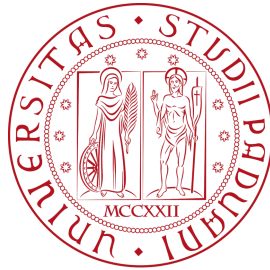


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in

Statistica per l'Economia e l'Impresa



## **Combinazione di previsioni: un confronto tra approcci differenti**

Relatrice: Prof.ssa Luisa Bisaglia  
Dipartimento di Scienze Statistiche

Laureanda: Margherita Palomba  
Matricola n. 1219560

Anno Accademico 2021/2022



# Indice

<b>Elenco degli acronimi</b>	<b>4</b>
<b>1 Introduzione</b>	<b>7</b>
1.1 La combinazione previsiva . . . . .	7
1.2 Combinare per l'adattamento . . . . .	9
1.3 Combinare per il miglioramento . . . . .	10
1.4 I due obiettivi della combinazione previsiva . . . . .	11
<b>2 Adattamento o Miglioramento</b>	<b>13</b>
2.1 Definizioni . . . . .	14
2.2 AFTER come rappresentante dei metodi CFA . . . . .	15
2.3 CFA vs. CFI . . . . .	17
2.3.1 Simulazioni illustrative . . . . .	17
2.3.2 Esempi teorici . . . . .	19
<b>3 Il metodo AI-AFTER</b>	<b>23</b>
3.1 La procedura del test statistico . . . . .	24
3.1.1 Come scegliere le combinazioni . . . . .	26
3.2 Utilizzo dell'algoritmo AFTER . . . . .	26
3.2.1 Una misura di salvaguardia . . . . .	28
3.3 Il limite del rischio previsivo . . . . .	28
<b>4 Studi di simulazione</b>	<b>31</b>
4.1 Esempi con la regressione lineare . . . . .	32
4.1.1 Scenario di adattamento . . . . .	32
4.1.2 Scenario di miglioramento . . . . .	33

4.2	Esempi con le serie storiche . . . . .	33
4.2.1	Scenario di adattamento . . . . .	34
4.2.2	Scenario di miglioramento . . . . .	34
4.3	Risultati e performance previsiva . . . . .	35
<b>5</b>	<b>Applicazione su dataset M4</b>	<b>41</b>
5.1	Dati e metodi previsivi . . . . .	42
5.2	Risultati . . . . .	43
<b>6</b>	<b>Conclusioni</b>	<b>47</b>
	<b>Ringraziamenti</b>	<b>49</b>
	<b>Bibliografia</b>	<b>51</b>

# Elenco degli acronimi

<b>BG o BG<sub>1</sub></b>	Metodo di Bates-Granger per la combinazione
<b>BMA</b>	Bayesian Model Averaging, media del modello bayesiano
<b>CFA</b>	Combining For Adaptament, combinazione per l'adattamento
<b>CFI</b>	Combining For Improvement, combinazione per il miglioramento
<b>CLR</b>	Conditional Linear Regression, regressione lineare condizionata
<b>D-M</b>	Test di Diebold - Mariano
<b>LR</b>	Linear Regression, regressione lineare
<b>MSFE</b>	Mean Squared Forecast Error, errore quadratico medio di previsione
<b>OLS</b>	Ordinary Least Squares, minimi quadrati ordinari
<b>PGD</b>	Processo Generatore dei Dati
<b>SA</b>	Simple average, media semplice



# Capitolo 1

## Introduzione

L'analista, al quale viene affidato il compito di trovare una soluzione ad un problema previsivo, spesso può avere a disposizione diverse previsioni per la stessa serie osservata a seconda del modello previsivo utilizzato. Tale modello può avere una struttura nota o può essere generato da un meccanismo sconosciuto. Per utilizzare queste previsioni, che concorrono tra di loro con il fine di prevedere il più accuratamente possibile i valori futuri della serie, l'analista può decidere o di selezionare una delle procedure previsive che sembra performare adeguatamente, o di combinare queste previsioni "candidate" in un qualche modo. Questo studio si pone l'obiettivo di esaminare a fondo questi due differenti approcci previsivi, per riuscire a confrontarli e a valutare oggettivamente quali metodi di combinazione delle previsioni sono in grado di raggiungere *performance* ottimali.

### 1.1 La combinazione previsiva

Molte ricerche mostrano che la combinazione delle previsioni migliora l'accuratezza rispetto all'utilizzo di singole previsioni e, di conseguenza, si ritiene che in molti contesti previsivi la combinazione porti a migliori previsioni finali. Ciò è stato, nel corso degli ultimi decenni, oggetto di numerosi studi, ed è stato dimostrato come combinare le previsioni porti ad un particolare vantaggio, che non è, come si potrebbe pensare, che la migliore tra le combinazioni possibili si comporti meglio della migliore tra le possibili previsioni individuali, ma che, nella pratica, combinare le previsioni ha un rischio meno elevato rispetto a selezionare un metodo di previsione individuale (Hibon e Evgeniou,

2005). Molto impegno è stato profuso affinché la combinazione previsiva potesse includere sia l'approccio di tipo frequentista che l'approccio di tipo Bayesiano, cercando così di considerare la probabilità di un evento sia da un punto di vista oggettivo, tramite le informazioni contenute nel campione osservato, sia da un punto di vista soggettivo, affidandosi ad un'informazione contenuta "a priori" nella serie. Molti metodi frequentisti progettano ed impiegano criteri basati sulle *performance* dei previsori per stimare pesi teoricamente corretti così da riuscire, tramite una combinazione lineare, a raggiungere *performance* previsive potenzialmente superiori a quelle di qualsiasi altra previsione proveniente da un singolo modello; esempi nella letteratura previsiva più recente sono i lavori di Cheng e Hansen (2015) e Zhang et al. (2016, 2018).

La combinazione previsiva è stata introdotta da Bates e Granger (1969) come una più che valida alternativa all'utilizzo di un unico metodo previsivo. In tale lavoro gli autori proponevano di impiegare il criterio di minimizzazione dell'errore quadratico medio per stimare i pesi ottimali per la combinazione tramite la matrice di varianza dell'errore di previsione. Successivamente, anche Granger e Ramanathan (1984) formularono diversi *frameworks* di regressione lineare per la ricerca di pesi ottimali e con il tempo furono introdotti altri metodi basati sull'idea di trovare un metodo di combinazione in grado di portare a migliori *performance* predittive, come, per esempio, citando alcuni metodi computazionali utilizzati anche nel campo previsivo, *bootstrapping*, *bagging*, *stacking* e *boosting* (Breiman, 1996, Breiman, 1998; LeBlanc e Tibshirani, 1996; Schapire et al., 1998).

Definire i criteri fondamentali per selezionare le previsioni da combinare, rimane una sfida per la ricerca nel campo della valutazione dei previsori. Questa selezione dei previsori implica il confronto di un insieme di previsioni con i loro corrispondenti valori effettivamente osservati *ex-post*, e il quantificare la perdita associata all'errore di previsione tramite una determinata funzione di perdita. È possibile utilizzare una varietà di misure di errore per valutare le prestazioni dei previsori, tra cui, ovviamente, l'errore quadratico medio di previsione (MSFE), l'errore medio assoluto (MAE), l'errore assoluto medio percentuale (MAPE) ed un insieme di misure ad esse correlate (Armstrong e Collopy, 1992; Hyndman e Koehler, 2006; Makridakis e Hibon, 2000). Con la popolarità dei metodi di combinazione basati sulla *performance* previsiva, è stato portato alla luce che i pesi stimati con metodi più sofisticati possono spesso deviare dall'obiettivo di trovare pesi teoricamente ottimali a causa dell'errore di stima e dell'incertezza che esso



comporta (e.g., Claeskens et al., 2016; Smith e Wallis, 2009); questo effetto importante, insieme ad altri dovuti alla discontinuità strutturale ed alle nuove informazioni (Lahiri et al., 2017; Qian et al., 2019), contribuisce al fenomeno chiamato *"forecast combination puzzle"* (Hendry e Clements, 2004; Stock e Watson, 2004) secondo il quale, nella pratica, la *performance* di una semplice media aritmetica o delle sue varianti può superare quella di alternative più sofisticate. Infatti, è stato evidenziato (Genre et al., 2013) che esiste solo un modesto numero di casi in cui si debba considerare combinazioni diverse dalla media semplice per sintetizzare le informazioni contenute nei dati e che l'aumento di variabilità delle stime suggerisce di scegliere con attenzione altri metodi.

## 1.2 Combinare per l'adattamento

Diversamente dalle strategie di combinazione che mirano direttamente a trovare i pesi ottimali per ottenere risultati potenzialmente migliori rispetto a quelli delle singole previsioni provenienti da modelli diversi, esiste una classe di metodi che aggiorna ricorsivamente i pesi utilizzati per la combinazione tramite schemi di riponderazione (ad esempio Lahiri et al., 2017; Yang, 2004). Questi metodi, in genere, hanno un obiettivo meno ambizioso ed un unico scopo: eguagliare la *performance* della migliore previsione candidata, con la promessa di ottenere le stime dei pesi con un costo minore. Questi procedimenti possono quindi essere interpretati come una selezione dinamica del modello, dove la combinazione tende a mettere tutto il peso sul miglior predittore disponibile, ogni volta che esso viene individuato. Questa classe di metodi di combinazione è chiamata *combinazione per l'adattamento* (CFI, combining for adaptament).

Yang (2004) ha sviluppato uno di questi metodi chiamato *AFTER* (*Aggregated Forecast Through Exponential Reweighting*), che sarà discusso approfonditamente nel Capitolo 2, e ha mostrato il costo teoricamente maggiore del rischio di previsione dato dal tentativo di ottenere previsioni migliori rispetto a quando si mira solo a raggiungere del migliore tra i candidati originali. Lahiri et al. (2017) hanno mostrato che il metodo AFTER tende ad imporre al candidato con le migliori prestazioni un peso asintoticamente unitario, sotto alcune deboli condizioni. Oltre ai summenzionati approcci frequentisti, i metodi bayesiani per la media dei modelli (Forte et al., 2018; Hoeting et al., 1999a; Steel, 2011) sono in linea con lo scopo di adattarsi al miglior previsore tra i candidati originali, poiché spesso si assume che il processo di generazione dei dati corrisponda ad

uno tra i modelli previsionali candidati (De Luca et al., 2018), purchè questi modelli siano noti.

Sebbene diversi metodi di combinazione abbiano diversi obiettivi, poiché nella pratica non sono necessariamente disponibili né le informazioni sui processi generatori dei dati né i modelli statistici sottostanti le previsioni candidate, di solito non è noto a priori ad un analista se sia fattibile ottenere una *performance* migliore dalla combinazione delle previsioni, piuttosto che selezionandone una dall'insieme delle candidate originarie. Ciò potrebbe portare ad una scelta impropria dei metodi di combinazione e, di conseguenza, a prestazioni previsive indesiderate; ad esempio, applicando ciecamente il metodo AFTER ci aspettiamo che questo performerà peggio se esiste una combinazione lineare, ottenuta tramite un metodo di ottimizzazione dei pesi, che è più performante rispetto a ognuno dei candidati originari.

### 1.3 Combinare per il miglioramento

Per uno sperimentatore può essere quindi meno rischioso utilizzare, ad esempio, una media dei modelli previsivi candidati, piuttosto che fare affidamento su un singolo modello; partendo dall'idea che ogni modello fornisce informazioni sui parametri di interesse, si può calcolare una media ponderata delle stime condizionate di tutti i possibili modelli per combinare le previsioni originali. Nonostante i problemi riguardo quale metodo di combinazione utilizzare e come svilupparne di nuovi per ottenere prestazioni migliori, è stato confermato dalla ricerca empirica e teorica che una combinazione delle previsioni candidate disponibili può portare ad importanti benefici e spesso produce risultati migliori rispetto alla selezione di una singola previsione candidata (ad es. Kourentzes et al., 2019; Stock e Watson, 2003; Yang, 2003). In particolare ci tengo a menzionare la recente *forecast competition*, denominata la *M4-Competition* (Makridakis et al., 2020). Questa competizione aveva come scopo quello di imparare dall'evidenza empirica sia come migliorare l'accuratezza delle previsioni, sia come l'utilizzo di tale apprendimento possa far avanzare la teoria e la pratica della previsione. I risultati della gara hanno mostrato che la stragrande maggioranza dei metodi previsivi più accurati sono combinazioni, ed è, quindi, molto probabile, che le combinazioni previsive diventeranno con il tempo i metodi di previsione predefiniti e *benchmarks* nella letteratura previsionale (Radchenko et al., 2021).

A contribuire in particolar modo ai benefici della combinazione concorrono due ragioni. La prima è che identificare quale procedura di previsione sia la migliore tra quelle candidate comporta spesso una sostanziale incertezza, infatti, a seconda del rumore presente nella serie nel periodo di valutazione delle previsioni, sono diversi i candidati che possono avere buone possibilità di essere selezionati come "il migliore" durante la procedura di selezione, e spesso questo approccio porta a previsioni post-selezione con varianza elevata. In secondo luogo, potrebbero essere disponibili diverse informazioni predittive per diversi previsori; in questa situazione, una combinazione delle previsioni candidate ha il potenziale per avere delle *performance* migliori anche di quelle della miglior procedura tra quelle candidate originariamente, grazie alla condivisione delle informazioni tra le previsioni. Esiste, infatti, la possibilità che nessuno tra i previsori originali performi adeguatamente, e che per questo vengano denominati, in alcuni casi, "*weak learners*", ma che sia comunque possibile combinar loro per trovare un'unica previsione che abbia persino prestazioni migliori di quelle del miglior candidato originale. Yang (2004) ha formalmente distinto questo obiettivo, denominandolo *combinazione per il miglioramento* (CFA, *combining for improvement*), da quello precedentemente discusso.

## 1.4 I due obiettivi della combinazione previsiva

L'obiettivo della combinazione per l'adattamento è quello di ottenere la migliore prestazione tra quelle delle procedure candidate originariamente, riducendo la varianza introdotta dall'incertezza della selezione. Combinare per il miglioramento, d'altra parte, mira a ottenere *performance* più elevate anche della procedura migliore tra le candidate originarie cercando una combinazione ottimale di queste e stimando direttamente i pesi ottimali utilizzando criteri basati sulle prestazioni. A seconda della natura dei dati che generano il processo e delle previsioni candidate, uno dei due obiettivi può essere preferibile all'altro. Adottare un approccio per l'adattamento quando il miglioramento è più appropriato, può portare ad evitare di condividere informazioni potenzialmente utili e non riuscire così ad ottenere previsioni migliori rispetto a quelle del candidato previsivo originale; d'altra parte, adottare un approccio per il miglioramento quando l'adattamento è più appropriato può comportare un aumento del rischio previsivo a causa del più pesante costo del metodo CFI, dovuto al suo più ambizioso obiettivo. Nei prossimi capitoli saranno affrontati degli esempi sia teorici che pratici per comprendere meglio in quali casi perseguire questi due determinati obiettivi.

Con lo scopo di individuare una procedura basata su un test che fosse in grado di valutare se sia possibile sovra-performare le prestazioni del migliore previsore individuale e quindi evidenziare il collegamento e la relazione tra CFA e CFI, Qian et al. (2022) hanno ideato un nuovo metodo per riuscire a cogliere i vantaggi della combinazione in entrambi i casi, avendo i dati e le previsioni candidate a portata di mano. Nello specifico la loro proposta si basa sul metodo AFTER (Yang, 2004), ideato per perseguire l'obiettivo di combinare per l'adattamento e che garantisce di performare tanto bene quanto (piuttosto che significativamente meglio di) il miglior candidato previsivo, senza sapere in anticipo quale esso sia.

Viene quindi da loro proposta un'estensione del metodo AFTER in grado di combinare sia per l'adattamento (A, *adaptement*) che per il miglioramento (I, *improvement*), così che questo algoritmo si possa adattare anche al secondo obiettivo quando questa strategia risulta essere significativamente migliore. Tale nuovo metodo è quindi chiamato AI-AFTER ed è stato anche implementato dagli autori in un pacchetto R chiamato *AIafter*, disponibile su <https://github.com/weiqian1/AIafter>. Il loro lavoro, focalizzato sull'affrontare i due diversi obiettivi di CFA e CFI per la combinazione di previsioni puntuali, rappresenta il cuore di questa tesi, in quanto costituisce un metodo originale in grado di inglobare differenti approcci previsivi ed è di spiccato interesse verificarne l'effettiva efficacia.

Il resto della tesi è strutturato come segue. Nel Capitolo 2 saranno definiti formalmente gli obiettivi della combinazione per l'adattamento e per il miglioramento e presentate simulazioni illustrative ed esempi teorici che favoriscano uno degli obiettivi combinatori, così da comprendere meglio come opera l'algoritmo di AI-AFTER nella sua prima fase. Nel Capitolo 3 ci sarà una descrizione dettagliata del test statistico utilizzato nella prima fase dell'algoritmo per verificare la possibilità di combinare per il miglioramento e, successivamente, della seconda fase del metodo AI-AFTER per la combinazione delle previsioni sia per l'adattamento che per il miglioramento. Una volta spiegato dal punto di vista teorico, il metodo AI-AFTER verrà messo alla prova tramite degli studi di simulazione nel Capitolo 4 ed infine, tramite un'applicazione con dati reali di serie storiche selezionate dal *dataset* della Competizione M4, nel Capitolo 5. Nel Capitolo 6 verranno fornite delle osservazioni conclusive per evidenziare i risultati ottenuti nel corso del lavoro.

## Capitolo 2

# Adattamento o Miglioramento

I due obiettivi di combinare per l'adattamento o per il miglioramento (CFA vs CFI) possono essere spiegati meglio tramite il rischio previsivo che mirano a raggiungere. L'accuratezza delle previsioni è un fattore critico per, tra le altre cose, ridurre i costi e fornire un migliore servizio per chi ne farà uso (Makridakis e Hibon, 2000). Le informazioni disponibili per migliorare tale accuratezza in specifiche situazioni, non sono ancora utilizzate come sarebbe opportuno e la conseguenza che ne deriva è una perdita di opportunità che a volte si può considerare molto grave. Studi empirici nel campo della previsione hanno confrontato l'accuratezza delle previsioni *post-sample* dei vari metodi in modo tale che le loro *performance* potessero essere determinate in maniera quantificabile ed oggettiva. In particolare, anche le M-Competition si basano su questo tipo di studi empirici, confrontando le *performance* di un grandissimo numero di metodi per il trattamento delle serie storiche e quelle delle loro combinazioni.

In una loro pubblicazione, Hibon e Evgeniou (2005) evidenziano e dimostrano quattro importanti ipotesi riguardanti la combinazione di previsioni: quando la stessa combinazione (o singolo predittore) viene utilizzata per tutte le serie prese in esame, il migliore tra tutti i metodi di combinazione non è mediamente più accurato del miglior metodo individuale; quando una combinazione diversa (o singolo predittore) viene utilizzata per ciascuna delle serie prese in esame, il migliore tra tutti i metodi di combinazione non è mediamente più accurato del miglior metodo individuale; la *performance* predittiva della combinazione selezionata non è migliore di quella del metodo individuale selezionato; il rischio, misurato come differenza di *performance post-sample* tra il modello selezionato e la migliore possibilità predittiva, nel caso di selezione di un singolo metodo non è su-

periore al rischio nel caso di selezione di una combinazione di previsioni. D'altra parte, quando non si conosce a priori quale sia il metodo di previsione individuale migliore, la selezione di una combinazione, nella pratica, porta ad una scelta che ha mediamente prestazioni significativamente migliori rispetto a quella che potrebbe avere un metodo individuale. Quindi, il vantaggio delle combinazioni non è che le migliori tra esse sono mediamente migliori delle migliori previsioni individuali, ma che la selezione di una combinazione è meno rischiosa della selezione di una singola previsione.

Verrà in seguito dimostrato che il rischio che si pone come obiettivo della combinazione per il miglioramento (ossia il rischio ottenuto usando i pesi ottimali per la combinazione) è sempre limitato superiormente dal rischio obiettivo della combinazione per l'adattamento (ossia il rischio del miglior candidato previsivo tra i candidati originali). D'altra parte, poichè nel rischio previsivo può essere aggiunto un costo extra più alto quando si combina per il miglioramento, l'adattamento può essere preferibile se ciò che si guadagna perseguendo l'obiettivo di sovraperformare il miglior candidato originale non è significativamente conveniente.

## 2.1 Definizioni

Il nostro obiettivo è quello di fare previsioni su una variabile casuale continua  $Y$ , i cui valori sono  $Y_1, Y_2, \dots$ , ai tempi  $i = 1, 2, \dots$  rispettivamente. Una procedura generale di previsione  $\delta$  è un meccanismo che produce le previsioni  $\hat{y}_{\delta,i}$  per  $i \geq 1$ . La procedura  $\delta$  può essere costruita sia su un modello statistico sia su altre informazioni, ma qui assumeremo che nè il modello nè altre informazioni riguardanti la procedura siano note. Per il problema di combinazione, iniziamo con una collezione di  $M$  procedure previsive  $\Delta = \{\delta_1, \delta_2, \dots, \delta_M\}$ , che corrispondono alle procedure previsive originariamente candidate.

Una procedura previsiva  $\psi$  è invece chiamata una combinazione di previsioni basata su  $\Delta$  se ognuna delle previsioni  $\hat{y}_{\psi,i}$  è una funzione misurabile di  $Y_1, \dots, Y_{i-1}$  e di  $\hat{y}_{\delta_j,l}$  per  $1 \leq l \leq i$  e  $1 \leq j \leq M$ . Sia  $\Psi$  una classe di procedure previsive combinatorie basata su  $\Delta$ . Data una  $\psi \in \Psi$ , sia  $R(\psi; n)$  il rischio previsivo derivato dalla funzione di perdita quadratica, quindi  $R(\psi; n) = \sum_{i=1}^n E(Y_i - \hat{y}_{\psi,i})^2$ . Denotiamo  $\psi^* = \psi^*(\Psi, n) = \operatorname{argmin}_{\psi \in \Psi} R(\psi; n)$  la scelta che minimizza il rischio previsivo  $R(\psi; n)$  tra tutte le  $\psi \in \Psi$ , e sia  $R(\Psi; n) = R(\psi^*; n)$  il minimo rischio ottenibile dalla classe  $\Psi$ .

**Definizione 1.** Sia  $\Psi_0 = \Delta$  l'insieme delle previsioni dei candidati originali. Un metodo di combinazione delle previsioni che combina per l'adattamento (CFA) mira a raggiungere il rischio previsivo  $R(\Psi_0; n)$  e ambisce a performare, in termini di rischio previsivo, all'incirca tanto bene quanto il miglior candidato originale contenuto in  $\Psi_0$ .

**Definizione 2.** Sia  $\Psi_L$  l'insieme di tutte le combinazioni lineari delle previsioni dei candidati originali. In altre parole, ogni membro  $\psi$  di  $\Psi_L$  produce previsioni aventi la forma:

$$\hat{y}_{\psi,i} = \sum_{j=1}^M w_{i,j} \hat{y}_{\delta_j,i}, \quad i \geq 1 \quad (2.1)$$

dove i vari  $w_{i,j}$  sono i pesi (o coefficienti della combinazione). Un metodo di combinazione previsiva che combina per il miglioramento (CFI) mira a raggiungere il rischio previsivo  $R(\Psi_L; n)$  e stima i pesi ottimali per la combinazione per minimizzare il rischio previsivo tramite dei criteri basati sulla performance previsiva; così un metodo di combinazione per il miglioramento ambisce a performare, in termini di rischio previsivo, all'incirca tanto bene quanto la combinazione lineare ottimale dei candidati originali.

È anche possibile costruire classi ulteriormente allargate estendendo l'insieme di previsioni candidate da  $\Psi_0$  a  $\Psi_L$ ; per esempio, uno potrebbe considerare ancora altre combinazioni lineari (Wang et al., 2014) o combinare le previsioni non linearmente.

La maggior parte dei metodi di combinazione previsiva hanno ambito ad oggi a combinare per il miglioramento, a partire dal lavoro seminale di Bates e Granger (1969), che aspiravano a derivare i pesi ottimali per la particolare situazione in cui si avevano due previsioni candidate nella 2.1, e il lavoro di Granger e Ramanathan (1984), che invece prendeva in considerazione situazioni più generali per la stima dei pesi ottimali tramite la regressione lineare. Molte ricerche successive hanno tentato di derivare i pesi ottimali per la combinazione 2.1 (e.g. Claeskens et al., 2016; Hansen, 2008; Hsiao e Wan, 2014; come molti altri), e ciò che tutti questi metodi hanno in comune è l'obiettivo di raggiungere  $R(\Psi_L; n)$ , ossia il minimo rischio possibile in  $\Psi_L$ , sotto determinate condizioni.

## 2.2 AFTER come rappresentante dei metodi CFA

I metodi che combinano per l'adattamento tipicamente non usano esplicitamente la covarianza tra le previsioni, ma costruiscono i pesi della combinazione basandosi sui "meriti" di ogni candidato come predittore. La media del modello bayesiano (BMA, *bayesian*

*model averaging*; e.g. Hoeting et al., 1999b) è un esempio di un metodo di combinazione per l'adattamento in ambito parametrico. Il metodo AFTER, sviluppato da Yang, 2004, è un metodo flessibile utilizzato per combinare diversi tipi di previsioni con lo scopo di adattamento e che, per la stima dei pesi, guarda tutte le *performance* passate di ciascuna previsione candidata. Nel resto del lavoro AFTER verrà utilizzato come il metodo rappresentativo per la combinazione per l'adattamento e verrà descritto brevemente l'algoritmo utilizzato per la stima dei suddetti pesi. In particolare, seguendo le precedenti notazioni, supponiamo che le procedure candidate siano  $\{\delta_1, \delta_2, \dots, \delta_M\}$ . Nell'istante temporale  $i$ , il peso assegnato dall'algoritmo a  $\delta_j$  è generato basandosi sulle sue precedenti performance predittive e corrisponde a:

$$w_{i,j} = \frac{\left( \prod_{t=1}^{i-1} -\hat{\sigma}_{j,t} \right) \exp\left(-\lambda \sum_{t=1}^{i-1} \phi\left(\frac{Y_i - \hat{y}_{\delta_j,t}}{\hat{\sigma}_{j,t}}\right)\right)}{\sum_{j'=1}^M \left( \prod_{t=1}^{i-1} \hat{\sigma}_{j',t} \right) \exp\left(-\lambda \sum_{t=1}^{i-1} \phi\left(\frac{Y_i - \hat{y}_{\delta_{j'},t}}{\hat{\sigma}_{j',t}}\right)\right)} \quad (2.2)$$

dove  $\hat{\sigma}_{j,t}$  corrisponde alla deviazione standard campionaria degli errori previsivi di  $\delta_j$  precedenti rispetto all'osservazione di  $Y_i$ , e  $\phi(\cdot)$  è la funzione di perdita dell'errore quadratico  $\phi(x) = x^2$ . Le previsioni candidate verranno poi combinate con la seguente formula:  $\hat{y}_n^* = \sum_{j=1}^M w_{n,j} \hat{y}_{n,j}$ . Si noti inoltre che tutti i pesi  $w_{n,j}$  all'istante temporale  $n \geq 1$  devono sempre sommare ad 1 e dipendono solo dalle previsioni e realizzazioni passate della serie osservata. Inoltre, va anche evidenziato come, per ogni osservazione aggiuntiva, i pesi nelle previsioni candidate vengano aggiornati ricorsivamente.

Come spiegato nella Definizione 1, non è necessario che un metodo di combinazione previsiva abbia tutti i pesi, tranne uno, uguali esattamente a zero, ma che la *performance* di una combinazione previsiva per l'adattamento in termini di rischio sia almeno buona quanto quella del miglior candidato contenuto in  $\Psi_0 = \Delta$ . In questo senso, il metodo AFTER precedentemente descritto è stato progettato per l'adattamento: il rischio delle previsioni calcolate con AFTER può essere esplicitamente confrontato con  $R(\Psi_0; n)$ , il che mostra che, sotto alcune deboli condizioni, il rischio di AFTER è sicuramente limitato superiormente da  $R(\Psi_0; n)$ , con una piccola penalità additiva. All'interno della famiglia delle CFA, la strategia è quella di assegnare al miglior candidato (o ai migliori candidati) il peso maggiore e così avvicinarsi alla *performance* del candidato più forte, poichè le informazioni contenute nella sua previsione vengono considerate quelle più attendibili (Lahiri et al., 2017).



## 2.3 CFA vs. CFI

Ricordando che  $R(\Psi_L; n) \leq R(\Psi_0; n)$ , combinare per il miglioramento riesce a raggiungere un minor rischio ed ha, quindi, un obiettivo più ambizioso rispetto a combinare per l'adattamento; tuttavia,  $R(\Psi_L; n)$  è anche più difficile da raggiungere, portando quindi la combinazione per il miglioramento a dover sopportare un costo extra più alto della combinazione per l'adattamento. Sebbene le prime ricerche nel campo fossero riuscite a derivare dei pesi ottimali per la combinazione 2.1 sotto determinate condizioni, la difficoltà e l'instabilità della stima dei pesi possono portare a *performance* empiriche sotto-ottimali per i metodi che perseguono l'obiettivo del miglioramento (Smith e Wallis, 2009; Claeskens et al., 2016). In aggiunta, quando una o più delle previsioni candidate originariamente carpiscono correttamente il reale processo generatore dei dati, o generano, comunque, previsioni molto accurate, il costo che viene pagato effettuando la ricerca nella classe più grande  $\Psi_L$  sarà maggiore rispetto a ciò che si guadagna ottenendo  $R(\Psi_L; n)$  piuttosto che  $R(\Psi_0; n)$ . Nella prossima Sezione (2.3.1) verranno presentate due simulazioni per un'illustrazione empirica, seguite da due esempi teorici nella Sezione 2.3.2 con ulteriori spiegazioni.

### 2.3.1 Simulazioni illustrative

Consideriamo due simulazioni illustrative che ci aiutino ad apprezzare le differenti *performance* dei metodi CFA e CFI.

**Simulazione 1.** Consideriamo come processo generatore dei dati per  $Y$ , un modello di regressione lineare con tre variabili:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \quad (2.3)$$

dove  $\epsilon_i, X_{i1}, X_{i2}$  e  $X_{i3}$  sono variabili  $N(0, 1)$  indipendenti ed identicamente distribuite (i.i.d.) e i veri parametri del modello sono  $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, b_0, b_0, b_0)$ . Consideriamo i tre seguenti modelli per generare le previsioni candidate, dove le previsioni di (2.4) e di (2.5) non aggiungono alcuna informazione aggiuntiva sul processo generatore dei dati rispetto al modello (2.6):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \quad (2.4)$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \quad (2.5)$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \quad (2.6)$$

Per verificare se questo scenario favorisca o meno la combinazione per l'adattamento, si effettua una simulazione per valutare la performance di AFTER, che è progettato per CFA. Per il confronto, verranno anche considerati i metodi di Bates-Granger (BG o BG<sub>1</sub>) e la regressione lineare (LR), che invece sono progettati per CFI con il fine di stimare i pesi ottimali (Bates e Granger, 1969; Granger e Ramanathan, 1984). Per il vero parametro  $b_0$  del processo generatore di dati (2.3) verrà utilizzata la sequenza decrescente di valori che spaziano tra il 2.0 e lo 0.2 ad intervalli regolari di 0.2, così da avere un totale di 10 diverse serie simulate. Brevemente, ogni serie avrà un totale di  $n = 70$  osservazioni, di cui le prime  $n_{train} = 50$  sono utilizzate solo per la stima dei parametri iniziali dei modelli dei previsori candidati, e le successive  $n_{eval} = 20$  per quantificare il rischio previsivo. Ogni metodo combinatorio viene valutato usando l'errore quadratico medio di previsione (MSFE, mean squared forecast error) relativo al MSFE della media aritmetica semplice (SA, simple average) dei modelli previsivi, così che il MSFE della SA corrisponda sempre al valore di base, ossia 1.0 (mostrato dalla linea tratteggiata nelle Figure 2.1 e 2.2). Gli MSFE medi relativi ottenuti dalle simulazioni sono sintetizzati in Figura 2.1, dove si può vedere che AFTER performa più favorevolmente di LR qualsiasi sia il vero valore di  $b_0$  e di BG quando  $b_0$  assume valori più elevati, quindi, in questo scenario, è appropriato mirare all'obiettivo di CFA. AFTER gode di un rischio previsivo significativamente minore rispetto alle alternative che combinano per il miglioramento quando  $b_0$  è relativamente grande ed ha una *performance* meno buona ma comunque confrontabile con BG e con SA quando il segnale diventa più debole e il rischio previsivo è dominato dalla varianza dell'errore casuale; d'altra parte, LR performa sempre peggio del metodo AFTER e ciò può essere attribuito all'alto costo derivato dalla stima degli errori e dall'incertezza della stima dei pesi ottimali.

**Simulazione 2.** Ora, invece, consideriamo un possibile scenario nel quale differenti previsioni candidate contengono diverse informazioni riguardo i dati. Il processo generatore di dati è lo stesso della Simulazione 1 (2.3), ma le previsioni candidate sono generate da modelli con variabili diverse:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \quad (2.7)$$

$$Y_i = \beta_0 + \beta_2 X_{i2} + \epsilon_i \quad (2.8)$$

$$Y_i = \beta_0 + \beta_3 X_{i3} + \epsilon_i \quad (2.9)$$

Per vedere se questo scenario favorisce la combinazione per il miglioramento, sarà svolto lo stesso esperimento e gli MSFE medi relativi saranno sintetizzati nella Figura 2.2. In

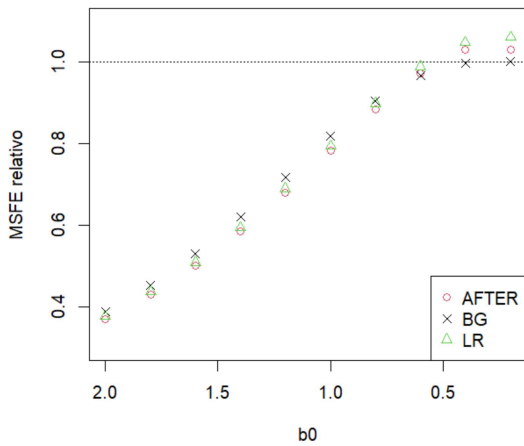


Figura 2.1: Performance previsa relativa dei metodi CFA vs. CFI, Simulazione 1.

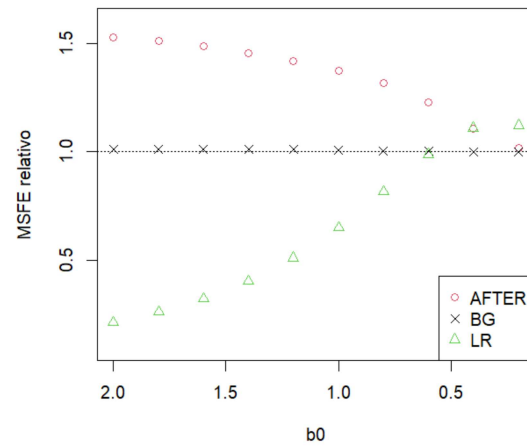


Figura 2.2: Performance previsa relativa dei metodi CFA vs. CFI, Simulazione 2.

netto contrasto con la Simulazione 1, AFTER performa scarsamente in confronto alle altre alternative CFI quando il valore  $b_0$  è relativamente grande; ciò non è sorprendente se si considera che il metodo AFTER è stato progettato per performare tanto bene quanto il miglior candidato previsivo, mentre, in questo scenario, una combinazione lineare ottimale dei tre candidati può ridurre di molto il rischio previsivo rispetto ai singoli candidati. Mentre BG ha all'incirca le stesse *performance* previsive di SA anche quando  $b_0$  assume valori più bassi, in questo caso LR performa nuovamente non adeguatamente, per via dell'instabile stima dei pesi, AFTER, invece, non sembra performare comunque significativamente meglio nè di BG nè di SA.

### 2.3.2 Esempi teorici

A partire dalle precedenti illustrazioni empiriche, che affrontano dei problemi ben specifici, ora saranno presentati due esempi teorici e di carattere più generale. Il primo includerà lo scenario dei modelli annidati, quindi la generalizzazione della Simulazione 1, mentre il secondo quello dei modelli detti "*weak learners*", ossia un insieme di predittori che individualmente non hanno una conoscenza adeguata del PGD, quindi la generalizzazione della Simulazione 2. Entrambi gli esempi saranno seguiti da una spiega-

zione riguardante quale sia l'obiettivo da favorire per la combinazione, a completamento delle osservazioni fatte nel corso del Capitolo.

**Esempio 1.** (*Modelli annidati*). In questo scenario le previsioni candidate sono quelle appartenenti alla stessa famiglia del modello originale, ma con diversi insiemi di variabili. Si consideri un processo generatore di dati dove:

$$Y_i = \beta_0 + \sum_{j=1}^{p_0} \beta_j X_{ij} + \epsilon_i \quad (2.10)$$

dove le  $\epsilon_i$  sono indipendenti ed identicamente distribuite (i.i.d.) con media 0 e sono disponibili  $p \geq p_0$  predittori. Una pratica comune di selezione delle variabili quando  $p$  è grande è quella di generare una sequenza di modelli annidati tramite una procedura *forward*. Si supponga che una procedura di selezione *forward* generi la seguente sequenza di modelli:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \\ &\dots \\ Y_i &= \beta_0 + \sum_{j=1}^{p_0} \beta_j X_{ij} + \epsilon_i \\ &\dots \\ Y_i &= \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \epsilon_i \end{aligned} \quad (2.11)$$

Quando  $n$  è abbastanza elevato, l'errore creato stimando i parametri del modello previsivo può considerarsi trascurabile e le previsioni generate dal modello (2.6) avranno un rischio pari a  $R(\Psi_0; n)$ . Inoltre, dato che nessuno degli altri modelli successivi nella sequenza 2.11 aggiunge ulteriori informazioni riguardo al PGD, ne deriva che  $R(\Psi_0; n) \simeq R(\Psi_L; n)$ . In questo caso, i metodi che combinano per l'adattamento e quelli che combinano per il miglioramento raggiungono lo stesso rischio, ma è ovviamente più complicato per i metodi per il miglioramento ottenere questo rischio a causa del costo extra per il bisogno di effettuare la ricerca del previsore migliore tra un insieme più grande di possibili previsioni combinate e per gli associati errori di stima dei pesi

ottimali (Yang, 2004). Questo esempio ci offre, quindi, uno scenario in cui è preferibile combinare per l'adattamento.

**Esempio 2.** ("*Weak learners*"). Si consideri un processo generatore di dati dove

$$Y_i = \beta_0 + \sum_{j \in \Omega} \beta_j X_{ij} + \epsilon_i \quad (2.12)$$

dove le  $\epsilon_i$  sono i.i.d. con media 0, sono disponibili  $p$  predittori e il vero insieme  $\Omega$  utilizzato di variabili è un sotto-insieme di  $\{1, \dots, p\}$ . Si supponga che le procedure previsive candidate abbiano la forma:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \epsilon_i \\ Y_i &= \beta_0 + \beta_2 X_{i2} + \epsilon_i \\ Y_i &= \beta_0 + \beta_3 X_{i3} + \epsilon_i \\ &\dots \\ Y_i &= \beta_0 + \beta_p X_{ip} + \epsilon_i \end{aligned} \quad (2.13)$$

Questa situazione può verificarsi se, per esempio, le  $p$  procedure previsive rappresentano previsioni con diversi insiemi informativi riguardanti il processo generatore di dati.  $X_{ip}$  può essere pensata come l'informazione disponibile al tempo  $i$  per il modello  $p$ . In questo scenario, un'appropriata combinazione delle procedure può potenzialmente performare molto meglio rispetto alla miglior procedura individuale; ciò significa che  $R(\Psi_L; n)$  può essere molto più basso di  $R(\Psi_0; n)$ . I metodi classici, il cui obiettivo è quello di combinare per il miglioramento e quindi ottenere una *performance* maggiore del miglior candidato individuale, diventano i favoriti in questi tipi di scenario.

Nei precedenti esempi, risulta chiaro quando CFA e CFI dovrebbero essere gli obiettivi favoriti data la conoscenza sia del processo generatore di dati che dei modelli previsivi candidati. Purtroppo però, senza questa conoscenza, nella pratica, può diventare complicato determinare quale sia il corretto obiettivo della combinazione. Mentre le informazioni e/o i modelli alla base di ciascuna delle procedure di previsione candidate possono essere sia conosciuti che non dall'analista, il PGD, in genere, non è mai noto. In aggiunta, l'insieme delle previsioni candidate a disposizione dall'analista potrebbe essere un mix di modelli statistici, algoritmi di *machine learning* e opinioni di esperti. In situazioni di questo tipo, non è chiaro quando sia meglio combinare per il miglioramento e quando per l'adattamento. Ad esempio, applicando a priori il metodo AFTER nell'Esempio 2

(e nella Simulazione 2), senza sapere che in quello scenario è favorito l'obiettivo del miglioramento, potrebbe portare a risultati previsionali sottoperformanti. Per affrontare il dilemma di quale sia l'obiettivo di combinazione favorevole nello scenario sottostante e, quindi, come scegliere un metodo di combinazione previsiva appropriato, nel Capitolo successivo verrà presentato un test statistico che misura l'evidenza del potenziale di combinazione per il miglioramento proposto da Qian et al. (2022). Le informazioni fornite da questo test consentiranno un'ulteriore comprensione del problema della combinazione e porteranno alla procedura di combinazione previsionale denominata AI-AFTER che performerà bene anche avendo solo a portata di mano i dati osservati e le previsioni candidate, senza conoscere i modelli alla base di queste. Studi di simulazione dettagliati che utilizzano varianti degli scenari mostrati negli Esempi 1 e 2, nonché esempi di serie temporali, nei quali sarà applicato AI-AFTER, saranno forniti nel Capitolo 4.

## Capitolo 3

# Il metodo AI-AFTER

Gli autori del metodo AI-AFTER (Qian et al., 2022), con il fine di comprendere il potenziale della combinazione per il miglioramento, hanno scelto di sviluppare un test di verifica d'ipotesi. Come spiegato precedentemente, la scelta di combinare per l'adattamento è una strategia relativamente conservatrice, quindi in questo caso si pone CFA come ipotesi nulla e CFI come ipotesi alternativa. Il test, di conseguenza, ci suggerirà di combinare per il miglioramento se i dati esaminati forniranno al test la prova che CFI è potenzialmente una strategia più utile. La scelta di combinare per il miglioramento è quella più sensata se, per le  $n$  osservazioni disponibili,  $R(\Psi_L; n)$  è significativamente minore di  $R(\Psi_0; n)$  e quindi una qualche combinazione lineare delle previsioni originali può performare più di tutte le singole previsioni candidate. Per una data classe  $\Psi$ ,  $R(\Psi; n) = R(\Psi^*; n)$ ; tuttavia, la stima di  $R(\Psi; n)$  coinvolge l'identificazione della miglior procedura  $\Psi^*$  e la stima del suo rischio. Il confronto successivo delle stime di  $R(\Psi_0; n)$  e  $R(\Psi_L; n)$  è equivalente a quello tra la previsione originale più performante e la procedura di combinazione CFI più performante. Per la classe  $\Psi_0$  è possibile stimare il rischio di ciascuna previsione originale calcolando il MSFE sui dati disponibili. Il più piccolo di questi rischi stimati sarà la nostra stima di  $R(\Psi_0; n)$ . Per quanto riguarda invece la stima di  $R(\Psi_L; n)$  bisogna decidere quali combinazioni lineari dei previsori si vogliono considerare, infatti, nonostante sotto determinate condizioni sia possibile derivare asintoticamente i pesi ottimali per  $\Psi_L$ , per campioni finiti i pesi ottimali sono solitamente ignoti e possono essere numerose le combinazioni che si possono provare per  $\Psi_L$ . Ovviamente è impossibile analizzare tutte le combinazioni lineari in  $\Psi_L$ , quindi, per stimare il rischio  $R(\Psi_L; n)$ , è possibile valutare un sottoinsieme finito  $\Psi \subset \Psi_L$ . Vogliamo che

$\Psi$  sia grande abbastanza da consentire l'esplorazione di diversi metodi per combinare le previsioni per il miglioramento e, nello specifico, saranno suggeriti due modi per definire la grandezza di  $\Psi$ : rendere il numero di metodi contenuti in  $\Psi$ , d'ora in avanti indicato con  $|\Psi|$ , approssimativamente pari al numero  $M$  di candidati originali (ad esempio tramite regressione lineare), così da mitigare la possibilità che avvenga il fenomeno dei "confronti multipli" che potrebbe dare un vantaggio scorretto ad uno dei due insiemi quando essi vengono confrontati empiricamente, o altrimenti includere in  $\Psi$  un numero di combinazioni nettamente maggiore di quello degli  $M$  candidati originali per avere maggiori possibilità di includere una combinazione che performi meglio, incorrendo però così ad un aumento dell'errore di I tipo a causa dei confronti multipli, per il quale verrà presentata nella sotto-Sezione 3.2.1 una misura di salvaguardia.

### 3.1 La procedura del test statistico

Verrà di seguito descritta la procedura che serve per determinare il potenziale del miglioramento avendo a disposizione le risposte osservate  $y_1, \dots, y_n$  e le corrispondenti previsioni candidate all'interno di una matrice  $X$  di dimensioni  $n \times M$ .

1. Selezionare una frazione  $\rho$  di osservazioni,  $n_0 = \lceil \rho n \rceil$ , che saranno usate per la stima dei pesi e la costruzione delle combinazioni previsive per le ultime osservazioni.
2. Decidere l'insieme  $\Psi$  di procedure di combinazioni tra una o più famiglie di combinazioni per il miglioramento. Esempi di combinazioni saranno dati della sotto-Sezione 3.1.1. Chiamiamo  $\tilde{M} = |\Psi|$  il numero di combinazioni in  $\Psi$  e  $\psi_1, \dots, \psi_{\tilde{M}}$  le particolari combinazioni.
3. Per ogni osservazione  $i = n_0 + 1, \dots, n$ , calcolare le  $\tilde{M}$  previsioni di  $Y_i$  applicando i metodi di combinazione in  $\Psi$  alle prime  $i - 1$  osservazioni.
4. Lo step precedente produce  $\tilde{M}$  combinazioni di previsioni per ognuna delle più recenti  $n_1 = (n - n_0)$  osservazioni. Per ogni  $j \in \{1, \dots, \tilde{M}\}$ , calcolare il MSFE empirico per  $\psi_j$  dalle più recenti  $n_1$  osservazioni. Identificare quindi la procedura  $\psi^* \in \Psi$  con il MSFE empirico più basso.
5. Come nello step precedente, identificare la procedura candidata originale  $\delta^* \in \Delta$  con il MSFE empirico più basso calcolato con le più recenti  $n_1$  osservazioni.



6. Confrontare l'errore di previsione di  $\psi^*$  con quello di  $\delta^*$  utilizzando il test di Diebold-Mariano (Diebold e Mariano, 1995) assumendo una funzione di perdita quadratica. L'ipotesi nulla del test è che le procedure  $\psi^*$  e  $\delta^*$  sono previsioni ugualmente accurate, l'ipotesi alternativa unilaterale è che  $\psi^*$  sia più accurata di  $\delta^*$ .

6'. Nel caso in cui il numero di procedure di combinazione delle previsioni sia molto più grande del numero delle previsioni candidate originarie ( $\tilde{M} \gg M$ ), il  $p$ -value ottenuto nello Step 6 potrebbe assumere un valore erroneamente prossimo allo zero a causa del fenomeno del confronto multiplo sopra-citato. Per correggerlo è possibile effettuare un test di randomizzazione che usa simulazioni per approssimare il reale valore del  $p$ -value dello Step 6. Il test di randomizzazione funziona come di seguito:

- a. Creare una matrice  $F$  di dimensioni  $n_1 \times (M + \tilde{M})$  delle previsioni originali e combinate per le più recenti  $n_1$  osservazioni.
- b. Eseguire ognuno dei seguenti passi  $N$  volte, dove  $N$  è un elevato numero di ripetizioni della randomizzazione:
  - i. Permutare casualmente le  $M + \tilde{M}$  colonne di  $F$ . Denominare le prime  $M$  colonne della matrice "previsioni originali" e le rimanenti  $\tilde{M}$  colonne "previsioni combinate".
  - ii. Eseguire gli Step 4-6 sugli insiemi di previsioni "originali" e "candidate".

Questo passo produce un insieme di  $N$   $p$ -value dal test di D-M.

- c. Il  $p$ -value del test di randomizzazione è la proporzione dei  $p$ -value ottenuti dallo Step 6'(c) che sono minori o uguali al  $p$ -value osservato nello Step 6.
7. Confrontare il  $p$ -value del test D-M con un livello di significatività  $\alpha$  scelto precedentemente. Un  $p$ -value minore di  $\alpha$  indica evidenza che almeno una delle procedure di combinazione in  $\Psi$  sia più accurata della miglior previsione candidata in  $\Delta$  poichè si ha evidenza empirica che  $R(\Psi_L; n) \leq R(\Psi_0; n)$ . In questo caso si sceglie di combinare per il miglioramento, altrimenti per l'adattamento.

### 3.1.1 Come scegliere le combinazioni

Come accennato precedentemente, lo Step 2 della procedura AI-After riguarda la natura e il numero di procedure combinate che possono essere incluse in  $\Psi$ . In pratica, si costruisce  $\Psi$  dalle seguenti combinazioni delle  $M$  previsioni originarie.

- **LASSO con AIC, BIC.** Vengono selezionate le soluzioni LASSO con i minori valori di AIC o BIC tra l'insieme di soluzioni.
- **Selezione stepwise con AIC, BIC.** Si applica l'algoritmo della selezione stepwise alle prime  $i$  osservazioni e vengono selezionati i modelli con i valori minori di AIC o BIC tra l'insieme di soluzioni.
- **Miglior insieme per ogni dimensione.** Le prime  $n_0$  osservazioni sono usate per selezionare il miglior modello per ogni dimensione compresa tra 1 e  $\min(M, n_0 - 1)$ . I coefficienti di questi modelli sono aggiornati dopo ogni osservazione  $i$ , ma l'insieme di variabili utilizzate in ciascun modello rimane invariato.
- **Regressione lineare condizionata (CLR).** Si considera una regressione lineare condizionata (CLR) delle previsioni originali senza un'intercetta in cui i coefficienti della regressione stimati sommano a 1.
- **Bates-Granger ( $BG_{0.9}$ ,  $BG_1$ ).** Si utilizza lo schema di pesi descritto da Bates e Granger, 1969.
- **Combinazioni Naïve (SA, MD, TM).** Si considerano tre combinazioni di previsioni molto semplici che non prendono in considerazione la *performance* delle singole combinazioni candidate: SA, ossia la media semplice delle previsioni, MD, ossia la mediana delle previsioni, e TM, ossia una media "tagliata" delle previsioni calcolata escludendo il 5% più piccolo e il 5% più elevato.

Costruendo così  $\Psi$ , si trovano  $\tilde{M} = \min(M, n_0 - 1) + 10$  distinte combinazioni delle previsioni originali per ogni passo  $i$  in avanti.

## 3.2 Utilizzo dell'algoritmo AFTER

Il test d'ipotesi precedentemente descritto indica se, con i dati e le previsioni candidate a disposizione, si debba eseguire la combinazione per il miglioramento. Se è bassa l'evidenza che una qualsiasi combinazione di previsioni possa avere una *performance* migliore

della migliore previsione individuale, allora il test suggerisce semplicemente di mirare al rischio previsivo del miglior previsore individuale, cioè di combinare per l'adattamento. Come accennato nella Sezione 2.2, il metodo AFTER può essere considerato un buon rappresentante dei metodi CFA poichè è in grado di fornire un'ottima protezione contro l'incertezza nella selezione del modello e di adattarsi automaticamente ai cambiamenti nel tempo che possono avvenire nel processo generatore dei dati e quindi alle relative prestazioni del previsore (Zou e Yang, 2004).

Se il test di ipotesi indica di preferire uno scenario di combinazione per il miglioramento, significa che siamo stati in grado di generare una o più combinazioni di previsioni che performano meglio di tutti i previsori originali. In questo caso, il nostro obiettivo dovrebbe essere quello di eguagliare la *performance* della migliore previsione combinata. In altre parole, questo può essere pensato come uno scenario di adattamento in cui le combinazioni di previsioni sono considerate come previsioni candidate, di conseguenza data questa nuova prospettiva è necessario proporre una nuova versione di AFTER per raggiungere il potenziale del miglioramento.

La discussione precedente suggerisce che AFTER, applicato alle corrette procedure di previsione, può essere utilizzato per combinare efficacemente previsioni sia con lo scopo di adattamento (A) che per quello di miglioramento (I), e questa nuova versione dell'algoritmo è stata per questo denominata "AI-AFTER" (Qian et al., 2022). Di seguito verranno descritti i successivi passi dell'algoritmo, a partire dallo Step 7 della Sezione 3.1, nel quale il *p-value* del test d'ipotesi effettuato ha determinato quale è l'insieme di previsori a cui applicare il metodo AFTER.

8. Se il *p-value* è maggiore del livello  $\alpha$  precedentemente scelto, si accetta l'ipotesi nulla e si conclude che CFA è la strategia più opportuna, per cui si applica l'algoritmo AFTER alle risposte e previsioni originali, rispettivamente  $Y$  e  $X$ . Altrimenti, si applica AFTER alla matrice  $\tilde{X}$  di dimensioni  $n_1 \times \tilde{M}$  di combinazioni di previsioni (e ai corrispondenti  $n_1$  valori osservati più recenti) prodotta precedentemente durante la procedura. Le previsioni  $\hat{y}_{1,i}$  ( $i = n_0 + 1, \dots, n$ ) saranno quindi ottenute in base alla direzione combinatoria (CFA o CFI) determinata in precedenza, e il corrispondente vettore di pesi per la combinazione ( $\mathbb{R}^M$  o  $\mathbb{R}^{\tilde{M}}$ ) è generato per le successive previsioni.

La *performance* di AFTER, inoltre, trae spesso beneficio dall'utilizzo di alcune poche  $n_{init}$  osservazioni come "burn-in". Ciò significa che le prime  $n_{init}$  osservazioni prodotte

da AFTER assegnano pesi equivalenti a tutti i candidati, perchè la *performance* relativa del previsore non può essere stimata attendibilmente usando solo le prime osservazioni. Le  $n_{init}$  osservazioni iniziali vengono quindi utilizzate solo per stimare i pesi della combinazione per le osservazioni future e, di default,  $n_{init} = 5$

### 3.2.1 Una misura di salvaguardia

Come precedentemente discusso, il test d'ipotesi è integrato come passaggio informativo per decidere la direzione della combinazione (CFA o CFI) per portare potenzialmente ad una migliore *performance* previsiva. D'altra parte, poiché il test d'ipotesi è soggetto ad errori di I e/o di II tipo (in particolare quando la dimensione del campione è piccola), è stata ulteriormente introdotta una funzione di salvaguardia nell'algoritmo AI-AFTER con lo scopo di proteggere da questi possibili errori, senza necessariamente scendere a compromessi sulla *performance* previsionale (discussa nella Sezione 3.3). Nello specifico, è stato elaborato un composito Step aggiuntivo: AFTER è prima applicato ad un insieme espanso di previsioni originali e candidate per generare una previsione "di salvaguardia"  $\hat{y}_{2,i}$ ; successivamente, AFTER è applicato nuovamente trattando  $(\hat{y}_{1,i}, \hat{y}_{2,i})$  come le due previsioni candidate così da generare i corrispondenti pesi per la combinazione per le previsioni  $\hat{y}_{C,i}$ . Nella pratica:

9. Creare una matrice  $\tilde{X}$  di dimensioni  $n_1 \times (M + \tilde{M})$  di previsioni candidate concatenando  $\tilde{X}$  alle più recenti  $n_1$  righe di  $X$ . Applicare l'algoritmo AFTER alle più recenti  $n_1$  risposte e alla matrice  $\tilde{X}$  composta dalle  $M$  previsioni candidate e dalle  $\tilde{M}$  previsioni combinate; ciò genera la previsione "di salvaguardia"  $\hat{y}_{2,i}$  ( $i = n_0 + 1, \dots, n$ ) e il vettore dei pesi per la combinazione in  $\mathbb{R}^{M+\tilde{M}}$  per le prossime previsioni.
10. Applicare l'algoritmo AFTER nuovamente con  $(\hat{y}_{1,i}, \hat{y}_{2,i})$  come previsioni candidate. Ciò crea le previsioni  $\hat{y}_{C,i}$  e un vettore dei pesi per la combinazione in  $\mathbb{R}^2$  da usare per le future previsioni.

## 3.3 Il limite del rischio previsivo

Ora, si considerino nuovamente  $\delta_1, \dots, \delta_M$  come le procedure previsive candidate originariamente, le quali quindi producono  $\hat{y}_{\delta_j,i}$  dove  $i = 1, 2, \dots$ , e  $1 \leq j \leq M$ . Costruendo  $\Psi$ ,

si hanno  $\tilde{M} = |\Psi|$  distinte combinazioni di previsioni, le quali verranno singolarmente denotate con  $k_l$  ( $1 \leq \tilde{M}$ ). Per mostrare i limiti del rischio previsivo che vengono proposti dall'approccio AI-AFTER è necessario assegnare delle probabilità a priori: si fissi una costante  $p_0$  ( $0 < p_0 < 1$ ) così che ogni  $\delta_i$  ( $i = 1, \dots, M$ ) abbia a priori  $\pi = (1 - p_0)\frac{1}{M}$  e che ogni  $k_l$  ( $l = 1, \dots, M$ ) abbia a priori  $\pi = \frac{p_0}{M}$ . Si noti che le probabilità appena definite sommano ad 1. Sia  $N$  un orizzonte di previsione totale e si supponga che le Condizioni 6 e 8 in Yang, 2004, siano soddisfatte. Così si può subito ottenere la previsione finale combinata trovata da AI-AFTER che raggiunga prestazioni quasi ottimali, come mostrato nel Teorema 1.

**Teorema 1.** *Sia  $\delta_C$  la procedura di combinazione AI-AFTER. Allora esiste una costante  $\lambda$ , tratta da (2.2), tale che  $\sum_{i=n+1}^N E(Y_i - \hat{y}_{\delta_C, i})^2$  non sia maggiore del più piccolo dei seguenti valori:*

- $\log(M) + \log\left(\frac{2}{1-p_0}\right) + \inf_{1 \leq M} \left( \sum_{i=n+1}^N E(Y_i + \hat{y}_{\delta_j, i})^2 \right)$
- $\log(\tilde{M}) + \log\left(\frac{2}{p_0}\right) + \inf_{1 \leq \tilde{M}} \left( \sum_{i=n+1}^N E(Y_i + \hat{y}_{k_l, i})^2 \right)$

Dal Teorema 1, senza alcuna conoscenza a priori e in un'ottica universale, AI-AFTER non è peggiore né della (sconosciuta) miglior previsione individuale tra le originarie in  $\Psi_0$  né della miglior combinazione previsiva in  $\Psi$ , aggiungendo una penalità relativamente piccola (quando  $n$  è grande). Di conseguenza, se l'errore di previsione quadratico medio non converge a 0, risulterà che  $\overline{\lim}_{N \rightarrow \infty} \frac{\sum_{i=n+1}^N E(Y_i - \hat{y}_{\delta_j, i})^2}{R} \leq 1$ , dove  $R$  è il valore minimo tra  $\inf_{1 \leq M} \sum_{i=n+1}^N E(Y_i - \hat{y}_{\delta_j, i})^2$  e  $\inf_{1 \leq \tilde{M}} \sum_{i=n+1}^N E(Y_i - \hat{y}_{k_l, i})^2$ . Dunque, AI-AFTER è asintoticamente mai peggiore del miglior candidato originale e della miglior previsione combinata, e si adatta in modo intelligente: è conservativo (mirando all'obiettivo dell'adattamento) quando non c'è alcun vantaggio nel perseguire il miglioramento, ed è altrimenti aggressivo, mirando alla *performance* delle combinazioni di previsioni alternative per l'obiettivo del miglioramento.



## Capitolo 4

# Studi di simulazione

In questo Capitolo saranno presentati i risultati di alcuni studi di simulazione per differenti scenari, basati o sulla regressione lineare o sulle serie storiche, che possiamo considerare rivisitazioni delle simulazioni nella Sezione 2.3.1. Negli scenari basati sulla regressione lineare, nel processo generatore di dati è presente un grande numero di variabili concomitanti e vengono considerati  $M$  differenti modelli candidati. Negli scenari basati sulle serie storiche, i valori passati della variabile risposta sono utilizzati dai modelli candidati insieme ad una o due variabili concomitanti. In entrambi i casi, vengono presentati scenari di combinazioni di previsioni per l'adattamento o per il miglioramento. Nelle simulazioni  $\rho = 1/3$ , e sarà usato un livello di significatività  $\alpha = 0.1$ . Saranno confrontate le performances predittive di AI-AFTER con quelle di 5 concorrenti: il metodo semplice AFTER, i metodi CLR,  $BG_1$  e SA, descritti nella Sezione 3.1.1, e una combinazione eseguita tramite regressione lineare (LR).

In ciascuna simulazione descritta nel resto del Capitolo, 200 serie storiche indipendenti sono generate, ognuna di numerosità pari a  $n = 100$  o a  $n = 300$ . Ciascuna realizzazione include  $n_{train}$  dati iniziali (che sono usati unicamente per costruire le previsioni candidate e non sono disponibili all'analista), le  $n$  osservazioni e previsioni candidate (disponibili all'analista incaricato di combinare le previsioni), ed un conseguente numero  $n_{eval}$  di osservazioni, future alle prime  $n_{train}$ , usate per valutare i metodi di combinazione. In tutte le simulazioni, si hanno  $n_{train} = n - 20$  e  $n_{eval} = 20$ . Per ciascuna realizzazione, i metodi di combinazione sono valutati attraverso il loro errore di previsione quadratico medio (MSFE) calcolato sulle  $n_{eval}$  osservazioni. Nello specifico, il MSFE

per un metodo di combinazione  $\Delta$  in una data realizzazione  $j$  è

$$MSFE_j^\Delta = \frac{1}{20} \sum_{i=n_{train}+1}^{n_{train}+20} (y_{i,j} - \hat{y}_{i,j}^\Delta)^2 \quad (4.1)$$

dove  $y_{i,j}$  denota il valore per la  $i$ -esima osservazione nella  $j$ -esima realizzazione, e  $\hat{y}_{i,j}^\Delta$  è la previsione di  $y_{i,j}$  prodotta dall'applicazione del metodo  $\Delta$ . Per ciascuno scenario sarà mostrata una sintesi dei  $MSFE_j^\Delta/MSFE_j^{SA}$  dei diversi metodi  $\Delta$  con lo scopo di confrontare le *performance* di ciascun metodo di combinazione in relazione a quelle della media semplice delle previsioni candidate.

## 4.1 Esempi con la regressione lineare

In questa Sezione, il PGD è:

$$y_i = \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i \quad (4.2)$$

dove le  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  sono variabili indipendenti ed identicamente distribuite (i.i.d), normali multivariate con media  $\mathbf{0}$  e matrice di covarianza con struttura autoregressiva e generico elemento  $\sigma_{jk} = 0.5^{|j-k|}$ ,  $p = 25$ ,  $\beta = (5, 4, 3, 2, 1, 0, \dots, 0)$ . I termini di errore  $\epsilon_i$  sono anch'essi i.i.d. con distribuzione  $N(0, 4)$  ed indipendenti da  $\mathbf{x}_i$ . Verranno considerati sia uno scenario di adattamento che uno di miglioramento (denominati rispettivamente "OLS - adattamento" e "OLS - miglioramento"); il processo generatore di dati in entrambi gli scenari è lo stesso (4.2), ma le previsioni candidate disponibili all'analista sono diverse.

### 4.1.1 Scenario di adattamento

Le previsioni candidate nello scenario di adattamento, per  $1 \leq i \leq n$ , sono costruite nella seguente maniera:

$$\text{Previsore 1 : } \hat{y}_{i,1} = \hat{\beta}_{1,0} + \sum_{j=1}^5 \hat{\beta}_{1,j} x_{ij}, \quad (4.3)$$

$$\text{Previsore k, } 2 \leq k \leq 25 : \hat{y}_{i,k} = \hat{\beta}_{k,0} + \sum_{j=1}^p I(k,j) \hat{\beta}_{k,j} x_{ij}$$

Le stime dei parametri  $\hat{\beta}_{k,j}$  vengono ottenute attraverso il metodo dei minimi quadrati ordinari, trovate utilizzando un insieme di  $n_{train} = n - 20$  osservazioni di  $(y, \mathbf{x})$ , che



sono usate solo per stimare le previsioni candidate e non sono disponibili all'analista. Le variabili casuali  $I(k, j)$  sono Bernoulliane di parametro 0.5 indipendenti tra loro, dunque ogni covariata ha il 50% di probabilità di essere inclusa come variabile esplicativa in ciascuna previsione candidate  $k$ ,  $2 \leq k \leq 25$ . Siccome nessuna delle previsioni  $k$  per  $k > 1$  aggiunge alcuna informazione riguardante il processo generatore di dati a (4.3), in questo caso  $R(\Psi_0; n) = R(\Psi_L; n)$  per un campione  $n$  sufficientemente elevato, ad esempio  $n = 100$ . Così, l'insieme prima definito di previsori candidati rappresenta uno scenario in cui combinare per l'adattamento è l'obiettivo più appropriato.

### 4.1.2 Scenario di miglioramento

Ora, per lo stesso processo generatore di dati (4.2), vengono considerate le previsioni candidate  $k$ ,  $1 \leq k \leq 25$ , della forma

$$\text{Previsore } k : \hat{y}_{i,k} = \hat{\beta}_{k,0} + \hat{\beta}_k x_{ik} \quad (4.4)$$

Anche questa volta i parametri  $\hat{\beta}_k$  sono stimati utilizzando le prime  $n_{train} = n - 20$  osservazioni. In questo scenario, ogni previsione può apportare la propria incompleta informazione riguardo al PGD, ma se le previsioni fossero combinate nel corretto modo, la combinazione di esse può cogliere tutta l'informazione in 4.2 e quindi produrre una previsione più accurata.

## 4.2 Esempi con le serie storiche

Si consideri un processo autoregressivo (AR) con due variabili concomitanti. In contrasto con gli esempi riguardanti la regressione lineare visti precedentemente, in questo caso sia lo scenario di adattamento che quello di miglioramento considerano lo stesso insieme di previsioni candidate, è invece la natura del processo generatore di dati a determinare quando sia più appropriato combinare per l'adattamento o per il miglioramento. Si assume che i previsori candidati abbiano accesso ai precedenti valori della variabile risposta ed al massimo ad una delle due covariate. Nello specifico, ogni previsione candidate è

basata su un modello AR(X) tra i seguenti:

$$\text{Previsore 1 : } \hat{y}_{i,1} = \hat{\beta}_{0,1} + \hat{\gamma}_{1,1}y_{i-1}$$

$$\text{Previsore 2 : } \hat{y}_{i,2} = \hat{\beta}_{0,2} + \hat{\gamma}_{1,2}y_{i-1} + \hat{\gamma}_{2,2}y_{i-2}$$

$$\text{Previsore 3 : } \hat{y}_{i,3} = \hat{\beta}_{0,3} + \hat{\beta}_{1,3}x_{i1} + \hat{\gamma}_{1,3}y_{i-1}$$

$$\text{Previsore 4 : } \hat{y}_{i,4} = \hat{\beta}_{0,4} + \hat{\beta}_{2,4}x_{i2} + \hat{\gamma}_{1,4}y_{i-1}$$

$$\text{Previsore 5 : } \hat{y}_{i,5} = \hat{\beta}_{0,5} + \hat{\beta}_{1,5}x_{i1} + \hat{\gamma}_{1,5}y_{i-1} + \hat{\gamma}_{2,5}y_{i-2}$$

$$\text{Previsore 6 : } \hat{y}_{i,6} = \hat{\beta}_{0,6} + \hat{\beta}_{2,6}x_{i2} + \hat{\gamma}_{1,6}y_{i-1} + \hat{\gamma}_{2,6}y_{i-2}$$

Come nella Sezione 4.1, i coefficienti  $\hat{\beta}$  e  $\hat{\gamma}$  sono stimati utilizzando le prime  $n_{train} = n - 20$  osservazioni di  $(y, \mathbf{x})$ , i cui valori non sono disponibili all'analista. I valori di  $\mathbf{x}_i$  sono i.i.d., e le due covariate sono tra loro indipendenti ed entrambe di distribuzione  $N(0, 4)$ . Verranno presentati sia uno scenario di adattamento che uno di miglioramento (denominati rispettivamente "AR(X) - adattamento" e "AR(X) - miglioramento").

#### 4.2.1 Scenario di adattamento

Le sei previsioni candidate appena descritte vengono applicate per prevedere i valori  $y_i$  generati dal seguente processo:

$$y_i = 0.3y_{i-1} + 0.6y_{i-2} + \epsilon_i \quad (4.5)$$

dove le  $\epsilon_i$  sono anch'esse i.i.d.  $N(0, 4)$  come nella Sezione 4.1. In questo scenario il Previsore 2 rappresenta il vero modello e raggiunge, quindi, il minor rischio previsivo (per una numerosità  $n$  abbastanza elevata). I Previsori 5 e 6 includono entrambi i ritardi AR, ma ognuno di loro utilizza anche una covariata che non apporta alcuna informazioni riguardo il PGD, mentre i Previsori 1, 3 e 4 non includono il secondo ritardo di  $y$  nel loro modello. Per una numerosità campionaria  $n$  sufficientemente elevata, ad esempio  $n = 100$ , le *performance* del Previsore 2 non possono essere in alcun modo migliorate dalla combinazione di esso con altre previsioni; dunque, combinare per l'adattamento è considerato l'obiettivo più appropriato.

#### 4.2.2 Scenario di miglioramento

Il processo generatore di dati in questo caso usa entrambe le covariate ed è:

$$y_i = 0.3y_{i-1} + 0.6y_{i-2} + 2x_{i1} + 2x_{i2} + \epsilon_i \quad (4.6)$$

Tabella 4.1: Percentuale (su 200 realizzazioni) in cui AI-AFTER ha selezionato la Combinazione per il Miglioramento come obiettivo per la combinazione previsiva, considerando un livello di significatività  $\alpha = 0.1$ .

Processo generatore di dati	Numerosità campionaria	% rifiuti $H_0$
OLS - adattamento	$n = 100$	3%
	$n = 300$	0.5%
AR(X) - adattamento	$n = 100$	0%
	$n = 300$	0%
OLS - miglioramento	$n = 100$	99%
	$n = 300$	100%
AR(X) - miglioramento	$n = 100$	81%
	$n = 300$	100%

Le  $\epsilon_i$  sono nuovamente i.i.d.  $N(0, 4)$ . In questo caso nessuno dei sei previsori candidati usa entrambe le covariate, alcuni usano  $x_1$  e altri  $x_2$ . Dunque, la combinazione delle previsioni può far ottenere *performance* migliori rispetto a quelle delle previsioni individuali, grazie alla condivisione delle loro informazioni. In questo scenario, quindi, combinare per il miglioramento è considerato l'obiettivo più appropriato.

### 4.3 Risultati e performance previsiva

Per prima cosa valutiamo la *performance* della procedura test del metodo AI-AFTER descritta nella Sezione 3.1 per determinare l'obiettivo della combinazione.

La Tabella 4.1 mostra, per ognuno dei quattro scenari di simulazione con una numerosità campionaria fissa,  $n = 100$  o  $n = 300$ , la proporzione delle 200 realizzazioni in cui il test ha rifiutato  $H_0$  e, quindi, consigliato la combinazione per il miglioramento. La metà superiore della tabella mostra che in tutti i casi la frequenza di errori di I tipo è stata minore rispetto al livello di significatività fissato  $\alpha = 0.1$ . Ciò, infatti, è comprensibile osservando che sotto l'ipotesi nulla del test si dovrebbe avere  $R(\Psi_0; n) = R(\Psi_L; n)$ , mentre è noto che, in quei casi,  $R(\Psi_0; n) < R(\Psi_L; n)$  perchè la *performance* prevista dal vero modello, che è presente tra i previsori candidati, è migliore rispetto alla *performance* prevista da un qualsiasi metodo di combinazione in  $\Psi_L$ . In ognuno dei quattro scenari

in cui la combinazione per l'adattamento è appropriata (e combinare per il miglioramento porta ad un rischio maggiore senza trarne alcun vantaggio), il test ha rifiutato  $H_0$  e consigliato la combinazione per il miglioramento al massimo il 3% delle volte. La metà inferiore della Tabella 4.1 mostra che il test di AI-AFTER funziona bene anche nel rivelare il potenziale del miglioramento, specialmente quando  $n = 300$ , con un tasso di rifiuti di  $H_0$  del 100% sia nello scenario OLS che AR(X) per il miglioramento. Come nella maggior parte dei test d'ipotesi, il test è meno potente quando  $n$  è più piccolo, ad esempio, quando  $n = 100$  nello scenario AR(X)-miglioramento,  $H_0$  viene rifiutata l'81% delle volte. Questi risultati possono essere osservati tramite la distribuzione del  $p$ -value mostrata dalle Figure 4.1 e 4.2. I risultati in questi esempi suggeriscono che il test riesce efficacemente a controllare l'errore di I tipo quando il vero processo generatore di dati è rappresentato da uno dei candidati e, dunque, è necessario combinare per l'adattamento. Quando è necessario combinare per il miglioramento, il test è stato più efficace ed informativo quando era in presenza di un campione più numeroso, ma potrebbe essere meno potente (come ci si aspetta da ogni test) se si utilizza una minore dimensione campionaria.

Per esaminare la robustezza delle *performance* delle previsioni del metodo AI-AFTER, le abbiamo comparate a quelle dei metodi AFTER,  $BG_1$ , LR, CLR e SA. Per ogni metodo  $\Delta$ , le Figure 4.3 e 4.4 mostrano la distribuzione empirica degli  $MSFE^\Delta$  relativi al  $MSFE^{SA}$  per le 200 realizzazioni. In linea con ciò che ci si aspettava, AI-AFTER riesce a seguire le tracce di AFTER negli scenari di adattamento mentre, negli scenari di miglioramento, performa competitivamente contro gli altri metodi di combinazione previsiva considerando diverse dimensioni campionarie e offre un significativo miglioramento rispetto ad AFTER. Complessivamente, i risultati delle simulazioni mostrano che AI-AFTER è dotato di una strategia di combinazione previsiva che è aggressiva quando il guadagno di CFI è alto ed è conservativa quando una combinazione delle previsioni non riesce a superare di molto (oppure affatto) il singolo candidato più performante.

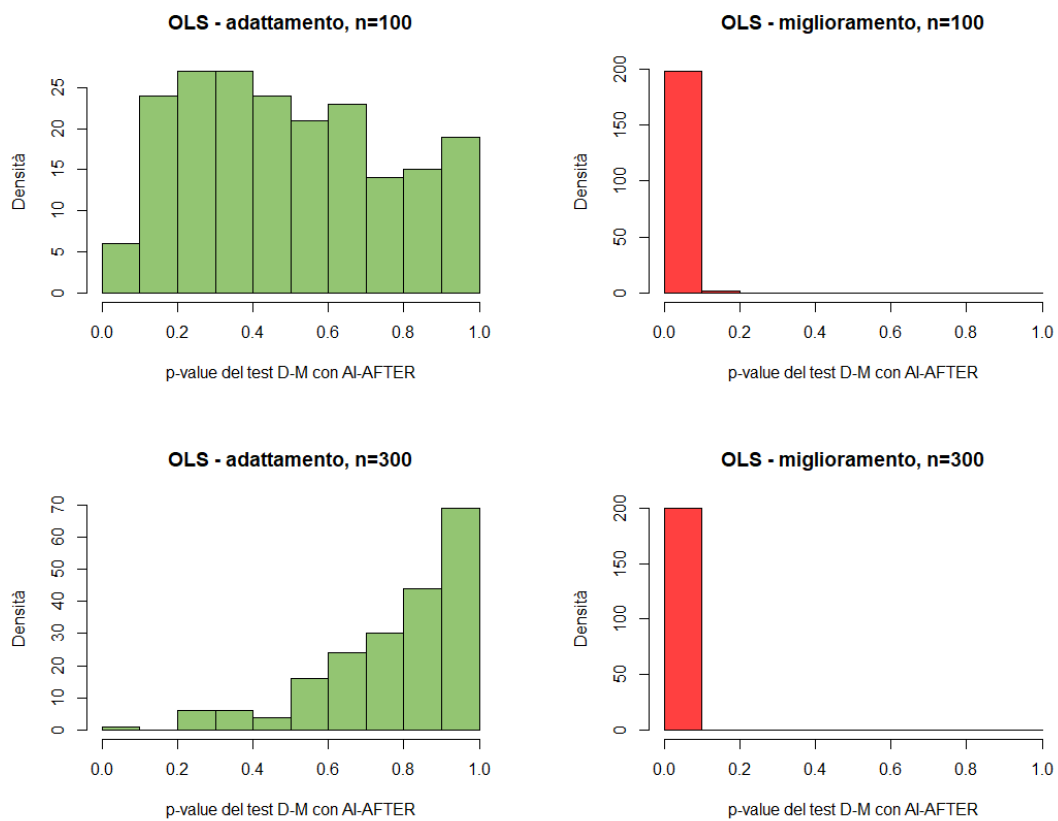


Figura 4.1: Esempi con regressione lineare: distribuzione del p-value del test AI-AFTER.

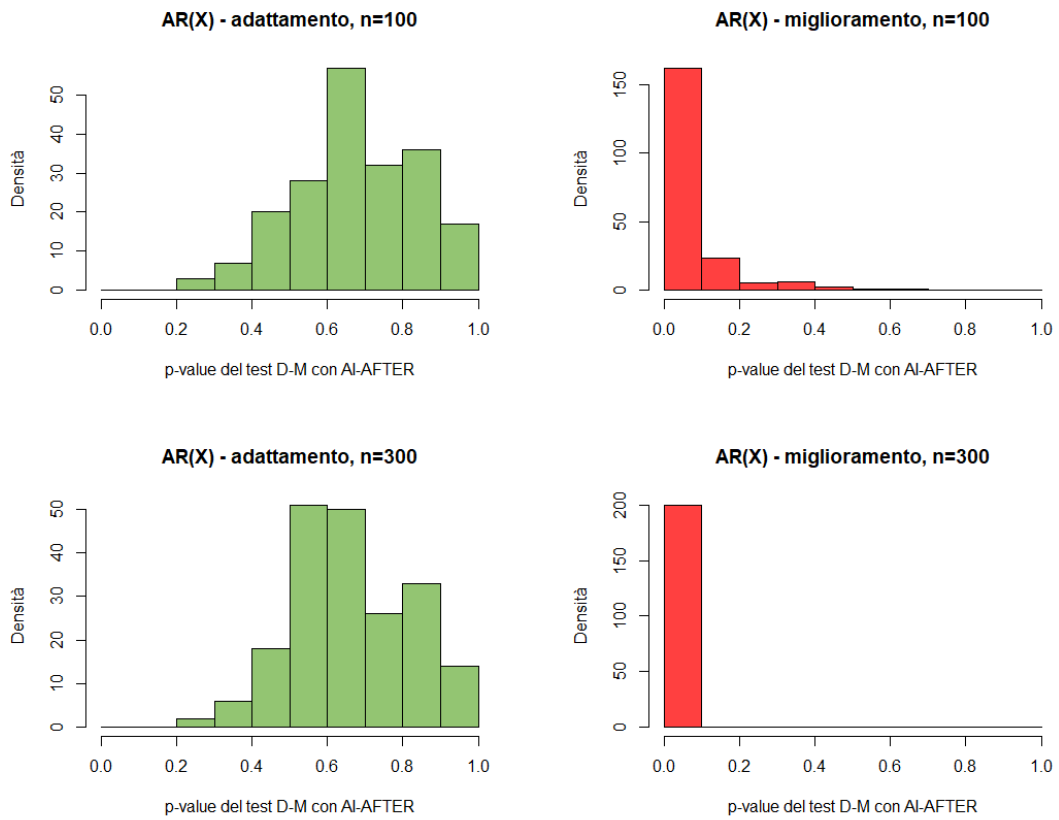


Figura 4.2: Esempi con serie storiche: distribuzione del p-value del test AI-AFTER.

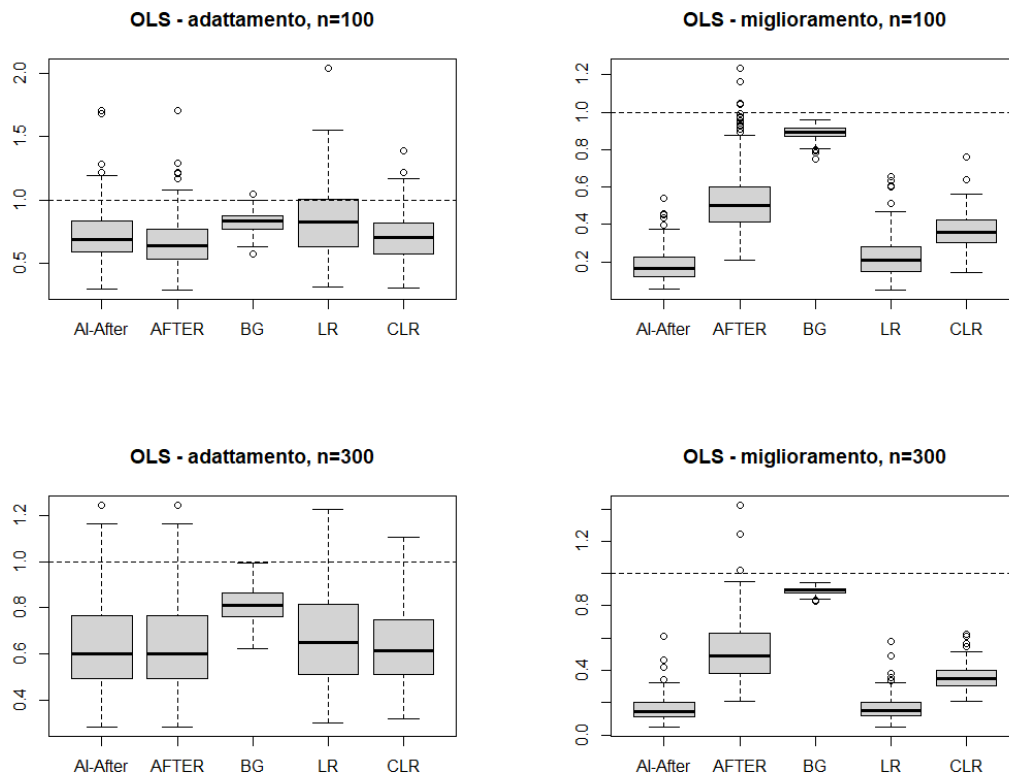


Figura 4.3: Esempi con regressione lineare: ogni boxplot mostra  $MSFE_j^\Delta / MSFE_j^{SA}$  per 200 realizzazioni.

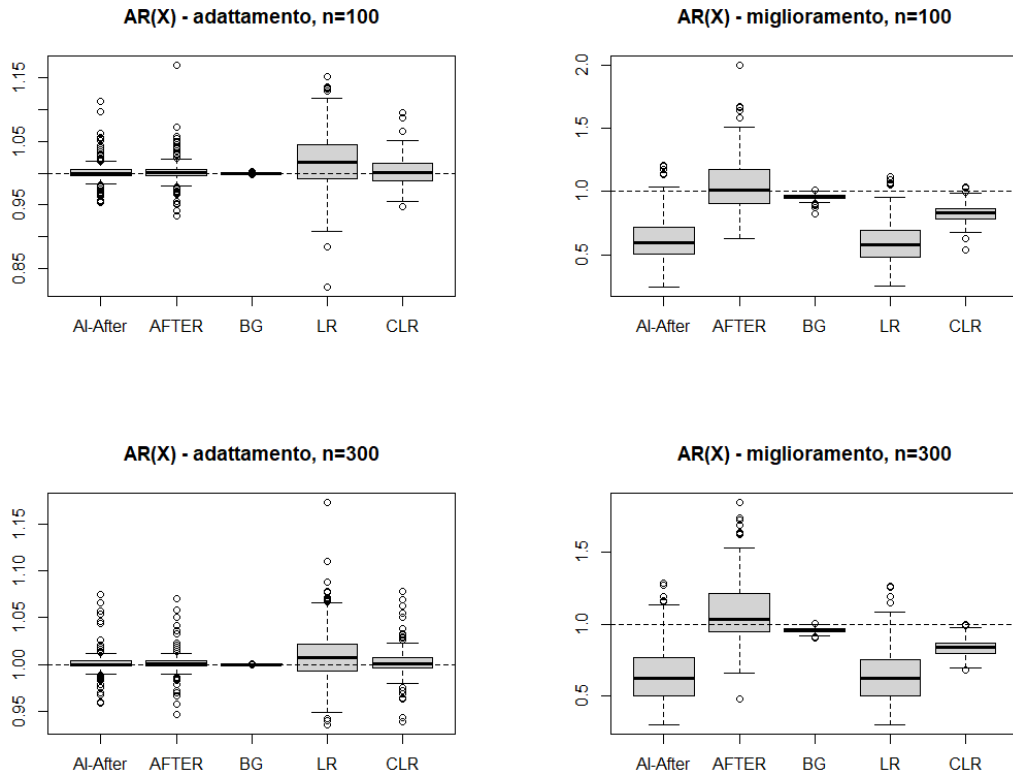


Figura 4.4: Esempi con serie storiche: ogni boxplot mostra  $MSFE_j^\Delta / MSFE_j^{SA}$  per 200 realizzazioni.



## Capitolo 5

# Applicazione su dataset M4

La competizione M4, svolta a seguito delle precedenti tre gare M (che hanno avuto luogo, rispettivamente, nel 1982, 1993 e 1998, Hyndman, 2020), si è conclusa a mezzanotte del 31 Maggio 2018, ed ha visto sfidarsi oltre sessanta gruppi di ricerca che hanno presentato i loro risultati previsivi calcolati basandosi su diversi metodi di valutazione. Lo scopo di questa competizione era di replicare i risultati delle competizioni precedenti ed ampliarne alcuni aspetti, tra cui, il più importante, il significativo incremento del numero delle serie storiche considerate. Il dataset è costituito da 100.000 serie temporali annuali, trimestrali, mensili e di altro tipo (incluse serie di dati ad alta frequenza come quelle settimanali, giornaliere ed orarie), che riguardano l'ambito Micro o Macro economico, Industriale, Finanziario, Demografico o di altro genere, selezionate da un *database* chiamato "ForeDeCk" costruito presso la National Technical University of Athens (NTUA) contenente 900.000 serie temporali.

Come già menzionato nel Capitolo 1, il risultato più importante di questa competizione è stato quello di (ri)confermare i vantaggi della combinazione di più previsioni (Makridakis et al., 2018), infatti *"La combinazione dei metodi (previsivi) è stata la regina della M4. Dei 17 metodi più accurati, 12 erano "combinazioni" di approcci prevalentemente statistici."* Questa conclusione implica che, dal punto di vista empirico, quasi nessun metodo singolo riesce a catturare adeguatamente i modelli per serie storiche, mentre una combinazione di questi metodi, ognuno dei quali cattura una diversa componente del modello reale, è più accurata perchè, attraverso la media ponderata, è in grado di ridurre gli errori dei singoli metodi.

## 5.1 Dati e metodi previsivi

Con il fine di valutare le *performance* del metodo di combinazione tramite un'applicazione su dati reali, gli ideatori del metodo AI-AFTER hanno selezionato 13 serie storiche trimestrali relative agli indici del PIL e IP di alcuni Paesi. Seguendo le loro tracce, ho estrapolato alcune serie appartenenti al dataset della M4-Competition. Questi dati erano stati resi disponibili dal 31 dicembre 2017, originariamente sul sito web della M4 (<https://www.m4.unic.ac.cy/the-dataset/>) e, successivamente, tramite il pacchetto R *M4comp2018* (Montero-Manso et al., 2018). In aggiunta, alla maggior parte dei partecipanti è stato richiesto di riportare il codice usato per generare le loro previsioni su GitHub (<https://github.com/M4Competition/M4-methods>), insieme ad una descrizione dettagliata dei loro metodi. In particolare, tra le serie trimestrali con un numero di osservazioni maggiore di 60, ne ho selezionate casualmente due per ogni ambito, per un totale, quindi, di 12 serie, ognuna di numerosità differente e pari a  $n = 67$  e  $n = 63$  (Macro-economica),  $n = 85$  e  $n = 69$  (Micro-economica),  $n = 115$  e  $n = 94$  (Demografica),  $n = 77$  e  $n = 63$  (Industriale),  $n = 118$  e  $n = 119$  (Finanziaria) e  $n = 70$  e  $n = 87$  (Altro tipo).

Durante la competizione sono stati utilizzati alcuni metodi come *benchmarks* per confrontare le previsioni e rilevare i fattori in grado di portare a migliori *performance* previsive (Makridakis et al., 2020). Inoltre, è noto che la maggior parte dei partecipanti che ha optato per la combinazione delle previsioni ha incluso tra le candidate quelle stimate con i metodi statistici *benchmarks* della M4, ottenendo da questa combinazione risultati più accurati rispetto a quelli derivati dai metodi individuali (Makridakis et al., 2018). Quindi, come previsioni candidate, ho scelto quelle prodotte da tali metodi *benchmarks*, nello specifico:

- **Naïve 1** Modello Random Walk ( $ARIMA(0, 1, 0)$ ).
- **Naïve S** Modello con Random Walk stagionale ( $ARIMA(0, 0, 0)(0, 1, 0)_4$ ).
- **SES** Modello di lisciamiento esponenziale semplice.
- **Holt** Modello di lisciamiento esponenziale tramite il metodo lineare di Holt.
- **Damped** Modello di lisciamiento esponenziale che assume un trend "damped".
- **Theta** Metodo previsivo Theta.

- **Comb** Media aritmetica semplice delle previsioni prodotte dai metodi SES, Holt e Damped.
- **ETS** Automatica selezione del miglior modello di lisciamiento esponenziale, tramite i criteri d'informazione.
- **ARIMA** Automatica selezione del miglior modello ARIMA, tramite i criteri d'informazione.

Nella competizione, per le serie trimestrali, venivano richieste le previsioni  $h = 8$  passi in avanti; dunque, ho utilizzato questo come orizzonte previsivo per la costruzione delle previsioni candidate.

## 5.2 Risultati

Come nel Capitolo 4, ho deciso di confrontare le *performance* previsive dei metodi AI-AFTER, AFTER, BG<sub>1</sub>, LR, CLR e SA. L'accuratezza di ciascun metodo di combinazione  $\Delta$ , in termini di  $MSFE^\Delta/MSFE^{SA}$ , è calcolato basandondosi sui  $n_{eval} = 8$  valori finali della serie osservata, con un livello di significatività  $\alpha = 0.1$  per il test di Diebold-Mariano eseguito dal metodo AI-AFTER.

Le *performance* relative di ciascun metodo di combinazione previsiva per la previsione dei valori futuri delle serie storiche sono riportate nella Tabella 5.1, insieme ai *p-value* dei test AI-AFTER in ciascun caso esaminato.

Per prima cosa, è interessante notare che, nella maggioranza dei casi, il test AI-AFTER, fissando il livello  $\alpha = 0.1$ , accetta l'ipotesi nulla e, quindi, suggerisce che gli scenari previsivi presi in considerazione dai dataset appartengono alla categoria dell'adattamento previsivo, per l'esattezza viene scelta la combinazione per l'adattamento in 8 casi e quella per il miglioramento nei restanti 4. Questo risultato è opposto rispetto a quello raggiunto da Qian et al., 2022, nella cui applicazione il test consigliava quasi sempre la combinazione per il miglioramento. Ciò, tuttavia, è ragionevole alla luce della forte influenza che hanno non solo i valori osservati, ma anche i modelli previsivi candidati scelti. In particolare, scegliendo, come nell'applicazione attuale, modelli previsivi che possono essere "inglobati" da altri appartenenti all'insieme dei previsori candidati, il test può giudicare non significativamente rilevante l'informazione trasmessa da alcune previsioni, questo anche alla luce dei risultati ottenuti nel Capitolo 4, dove abbiamo visto

Tabella 5.1: MSFE dei metodi di combinazione delle previsioni, relativi a quello del metodo SA, per le 12 serie storiche considerate. Orizzonte previsivo di  $h=8$  periodi (2 anni).

	Macro	Micro	Demografica	Industriale	Finanziaria	Altro
AI p-value	0.603	0.433	0.321	0.572	0.096	0.060
AI-AFTER	0.787	0.035	0.853	1.029	0.997	1.211
AFTER	0.918	0.235	0.870	1.062	1.005	1.378
BG <sub>1</sub>	0.977	0.087	0.951	1.010	0.984	1.057
LR	1.607	0.041	1.233	5.181	1.087	8.576
CLR	0.838	0.034	0.855	1.059	0.955	1.236
AI p-value	0.059	0.969	0.782	0.078	0.411	0.991
AI-AFTER	0.584	0.505	0.074	1.109	0.372	0.146
AFTER	0.781	0.622	0.074	1.654	0.372	0.147
BG <sub>1</sub>	0.964	1.019	0.189	1.007	0.859	0.696
LR	0.466	0.974	0.019	0.849	55.602	0.165
CLR	0.878	0.787	0.077	0.935	0.468	0.147

che, quando la numerosità campionaria non è elevata, il test AI-AFTER tende ad essere più conservativo del dovuto in alcuni casi. Nonostante ciò, AI-AFTER continua a fornire risultati in termini di MSFE migliori (o almeno pari, ma solamente in 2 casi) a quelli di AFTER, confermando che la misura di salvaguardia adottata nel metodo AI-AFTER raggiunge effettivamente il risultato di garantire le *performance* previsive desiderate.

AI-AFTER, però, non performa competitivamente solo contro l'originale metodo AFTER, ma si dimostra una più che valida alternativa anche rispetto agli altri metodi di combinazione. Infatti, nelle applicazioni di tali metodi alle 12 serie storiche prese in considerazione, si classifica tra il primo e il secondo posto, in termini di MSFE relativo, ben 10 quando si considerano tutte le serie e 8 volte su 8 quando si considerano quelle in cui è stata selezionata la combinazione per l'adattamento; gli altri metodi, contrariamente ad AI-AFTER, non sono flessibili alle esigenze che vengono poste dallo scenario previsivo con cui si lavora e, di conseguenza, non riescono sempre a performare adeguatamente.

Infine, è stato possibile studiare in questa applicazione il fenomeno che abbiamo in

precedenza chiamato "*forecast combination puzzle*", così da capire se la media aritmetica semplice delle previsioni candidate (SA) potesse essere sufficientemente accurata per questi problemi previsivi. Dall'analisi condotta è apparso che, tramite l'applicazione dei metodi di combinazione previsiva, come il metodo AI-AFTER, la *performance* del metodo SA può essere (spesso anche sostanzialmente) migliorata, come è avvenuto nella quasi totalità dei casi considerati (11 casi su 12); questo sembra essere in linea con le recenti scoperte secondo cui alcuni metodi di combinazione più complicati possono in media sovraperformare significativamente quelli più semplici (Makridakis et al., 2020).



## Capitolo 6

# Conclusioni

Il presente studio si è posto l'obiettivo di approfondire i due principali approcci alla combinazione previsiva, con il fine di riuscire ad inserirsi all'interno dell'attivo dibattito che li riguarda.

Dai risultati ottenuti, in particolar modo durante gli studi di simulazione, è emerso chiaramente che le piene potenzialità della combinazione delle previsioni possono essere sfruttate solo se si è in grado di dedurre quale sia l'obiettivo più appropriato da raggiungere e di applicare di conseguenza il metodo di combinazione più adatto. Ciò non è possibile, quindi, affidandosi a procedimenti prefissati senza tenere conto dei rischi e dei costi che potrebbero comportare in determinate situazioni.

Non essendo nella pratica, tuttavia, in grado di conoscere a priori tutti i dettagli dello scenario previsivo in cui si adopera, come il processo generatore di dati o le previsioni candidate, la scelta del metodo di combinazione da utilizzare può risultare particolarmente critica e necessita di maggiori misure precauzionali.

Oltre agli studi di simulazione, anche l'applicazione sul dataset della M4-Competition ha consentito di mettere alla prova il metodo AI-AFTER, il quale sicuramente spicca per *performance* rispetto ai metodi che perseguono uno solo dei due scopi della combinazione. Grazie al suo approccio, infatti, AI-AFTER non solo ha dimostrato di essere in grado di determinare l'obiettivo più appropriato per la combinazione, ma è riuscito anche a combinare le previsioni in modo intelligente per raggiungere automaticamente l'obiettivo corretto così che le sue previsioni finali performassero in modo ottimale in entrambi gli scenari. Questo risultato è in linea con quelli evidenziati dagli ideatori del metodo, tuttavia va notato che il lavoro si è concentrato sulla situazione in cui gli errori di

previsione sono stazionari ed il rischio è calcolato utilizzando come funzione di perdita l'errore quadratico medio; sarebbe sicuramente interessante, in futuro, eseguire studi teorici e numerici sulle prestazioni relative della combinazione per l'adattamento o per il miglioramento sotto altre funzioni di perdita, non stazionarietà o in presenza di rotture strutturali.

Nel campo della previsione statistica ci sono ancora discussioni aperte riguardo quale approccio sia più vantaggioso perseguire tra quello per l'adattamento e quello per il miglioramento. Tuttavia, i risultati dello studio indicano che le metodologie che guardano ad entrambi con occhio oggettivo e ne studiano contemporaneamente l'efficacia, come il metodo AI-AFTER, sono quelle da favorire, in quanto portano ad un'accuratezza delle previsioni contro la quale è spesso difficile, per gli altri metodi, competere.



# Ringraziamenti

A chiusura di questo lavoro, vorrei dedicare questo spazio alle persone che mi hanno sostenuto ed aiutato durante questo percorso.

In primo luogo vorrei ringraziare la Professoressa Luisa Bisaglia, che mi ha seguita, durante la stesura della tesi, con immensa disponibilità e che mi ha trasmesso, durante le sue lezioni, una forte passione per la sua materia.

Il ringraziamento più grande va ai miei genitori, che con la loro guida e il loro incoraggiamento mi hanno sempre spronata ad essere la versione migliore di me stessa, e a mio fratello, che mi ha dato un motivo in più per esserlo.

Ringrazio tutte le persone a me più care che da sempre credono in me e che restando al mio fianco mi trasmettono forza ed affetto, con le quali ho festeggiato i traguardi raggiunti e condiviso momenti di crescita.

Ed infine, dedico questa tesi anche a me stessa, per aver dimostrato una tenacia che non credevo di avere e per non aver mai rinunciato a ciò che ritenevo importante.

Margherita Palomba



# Bibliografia

- Armstrong, J. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1), 69–80.
- Bates, J. & Granger, C. (1969). The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4), 451–468.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 26(2), 123–140.
- Breiman, L. (1998). Arcing Classifiers. *The Annals of Statistics*, 26(3), 801–824.
- Cheng, X. & Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2), 280–293.
- Claeskens, G., Magnus, J., Vasnev, A. & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3), 754–762.
- De Luca, G., Magnus, J. & Peracchi, F. (2018). Weighted-average least squares estimation of generalized linear models. *Journal of Econometrics*, 204(1), 1–17.
- Diebold, F. & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263.
- Forte, A., Garcia-Donato, G. & Steel, M. (2018). Methods and Tools for Bayesian Variable Selection and Model Averaging in Normal Linear Regression. *International Statistical Review*, 86(2), 237–258.
- Genre, V., Kenny, G., Meyler, A. & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1), 108–121.
- Granger, C. & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), 197–204.

- Hansen, B. (2008). Least-squares forecast averaging. *Journal of Econometrics*, 146(2), 342–350.
- Hendry, D. & Clements, M. (2004). Pooling of forecasts. *Econometrics Journal*, 7(1), 1–31.
- Hibon, M. & Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1), 15–24.
- Hoeting, J., Madigan, D., Raftery, A. & Volinsky, C. (1999a). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statistical Science*, 14(4), 382–417.
- Hoeting, J., Madigan, D., Raftery, A. & Volinsky, C. (1999b). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statistical Science*, 14(4), 382–417.
- Hsiao, C. & Wan, S. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178, 294–309.
- Hyndman, R. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*, 36(1), 7–14.
- Hyndman, R. & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Kourentzes, N., Barrow, D. & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, 209, 226–235.
- Lahiri, K., Peng, H. & Y., Z. (2017). Online learning and forecast combination in unbalanced panels. *Econometric Reviews*, 36(1-3), 257–288.
- LeBlanc, M. & Tibshirani, R. (1996). Combining Estiamates in Regression and Classification. *Journal of the American Statistical Association*, 91(436), 1641–1650.
- Makridakis, S. & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476.
- Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808.

- Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.
- Montero-Manso, P., Netto, C. & Talagala, T. (2018). M4comp2018: Data from the M4 Competition. *R package version 0.1. 0*.
- Qian, W., Rolling, C., Cheng, G. & Yang, Y. (2022). Combining forecasts for universally optimal performance. *International Journal of Forecasting*, *38*(1), 193–208.
- Qian, W., Rolling, C. & Cheng, Y., G. and Yang. (2019). On the Forecast Combination Puzzle. *Econometrics*, *7*(3).
- Radchenko, P., Vasnev, A. & Wang, W. (2021). Too similar to combine? On negative weights in forecast combination. *International Journal of Forecasting*.
- Schapire, R., Freund, Y., Bartlett, P. & Lee, W. (1998). Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *The Annals of Statistics*, *26*(5), 1651–1686.
- Smith, J. & Wallis, K. (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.
- Steel, M. (2011). Bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, *200*, 30–41.
- Stock, J. & Watson, M. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.
- Stock, J. & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.
- Wang, Z., Paterlini, S., Gao, F. & Yang, Y. (2014). Adaptive minimax regression estimation over sparse lq-hulls. *Journal of Machine Learning Research*, *15*, 1675–1711.
- Yang, Y. (2003). Regression with multiple candidate models: Selecting or mixing? *Statistica Sinica*, *13*(3), 783–809.
- Yang, Y. (2004). Combining forecasting procedures: Some theoretical results. *Econometric Theory*, *20*(1), 176–222.
- Zhang, X., Yu, D., Zou, G. & Liang, H. (2016). Optimal Model Averaging Estimation for Generalized Linear Models and Generalized Linear Mixed-Effects Models. *Journal of the American Statistical Association*, *111*(516), 1775–1790.

- Zhang, X. & Yu, J. (2018). Spatial weights matrix selection and model averaging for spatial autoregressive models. *Journal of Econometrics*, 203(1), 1–18.
- Zou, H. & Yang, Y. (2004). Combining time series models for forecasting. *International Journal of Forecasting*, 20(1), 69–84.