono scalati, di fatto rendendo possibile la media attraverso le serie sia dello stesso livello che tra livelli diversi, inoltre dato che la media minimizza la somma degli scarti al quadrato, il MASE è appropriato per misurare l'accuratezza delle previsioni coerenti che vengono generate dai diversi metodi, mentre l'AMSE risulta uno strumento efficace per misurare per la distorsione delle previsioni, Spiliotis et al. (2021) [9].

Tutti i risultati proposti in seguito sono ottenuti attraverso l'algoritmo riportato nell'appendice.

## 5.2 Risultati

Prima di iniziare con le valutazioni dettagliate dei risultati ottenuti, è bene sottolineare che anche se i metodi che implementano il *Minimum Trace Reconciliation* sono i metodi più all'avanguardia nel campo della previsione per serie storiche gerarchiche e raggruppate, anche i metodi meno complessi dal punto di vista teorico come i metodi *bottom-up* e *top-down* risultano quasi sempre competitivi e, in alcune circostanze, dei buonissimi indicatori con cui confrontare metodi di previsione.

TABELLA 5.2: Performance dei vari metodi di riconciliazione delle previsioni per il dataset "TourismData v4" dopo aver applicato un approccio di valutazione a finestra scorrevole (media di 36 valutazioni di previsioni con orizzonte di previsione uguale a 12). Le valutazioni vengono misurate in termini di MASE e AMSE, stimati sia per livello della gerarchia sia media complessiva.

Metodo	Livello 0	Livello 1	Livello 2	Livello 3	Media
MASE					
BU	1.518	1.061	0.878	0.724	1.045
TD	1.739	1.175	0.908	0.742	1.141
OLS	1.704	1.215	0.955	0.810	1.171
WLS <sub>u</sub>	<b>1.485</b>	1.045	0.868	0.720	1.029
MinT SS	1.545	1.084	0.889	0.745	1.066
MinT Shrink	1.545	<b>0.999</b>	<b>0.807</b>	<b>0.680</b>	<b>1.008</b>
AMSE					
BU	1.226	0.699	0.502	0.374	0.700
TD	1.590	0.877	0.563	0.426	0.864
OLS	1.552	0.980	0.686	0.525	0.936
WLS <sub>u</sub>	1.232	0.724	0.518	0.380	0.713
MinT SS	1.349	0.808	0.581	0.436	0.793
MinT Shrink	<b>0.894</b>	<b>0.562</b>	<b>0.418</b>	<b>0.330</b>	<b>0.551</b>

La tabella 5.2 riassume l'accuratezza previsiva dei metodi di riconciliazione selezionati in questo studio in termini di accuratezza (MASE) e di distorsione rispetto ai veri valori (AMSE). La prima colonna presenta il metodo usato per riconciliare le previsioni generate in maniera indipendente per ogni serie della gerarchia, mentre le restanti colonne indicano il livello di aggregazione. Per trovare un valore per livello di aggregazione si è usata una media aritmetica tra i valori dello stesso livello. L'ultima colonna comprende una misura riassuntiva per tutti i livelli, la media delle misure riportate nella tabella. Questo procedimento risulta corretto anche dal punto di vista logico visto che si sono scelte delle misure indipendenti dalla scala dei dati.

I risultati per i dati sul flusso turistico australiano mostrano come l'approccio migliore per la riconciliazione sia, in media e in tutti i livelli tranne che per il livello del totale, il MinT *Shrink*, ovvero lo stimatore che contrae le entrate fuori dalla diagonale della matrice di

varianza-covarianza degli errori di previsione verso lo zero. Se ci si concentra sul MASE in particolare, il metodo MinT Shrink è più accurato in media del 12% e del 4% rispetto ai metodi di riconciliazione più tradizionali, cioè i metodi *top-down* e *bottom-up*. Se si guarda per livello, anche se le percentuali di accuratezza differiscono da quella in media, si ha quasi sempre un netto guadagno da parte del MinT Shrink, questo fatto può essere ricondotto in maniera quasi certa a come questi approcci di riconciliazione sono costruiti. Infatti, i metodi TD e BU utilizzano solamente le informazioni di uno dei livelli della gerarchia, omettendo tutte quelle contenute negli altri livelli, problema per cui l'approccio a traccia minima è la naturale soluzione.

Come ci aspettavamo, il metodo OLS non è molto performante, questo è dovuto agli assunti di cui ha bisogno per poter essere definito, assunti che non sono quasi mai verificati, se non nelle simulazioni o in casi molto banali.

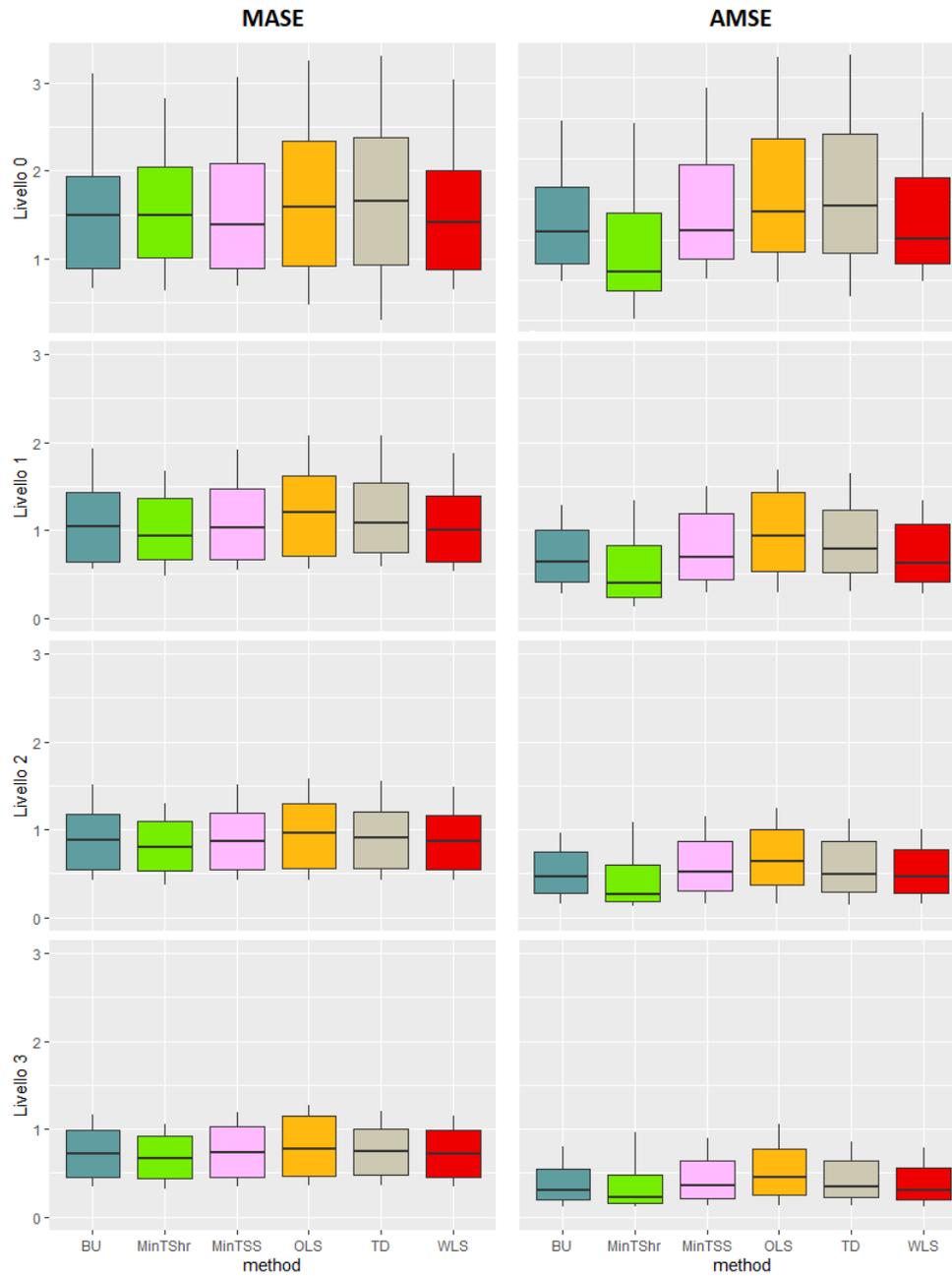
Anche se il metodo  $WLS_u$  risulta il metodo più accurato per il Totale, sempre considerando solamente il MASE, questo approccio risulta lo stesso meno performante del metodo MinT Shrink in media del 2.2%. L'utilizzo dei minimi quarati pesati conferma, in ogni caso, quanto detto da Wickamasurya et al. (2019) [5], cioè che anche questo metodo riesce a migliorare l'accuratezza rispetto alle previsioni di base generate in maniera indipendente per ogni serie della gerarchia.

Per quanto riguarda MinT SS, cioè il metodo che utilizza la matrice di varianza e covarianza strutturale, nonostante ottenga buoni risultati, non riesce comunque a performare meglio del MinT Shrink, poiché quest'ultimo incorpora informazioni anche sulla struttura di correlazione delle serie nella gerarchia, mentre il primo dipende solamente dalla struttura di aggregazione.

Per quanto riguarda le informazioni fornite circa la distorsione delle stime riconciliate, si possono fare considerazioni analoghe. Il metodo che riesce ad ottenere i risultati più positivi il MinT Shrink, seguito dal metodo BU, dal metodo  $WLS_u$ , dal MinT SS e infine OLS e TD. È interessante far notare come il metodo TD soffra di una distorsione molto maggiore al Livello 0, punto di partenza del metodo per la costruzione delle previsioni coerenti, indice che le previsioni generate per quel livello sono abbastanza distorte dal loro vero valore. Nonostante questo problema e il fatto che le proporzioni usate per disaggregare il valore iniziale non siano ottenute attraverso un metodo particolarmente sofisticato, la distorsione delle previsioni per i livelli successivi si riduce notevolmente, fino a risultare simile a quella del metodo MinT SS.

La più importante evidenza che questi risultati portano in ogni caso è che se si tiene conto delle informazioni di tutte le serie della gerarchia (informazioni cross-sezionali) le previsioni che si ottengono dai metodi di riconciliazione risultano essere più accurate e meno distorte, in media e per ogni livello della gerarchia. Da questo è chiaro che bisogna sempre preferire un metodo che implementi la riconciliazione ottimale delle previsioni rispetto a un metodo che utilizzi informazioni parziali, contenute in un solo livello della gerarchia.

FIGURA 5.2: Accuratezza dei metodi per la riconciliazione delle previsioni per metodo utilizzato e per livello di aggregazione.



### 5.3 Conclusioni

Da quando Hyndman et al. (2011) [11] introdussero per la prima volta l'approccio a combinazione ottimale, molti degli sforzi nel campo della ricerca sui problemi di previsioni per serie storiche gerarchiche ed a gruppi si sono canalizzati in questa particolare metodologia. Nonostante i metodi tradizionali non debbano essere abbandonati, perché ancora buonissimi indicatori di qualità delle previsioni e realizzabili con costi computazionali praticamente trascurabili, è stato chiaro fin da subito che incorporare informazioni di altri livelli e soprattutto sulle relazioni esistenti tra le serie della gerarchia, avrebbe permesso di ottenere dei vantaggi dal punto di vista dell'accuratezza e dei risultati.

Il secondo grande passo per l'evoluzione di queste tecniche di riconciliazione è stato fatto da Wickramasuriya et al. (2019) [5], quando dimostrarono che le previsioni ottenute attraverso metodi di riconciliazione garantivano risultati accurati almeno quanto quelli ottenuti utilizzando previsioni ottenute in maniera indipendente per ogni serie della gerarchia, a scapito dei metodi per la generazione di previsioni per serie gerarchiche tradizionali come *bottom-up* e *top-down*. La grande innovazione, tuttavia, fu il nuovo approccio per la riconciliazione ottimale che presentarono insieme a questi risultati. Chiamato a traccia minima, questo loro metodo non solo riusciva a superare il problema di identificabilità di cui soffriva la soluzione precedentemente proposta da Hyndman et al., ma forniva sia una rappresentazione alternativa, una computazionalmente meno costosa, dello stimatore, sia nuovi metodi per la stima della matrice di varianza e covarianza degli errori di previsione.

Da quando questi metodi sono stati pubblicati la riconciliazione gerarchica delle previsioni è diventato uno strumento noto e molto potente per la generazione di previsioni coerenti per collezioni dati. Citando direttamente dal lavoro di di Hollyman et al. (2021) [2], "*Queste tecniche sono ampiamente applicabili in diversi settori economici e meritano di diventare degli strumenti importanti nella "cassetta degli attrezzi" dei problemi di previsione*", e ancora "*Il guadagno (in accuratezza e non distorsione, ndr) di queste metodologie di riconciliazione è particolarmente essenziale quando c'è più bisogno: quando le previsioni esistenti sono di qualità povera*". Tutte queste caratteristiche positive sono infatti frutto di prove empiriche di cui la letteratura in merito è ben documentata, ed è stato dimostrato anche dall'esperimento di valutazione dei metodi di riconciliazione proposto sul flusso turistico australiano.

In questo specifico caso, l'attenzione è stata posta sui problemi di riconciliazione cross-sezionale, ma l'accuratezza dei risultati ottenuti usando la riconciliazione ottimale delle previsioni non cambiano nel caso in cui ci si trovi a dover risolvere un problema di previsione cross-temporale, ovvero quando si vuole che le previsioni generate siano coerenti a diversi orizzonti di tempo. Anche in questi casi il MinT riesce a fornire i risultati migliori, catturando le interconnessioni tra le diverse serie storiche.

È stato inoltre descritta una diversa direzione degli sforzi della ricerca sui problemi di previsione, la non negatività delle previsioni. L'approccio a traccia minima, infatti, soffre di

---

un problema, e cioè che può generare delle previsioni negative, anche quando le previsioni base di partenza sono tutte positive. In campi economici, come per il flusso di turismo australiano, dove valori negativi sono insensati, questo risulta un problema. La soluzione proposta da Wickramasuriya et al. (2020) [13] è quella di imporre un vincolo di non negatività durante la fase di stima. Il problema così risulta risolto dal punto di vista teorico ma ne sorge un altro dal punto di vista pratico: il *tradeoff* tra costo computazionale e accuratezza delle previsioni. L'aggiunta del vincolo inevitabilmente appesantisce il problema, tuttavia ci sono diversi metodi che possono essere applicati per ridurre la complessità computazionale dell'algoritmo, a patto di aggiungere della distorsione nelle stime finali delle previsioni. È innegabile dire, però, che nonostante ci sia della distorsione nelle previsioni, un insieme di previsioni sensate dal punto di vista della non negatività e coerenti risulta essere molto utile negli ambiti manageriali dove prendere decisioni informate risulta essere uno degli aspetti più critici.



## 6. Appendice

### 6.1 Teorema A.1

*Dimostrazione.* La formula obiettivo può essere vista come,

$$tr[\mathbf{S}\mathbf{P}\mathbf{W}_h\mathbf{P}'\mathbf{S}'] = tr[\mathbf{S}'\mathbf{S}\mathbf{P}\mathbf{W}_h\mathbf{P}']$$

Usando il Lemma 1 di Wang et al. (1986), visto che sia  $\mathbf{S}'\mathbf{S}$  che  $\mathbf{P}\mathbf{W}_h\mathbf{P}'$  sono simmetriche e, rispettivamente, definita e semidefinita positive, abbiamo che

$$tr[\mathbf{S}'\mathbf{S}\mathbf{P}\mathbf{W}_h\mathbf{P}'] \geq \lambda_{\min}(\mathbf{S}'\mathbf{S})tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'] \geq \lambda_{\min}(\mathbf{C}'\mathbf{C} + \mathbf{b}\mathbf{I})tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'],$$

Dove  $\lambda_{\min}(A)$  corrisponde all'autovalore minimo della matrice  $\mathbf{A}$  e  $\mathbf{S} = \begin{bmatrix} \mathbf{C} \\ \mathbf{I}_n \end{bmatrix}$ . Applicando le disuguaglianze di Weyl (Horn & Johnson, 1990), usando  $\mathbf{C}'\mathbf{C}$  e  $\mathbf{I}$  entrambe simmetriche, la prima semidefinita positiva e la seconda definita positiva, ottengo che

$$\lambda_{\min}(\mathbf{C}'\mathbf{C} + \mathbf{I})tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'] \geq [\lambda_{\min}(\mathbf{C}'\mathbf{C}) + \lambda_{\min}(\mathbf{I})]tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'] \geq tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'],$$

Usando il fatto che tutti gli autovalori della matrice identità sono unitari. Specificando il vincolo che  $\mathbf{P}\mathbf{S} = \mathbf{I}$  il problema di minimizzazione si può esprimere come

$$\min_{\mathbf{P}} tr[\mathbf{P}\mathbf{W}_h\mathbf{P}'] \quad \text{soggetto a } \mathbf{P}\mathbf{S} = \mathbf{I}$$

Se  $\mathbf{W}_h$  è definita positiva, e definiamo  $\mathbf{L} = \mathbf{W}_h^{-\frac{1}{2}}\mathbf{S}$  e  $\mathbf{H} = \mathbf{P}\mathbf{W}_h^{\frac{1}{2}}$ , allora il problema viene tradotto in

$$\min_{\mathbf{H}} tr[\mathbf{H}\mathbf{H}'] \quad \text{soggetto a } \mathbf{H}\mathbf{L} = \mathbf{I}$$

La soluzione di questo problema è nota e si può derivare usando l'inversa generalizzata di Moore-Penrose di  $\mathbf{L}$  (Penrose, 1956). Sappiamo inoltre che  $\mathbf{L}$  è a rango pieno, e che possiamo scrivere  $\mathbf{H} = (\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'$ , questo ci permette di scrivere

$$\mathbf{P} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$$

Poiché la soluzione attraverso l'uso del MinT può essere rappresentato similmente ad uno stimatore usando i Minimi Quadrati Generalizzati, possiamo riformulare il problema della minimizzazione della traccia attraverso un problema di minimi quadrati soggetto ad un vincolo di uguaglianza lineare, che si traduce in

$$\min_{\tilde{\mathbf{y}}_h} \frac{1}{2} [\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h]' \mathbf{W}_h [\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h] \quad \text{soggetto a } \mathbf{U}' \tilde{\mathbf{y}}_h = 0$$

o equivalente può cambiare il vincolo, imponendo che  $\tilde{\mathbf{y}}_h = \mathbf{S} \tilde{\mathbf{b}}_h$ , con  $\tilde{\mathbf{b}}_h$  il vettore delle previsioni per le serie al *bottom-level* della gerarchia e  $\mathbf{U}' = [\mathbf{I}_{m^*} | -\mathbf{C}']$ .

La funzione Lagrangiana per minimizzare il problema è definita come

$$\mathcal{L}[\tilde{\mathbf{y}}_h, \lambda] = \frac{1}{2} [\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h]' \mathbf{W}_h^{-1} [\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h] - \lambda' \mathbf{U}' \tilde{\mathbf{y}}_h$$

Dove  $\lambda$  è un vettore di moltiplicatori di Lagrange.

La condizione di primo ordine necessaria per cui abbiamo che  $\tilde{\mathbf{y}}_h$  è la soluzione del problema di minimizzazione richiede che esista  $\tilde{\lambda}$ , un vettore tale per cui si riesce a soddisfare il sistema di equazioni

$$\begin{bmatrix} \mathbf{W}_h^{-1} & \mathbf{U} \\ \mathbf{U}' & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_h \\ -\tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_h^{-1} \hat{\mathbf{y}}_T(h) \\ 0 \end{bmatrix}.$$

Assumendo che la matrice  $\mathbf{W}_h^{-1}$  è definita positiva la soluzione si trova ponendo

$$\begin{aligned} \tilde{\mathbf{y}}_h &= [\mathbf{I} - \mathbf{W}_h \mathbf{U} (\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}'] \hat{\mathbf{y}}_h \\ \text{e } \tilde{\lambda} &= -(\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}' \hat{\mathbf{y}}_h \end{aligned}$$

Le stime riconciliate del *bottom-level* sono quindi calcolate come

$$\tilde{\mathbf{b}}_h = \left[ [0_{n \times m^*} \mid \mathbf{I}_n] - [0_{n \times m^*} \mid \mathbf{I}_n] \mathbf{W}_h \mathbf{U} (\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}' \right] \hat{\mathbf{y}}_T(h) \mathbf{P} \hat{\mathbf{y}}_h$$

Per cui la rappresentazione alternativa di  $\mathbf{P}$  diventa

$$\mathbf{P} = \mathbf{J} - \mathbf{J} \mathbf{W}_h \mathbf{U} (\mathbf{U}' \mathbf{W}_h \mathbf{U})^{-1} \mathbf{U}'$$

dove  $\mathbf{J} = [0_{n \times m^*} \mid \mathbf{I}_n]$ . □

## 6.2 Algoritmo rolling-origin per la valutazione dei metodi di riconciliazione

```

#Assunto BottomOsservation come le osservazioni delle
bottom time series
#Generazione della matrice C (e quindi S' = [C | I])
OnlyData <- BottomOsservation[,3:78]
OnlyData %>%
  colnames() %>%
  as.list() -> BottomNames

name_bts <- data.frame(
  State = c(BottomNames %>%
    gsub('.{2}$', '',.)),
  Zones = c(BottomNames %>%
    gsub('^.', '',.) %>%
    gsub('.$', '',.)),
  Region = c(BottomNames %>%
    gsub('^.{2}', '',.)))

C <- Cmatrix( ~ State/Zones/Region, name_bts, sep = "")

#Data frame per contenere gli errori ai vari istanti di
previsione per livello
MASEByLevel <- data.frame(Level0 = double(),
  Level1 = double(),
  Level2 = double(),
  Level3 = double())

AMSEByLevel <- data.frame(Level0 = double(),
  Level1 = double(),
  Level2 = double(),
  Level3 = double())

```

```

h <- 12 #orizzonte di previsione dall'ultima
      osservazione del dataset di train

for(t in 0:1) {
  #Divisione iterativa del dataset in una parte
  #di train e una parte di test
  inds <- partition(BottomOsservazione$AAA,
                    p = c(train = 1-t/nrow(AT), test = t/nrow(AT))
                    , type = "blocked")

  train_bts <- BottomOsservazione[inds$train, 3:78]
  sapply(1:nrow(train_bts), function(i)
as.matrix(C) %*% as.matrix(t(train_bts[i, ])) ) %>%
  t() -> train_uts
  colnames(train_uts) <- C@Dimnames[[1]]
  hts_train <- cbind(train_uts, train_bts)

  test_bts <- BottomOsservazione[inds$test, 3:78]
  sapply(1:nrow(test_bts), function(i)
as.matrix(C) %*% as.matrix(t(test_bts[i, ])) ) %>%
  t() -> test_uts
  colnames(test_uts) <- C@Dimnames[[1]]
  hts_test <- cbind(test_uts, test_bts)

  #Aggiornamento dei modelli ad ogni passo, in funzione
  al
  #dataset di train e di test
  fitted <- lapply(1:ncol(hts_train),
                  function(i) auto.arima(hts_train[,i]))

  if(h > nrow(hts_test)) {
    h <- h - 1
  }

  #Generazione delle previsioni base, per ogni serie del
  modello
  basef <- lapply(1:ncol(hts_train), function(i)
    forecast(hts_train[,i],

```

```

        h, model = fitted[[i]], level = 0.95)$mean)
sapply(1:h, function(i) as.double(basef[[1]][i])) %>%
  as.matrix() -> M_basef
for(j in 2:ncol(hts_train)) {
  M_basef <- cbind(M_basef,
    sapply(1:h, function(i)
      as.double(basef[[j]][i])))
}
colnames(M_basef) <- colnames(hts_test)

#Costruzione matrice di residui
Mres <- lapply(1:ncol(hts_train), function(i)
  forecast(hts_train[,i], h, model = fitted[[i]],
    level = 0.95)$residuals)

sapply(1:nrow(hts_train), function(i) as.double(Mres
  [[1]][i])) %>%
  as.matrix() -> RES

for(j in 2:ncol(hts_train)) {
  RES <- cbind(RES, sapply(1:nrow(hts_train),
    function(i) as.double(Mres[[j]][i])))
}

##Riconciliazione con metodi differenti
#RICONCILIAZIONE BOTTOM-UP
M_basef %>%
  htsrec(., comb = "bu", C = C) -> cs_bu

mase_bu <- sapply(1:ncol(cs_bu$recf),
  function(i) mase(actual = hts_test
    [1:h,i],
    predicted = cs_bu$recf[,i], h = t,
    train = hts_train[,i], s = 12))

amse_bu <- sapply(1:ncol(cs_bu$recf),
  function(i) amse(actual = hts_test
    [1:h,i],
    predicted = cs_bu$recf[,i], h = t,
    train = hts_train[,i], s = 12))

```

```

MASEByLevel <- rbind(MASEByLevel, c(mase_bu[1], mean(
  mase_bu[2:8]),
  mean(mase_bu[9:29]), mean(mase_bu
    [30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_bu[1], mean(
  amse_bu[2:8]),
  mean(amse_bu[9:29]), mean(amse_bu
    [30:105])))

#RICONCILIAZIONE TOP-DOWN
# average historical proportions
props <- colMeans(hts_train[1:nrow(hts_train),-c(1:29)]/
  hts_train[1:nrow(hts_train),1])
M_basef[,1] %>%
  tdrec(., C = C, weights = props) -> cs_td

mase_td <- sapply(1:ncol(cs_td),
  function(i) mase(actual = hts_test
    [1:h,i],
    predicted = cs_td[,i], h = t,
    train = hts_train[,i], s = 12))

amse_td <- sapply(1:ncol(cs_td),
  function(i) amse(actual = hts_test
    [1:h,i],
    predicted = cs_td[,i], h = t,
    train = hts_train[,i], s = 12))

MASEByLevel <- rbind(MASEByLevel, c(mase_td[1], mean(
  mase_td[2:8]),
  mean(mase_td[9:29]), mean(mase_td
    [30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_td[1], mean(
  amse_td[2:8]),
  mean(amse_
    td[9:29]), mean(amse_td[30:105])))

```

```
#OLS
M_basef %>%
  htsrec(., comb = "ols", C = C) -> cs_ols

mase_ols <- sapply(1:ncol(cs_ols$recf),
  function(i) mase(actual = hts_
    test[1:h,i], predicted = cs_
    ols$recf[,i], h = t, train =
    hts_train[,i], s = 12))

amse_ols <- sapply(1:ncol(cs_ols$recf),
  function(i) amse(actual = hts_
    test[1:h,i], predicted = cs_
    ols$recf[,i], h = t, train =
    hts_train[,i], s = 12))

MASEByLevel <- rbind(MASEByLevel, c(mase_ols[1], mean(
  mase_ols[2:8]), mean(mase_ols[9:29]), mean(mase_ols
  [30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_ols[1], mean(
  amse_ols[2:8]), mean(amse_ols[9:29]), mean(amse_ols
  [30:105])))

#WLS
M_basef %>%
  htsrec(., comb = "wls", C = C, res = RES) -> cs_wls

mase_wls <- sapply(1:ncol(cs_wls$recf),
  function(i) mase(actual = hts_test
    [1:h,i], predicted = cs_wls$recf
    [,i], h = t, train = hts_train[,
    i], s = 12))

amse_wls <- sapply(1:ncol(cs_wls$recf),
  function(i) amse(actual = hts_test
    [1:h,i], predicted = cs_wls$recf
    [,i], h = t, train = hts_train[,
    i], s = 12))
```

```

MASEByLevel <- rbind(MASEByLevel, c(mase_wls[1], mean(
  mase_wls[2:8]), mean(mase_wls[9:29]), mean(mase_wls
  [30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_wls[1], mean(
  amse_wls[2:8]), mean(amse_wls[9:29]), mean(amse_wls
  [30:105])))

#MINT STRUCTURAL SCALING
M_basef %>%
  htsrec(., comb = "struc", C = C, res = RES) -> cs_
  minshr

mase_minshr <- sapply(1:ncol(cs_minshr$recf),
  function(i) mase(actual = hts_
    test[1:h,i], predicted = cs_
    minshr$recf[,i], h = t, train
    = hts_train[,i], s = 12))

amse_minshr <- sapply(1:ncol(cs_minshr$recf),
  function(i) amse(actual = hts_
    test[1:h,i], predicted = cs_
    minshr$recf[,i], h = t, train
    = hts_train[,i], s = 12))

MASEByLevel <- rbind(MASEByLevel, c(mase_minshr[1],
  mean(mase_minshr[2:8]), mean(mase_minshr[9:29]),
  mean(mase_minshr[30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_minshr[1],
  mean(amse_minshr[2:8]), mean(amse_minshr[9:29]),
  mean(amse_minshr[30:105])))

```

```

#MINT SHRINK
M_basef %>%
  htsrec(., comb = "shr", C = C, res = RES) -> cs_
  minshr

mase_minshr <- sapply(1:ncol(cs_minshr$recf),
  function(i) mase(actual = hts_
    test[1:h,i],
    predicted = cs_minshr$recf[,i],
    h = t,
    train = hts_train[,i], s = 12))

amse_minshr <- sapply(1:ncol(cs_minshr$recf),
  function(i) amse(actual = hts_
    test[1:h,i],
    predicted = cs_minshr$recf[,i],
    h = t,
    train = hts_train[,i], s = 12))

MASEByLevel <- rbind(MASEByLevel, c(mase_minshr[1],
  mean(mase_minshr[2:8]), mean(mase_
    minshr[9:29]),
  mean(mase_minshr[30:105])))
AMSEByLevel <- rbind(AMSEByLevel, c(amse_minshr[1],
  mean(amse_minshr[2:8]), mean(amse_
    minshr[9:29]),
  mean(amse_minshr[30:105])))
}

#A questo punto si ritorna le matrice con i valori del
MASE e AMSE e puo' iniziare la fase di
interpretazione dei risultati

```



## 6.3 Ripartizione geografica dell'Australia

TABELLA 6.1: Wickamasurya et al. (2019) [5]. Le righe non numerate sono escluse dal dataset.

Livello serie	Nome	Etichetta	Livello serie	Nome	Etichetta
<i>L0: Totale</i>			<i>Continuazione</i>		
1	Australia	Totale	49	Gippsland	BCB
<i>L1: Stati</i>			50	Phillip Island	BCC
2	NSW	A	51	Central Murray	BDA
3	VIC	B	52	Goulburn	BDB
4	QLD	C	53	High Country	BDC
5	SA	D	54	Melbourne East	BDD
6	WA	E	55	Upper Yarra	BDE
7	TAS	F	56	Murray East	BDF
8	NT	G	57	Malle	BEA
<i>L2: Zone</i>			58	Wimmera	BEB
9	Metro NSW	AA	59	Western Grampians	BEC
10	Nth Coast NSW	AB	60	Bendigo Loddon	BED
	Sth Coast NSW	AC	61	Macedon	BEE
11	Sth NSW	AD	62	Spa Country	BEF
12	Nth NSW	AE	63	Ballarat	BEG
	ACT	AF	64	Central Highlands	BEH
13	Metro VIC	BA	65	Gold Coast	CAA
	West Coast VIC	BB	66	Brisbane	CAB
14	East Coast VIC	BC	67	Sunshine Coast	CAC
15	Nth East VIC	BD	68	Central Queensland	CBA
16	Nth West VIC	BE	69	Bundaberg	CBB
17	Metro QLD	CA	70	Fraser Coast	CBC
18	Central Coast QLD	CB	71	Mackay	CBD
19	Nth Coast QLD	CC	72	Whitsundays	CCA
20	Inland QLD	CD	73	Northern	CCB
21	Metro SA	DA	74	Tropical North Queensland	CCC
22	Sth Coast SA	DB	75	Darling Downs	CDA
23	Inland SA	DC	76	Outback	CDB
24	West Coast SA	DD	77	Adelaide	DAA
25	West Coast WA	EA	78	Barossa	DAB
	Nth WA	EB	79	Adelaide Hills	DAC
	Sth WA	EC	80	Limestone Coast	DBA
	Sth TAS	FA	81	Fleurieu Peninsula	DBB
26	Nth East TAS	FB	82	Kangaroo Island	DBC
27	Nth West TAS	FC	83	Murraylands	DCA
28	Nth Coast NT	GA	84	Riverland	DCB
29	Central NT	GB	85	Clare Valley	DCC
<i>L3: Regioni</i>			86	Flinders Range and Outback	DCD
30	Sydney	AAA	87	Eyre Peninsula	DDA
31	Central Coast	AAB	88	Yorke Peninsula	ddb
32	Hunter	ABA	89	Australia's Coral Coast	EAA
33	North Coast NSW	ABB	90	Experience Perth	EAB
34	South Coast	ACA	91	Australia's South West	EAC
35	Snowy Mountains	ADA	92	Australia's North West	EBA
36	Capital Country	ADB	93	Australia's Golden Outback	ECA
37	The Murray	ADC	94	Hobart and the South	FAA
38	Riverina	ADD	95	East Coast	FBA
39	Central NSW	AEA	96	Launceston, Tamar and the North	FBB
40	New England North West	AEB	97	North West	FCA
41	Outback NSW	AEC	98	Wilderness West	FCB
42	Blue Mountains	AED	99	Darwin	GAA
43	Canberra	AFA	100	Kakadu Arnhem	GAB
44	Melbourne	BAA	101	Katherine Daly	GAC
45	Peninsula	BAB	102	Barkly	GBA
46	Geelong	BAC	103	Lasseter	GBB
47	Western	BBA	104	Alice Springs	GBC
48	Lakes	BCA	105	MacDonnell	GBD



## Bibliografia

- [1] R. J. Hyndman e A. J. Lee e E. Wang. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics Data Analysis*, 97:16–32, 2016. ISSN 0167-9473. doi: <https://doi.org/10.1016/j.csda.2015.11.007>. URL <https://www.sciencedirect.com/science/article/pii/S016794731500290X>.
- [2] R. Hollyman e F. Petropoulos e M. E. Tipping. Understanding forecast reconciliation. *European Journal of Operational Research*, 294(1):149–160, 2021. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2021.01.017>. URL <https://www.sciencedirect.com/science/article/pii/S0377221721000199>.
- [3] N. Kourentzes e G. Athanasopoulos. Cross-temporal coherent forecasts for australian tourism. *Annals of Tourism Research*, 75:393–409, 2019. ISSN 0160-7383. doi: <https://doi.org/10.1016/j.annals.2019.02.001>. URL <https://www.sciencedirect.com/science/article/pii/S0160738319300167>.
- [4] R. Hyndman e G. Athanasopoulos e C. Bergmeir e G. Caceres e L. Chhay e M. O’Hara-Wild e F. Petropoulos e S. Razbash e Earo Wang e F. Yasmeeen. *forecast: Forecasting functions for time series and linear models*, 2022. URL <https://pkg.robjhyndman.com/forecast/>. R package version 8.16.
- [5] S. L. Wickramasuriya e G. Athanasopoulos e R. J. Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819, 2019. doi: [10.1080/01621459.2018.1448825](https://doi.org/10.1080/01621459.2018.1448825). URL <https://doi.org/10.1080/01621459.2018.1448825>.
- [6] Lawson C. L. e Hanson R.J. *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974. ISBN 0-13-822585-0.
- [7] C. W. Gross e J. E. Sohl. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9(3):233–254, 1990. doi: <https://doi.org/10.1002/for.3980090304>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980090304>.
- [8] B. J. Dangerfield e J. S. Morris. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, 8(2):233–241, 1992. ISSN 0169-2070. doi: [https://doi.org/10.1016/0169-2070\(92\)90121-O](https://doi.org/10.1016/0169-2070(92)90121-O). URL <https://www.sciencedirect.com/science/article/pii/0169207092901210>.
- [9] E. Spiliotis e M. Abolghasemi e R. J. Hyndman e F. Petropoulos e V. Assimakopoulos. Hierarchical forecast reconciliation with machine learning. *Applied Soft*

- Computing*, 112:107756, 2021. ISSN 1568-4946. doi: <https://doi.org/10.1016/j.asoc.2021.107756>. URL <https://www.sciencedirect.com/science/article/pii/S1568494621006773>.
- [10] R. Koenker e Pin Ng. SparseM: A sparse matrix package for r. *The R Manuals*, 2003. URL <https://cran.r-project.org/web/packages/SparseM>.
- [11] R. J. Hyndman e R. A. Ahmed e G. Athanasopoulos e Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics Data Analysis*, 55(9):2579–2589, 2011. ISSN 0167-9473. doi: <https://doi.org/10.1016/j.csda.2011.03.006>. URL <https://www.sciencedirect.com/science/article/pii/S0167947311000971>.
- [12] G. Athanasopoulos e R. A. Ahmed e R. J. Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166, 2009. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2008.07.004>. URL <https://www.sciencedirect.com/science/article/pii/S0169207008000691>.
- [13] S. L. Wickramasuriya e R. J. Hyndman e B. A. Turlach. Optimal non-negative forecast reconciliation. *Statistics and Computing*, 30:1167–1182, 2020. doi: <https://doi.org/10.1007/s11222-020-09930-0>. URL <https://link.springer.com/article/10.1007/s11222-020-09930-0>.
- [14] G. Athanasopoulos e R. J. Hyndman e N. Kourentzes e Fotios Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, 2017. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2017.02.046>. URL <https://www.sciencedirect.com/science/article/pii/S037722171730191>.
- [15] A. B. Schwarzkopf e R. J. Teresine e J. S. Morris. Top-down versus bottom-up forecasting strategies. *International Journal of Production Research*, 26(11):1833–1843, 1988. doi: [10.1080/00207548808947995](https://doi.org/10.1080/00207548808947995). URL <https://doi.org/10.1080/00207548808947995>.
- [16] E. Shlifer e R. W. Wolff. Aggregation and proration in forecasting. *Management Science*, 25(6):594–603, 1979. doi: [10.1287/mnsc.25.6.594](https://doi.org/10.1287/mnsc.25.6.594). URL <https://doi.org/10.1287/mnsc.25.6.594>.
- [17] S. Bonettini e R. Zanella e L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 25(1):015002, nov 2008. doi: [10.1088/0266-5611/25/1/015002](https://doi.org/10.1088/0266-5611/25/1/015002). URL <https://doi.org/10.1088/0266-5611/25/1/015002>.
- [18] D. Girolimetto e T. Di Fonzo. *FoReco: Point Forecast Reconciliation*, 2022. URL <https://cran.r-project.org/package=FoReco>.

- 
- [19] C. Lanczos. Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49:33–53, 1952. URL <http://dx.doi.org/10.6028/jres.049.006>.
- [20] L. J. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450, 2000. ISSN 0169-2070. doi: [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0). URL <https://www.sciencedirect.com/science/article/pii/S0169207000000650>. The M3-Competition.