

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



**Il mercato elettrico italiano.  
Modelli statistici di previsione dei prezzi e della  
domanda.**

Relatore Prof. Luigi Grossi  
Dipartimento di Scienze Statistiche

Laureanda Sara Meneghetti  
Matricola 2022559

Anno Accademico 2022/2023



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Mercato energetico</b>	<b>5</b>
1.1 Liberalizzazione del mercato . . . . .	5
1.2 Mercato energetico italiano . . . . .	6
1.2.1 Funzionamento del mercato energetico italiano . . . . .	6
1.2.2 Zone di mercato . . . . .	8
1.3 Caratteristiche del mercato energetico italiano . . . . .	9
1.3.1 Picchi di prezzo . . . . .	9
1.3.2 Stagionalità . . . . .	10
1.3.3 Incontro di domanda e offerta . . . . .	10
<b>2 Metodologie</b>	<b>13</b>
2.1 Notazione . . . . .	14
2.2 Modelli . . . . .	15
2.2.1 Modello lineare autoregressivo . . . . .	15
2.2.2 Support vector machine . . . . .	17
2.2.3 Foresta casuale . . . . .	19
2.2.4 Gradient boosting . . . . .	20
2.3 Misure di importanza delle variabili . . . . .	22
2.4 Riconciliazione delle previsioni . . . . .	23
2.4.1 Approcci . . . . .	24
2.4.2 Notazione . . . . .	26
2.4.3 Riconciliazione ottimale puntuale . . . . .	30
2.5 Criteri di valutazione . . . . .	34
<b>3 Analisi empiriche</b>	<b>37</b>
3.1 Presentazione dei dati . . . . .	37
3.2 Risultati . . . . .	41
3.2.1 Analisi dei prezzi . . . . .	41
3.2.1.1 Performance previsionale dei modelli . . . . .	41
3.2.1.2 Importanza delle variabili . . . . .	44
3.2.2 Analisi dei consumi . . . . .	49
3.2.2.1 Performance previsionale dei modelli . . . . .	49

---

3.2.2.2	Riconciliazione delle previsioni . . . . .	52
<b>Conclusioni</b>		<b>54</b>
<b>Appendice</b>		<b>59</b>
A	Analisi dei prezzi: risultati basati sull'indice MAE . . . . .	59
B	Grafici di confronto delle previsioni dei prezzi . . . . .	62
C	Analisi dei consumi: risultati basati sull'indice MAE . . . . .	78
D	Riconciliazione delle previsioni: risultati basati sull'indice MAE . . . . .	81
<b>Bibliografia</b>		<b>85</b>





# Introduzione

La previsione dei prezzi e della domanda di elettricità è un argomento ampiamente esplorato negli ultimi due decenni a seguito del processo di liberalizzazione dei mercati energetici avviato alla fine del secolo scorso (Weron, 2014; Nowotarski & Weron, 2018). Gli attori che operano nel mercato elettrico come produttori, gestori delle reti, *traders*, grossisti e consumatori di grandi dimensioni sono interessati ai prezzi futuri e ai consumi del mercato energetico in quanto questi ultimi, tra loro interconnessi, influenzano le loro decisioni. In quest'ambito assumono quindi un ruolo rilevante la modellazione e la previsione delle serie storiche dei prezzi e dei consumi di energia elettrica.

Il mercato elettrico italiano (IPEX) è uno dei mercati maggiormente studiati al mondo per diversi motivi (Grossi & Nan, 2019). In primo luogo, IPEX rappresenta uno dei mercati elettrici più trasparenti al mondo (Fianu et al., 2022). Infatti, i dati sulla quantità e sul prezzo dell'offerta sono resi disponibili con un ritardo di due settimane contenenti informazioni dettagliate relative al gestore e al nome dell'impianto. In secondo luogo, IPEX è un tipico mercato zonale in cui si osserva un prezzo nazionale unico (PUN) quando non si verifica una scissione del mercato. In tutti i restanti periodi il PUN è ottenuto come media ponderata dei prezzi zionali osservati nelle zone reali e virtuali in cui è suddiviso il Paese. La struttura zonale dell'IPEX consente a ricercatori e professionisti di esplorare i principali vantaggi e svantaggi dell'integrazione dei mercati. Quest'ultimo rappresenta uno degli obiettivi perseguiti dall'Unione Europea (Grossi & Nan, 2019). L'impatto di possibili eventi di congestione della rete sui prezzi e la volatilità osservata sul mercato zonale italiano (Gianfreda & Grossi, 2012) potrebbero essere considerati, infatti, un esperimento su piccola scala degli effetti attesi sul mercato europeo integrato. La maggior parte degli articoli che si occupano di previsioni energetiche si basa sulla stima di modelli autoregressivi lineari (Grossi et al., 2022). Questi ultimi rappresentano il modo più semplice per approssimare il processo generatore dei dati delle serie temporali. Essi sono inoltre facilmente interpretabili e possono essere stimati da qualsiasi software statistico non essendo computazionalmente intensivi.

Tuttavia i modelli di apprendimento automatico hanno recentemente attirato grande attenzione in letteratura (Lago et al., 2018; Ghodducci et al., 2019; Lucas et al., 2020; Schnürch & Wagner, 2020). Le principali ragioni alla base di questo interesse sono l'elevata flessibilità e la possibilità di inserire un ampio numero di variabili esplicative (Marcjasz et al., 2019). Non risulta chiaro però se i metodi di *machine learning* per la previsione dei prezzi e dei consumi del mercato energetico abbiano prestazioni migliori rispetto a modelli più semplici. Alcuni risultati in letteratura dimostrano infatti che questi ultimi hanno prestazioni a volte anche migliori rispetto a modelli *black-box*.

Il presente lavoro di tesi si propone di confrontare, quindi, le performance predittive di tre modelli di apprendimento automatico (*Support Vector Machines* (SVM), *Random Forest* (RF) e *Gradient Boosting* (GB)) con quella di un modello autoregressivo lineare con variabili esogene (ARX), con e senza penalizzazione LASSO, utilizzando i dati osservati sul mercato elettrico italiano (IPEX).

Un primo obiettivo dell'analisi è prevedere i prezzi dell'energia elettrica attraverso queste metodologie. L'insieme dei regressori utilizzati in questo tipo di modelli è però estremamente variabile. Si rendono quindi necessari criteri per la selezione e il *ranking* dei regressori più predittivi. Oltre alla valutazione della performance previsive dei modelli, il presente lavoro si propone di valutare anche l'influenza delle variabili esplicative utilizzate nell'analisi dei prezzi, estendendo un indice d'importanza originariamente proposto nell'ambito dei modelli RF.

È inoltre interessante prevedere la domanda, applicando anche in questa seconda analisi sia i modelli lineari, sia i modelli più complessi citati in precedenza.

I dati in serie storiche sono caratterizzati da una naturale gerarchia temporale. In questi casi è ragionevole richiedere che le previsioni rispettino i vincoli di aggregazione derivanti da tale struttura, cioè che siano coerenti (Girolimetto, 2020). Risulta quindi utile approfondire la tecnica di riconciliazione temporale delle previsioni (*Forecast Reconciliation*, FR). L'efficienza di quest'ultima viene valutata applicandola alle previsioni dei consumi energetici.

Il presente elaborato è organizzato come segue: il Capitolo 1 descrive il processo di liberalizzazione da un regime di monopolio ad uno di concorrenza nelle fasi di produzione e vendita della filiera energetica. In questo capitolo viene descritto, inoltre, il mercato energetico italiano, delineandone il funzionamento, la suddivisione in zone e le principali caratteristiche. Nel Capitolo 2 vengono presentate le metodologie utilizzate. Vengono descritti i modelli adattati per la previsione dei prezzi e della domanda. Sono poi presentate le misure di importanza delle variabili attraverso le quali viene valutata l'influenza di ciascun regressore nell'analisi dei prezzi. Nel capitolo vengono inoltre



descritte le diverse procedure di riconciliazione delle previsioni presenti in letteratura. Infine sono presentati i criteri di valutazione utilizzati per confrontare i risultati ottenuti. Il Capitolo 3 riporta le analisi effettuate sui prezzi e sulla domanda confrontando le performance predittive dei modelli e valutando l'efficacia delle metodologie utilizzate.



# Capitolo 1

## Mercato energetico

### 1.1 Liberalizzazione del mercato

Negli anni Novanta del secolo scorso nel settore dell'energia ebbe luogo una trasformazione senza precedenti. Le cause di questa "rivoluzione" erano già emerse negli anni immediatamente precedenti, nei quali il progresso tecnologico e il mutare degli obiettivi politici e sociali a livello globale avevano determinato un dibattito economico. Il risultato fu l'idea che per produrre energia elettrica il modo migliore fosse ricorrere ad una forma di mercato di libera concorrenza.

Il graduale passaggio da un regime di monopolio ad uno di concorrenza, nelle fasi di produzione e vendita della filiera energetica, fu un processo che venne definito "liberalizzazione" del mercato dell'energia.

Nonostante lievi differenze nelle motivazioni ideologiche e politiche fra i vari paesi, in quegli anni si diffuse la convinzione che la liberalizzazione, riuscita con successo in altri settori, potesse essere duplicata anche in quello dell'energia (Weron, 2007). Era convinzione diffusa in quel periodo, infatti, che la libera concorrenza avrebbe portato ad un incremento dell'efficienza del settore elettrico e ad un aumento della competitività degli stati. I primi esperimenti in alcuni paesi europei ebbero risultati positivi in termini di costi inferiori e maggior possibilità di scelta per i consumatori.

Questo processo, cioè l'abolizione dei diritti esclusivi nel mercato dell'energia elettrica detenuti da un'unica azienda di Stato, fu giustificato con l'apertura a nuovi soggetti e a nuove tecnologie di produzione.

Oltre all'entrata di nuovi concorrenti e alla nascita di nuovi operatori (grossisti, produttori, distributori e consorzi), la liberalizzazione determinò anche altri cambiamenti.

L'efficienza del settore dell'energia non fu più affidata alla ricerca sistematica della corrispondenza dei costi marginali ai prezzi, ma all'operare di meccanismi di mercato e a forme di regolazione dello stesso più snelle con un orientamento del pubblico servizio (obblighi di fornitura, compatibilità ambientale ecc.) e delle tariffe verso formulazioni più semplici.

A partire dal 1991, questo progressivo processo di liberalizzazione iniziò anche in Italia. Fino a quel momento, infatti, ENEL (Ente Nazionale per l'energia Elettrica) deteneva la titolarità esclusiva delle "attività di produzione, importazione, esportazione, trasporto, trasmissione, distribuzione e vendita dell'energia elettrica sul territorio nazionale". Questa trasformazione fu motivata dal continuo crescere della domanda di energia elettrica e dal dubbio che ENEL non sarebbe stato in grado di soddisfarla a pieno.

Il trentennio di monopolio ENEL terminò nel 1992, ma una vera e propria liberalizzazione iniziò soltanto nel 1999 con l'entrata in vigore del Decreto Bersani, che rese completamente libere le attività di produzione, importazione, esportazione, acquisto e vendita di energia elettrica. Infine il Decreto Bersani Bis concluse il processo nel 2007, dando la possibilità anche ai clienti domestici di scegliere liberamente il proprio fornitore di energia.

## **1.2 Mercato energetico italiano**

La filiera del mercato energetico italiano è un sistema articolato e complesso, che si compone di diverse fasi quali la generazione, la trasmissione e il dispacciamento, la distribuzione e la vendita ai clienti finali. É pertanto necessario analizzare la struttura, il funzionamento e le principali caratteristiche del mercato.

### **1.2.1 Funzionamento del mercato energetico italiano**

Il mercato elettrico italiano, al fine di promuovere la competizione nelle attività di produzione e di compravendita di energia elettrica, si conforma ai principi di obiettività, neutralità e trasparenza.

La Figura 1.1 descrive la struttura del mercato italiano.

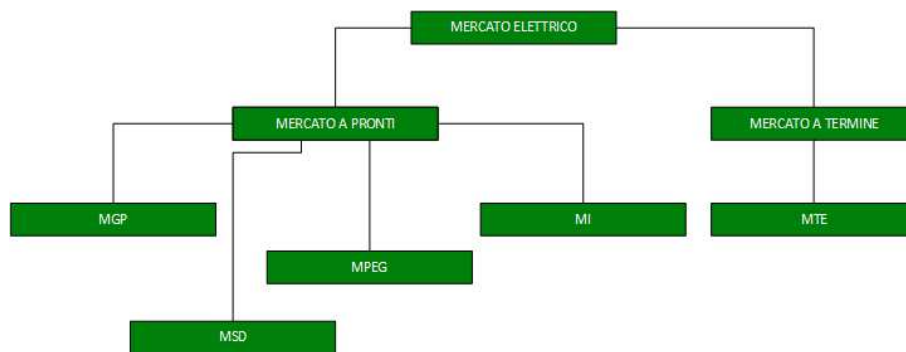


FIGURA 1.1: Organizzazione del mercato elettrico italiano (Fonte: GME S.p.A)

Questo si articola nel **Mercato elettrico a pronti (MPE)** e nel **Mercato elettrico a termine (MTE)**.

Il primo, anche denominato **mercato spot**, è suddiviso in quattro sezioni:

- **Mercato del Giorno Prima (MGP)** : sede di negoziazione delle offerte di acquisto e vendita di energia elettrica per ciascuna ora del giorno successivo, a cui possono partecipare tutti gli operatori elettrici;
- **Mercato Infragiornaliero (MI)**: permette agli operatori di apportare modifiche ai programmi definiti nel Mercato del Giorno Prima tramite ulteriori Offerte di acquisto o vendita;
- **Mercato dei prodotti giornalieri (MPEG)**: sede per la negoziazione dei prodotti giornalieri con obbligo di consegna dell'energia;
- **Mercato per il Servizio di Dispacciamento (MSD)**: sede di negoziazione delle Offerte di vendita e di acquisto di servizi di dispacciamento.

Il luogo virtuale in cui avviene l'incontro tra domanda e offerta per la compravendita dell'energia elettrica all'ingrosso è la **Borsa elettrica italiana (Italian Power Exchange - IPEX)**. Quest'ultima è uno dei mercati elettrici più trasparenti al mondo (Fianu et al., 2022). I microdati sulla quantità e sul prezzo dell'offerta sono resi disponibili con un ritardo di due settimane e contengono informazioni dettagliate relative al gestore e al nome dell'impianto. La gestione economica della borsa elettrica è affidata al Gestore dei mercati energetici.

Il **GME (Gestore Mercati Energetici)** è una società per azioni, controllata dal GSE (Gestore dei Servizi Energetici), alla quale è affidata la gestione economica del mercato

elettrico secondo criteri di trasparenza e obiettività. Esso, promuovendo la concorrenza tra i produttori, assicura la disponibilità di un adeguato livello di Riserva di potenza. In particolare il GME gestisce il Mercato del giorno prima dell'energia (MGP), il Mercato infragiornaliero (MI), il Mercato dei prodotti giornalieri (MPEG), il Mercato dei servizi di dispacciamento (MSD) e il Mercato a Termine Elettrico (MTE). Il GME rappresenta l'autorità di controllo centralizzato su tutti gli impianti di produzione e ne assicura il funzionamento nelle condizioni di massima sicurezza. Garantisce così la continuità e la qualità del servizio.

### 1.2.2 Zone di mercato

Il mercato elettrico italiano è suddiviso in zone. Fino al 31 dicembre 2020 queste erano Nord (NOR), Centro-Nord (CNOR), Centro-Sud (CSOU), Sud (SOU), Sicilia (SIC) e Sardegna (SAR). Da gennaio 2021 è stata modificata la struttura geografica del mercato: la regione Umbria è stata spostata dalla zona Centro-Nord alla zona Centro-Sud ed è stata introdotta una nuova macrozona, formata da una sola regione, la Calabria, che è stata separata dalla macrozona Sud. La distribuzione geografica fino a dicembre 2020 e da gennaio 2021 viene illustrata in Figura 1.2.

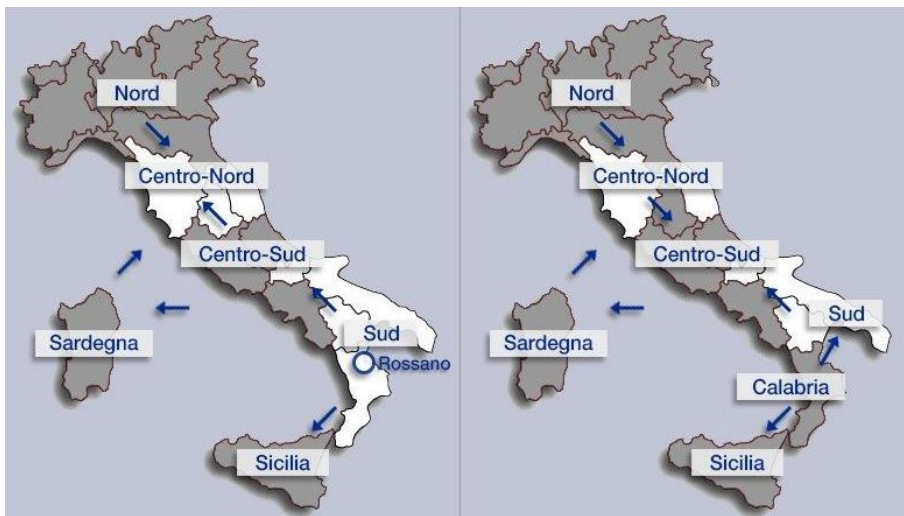


FIGURA 1.2: A sinistra le zone del mercato elettrico in vigore fino al 31 dicembre 2020 (con il polo di produzione di Rossano, in provincia di Catanzaro), a destra la nuova configurazione zonale dal 1 gennaio 2021 (Fonte: TERNA S.p.A)

Questa modifica, apparentemente tecnica, ha prodotto progressi significativi in termini di efficienza del mercato elettrico, oltre a migliorare l'integrazione e l'omogeneità con il mercato europeo. L'obiettivo di questa diversificazione è quello di cogliere correttamente l'andamento dei principali flussi di potenza a seconda delle condizioni di domanda e offerta, evitando i cosiddetti "colli di bottiglia" nella capacità di trasporto della rete. IPEX è un tipico mercato zonale in cui un prezzo di mercato nazionale unico (PUN) è ottenuto come media ponderata dei prezzi zionali osservati nelle zone reali e virtuali in cui è suddiviso il Paese. La struttura zonale dell'IPEX consente a ricercatori e professionisti di esplorare i principali vantaggi e svantaggi dell'integrazione dei mercati, che è un tema caldo in vista dell'integrazione dei mercati energetici europei perseguita dall'Unione Europea (Grossi & Nan, 2019). L'impatto di possibili eventi di congestione della rete sui prezzi e la volatilità osservata sul mercato zonale italiano potrebbero essere considerati come un esperimento su piccola scala degli effetti attesi sul mercato europeo integrato (Gianfreda & Grossi, 2012).

## 1.3 Caratteristiche del mercato energetico italiano

È necessario ora esaminare le cosiddette "regolarità empiriche" del mercato energetico italiano. Molte delle caratteristiche presentate sono universali, nel senso che sono condivise dalla maggior parte dei mercati spot dell'energia nel mondo (Weron, 2007).

### 1.3.1 Picchi di prezzo

Una delle caratteristiche principali dei mercati elettrici sono le repentine e generalmente imprevedibili variazioni estreme dei prezzi spot, note come salti o **picchi**. In un brevissimo periodo di tempo il prezzo del sistema può aumentare notevolmente e poi tornare al livello precedente.

L'intensità del picco non è omogenea nel tempo. I picchi sono particolarmente elevati durante le ore di punta, cioè intorno alle 9 e alle 18 nei giorni lavorativi e durante i periodi di forte consumo.

La natura "spigolosa" dei prezzi spot rappresenta l'effetto della non immagazzinabilità dell'energia elettrica. L'elettricità da consegnare ad un'ora specifica non può essere sostituita con l'elettricità disponibile poco dopo o poco prima. Dal momento che attualmente non c'è una tecnologia efficiente per immagazzinare grandi quantità di energia, quest'ultima deve essere consumata contestualmente alla sua produzione. Pertanto, le fluttuazioni di carico estreme, causate da condizioni meteorologiche avverse spesso in combinazione con interruzioni di produzione o guasti di trasmissione, possono portare

a picchi di prezzo. Quest'ultimi sono normalmente di breve durata e non appena il fenomeno meteorologico o l'interruzione del servizio è terminato, i prezzi tornano ad un livello normale.

### 1.3.2 Stagionalità

È noto che la domanda di energia elettrica presenta fluttuazioni stagionali. Queste si verificano principalmente a causa del cambiamento delle condizioni climatiche, come la temperatura, o del mutamento del numero di ore giornaliere di luce.

Va sottolineato, tuttavia, che a volte anche l'offerta di energia elettrica evidenzia variazioni stagionali della produzione. La produzione delle unità idroelettriche, ad esempio, è fortemente dipendente dalle precipitazioni e dallo scioglimento delle nevi, che varia da stagione a stagione.

La stagionalità può essere anche infragiornaliera, settimanale o mensile. La domanda di elettricità, ad esempio, non è uniforme per tutta la settimana: è maggiore durante l'orario di lavoro nei giorni feriali ed è bassa durante la notte e i fine settimana (a causa della scarsa attività industriale).

Queste fluttuazioni della domanda e dell'offerta si traducono in un comportamento stagionale dei prezzi dell'elettricità.

### 1.3.3 Incontro di domanda e offerta

Per comprendere meglio le caratteristiche principali del mercato energetico italiano è necessario analizzare la domanda e l'offerta che inevitabilmente influenzano l'andamento dei prezzi.

Il ruolo di una Borsa elettrica è quello di determinare giornalmente un prezzo di compensazione (o equilibrio) del mercato (MCP), che viene annunciato pubblicamente e comunemente noto come **prezzo spot**. Il processo di determinazione prevede che le offerte valide e congrue ricevute vengano ordinate per prezzo crescente in una curva di offerta aggregata e le offerte di acquisto vengano ordinate per prezzo decrescente in una curva di domanda aggregata. L'intersezione delle due curve determina il prezzo di equilibrio e la quantità complessivamente scambiata, come illustrato in Figura 1.3.

Il prezzo spot, risultante dall'intersezione tra domanda e offerta, non è molto sensibile agli spostamenti della domanda quando questa è bassa. Lo stack di fornitura è, infatti, in genere piatto nella regione a bassa domanda. Tuttavia, quando la domanda è alta e una frazione maggiore di energia proviene da fonti costose, anche un piccolo aumento nei consumi può costringere i prezzi a salire vertiginosamente.



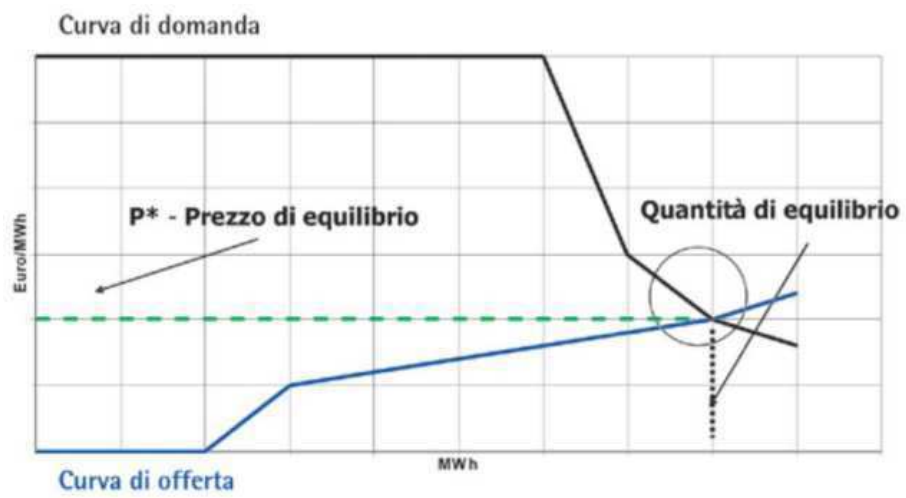


FIGURA 1.3: Determinazione del prezzo di equilibrio (Fonte: ENEA)



# Capitolo 2

## Metodologie

Nel seguente capitolo vengono presentati i metodi e i modelli utilizzati per la previsione dei prezzi e della domanda di energia elettrica. Si considera come modello di riferimento quello autoregressivo lineare con variabili esogene, con e senza interazioni tra predittori. La scelta del *benchmark* viene determinata da diversi fattori. In primo luogo, i modelli lineari autoregressivi rappresentano il modo più semplice per approssimare il processo generatore dei dati delle serie temporali. In secondo luogo, i parametri possono essere opportunamente interpretati in quanto misurano la correlazione parziale di ciascun regressore con la variabile dipendente. I modelli lineari, infine, non sono computazionalmente intensivi e possono essere stimati utilizzando qualsiasi software statistico. Per questi motivi la maggior parte degli articoli che si occupano di previsioni energetiche si basa sulla stima di modelli autoregressivi lineari (Grossi et al., 2022).

Vengono poi descritti alcuni modelli non parametrici tratti dalla letteratura sull'apprendimento automatico. Un approccio non parametrico, a differenza di uno parametrico, non prevede alcuna ipotesi esplicita sulla forma funzionale della funzione  $f$  da stimare. Viene stimata la funzione che si avvicina il più possibile alle osservazioni. Assumere una forma parametrica semplifica il problema poiché i modelli sono facilmente interpretabili e richiedono meno osservazioni rispetto a quelli non parametrici. Tuttavia, se il modello non è correttamente ipotizzato la stima non sarà accurata (Hastie et al., 2013). L'utilizzo di modelli più flessibili che possono adattarsi a diverse forme funzionali possibili di  $f$  comporta la stima di un maggior numero di parametri e l'elevata complessità del modello può portare ad un problema di sovradattamento (*overfitting*). Questo fenomeno si verifica perché il modello segue troppo i dati scambiando piccole regolarità locali dovute al caso per un comportamento sistematico (Azzalini & Scarpa, 2009).

I modelli non parametrici che vengono discussi di seguito sono *support vector machine*, foreste casuali e *gradient boosting*, i quali acquisiscono automaticamente le interazioni

rilevanti tra le variabili.

Dato l'elevato numero di predittori, i modelli autoregressivi vengono stimati anche utilizzando la regolarizzazione LASSO. Quest'ultima migliora le previsioni in presenza di un elevato numero di regressori e opera una selezione delle variabili più rilevanti.

Uno degli obiettivi del presente lavoro di tesi è analizzare, infatti, l'influenza dei predittori nel prevedere i prezzi dell'energia elettrica. Vengono quindi descritte delle misure che permettono di ordinare le variabili in termini di importanza. Vengono poi presentati i criteri di valutazione utilizzati per operare il confronto tra i modelli adattati.

Viene infine esposta la tecnica di riconciliazione delle previsioni. Quest'ultima permette di ottenere delle previsioni coerenti, ovvero che rispettano i vincoli di aggregazione derivanti dalla struttura gerarchica temporale dei dati storici.

## 2.1 Notazione

Di seguito viene formalizzato il problema e specificata la notazione adottata seguendo Grossi et al., 2022.

Sia  $Y_t$  il processo generatore dei dati (*Data Generating Process*, DGP) che ha prodotto la serie storica univariata  $\{y_1, y_2, \dots, y_T\}$ . Si assume che l'incognita DGP sia una funzione possibilmente non lineare  $f(\cdot)$  di  $p$  stati passati di  $Y_t$  e  $m$  variabili esogene più un errore casuale.

Il DGP può essere formalizzato come:

$$y_t = f(\mathbf{r}_{t-1}, \mathbf{z}_t) + a_t \quad (2.1)$$

dove  $a_t$  è il termine di disturbo,  $\mathbf{r}_{t-1} = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$  è il vettore dei ritardi e  $\mathbf{z}_t$  è il vettore di variabili esogene disponibili al tempo  $t$ . Le serie temporali osservate di lunghezza  $T$ ,  $\{y_1, y_2, \dots, y_T\}$ , possono essere riorganizzate in un insieme di  $T - p$  coppie di tipo  $\{(y_t, \mathbf{r}_{t-1})\}_{t=p+1}^T$ , producendo la seguente matrice

$$Y = \begin{bmatrix} y_{p+1} & y_p & \dots & y_2 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_t & y_{t-1} & \dots & y_{t-p+1} & y_{t-p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_T & y_{T-1} & \dots & y_{T-p+1} & y_{T-p} \end{bmatrix}$$

che forma le prime  $p + 1$  colonne della matrice di dati  $D = [Y : Z]$  di dimensione  $(T - p) \times (p + m + 1)$ , dove  $Z$  è la matrice  $(T - p) \times m$  delle variabili esogene.

## 2.2 Modelli

Di seguito vengono introdotti i modelli che saranno utilizzati per prevedere i prezzi e i consumi di energia elettrica: modelli autoregressivi lineari, modelli autoregressivi con regolarizzazione LASSO, *support vector machine*, foreste casuali e *gradient boosting*.

### 2.2.1 Modello lineare autoregressivo

Viene preso in considerazione come modello di riferimento quello Autoregressivo con variabili esogene (ARX), determinato dalla specificazione della funzione  $f(\cdot)$  in (2.1). I modelli autoregressivi sono impiegati in diversi ambiti e, come detto, sono spesso usati come *benchmark* per approcci più sofisticati. La formulazione generale di un ARX( $p, g_1, \dots, g_m$ ) con  $m$  variabili esogene è la seguente:

$$y_t = c + \sum_{i=1}^p \phi y_{t-i} + \sum_{i=1}^m \Lambda^i(B) z_t^i + a_t \quad (2.2)$$

dove  $a_t$  è il termine di errore con media zero e  $\Lambda^i(B) = \lambda_0^i + \lambda_1^i B + \dots + \lambda_{g_i}^i B^{g_i}$ , con  $B$  operatore ritardo. Si tratta di una formulazione generale che consente alle variabili esogene di essere ritardate. L'equazione (2.2) implica un modello lineare in cui i regressori sono l'insieme dei ritardi di  $y_t$ ,  $\mathbf{r}_{t-1}$ , e le  $m$  variabili esogene  $\mathbf{z}_t = [z_{1t}, z_{2t}, \dots, z_{mt}]$ . Quindi, per il generico intervallo di tempo  $t$ , con  $t = p + 1, p + 2, \dots, T$ , i regressori sono  $\mathbf{x}_t = \{\mathbf{r}_{t-1}, \mathbf{z}_t\}$ .

Il modello specificato nell'equazione (2.2) contiene un gran numero di covariate. Risulta quindi necessario ridurre la varianza delle stime dei parametri e controllare l'*overfitting* adattando una versione regolarizzata dello stesso modello. Viene di conseguenza preso in considerazione il modello ARX con regolarizzazione LASSO, indicato come modello ARX-L (Tibshirani, 1996; Hastie et al., 2009).

La procedura LASSO è uno degli approcci più popolari per la selezione del modello nell'ambito di modelli lineari e lineari generalizzati. Esso rientra nei cosiddetti metodi di regolarizzazione (*shrinkage*): attraverso l'uso di vincoli e la riduzione dei coefficienti stimati verso lo zero è possibile ridurre sostanzialmente la varianza delle stime a costo di un trascurabile aumento della distorsione. Questo metodo di regolarizzazione ha il vantaggio di eseguire simultaneamente la selezione e la stima del modello. Esso si è dimostrato efficace anche in contesti dimensionali elevati dove la dimensione dello spazio

dei parametri cresce con la dimensione del campione  $n$ . In un contesto di modellazione autoregressiva le caratteristiche del LASSO diventano particolarmente vantaggiose in quanto sia l'ordine del processo che i corrispondenti coefficienti autoregressivi possono essere stimati simultaneamente. Tale procedura gode inoltre di buone proprietà asintotiche. In particolare la consistenza della selezione del modello, la consistenza della stima e la coerenza della previsione valgono se il ritardo massimo  $p$  cresce con  $n$  (Nardi & Rinaldo, 2011).

Le stime dei coefficienti si ottengono minimizzando

$$\frac{1}{2} \sum_{t=p+1}^T (y_t - \mathbf{x}_t^\top \boldsymbol{\beta} + \beta_0)^2 + \lambda \|\boldsymbol{\beta}\|_1 = RSS + \lambda \|\boldsymbol{\beta}\|_1 \quad (2.3)$$

Il termine di penalizzazione introdotto permette di ridurre le stime dei coefficienti verso lo zero e, per un parametro di regolarizzazione  $\lambda$  sufficientemente grande, alcune stime vengono poste esattamente uguali a zero. Si noti che se  $\lambda \rightarrow 0$  viene fornito lo stimatore ai minimi quadrati.

La Figura 2.1 illustra la funzione di errore dello stimatore ai minimi quadrati  $\hat{\beta}$  e la regione corrispondente al vincolo del LASSO (area blu),  $|\beta_1| + |\beta_2| \leq s$ , con  $s$  parametro di regolazione (parametrizzazione di  $\lambda$ ). Se  $s$  è sufficientemente grande, cioè  $\lambda = 0$ , la regione delimitata dal vincolo conterrà  $\hat{\beta}$  e le stime della regressione LASSO saranno le stesse della regressione ai minimi quadrati. Le ellissi centrate attorno a  $\hat{\beta}$  rappresentano regioni con RSS costante. In altre parole, tutti i punti su una data ellisse hanno un valore comune di RSS. Man mano che le ellissi si allontanano dalle stime dei coefficienti l'RSS aumenta.

La soluzione al problema di minimo è data dal punto in cui un'ellisse contatta la regione del vincolo. Quest'ultima presenta dei punti angolosi in corrispondenza degli assi. Di conseguenza alcune stime sono poste esattamente uguali a zero. La penalizzazione LASSO permette in questo modo, data la natura del vincolo, di operare una selezione delle variabili. In Figura 2.1 l'intersezione si verifica in  $\beta_2 = 0$  quindi il modello risultante includerà solo  $\beta_1$ .

Si considera inoltre una seconda modifica del modello ARX, la quale tiene conto non solo dei regressori, ma anche delle interazioni tra ciascuna coppia di essi. Si esegue sempre una selezione di variabili esplicative attraverso la regressione dei minimi quadrati lineari regolati da LASSO.

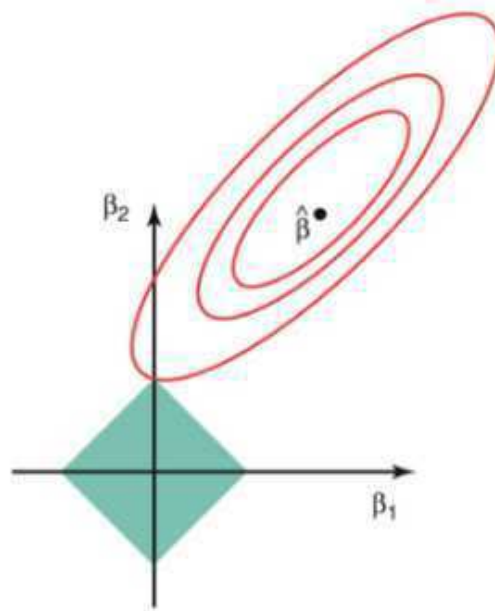


FIGURA 2.1: Funzione di errore dello stimatore ai minimi quadrati e regione corrispondente al vincolo del LASSO (Fonte: Hastie, T., R. Tibshirani, and J. Friedman (2009). *Introduzione all'apprendimento statistico con applicazioni in R*)

## 2.2.2 Support vector machine

*Support Vector Machine* (SMV) è un modello di *machine learning* nato per la classificazione (Cortes & Vapnik, 1995) e poi adattato alla regressione (Drucker et al., 1996; Vapnik & Chapelle, 2000).

SMV rappresenta una generalizzazione di un classificatore semplice e intuitivo chiamato classificatore a margine massimo. Quest'ultimo è basato sul concetto di iperpiano separante ottimale. L'idea è quella di trovare una superficie di separazione delle osservazioni che minimizzi la sua distanza dal punto più vicino in ciascuna classe. Associate a questa retta di massima separazione si individuano altre due rette, parallele alla prima e passanti per il punto più prossimo di ciascuna delle due classi. I punti su tali bande di separazione sono detti vettori di supporto. In questo modo l'iperpiano viene definito sulla base di un sottinsieme potenzialmente piccolo di osservazioni.

Il classificatore a margine massimo è un approccio naturale per la classificazione. Tuttavia, quest'ultimo non può essere sempre utilizzato in quanto spesso le classi non risultano perfettamente separabili da un iperpiano.

Questo classificatore può essere esteso con l'obiettivo di avere maggiore robustezza rispetto alle singole osservazioni e una migliore classificazione della maggior parte di esse. In questo contesto viene usato il concetto di margine morbido il quale prevede che alcune

osservazioni si trovino sul lato errato del margine, o addirittura sul lato errato dell'iperpiano. Questa generalizzazione è nota come classificatore *support vector* e può essere applicata in una gamma più ampia di casi. Più il margine è ampio, più osservazioni vengono classificate erroneamente e minore è l'adattamento ai dati. Questo equivale ad ottenere un classificatore potenzialmente più distorto ma che allo stesso tempo può presentare una varianza inferiore, controllando quindi l'*overfitting*. In questo caso le osservazioni che influiscono sul classificatore, i cosiddetti vettori di supporto, sono quelle che si trovano sul margine e dal lato sbagliato di esso. Questo significa che il classificatore è abbastanza robusto rispetto al comportamento di osservazioni che sono lontane dall'iperpiano.

Un'ulteriore estensione è SVM, il quale permette di trattare i casi di confini non lineari tra le classi. Per affrontare il problema si considera l'ampliamento dello spazio delle variabili utilizzando funzioni dei predittori.

Si ha la seguente funzione di perdita regolarizzata:

$$\sum_{t=p+1}^T V_{\epsilon}(y_t - f(\mathbf{x}_t)) + \frac{\lambda}{2} \|\beta\|_2^2 \quad (2.4)$$

con

$$V_{\epsilon}(e) = \begin{cases} 0 & \text{se } |e| < \epsilon \\ |e| - \epsilon & \text{altrimenti} \end{cases} \quad (2.5)$$

È possibile approssimare  $f(\cdot)$  utilizzando trasformazioni delle variabili esplicative:

$$f(\mathbf{x}) \approx \beta_0 + \sum_{s=1}^S \beta_s h_s(\mathbf{x}) \quad (2.6)$$

dove  $h_s(\cdot)$  sono funzioni di base e  $S$  è il numero di funzioni di base utilizzate. La funzione stimata risulta

$$\hat{f}(\mathbf{x}) \approx \hat{\beta}_0 + \sum_{s=1}^S \beta_s \langle h_s(\mathbf{x}_i), h_s(\mathbf{x}_j) \rangle \quad (2.7)$$

dove  $\langle \cdot, \cdot \rangle$  indica il prodotto interno. I regressori  $\mathbf{x}_i$  entrano in questa formulazione solo attraverso i prodotti interni della forma  $\langle h_s(\mathbf{x}_i), h_s(\mathbf{x}_j) \rangle$ . Analogamente quando  $f(\cdot)$  è lineare, si può mostrare che i regressori  $\mathbf{x}_i$  entrano nella soluzione di minimizzazione solo attraverso i prodotti interni

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \mathbf{x}_i^{\top} \mathbf{x}_j \quad (2.8)$$



La specificazione delle funzioni di base può quindi avvenire attraverso la funzione *kernel*

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle h_s(\mathbf{x}_i), h_s(\mathbf{x}_j) \rangle \quad (2.9)$$

che calcola i prodotti interni nello spazio delle variabili trasformate. In questa applicazione viene utilizzato il *kernel* radiale:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (2.10)$$

Il vantaggio di usare un *kernel*, piuttosto che allargare lo spazio delle variabili usando funzioni delle variabili originali, è computazionale. Usando i *kernel* non è necessario lavorare esplicitamente nello spazio delle variabili allargato. Per il *kernel* radiale lo spazio delle variabili è implicito e di dimensione infinita.

Ci si aspetta che SVM sia in grado di catturare la dipendenza reciproca tra i pattern settimanali e annuali in quanto l'espansione attraverso le funzioni di base introduce interazioni tra i regressori.

### 2.2.3 Foresta casuale

Un altro approccio di *machine learning* considerato è la foresta casuale (Random Forest (RF); Breiman, 2001).

Il modello RF è una modifica di una procedura denominata *bagging* o *bootstrap aggregation*: si adattano molti alberi ramificati a versioni *bootstrap* dei dati e si effettua una media delle previsioni ottenute.

Un albero decisionale (Breiman et al., 1984) è un modello di apprendimento automatico che partiziona ricorsivamente lo spazio dei dati generato dalle variabili esplicative in regioni distinte e non sovrapposte, identificate dai nodi. L'albero viene fatto crescere procedendo quindi per suddivisioni successive, selezionando quella che porta maggior guadagno in termini di abbassamento della devianza. Il processo continua fino a quando non viene raggiunto un criterio di arresto, ad esempio quando tutte le foglie, ovvero i nodi terminali, hanno numerosità al di sotto di una soglia minima prefissata. L'albero viene poi potato e quindi ridotto progressivamente secondo un ulteriore criterio.

Dal momento che nell'algoritmo di crescita ogni suddivisione è condizionata alla precedente, gli alberi sono modelli instabili rispetto a piccole perturbazioni dei dati. Il *bagging* quindi, nel processo di mediare le previsioni ottenute dall'adattamento di modelli più o meno indipendenti tra loro, riduce la varianza senza aumentare la distorsione, affrontando direttamente il problema del sovradattamento. L'aggregazione di diversi alberi

decisionali permette inoltre di aumentare l'accuratezza predittiva.

A differenza del *bagging*, RF considera diversi sottoinsiemi delle variabili esplicative ad ogni iterazione. La procedura consiste quindi nel selezionare in modo casuale, ad ogni nodo di un albero, un numero  $F$  di variabili esplicative che vengono ispezionate per trovare il punto di suddivisione ottimale secondo il criterio di crescita utilizzato. Oltre ad evitare problemi di sovraddattamento, questo modello presenta un vantaggio di velocità computazionale rispetto al *bagging*.

I parametri di regolazione del modello sono il numero di alberi  $N_t$  che formano la foresta e il numero di covariate  $F$  da selezionare in ciascun nodo. Si può dimostrare che l'errore globale converge ad una soglia inferiore al crescere di  $N_t$  e che non si presentano problemi di sovraddattamento quando vengono aggiunti ulteriori alberi. Scegliendo quindi un numero sufficientemente alto, l'errore di previsione che ne deriva non sarà molto distante dal suo limite inferiore. Il numero di variabili esplicative  $F$ , invece, viene generalmente mantenuto costante in tutti i nodi. Una scelta tipica è  $k/3$ , dove  $k$  è il numero di variabili esplicative considerato.

RF è originato al di fuori dell'ambito delle serie temporali. Si rende quindi necessario motivarne l'applicazione in questo contesto. Il *bootstrap* ordinario è adatto per osservazioni indipendenti ed identicamente distribuite e non per le serie temporali. Tuttavia, nel caso di serie temporali si presume che tutta la memoria del passato relativa a  $y_t$ , richiesta per una previsione un passo in avanti, sia conservata nei vettori ritardati  $\mathbf{r}_{t-1}$ . Sotto questa ipotesi le unità da campionare non sono le singole  $y_t$ , con le relative variabili esogene, bensì le righe della matrice  $\mathbf{D}$ . In alternativa, una procedura che può essere applicata ad osservazioni dipendenti è il *Moving Block Bootstrap* (MBB; Kunsch, 1989) che consiste nel ricampionamento di blocchi di osservazioni consecutive con reinserimento. Ci si può quindi limitare al ricampionamento delle righe della sottomatrice  $\mathbf{Y}$ , dove la lunghezza del blocco è uguale a  $p + 1$  e i blocchi sovrapposti sono i vettori  $[y_t, \mathbf{r}_{t-1}]$ .

### 2.2.4 Gradient boosting

Il *gradient boosting* (Friedman, 2001) è un modello di *machine learning* che combina la tecnica del *boosting* con la tecnica di discesa del gradiente (*gradient descent*). Il *boosting*, come la foresta casuale, consiste nella combinazione di diversi alberi. L'idea è quella di assegnare maggior peso alle osservazioni che sono state mal classificate con il fine di far lavorare maggiormente il classificatore su questi punti. Il *boosting* rappresenta un approccio che "impara lentamente" facendo crescere gli alberi in sequenza: ogni albero viene fatto crescere utilizzando le informazioni provenienti dagli alberi cresciuti

in precedenza. Si mediano quindi tanti alberi cresciuti su versioni pesate dei dati di stima originali.

Considerando una funzione di perdita  $L(y, f(\mathbf{x}))$ , si vuole minimizzare

$$J = \sum_i L(y, f(\mathbf{x})) \quad (2.11)$$

Si può mostrare che è possibile interpretare i residui come gradienti negativi

$$\frac{\partial J}{\partial f(\mathbf{x}_i)} = \frac{\partial \sum_i L(y, f(\mathbf{x}))}{\partial f(\mathbf{x}_i)} = \frac{\partial L(y, f(\mathbf{x}))}{\partial f(\mathbf{x}_i)} = f(\mathbf{x}_i) - y_i \quad (2.12)$$

$$-g(\mathbf{x}_i) = - \left[ \frac{\partial L(y, f(\mathbf{x}))}{\partial f(\mathbf{x}_i)} \right] = y_i - f(\mathbf{x}_i) \quad (2.13)$$

La tecnica del *gradient descent* minimizza una funzione muovendosi nella direzione opposta al gradiente. A partire da una soluzione iniziale  $\theta_0$  si procede iterativamente con aggiornamento

$$\theta_{m+1} = \theta_m - \rho \frac{\partial J}{\partial \theta_m} \quad (2.14)$$

In breve l'algoritmo può essere descritto come segue.

Dopo aver inizializzato il modello con un valore costante

$$f_0(\mathbf{x}) = \arg \min_{\gamma} \frac{1}{n} \sum_{i=1}^n L(y_i, \gamma) \quad (2.15)$$

ad ogni iterazione  $1 < m < M$  si calcolano i gradienti negativi. Si stima un albero di regressione  $h_m(\mathbf{x})$  su  $-g(\mathbf{x}_i)$  assegnando regioni terminali  $R_{jm}$ ,  $j = 1, \dots, J_m$ . Si calcola

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{\mathbf{x}_i \in R_{jm}} L(y_i, f_{m-1}(\mathbf{x}_i) + \gamma) \quad (2.16)$$

e si aggiorna il modello

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \sum_{j=1}^{J_m} \gamma_{jm} I(\mathbf{x} \in R_{jm}) \quad (2.17)$$

Anche nel caso del *gradient boosting* è possibile introdurre delle tecniche di regolarizzazione per ridurre il rischio di sovradattamento. La regola di aggiornamento viene

modificata in questo modo

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \lambda \cdot \sum_{j=1}^{J_m} \gamma_{jm} I(\mathbf{x} \in R_{jm}) \quad (2.18)$$

Il parametro  $0 < \lambda < 1$  controlla il "tasso di apprendimento" della procedura di *boosting*. Esiste un *trade-off* tra  $\lambda$  e  $M$ : più piccoli sono i valori di  $\lambda$  maggiore è lo *shrinkage* e maggiore è  $M$ .

Il *gradient boosting* presenta diversi vantaggi: permette l'utilizzo di dati di natura mista, è robusto rispetto alla presenza di *outlier* nei dati di input, presenta una buona capacità previsiva ed è un modello interpretabile.

## 2.3 Misure di importanza delle variabili

Come già detto in precedenza, quest'analisi mira ad identificare le variabili importanti, cioè quelle che maggiormente influenzano le previsioni dei prezzi del mercato energetico. È possibile ottenere informazioni riguardanti il ruolo esplicativo svolto dai regressori tramite una misura di importanza delle variabili (*Variable Importance Measure*, VIM). In questo contesto si possono distinguere due tipologie di metodi: metodi *model-specific*, che utilizzano elementi specifici della struttura del modello considerato, e metodi *model-agnostic*, ovvero indipendenti dal modello (Biecek & Burzykowski, 2021).

Il metodo descritto di seguito rientra nella prima tipologia ed è applicato nel contesto di combinazione di alberi nel quale rientrano RF e GB. Seguendo Breiman et al., 1984, sia  $\{X_1, X_2, \dots, X_k\}$  l'insieme di variabili predittive per la variabile risposta  $Y$ . L'importanza di una variabile  $X_j \in \{X_1, X_2, \dots, X_k\}$  nello spiegare  $Y$  è definita come la diminuzione totale dell'eterogeneità di  $Y$  data dalla conoscenza di  $\{X_1, X_2, \dots, X_k\}$  quando lo spazio dei regressori è partizionato ricorsivamente. Da questo approccio deriva una misura di importanza delle variabili chiamata *Total Decrease in Node Impurity* (TDNI). Tale misura è ottenuta dalla somma di tutte le diminuzioni dell'indice di eterogeneità nei nodi dell'albero.

Data la diversa natura delle variabili che formano  $\{X_1, X_2, \dots, X_k\}$ , è necessario indagare se la natura di un regressore incide nella valutazione della sua importanza. Breiman et al., 1984, hanno osservato che le misure TDNI tendono a favorire le covariate con più valori: sono preferite, ad esempio, le variabili continue rispetto alle variabili categoriali. Occorre di conseguenza applicare una correzione che superi questo limite. In questo lavoro verrà applicata la correzione euristica proposta da Sandri & Zuccolotto, 2008, 2010, che viene riportata di seguito.

Sia  $\mathbf{X}$  la matrice ( $N \times p$ ) contenente gli  $N$  valori osservati delle  $p$  covariate  $X_1, \dots, X_p$ . Si genera un insieme di matrici  $\{Z_s^S\}_{s=1}^S$  permutando casualmente  $S$  volte le  $N$  righe di  $\mathbf{X}$ . Le colonne di  $Z$  sono chiamate "pseudocovariate". La permutazione di riga distrugge l'associazione esistente tra la variabile risposta  $Y$  e ciascuna pseudocovariata  $Z_i$ . Al contrario viene preservata l'associazione tra due pseudocovariate  $Z_i$  e  $Z_j$ , uguale all'associazione esistente tra  $X_i$  e  $X_j$ . Sandri & Zuccolotto, 2008, 2010, hanno dimostrato che questo metodo può ridurre efficacemente la distorsione dovuta a diversi livelli di misurazione delle covariate, producendo un corretto ordinamento delle stesse secondo la loro importanza.

La rilevanza dei predittori negli altri modelli implementati viene invece valutata utilizzando un metodo *model-agnostic* basato su permutazioni delle variabili (Biecek & Burzykowski, 2021). Quest'ultimo si basa sulla funzione di perdita  $L(\hat{y}, X, y)$ , la quale quantifica la bontà di adattamento del modello in uso, dove  $X$  indica la matrice che contiene i valori osservati delle covariate,  $y$  denota i valori osservati di  $Y$  e  $\hat{y}$  denota le previsioni corrispondenti. Questa funzione di perdita può essere il valore di log-verosimiglianza o qualsiasi altra misura delle prestazioni del modello, ad esempio l'errore quadratico medio.

Il metodo può essere riassunto nei seguenti passaggi:

1. Si calcolano le previsioni sull'insieme dei dati originali e la funzione di perdita  $L_0$  per ogni regressore  $X_j$ ;
2. Si modifica la matrice  $X$  permutando casualmente i valori della  $j$ -esima variabile, ottenendo una nuova  $X_j$ ;
3. Si calcolano le nuove previsioni  $\hat{y}_j$  sull'insieme dei dati permutati;
4. Si calcola la funzione di perdita  $L_j$ ;
5. Si quantifica l'importanza di  $X_j$  utilizzando direttamente  $L_j$  oppure calcolando  $vim_D^j = L_j - L_0$  o  $vim_R^j = L_j/L_0$ .

In generale questa procedura viene eseguita più volte per far sì che il risultato non dipenda da particolari permutazioni osservate. Di conseguenza la misura di importanza delle variabili finale è una misura di importanza media.

## 2.4 Riconciliazione delle previsioni

In diversi contesti le serie storiche possono presentare osservazioni che sono organizzate secondo una struttura gerarchica.

Una serie storica gerarchica è una serie storica multipla linearmente vincolata che comprende una raccolta di serie storiche che seguono una struttura di aggregazione (Panagiotelis et al., 2021). È ragionevole quindi, in questo contesto, richiedere che le previsioni rispettino i vincoli di aggregazione derivanti dalla struttura gerarchica, cioè che siano coerenti (*coherent*, Athanasopoulos et al., 2020).

La riconciliazione delle previsioni (*Forecast Reconciliation*, FR) è un processo mediante il quale le previsioni generate in modo indipendente da serie storiche gerarchiche vengono riconciliate per far sì che i livelli si "sommino" allo stesso modo dei dati storici. Di conseguenza viene rispettata la proprietà di coerenza. Tramite un approccio FR è quindi possibile prevedere le serie storiche individualmente preservando anche le relazioni all'interno della struttura di aggregazione (Hyndman & Athanasopoulos, 2021).

Dal punto di vista previsivo, considerando una situazione in cui esiste una serie "totale" composta da due costituenti, l'analogo processo crea due previsioni indipendenti per una data serie disaggregata. La prima è una previsione diretta, detta previsione di base. La seconda è una previsione indiretta ottenuta dalla previsione per la serie aggregata meno la previsione di altri componenti. Un vantaggio pratico dell'approccio FR, delineato in letteratura, è la sua applicabilità a qualsiasi insieme di previsioni di base.

Nel presente lavoro, data la natura dei dati analizzati, si prende in considerazione una struttura gerarchica temporale. Sono identificabili livelli di aggregazione temporali dati dalla frequenza di osservazione della serie. Quest'ultima influisce sui risultati del modello di stima sia in termini di adattamento che di previsione. In questo contesto l'approccio FR può quindi risultare molto utile. Le serie storiche osservate a diverse frequenze possono infatti contenere informazioni differenti attenuando o accentuando determinate caratteristiche: a frequenze basse emergono trend e cicli, mentre ad alte frequenze sarà più evidente un'eventuale componente stagionale e dinamiche di breve periodo. Una gerarchia di tipo temporale viene costruita a partire dall'aggregazione non sovrapposta delle osservazioni di una serie storica a intervalli regolari: ad esempio una serie storica trimestrale può essere aggregata a frequenza semestrale e annuale.

### 2.4.1 Approcci

Gli approcci di riconciliazione classici sono tre: l'approccio *Bottom-up*, il quale implica una previsione dal basso verso l'alto, l'approccio *Top-down*, che prevede al contrario una previsione dall'alto verso il basso e l'approccio *Middle-out*, che risulta dalla combinazione dei primi due.

La previsione dal basso verso l'alto (Dunn et al., 1976) comporta la previsione del livello

inferiore della gerarchia e successivamente l'aggregazione per ottenere stime per i livelli più alti. Il vantaggio principale di questo approccio è che non vi è perdita d'informazione a causa dell'aggregazione dal momento che le previsioni sono ottenute al livello più basso. Tuttavia, questo metodo ignora le relazioni tra le serie. Inoltre, le informazioni tendono a essere più rumorose ai livelli inferiori della gerarchia, il che può potenzialmente ridurre l'accuratezza complessiva delle previsioni (Di Fonzo & Girolimetto, 2022). I dati aggregati sono invece meno rumorosi e, come detto, possono evidenziare trend difficili da distinguere dalle sole serie di livello inferiore.

Nell'approccio *Top-down* (Gross & Sohl, 1990) viene previsto per primo il livello superiore della gerarchia. La previsione viene poi suddivisa per ottenere stime per i livelli inferiori utilizzando di solito proporzioni storiche o previste (Athanasopoulos et al., 2009). Tali proporzioni possono essere modellate in diversi modi, ma essenzialmente producono previsioni condizionate alla previsione del livello più alto.

L'approccio *Middle-out* è una combinazione degli approcci *Bottom-up* e *Top-down* in cui si ottiene la previsione dei livelli intermedi della gerarchia. Questi valori vengono poi utilizzati per calcolare le previsioni per i livelli superiori adottando l'approccio *Bottom-up* e per i livelli inferiori quello *Top-down*. Le tecniche tradizionali di riconciliazione delle previsioni presentano però delle forti debolezze:

- non riescono a sfruttare in modo efficiente i dati disponibili in quanto utilizzano previsioni su un numero limitato di serie;
- non considerano l'intrinseca struttura di correlazione della gerarchia, cioè non prendono in considerazione la correlazione tra le serie a ciascun livello;
- non consentono di ricavare facilmente degli intervalli di previsione.

In alternativa agli approcci classici, Hyndman et al., 2011, hanno introdotto un approccio di combinazione ottimale. In questo caso le serie storiche sono previste indipendentemente a tutti i livelli. Vengono prodotte previsioni di base incoerenti, le quali vengono poi trasformate utilizzando tutte le informazioni e le relazioni della gerarchia. Il risultato è ottenuto utilizzando un modello di regressione lineare e le nuove previsioni coerenti corrispondono alle somme ponderate delle previsioni di tutti i livelli. I pesi sono ottenuti dalla risoluzione di un sistema di equazioni, assicurando che siano soddisfatte le relazioni naturali tra i diversi livelli della gerarchia.

Questo approccio, proposto in un quadro trasversale, è stato poi esteso in un contesto temporale (Athanasopoulos et al., 2017, Nystrup et al., 2020) e cross-temporale (Kourentzes & Athanasopoulos, 2019, Di Fonzo & Girolimetto, 2022).

Hollyman et al., 2021, hanno esaminato per la prima volta il problema della riconciliazione delle previsioni da una prospettiva di combinazione delle previsioni (*Forecast Combination* (FC), Bates & Granger, 1969, Timmermann, 2006). Un approccio basato sulla combinazione previsionale ne eredita anche i vantaggi ed è stato ampiamente dimostrato che la FC, riducendo la varianza dell'errore, migliora l'accuratezza delle previsioni. In particolare Hollyman et al., 2021, hanno mostrato come generare previsioni imparziali e hanno proposto un approccio che permette di ottenere previsioni coerenti condizionali combinate (*Combined Conditional Coherent, CCC*).

Inoltre Panagiotelis et al., 2023, hanno esteso la riconciliazione di previsioni puntuali a previsioni probabilistiche, data la loro importanza in un contesto decisionale. L'intera distribuzione di probabilità dei valori futuri fornisce infatti una descrizione completa dell'incertezza associata alle previsioni (Gneiting & Katzfuss, 2014). L'obiettivo è costruire densità e trarre campioni da una previsione probabilistica riconciliata.

## 2.4.2 Notazione

Di seguito viene definito il problema utilizzando la notazione di Athanasopoulos et al., 2017. Si considera una serie storica  $\{x_t\}_{t=1}^T$  osservata alla più alta frequenza disponibile  $m$ . Dato un fattore  $k$  di  $m$ , si può costruire una versione temporalmente aggregata di  $x_t$  che abbia un periodo stagionale pari a  $M_k = m/k$  attraverso la somma non sovrapposta di  $k$  valori consecutivi. Per disporre di dati regolari, si assume che il numero totale delle osservazioni per  $x_t$ , denominato  $T$ , sia un multiplo di  $m$ . Viene fissato quindi  $T = N \times m$ , dove  $N$  è la lunghezza della serie osservata alla più bassa frequenza possibile. Le serie aggregate possono essere scritte come segue:

$$x_t^{[k]} = \sum_{t=(l-1)k+1}^{lk} x_t \quad (2.19)$$

dove  $l = 1, \dots, M_k$  e  $k \in \mathcal{K} = \{k_p, k_{p-1}, \dots, k_1\}$  è l'insieme dei  $p$  fattori di  $m$ , posti in ordine in modo tale che  $k_p = m$  e  $k_1 = 1$ .

Al fine di avere un unico indice comune per indicare le osservazioni ai diversi ordini di aggregazione temporale si definisce  $\tau$  come l'indice delle osservazioni al livello più aggregato, ossia  $x_\tau^{[m]}$ , con  $\tau = 1, \dots, N$ , è la serie al livello di aggregazione temporale più elevato. Fissato  $\tau$  e per  $k \neq m$ , è possibile raccogliere in un unico vettore di dimensione  $(M_k \times 1)$  ciascuna serie aggregata temporalmente:

$$\mathbf{x}_\tau^{[k]} = \left[ x_{M_k(\tau-1)+1}^{[k]} x_{M_k(\tau-1)+2}^{[k]} \dots x_{M_k\tau}^{[k]} \right]' \quad (2.20)$$



con  $\tau = 1, \dots, N$  e  $k \in \{k_{p-1}, \dots, k_2, 1\}$ .

In questo modo è possibile rappresentare il vettore di dimensione  $[(k^* + m) \times 1]$ , con  $k^* = \sum_{j=1}^{p-1} k_j$ , che raccoglie tutte le osservazioni come:

$$\mathbf{x}_\tau = \left[ x_\tau^{[m]} \mathbf{x}_\tau^{[k_{p-1}]} \dots \mathbf{x}_\tau^{[k_2]} \mathbf{x}_\tau^{[1]} \right]' = \left[ \mathbf{t}'_{x_\tau} \mathbf{x}_\tau^{[1]} \right]' \quad (2.21)$$

con  $\tau = 1, \dots, N$ ,  $\mathbf{t}_{x_\tau} = \left[ x_\tau^{[m]} \mathbf{x}_\tau^{[k_{p-1}]} \dots \mathbf{x}_\tau^{[k_2]} \right]$  vettore di dimensione  $(k^* \times 1)$  contenente i valori di tutte le serie aggregate temporalmente e  $\mathbf{x}_\tau^{[1]}$  vettore di dimensione  $(m \times 1)$  contenente le osservazioni al tempo  $\tau$  della serie storica a frequenza più alta.

Le relazioni di aggregazione che legano la serie con la frequenza più alta e quelle a frequenza minore possono essere rappresentate graficamente in una struttura gerarchica. Si consideri ad esempio una serie trimestrale: essa può essere aggregata in serie annuali e semestrali ( $\mathcal{K} = \{4, 2, 1\}$ ) e l'intero sistema è facilmente rappresentabile attraverso il diagramma ad albero in Figura 2.2.

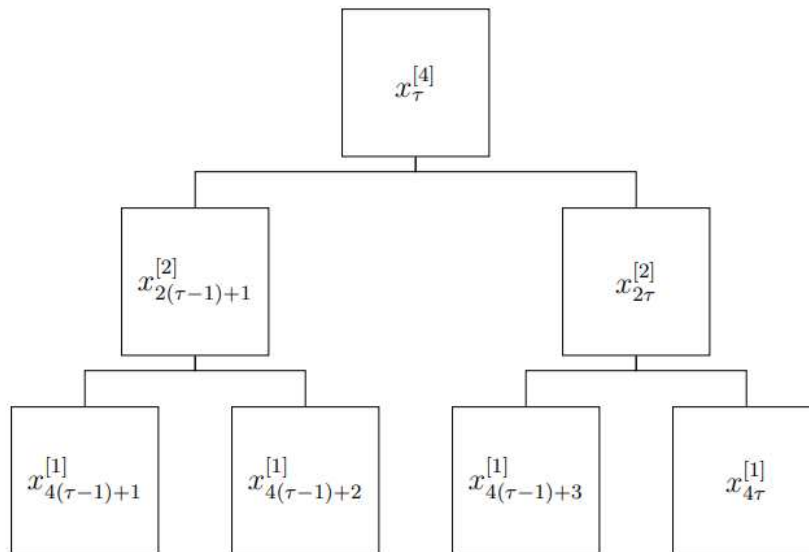


FIGURA 2.2: Gerarchia temporale per una serie storica trimestrale (Fonte: Girolimetto D. (2020). *Riconciliazione contemporanea, temporale e cross-temporale di previsioni di serie storiche.*)

Possiamo rappresentare la struttura gerarchica come segue:

$$\mathbf{x}_\tau = \mathbf{R}_1 \mathbf{x}_\tau^{[1]} \quad (2.22)$$

con  $\tau = 1, \dots, N$  e  $\mathbf{R}_1$  matrice di dimensione  $[(k^* + m) \times m]$  che incorpora i vincoli di aggregazione temporale e viene chiamata matrice somma temporale (*temporal summing matrix*). La matrice somma è formata da due blocchi:

$$\mathbf{R}_1 = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{I}_m \end{bmatrix} \quad (2.23)$$

in cui  $\mathbf{I}_m$  è una matrice identità e  $\mathbf{K}_1$  è la matrice di aggregazione temporale (*temporal aggregation matrix*) di dimensione  $(k^* \times m)$  che aggrega gli elementi di  $\mathbf{x}_\tau^{[1]}$  per ottenere i primi  $k^*$  elementi di  $\mathbf{x}_\tau$ , cioè  $\mathbf{t}_{x_\tau}$ :

$$\mathbf{t}_{x_\tau} = \mathbf{K}_1 \mathbf{x}_\tau^{[1]} \quad (2.24)$$

Questa espressione offre una maniera equivalente per rappresentare queste strutture gerarchiche:

$$\mathbf{Z}'_1 \mathbf{x}_\tau = \mathbf{0}_{(k^* \times 1)} \quad (2.25)$$

dove  $\mathbf{0}$  è un vettore  $(k^* \times 1)$  di zeri e  $\mathbf{Z}'_1$  è la matrice di dimensione  $[k^* \times (k^* + m)]$  dei vincoli di aggregazione (*zero constraints kernel matrix*) data da

$$\mathbf{Z}'_1 = [\mathbf{I}_{k^*} - \mathbf{K}_1] \quad (2.26)$$

Riprendendo il caso della serie storica trimestrale, in cui  $k^* = 3$  e  $m = 4$ , si ha:

$$\mathbf{R}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$e \quad \mathbf{Z}'_1 = \begin{bmatrix} 1 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}$$

Indicando con  $H$  l'orizzonte massimo di previsione per la serie aggregata con la frequenza più bassa, si generano  $M_k H$  previsioni per ogni livello di aggregazione  $k$ . Per ogni orizzonte  $h = 1, \dots, H$ , le previsioni di base possono essere raccolte in un vettore di

dimensione  $[(k^* + m) \times 1]$ :

$$\hat{\mathbf{x}}_h = \left[ \hat{x}_h^{[m]} \hat{\mathbf{x}}_h^{[k_{p-1}]} \dots \hat{\mathbf{x}}_h^{[k_2]} \hat{\mathbf{x}}_h^{[1]} \right]' \quad (2.27)$$

dove  $\hat{\mathbf{x}}_h^{[k]} = \left[ x_{M_k(h-1)+1}^{[k]} x_{M_k(h-1)+2}^{[k]} \dots x_{M_k h}^{[k]} \right]'$  è un vettore di lunghezza  $M_k$ . Queste quantità, ovviamente, non sono temporalmente coerenti, in quanto non sfruttano l'informazione gerarchica a disposizione. Secondo la proprietà di coerenza proposta da Athanopoulos et al., 2020, un set di previsioni  $h$  passi in avanti  $\tilde{\mathbf{x}}_h$  è detto coerente se  $\tilde{\mathbf{x}}_h \in \mathcal{S}$ , dove  $\mathcal{S}$  è un sottoinsieme di  $\mathbb{R}^{(k^*+m)}$  in cui valgono i vincoli di aggregazione espressi nell'equazione (2.25). Bisogna quindi operare in modo da allineare le previsioni.

Il generico vettore delle previsioni riconciliate  $\tilde{\mathbf{x}}_h$  può essere scritto come:

$$\tilde{\mathbf{x}}_h = \mathbf{R}_1 \mathbf{G} \hat{\mathbf{x}}_h \quad (2.28)$$

con  $h = 1, \dots, H$  e  $\mathbf{G}$  matrice (*mapping matrix* o matrice di riconciliazione) che trasforma  $\hat{\mathbf{x}}_h$  nelle previsioni riconciliate  $\mathbf{x}_h^{[1]}$ . In tutti gli approcci, siano essi *Top-down*, *Bottom-up*, *Middle-out* o "ottimali", essa svolge il ruolo di matrice di ponderazione.

Come detto, Hollyman et al., 2021, hanno proposto un approccio FR basato sulla combinazione previsionale mostrando come ottenere previsioni imparziali. Affinché le previsioni riconciliate siano imparziali deve valere

$$\mathbf{G} \mathbf{R}_1 = \mathbf{I}_{(k^*+m)} \quad (2.29)$$

(Wickramasuriya et al., 2019).

In particolare, viene proposto un approccio di riconciliazione dall'alto verso il basso il quale prevede di combinare previsioni riconciliate condizionate alle previsioni di base a ciascun livello e coerenti, cioè che si aggregano attraverso la gerarchia. Se le previsioni di base sono imparziali, anche le previsioni riconciliate in questo modo lo saranno. Secondo Bates & Granger, 1969, se le previsioni dirette non sono perfettamente correlate, una combinazione ponderata delle due beneficia di un effetto di diversificazione per il quale tale combinazione avrà una varianza di errore inferiore rispetto alle singole previsioni. Analogamente alla riconciliazione di una previsione puntuale, nell'impostazione probabilistica il punto di partenza è una previsione incoerente. Nella pratica si distinguono due modi per eseguire la riconciliazione probabilistica delle previsioni. Il primo è un metodo che coinvolge l'integrazione. Nel caso di distribuzioni ellittiche, inclusa la distribuzione gaussiana, questo metodo fornisce una soluzione elegante che coinvolge trasformazioni lineari di parametri di scala e posizione. In alternativa, una distribuzione può essere

riconciliata simulando dalla previsione di base incoerente. Viene poi riconciliato ogni vettore campionato come se fosse una previsione puntuale (Panagiotelis et al., 2023).

### 2.4.3 Riconciliazione ottimale puntuale

Dato il vettore  $\hat{\mathbf{x}}_h$  (2.27) di previsioni di base non distorte, estendendo il risultato di Athanasopoulos et al., 2017, è possibile adattare al caso temporale la soluzione mostrata da Wickramasuriya et al., 2019. L'espressione

$$\tilde{\mathbf{x}}_h = \mathbf{R}_1(\mathbf{R}_1' \boldsymbol{\Omega}_h^{-1} \mathbf{R}_1)^{-1} \mathbf{R}_1' \boldsymbol{\Omega}_h^{-1} \hat{\mathbf{x}}_h = \mathbf{R}_1 \mathbf{G} \hat{\mathbf{x}}_h \quad (2.30)$$

con  $\tilde{\mathbf{x}}_h$  vettore  $[(k^* + m) \times 1]$  dei valori riconciliati e  $\boldsymbol{\Omega}_h$  matrice  $[(k^* + m) \times (k^* + m)]$  definita positiva, è l'unica soluzione che minimizza la traccia della matrice di covarianza degli errori delle previsioni riconciliate

$$Var(\tilde{\mathbf{e}}_h | \mathbf{I}_h) = \mathbf{R}_1 \mathbf{G} \boldsymbol{\Omega}_h \mathbf{G}' \mathbf{R}_1' \quad (2.31)$$

sotto il vincolo di non distorsione  $\mathbf{R}_1 \mathbf{G} \mathbf{R}_1 = \mathbf{R}_1$ , con  $\tilde{\mathbf{e}}_h = \mathbf{x}_h - \tilde{\mathbf{x}}_h$  vettore degli errori delle previsioni riconciliate e  $\boldsymbol{\Omega}_h = Var(\hat{\mathbf{e}}_h | \mathbf{I}_T)$  con  $\hat{\mathbf{e}}_h = \mathbf{x}_h - \hat{\mathbf{x}}_h$  vettore degli errori delle previsioni di base.

La matrice  $\boldsymbol{\Omega}_h$  risulta difficile da stimare, soprattutto per  $h > 1$ . Per questo motivo nella letteratura vengono prese in considerazione approssimazioni semplificate tali per cui  $\boldsymbol{\Omega}_h = k_h \boldsymbol{\Omega}$  con  $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1$  e  $k_h$  costante di proporzionalità. Non è necessario che quest'ultima sia nota in quanto in (2.30) scompare per effetto di semplificazioni.

Alcune delle approssimazioni sfruttano i residui dei modelli utilizzati per ricavare le previsioni di base. Si considera il vettore  $(M_k \times 1)$  dei residui *in-sample* al tempo  $\tau$ :

$$\hat{\mathbf{e}}_\tau^{[k]} = \mathbf{x}_\tau^{[k]} - \hat{\mathbf{x}}_\tau^{[k]} \quad (2.32)$$

con  $\tau = 1, \dots, N$  e  $k \in \mathcal{K}$ .

Esso può essere organizzato in  $p$  matrici di dimensione  $(N \times M_k)$

$$\hat{\mathbf{E}}_x^{[k]} = \begin{bmatrix} (\hat{\mathbf{e}}_1^{[k]})' \\ \vdots \\ (\hat{\mathbf{e}}_\tau^{[k]})' \\ \vdots \\ (\hat{\mathbf{e}}_N^{[k]})' \end{bmatrix}, \quad k \in \mathcal{K},$$

che a loro volta possono essere raggruppate nella matrice  $[N \times (k^* + m)]$  dei residui *in-sample*

$$\hat{\mathbf{E}}_x = \left[ \hat{\mathbf{E}}_x^{[m]} \hat{\mathbf{E}}_x^{[k_p-1]} \dots \hat{\mathbf{E}}_x^{[k_2]} \hat{\mathbf{E}}_x^{[1]} \right]. \quad (2.33)$$

Di seguito vengono riportate alcune delle forme più utilizzate per approssimare  $\mathbf{\Omega}$ . Le prime proposte sono state formulate da Athanasopoulos et al., 2017 e Hyndman & Kourentzes, 2018 mentre le più recenti da Nystrup et al., 2020. Le matrici riportate ai punti 4, 5 e 6 fanno riferimento a strutture diagonali che ignorano la correlazione tra i livelli. Nell'ambito delle gerarchie temporali è d'interesse sfruttare anche l'informazione sulla dinamica delle serie a diverse frequenze. Nystrup et al., 2020 propongono quindi delle formulazioni che includono una struttura di autocorrelazione (matrici riportate ai punti 7, 8, 9 e 10). Il prefisso "t-" indica il fatto che si stanno considerando gerarchie temporali.

1. Identità (**t-ols**):

$$\hat{\mathbf{\Omega}}_{t-ols} = \mathbf{I}_{k^*+m}$$

Questa formulazione corrisponde alla soluzione dei minimi quadrati ordinari. È necessario sottolineare che assumere l'equivarianza con strutture gerarchiche temporali è restrittivo ed irrealistico. Questa opzione infatti, tecnicamente plausibile, non incontra il favore degli specialisti dell'argomento.

2. MinT-sam (**t-sam**):

$$\hat{\mathbf{\Omega}}_{t-sam} = \frac{1}{N} (\hat{\mathbf{E}}_x)' \hat{\mathbf{E}}_x$$

Questa matrice è ben definita se  $N > (k^* + m)$  altrimenti si rischiano problemi di non invertibilità. Inoltre per poter essere considerata la matrice di covarianza dei residui *in-sample* è necessario che la media per colonna di  $\hat{\mathbf{E}}_x$  sia nulla, cioè  $N^{-1} \sum_{\tau=1}^N \hat{e}_{\tau,l} = 0$  per  $l = 1, \dots, k^* + m$ . Se questo non si verifica  $\hat{\mathbf{\Omega}}_{t-sam}$  è la matrice dell'errore quadratico medio delle osservazioni (*sample Mean Square Error matrix*).

3. MinT-shr (**t-shr**):

$$\hat{\mathbf{\Omega}}_{t-shr} = \lambda (\mathbf{I}_{k^*+m} \odot \hat{\mathbf{\Omega}}_{t-sam}) + (1 - \lambda) \hat{\mathbf{\Omega}}_{t-sam}$$

dove  $\odot$  è il prodotto di Hadamard, elemento per elemento e  $0 \leq \lambda \leq 1$ . Se  $N < (k^* + m)$   $\hat{\mathbf{\Omega}}_{t-sam}$  non è utilizzabile per calcolare le previsioni riconciliate. Questa approssimazione si ottiene "restringendo" gli elementi di  $\hat{\mathbf{\Omega}}_{t-sam}$  fuori dalla

diagonale verso lo 0 attraverso il parametro di regolarizzazione  $\lambda$  (*shrinkage*). Questa quantità si calcola seguendo l'approccio di Ledoit & Wolf, 2004, sviluppato da Schäfer & Strimmer, 2005:

$$\lambda = \frac{\sum_{i \neq j} \widehat{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}^2} \quad (2.34)$$

dove  $\hat{r}_{ij}$  è l'elemento in posizione  $(i, j)$  della matrice di correlazione degli errori *in-sample*.

4. **Strutturale (t-*struc*):**

$$\hat{\mathbf{\Omega}}_{t-struc} = \text{diag}(\mathbf{R}_1 \mathbf{1}_m)$$

con  $\mathbf{1}_m$  vettore unitario di lunghezza  $m$ . Questa variante, proposta da Athanopoulos et al., 2017, dipende solo dalla struttura della gerarchia. A differenza degli OLS in questo caso si assume l'equivarianza degli errori solo al livello più disaggregato. Tale approssimazione risulta particolarmente utile nei casi in cui non si hanno a disposizione i residui.

5. *Hierarchy variance scaling* (t-*wlsh*):

$$\hat{\mathbf{\Omega}}_{t-wlsh} = \mathbf{I}_{k^*+m} \odot \hat{\mathbf{\Omega}}_{t-sam}$$

Si tratta di una matrice diagonale che contiene una stima diversa per la varianza ad ogni nodo, assumendo quindi eteroschedasticità tra ed entro i livelli della gerarchia.

6. *Series variance scaling* (t-*wlsv*):

$$\hat{\mathbf{\Omega}}_{t-wlsv} = \begin{bmatrix} (\hat{\sigma}^{[m]})^2 & 0 & \cdots & 0 \\ 0 & (\hat{\sigma}^{[k_{p-1}]} )^2 \mathbf{I}_{M_{k_{p-1}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\hat{\sigma}^{[1]})^2 \mathbf{I}_{M_1} \end{bmatrix},$$

$\hat{\mathbf{\Omega}}_{t-wlsv}$  è una matrice diagonale formata dalle stime delle varianze degli errori di ogni ordine di aggregazione  $(\hat{\sigma}^{[k]})^2$  con  $k \in \mathcal{K}$ .

7. Autocovariance (**t-acov**):

$$\hat{\mathbf{\Omega}}_{t-acov} = \begin{bmatrix} \hat{\mathbf{\Omega}}^{[m]} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Omega}}^{[k_{p-1}]} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{\Omega}}^{[k_2]} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \hat{\mathbf{\Omega}}^{[1]} \end{bmatrix},$$

dove le  $p$  matrici  $\hat{\mathbf{\Omega}}^{[k]}$ , ciascuna di dimensione  $(M_k \times M_k)$ , vengono calcolate come

$$\hat{\mathbf{\Omega}}^{[k]} = \frac{1}{N} \sum_{\tau=1}^N \hat{\mathbf{e}}_{\tau}^{[k]} (\hat{\mathbf{e}}_{\tau}^{[k]})' = \frac{1}{N} (\hat{\mathbf{E}}^{[k]})' \hat{\mathbf{E}}^{[k]}, \quad k \in \mathcal{K}.$$

Affinché  $\hat{\mathbf{\Omega}}_{t-acov}$  sia invertibile,  $N$  deve essere maggiore di  $m$ . Questo vincolo risulta essere meno restrittivo della condizione necessaria per la non singolarità di  $\hat{\mathbf{\Omega}}_{t-sam}$

 8. Structural AR(1) (**t-strar1**):

$$\hat{\mathbf{\Omega}}_{t-strar1} = \hat{\mathbf{\Omega}}_{t-struc}^{\frac{1}{2}} \mathbf{\Gamma} \hat{\mathbf{\Omega}}_{t-struc}^{\frac{1}{2}}$$

Per definire  $\mathbf{\Gamma}$  si considerano prima le matrici di Toeplitz  $(M_k \times M_k)$  costruite a partire dai coefficienti  $\rho_{[k]}$  di autocorrelazione del primo ordine dei residui per ciascun livello  $k \in \mathcal{K}$ :

$$\mathbf{\Gamma}^{[m]} = \mathbf{1}, \quad \mathbf{\Gamma}^{[k]} = \begin{bmatrix} 1 & \rho_{[k]} & \cdots & \rho_{[k]}^{M_k-1} \\ \rho_{[k]} & 1 & \cdots & \rho_{[k]}^{M_k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{[k]}^{M_k-1} & \rho_{[k]}^{M_k-2} & \cdots & 1 \end{bmatrix}, \quad k \in \mathcal{K}.$$

$\mathbf{\Gamma}$  non è altro che la matrice  $[(k^* + m) \times (k^* + m)]$  composta da  $p$  blocchi:

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Gamma}^{[k_{p-1}]} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Gamma}^{[1]} \end{bmatrix} \quad (2.35)$$

9. *Series AR(1) (t-sar1)*:

$$\hat{\Omega}_{t-sar1} = \hat{\Omega}_{t-wlsv}^{\frac{1}{2}} \Gamma \hat{\Omega}_{t-wlsv}^{\frac{1}{2}}$$

con  $\Gamma$  definita in (2.10).

10. *Hierarchy AR(1) (t-har1)*:

$$\hat{\Omega}_{t-har1} = \hat{\Omega}_{t-wlsh}^{\frac{1}{2}} \Gamma \hat{\Omega}_{t-wlsh}^{\frac{1}{2}}$$

che combina la *hierarchy variance scaling* con la matrice di autocorrelazione  $\Gamma$ .

## 2.5 Criteri di valutazione

Di seguito vengono descritti i criteri di valutazione utilizzati per confrontare i modelli in termini di accuratezza delle previsioni.

Si definisce  $\hat{y}_t$  la previsione della serie storica  $y_t$  e  $e_t = \hat{y}_t - y_t$  l'errore di previsione corrispondente.

La performance previsiva è valutata tramite la radice dell'errore quadratico medio (*Root Mean Squared Error*, RMSE)

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad (2.36)$$

e l'errore medio assoluto (*Mean Absolute Error*, MAE)

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| \quad (2.37)$$

É possibile dimostrare che l'MSE può essere scomposto come la somma di varianza e distorsione quadratica:

$$\begin{aligned} E\{(\hat{y}_t - y_t)^2\} &= E\{(\hat{y}_t \pm E(\hat{y}_t) - y_t)^2\} = \\ &= \{E(\hat{y}_t) - y_t\}^2 + var\{\hat{y}_t\} = \\ &= distorsione^2 + varianza \end{aligned} \quad (2.38)$$

Queste due quantità sono in contrasto: all'aumentare della varianza diminuisce la distorsione e viceversa. Metodi abbastanza flessibili presentano una varianza elevata e bassa distorsione, ma come già detto possono portare a problemi di sovradattamento.



Vi è quindi un *trade-off* tra varianza e distorsione di cui è necessario tenere conto.

L'aumento dell'indice di complessità del modello ne migliora le prestazioni rispetto ai dati su cui è stato costruito, ma questo non avviene se si considerano altri dati. Una strategia efficace, adottata nel presente lavoro di tesi, è utilizzare una parte dei dati che si hanno a disposizione per stimare il modello e la restante parte per verificarne l'adeguatezza. Questa tecnica risulta utile per trovare un compromesso tra varianza e distorsione: consente di contenere, infatti, la varianza senza aumentare la distorsione.

La presenza di differenze significative tra le previsioni dei modelli è valutata, inoltre, attraverso il test Diebold-Mariano (DM) a una coda (Diebold & Mariano, 1995). Si considerano due serie storiche  $\hat{y}_{1t}$  e  $\hat{y}_{2t}$  che corrispondono alle previsioni ottenute dall'adattamento di due distinti modelli. Si definisce una funzione di perdita  $L(e_{it})$ ,  $i = 1, 2$ , sull'errore di previsione. Il test si basa sul calcolo della differenza tra le funzioni di perdita calcolate per le due previsioni

$$d_t = L(e_{1t}) - L(e_{2t}) = L(\hat{y}_{1t} - y_t) - L(\hat{y}_{2t} - y_t) \quad (2.39)$$

Le previsioni  $\hat{y}_{1t}$  e  $\hat{y}_{2t}$  sono ugualmente accurate se il valore atteso di questa differenza è nullo per ogni valore di  $t$ . L'ipotesi nulla in quest'applicazione è del tipo "la performance predittiva del primo modello è uguale o peggiore alla performance predittiva del secondo modello". Formalmente risulta la seguente struttura d'ipotesi:

$$\begin{cases} H_0 : E(L(e_{1t}) - L(e_{2t})) \geq 0 \\ H_1 : E(L(e_{1t}) - L(e_{2t})) < 0 \end{cases} \quad (2.40)$$

Sotto determinate condizioni si verifica che

$$\sqrt{h}(\bar{d} - \mu_d) \xrightarrow{d} N(0, \hat{\omega}_m(d_t)) \quad (2.41)$$

dove  $h$  è il numero di passi in avanti considerati nell'effettuare le previsioni,  $\bar{d} = \frac{1}{h} \sum_{t=1}^h d_t$  è la media campionaria della differenza  $d_t$ ,  $\hat{\omega}_m(d_t)$  è la varianza di lungo periodo di  $d_t$  e  $\mu = E(d_t)$ .

L'ipotesi nulla può essere verificata attraverso la seguente statistica test:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\hat{\omega}_m(d_t)}{h}}} = \frac{\bar{d}}{\sqrt{\frac{1}{h} \sum_{i=-m}^m \hat{\gamma}_i}} \xrightarrow{d} N(0, 1) \quad (2.42)$$

dove  $\hat{\gamma}_i$  sono le autocovarianze campionarie della serie  $d_t$ .

# Capitolo 3

## Analisi empiriche

### 3.1 Presentazione dei dati

I dati analizzati nel presente lavoro di tesi sono rilasciati dal Gestore dei Mercati Energetici (GME) e da TERNA (Rete Elettrica Nazionale S.p.A.). Quest'ultima ha un ruolo centrale nel sistema elettrico italiano. In qualità di TSO (*Transmission System Operator*) essa garantisce a tutti gli operatori il corretto funzionamento delle attività di trasmissione di energia elettrica sulla rete.

Le serie storiche considerate riguardano i prezzi e la domanda di energia del mercato elettrico italiano (IPEX) nelle zone Nord e Centro-Nord dall'1 gennaio 2017 al 31 dicembre 2020. Come esposto nel capitolo precedente una prima parte dei dati, le osservazioni fino all'anno 2019, è stata utilizzata per la stima dei modelli e la restante parte, le osservazioni riferite all'anno 2020, per verificarne l'adeguatezza. I dati hanno una frequenza oraria. Ogni giorno è quindi composto da 24 periodi, con 00:00-01:00 definito come ora 1. In questa analisi, seguendo una pratica diffusa in letteratura, ogni serie oraria è stata modellata singolarmente, eliminando così il problema della modellazione della periodicità intra-giornaliera. Inoltre, le due zone sono state considerate separatamente. La Figura 3.1 mostra la serie temporale dei prezzi relativi alla zona Nord in un piccolo orario (ora 12).

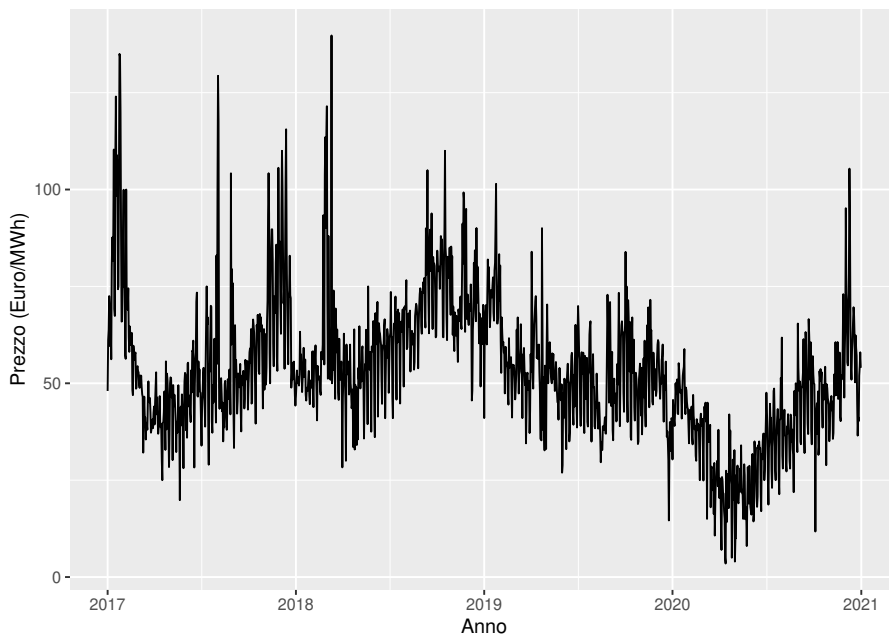


FIGURA 3.1: Serie storica giornaliera dei prezzi spot nella zona Nord alla 12° ora

È possibile osservare alcune regolarità dei prezzi dell'elettricità come l'andamento stagionale dovuto alle stagioni astronomiche e la presenza di alcuni picchi. Come descritto nel primo capitolo i mercati energetici sono infatti caratterizzati da repentine e generalmente imprevedibili variazioni estreme dei prezzi spot le cui cause possono essere diverse. Si osserva una stagionalità multiscala: oltre alle oscillazioni annuali, settimanali e giornaliere i cicli osservati sono strettamente correlati alla domanda. I prezzi sono sempre legati all'aumento o alla diminuzione improvvisa della domanda e dell'offerta di energia elettrica che può essere causata da eventi atmosferici estremi (giornate molto calde o molto fredde o rotture degli impianti). La Figura 3.2 mostra la serie temporale dei consumi relativi alla zona Nord in un piccolo orario (ora 12). È evidente l'andamento stagionale e le imprevedibili variazioni della domanda che il sistema elettrico italiano deve soddisfare.

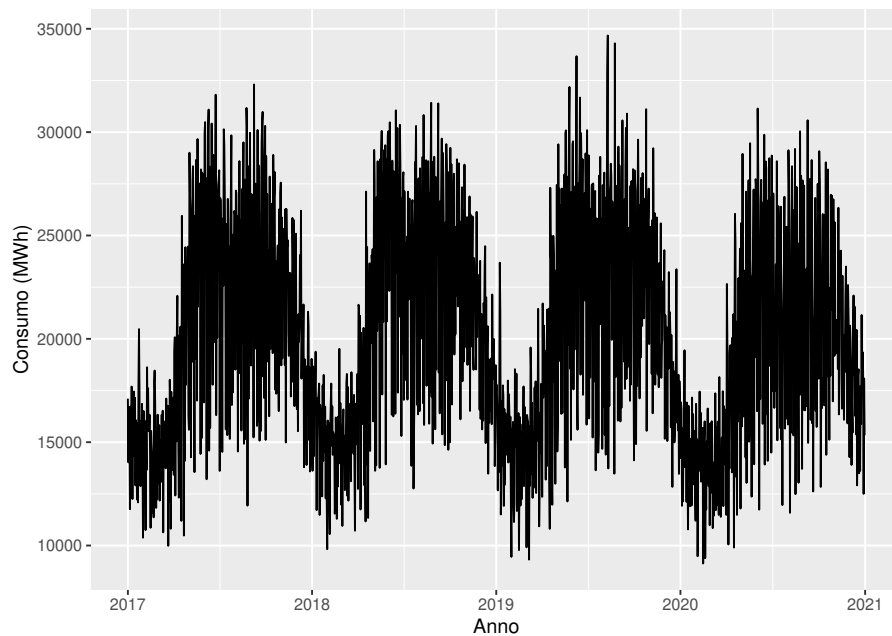


FIGURA 3.2: Serie storica giornaliera dei consumi nella zona Nord alla 12° ora

L'obiettivo dell'analisi è prevedere i prezzi zonali del mercato elettrico italiano implementando i modelli presentati nel capitolo precedente: modello autoregressivo lineare con variabili esogene (ARX), modello autoregressivo con regolarizzazione LASSO con e senza interazioni (ARX-L, ARX-L int.), *support vector machine* (SVM), foreste casuali (RF) e *gradient boosting* (GB).

Nell'adattamento dei modelli viene sfruttata l'informazione data dal prezzo al tempo  $t - 1, t - 2, \dots, t - 7$ , indicata come  $y_1, y_2, \dots, y_7$ , e una serie di regressori elencati di seguito:

- il giorno dell'anno;
- il giorno della settimana;
- un calendario fittizio delle festività;
- la domanda di energia elettrica prevista un giorno prima (fonte: Gestore dei Mercati Energetici (GME));
- la generazione eolica prevista un giorno prima (fonte: TERNA SpA);
- i consumi energetici del giorno prima (fonte: TERNA SpA).

Nell'adattamento dei modelli RF e GB il giorno dell'anno è una variabile discreta con valori da 1 a 365 per catturare il ciclo annuale. Per gli altri modelli invece l'informazione sulla periodicità stagionale è fornita da sinusoidi con periodo base 365, ovvero vengono aggiunti i seguenti 32 regressori:  $\cos(\omega_j t)$ ,  $\sin(\omega_j t)$  con  $\omega_j = 2\pi j/365$  e  $j = 1, 2, \dots, 16$ . Questa scelta è motivata dal fatto che RF e GB sono modelli basati su alberi che approssimano funzioni lisce partizionando ricorsivamente lo spazio dei dati, mentre i modelli lineari accettano trasformazioni lisce dei regressori. La scelta del numero di armoniche, pari a 16, è una scelta diffusa in letteratura. Nelle analisi di dati su prezzi italiani il numero di armoniche utilizzate è generalmente maggiore che per gli altri paesi. Ciò avviene poiché il tipico picco dei consumi di energia elettrica, e di conseguenza dei prezzi, in inverno e durante le vacanze estive in Italia richiede più sinusoidi per essere ben catturato (Grossi et al., 2022). Per indicare il giorno della settimana nei modelli lineari sono state inserite delle variabili dummy, mentre nei restanti modelli l'informazione è data da una variabile categoriale con sette classi.

Come descritto nel capitolo precedente i modelli adattati dipendono da alcuni iperparametri. Il parametro di regolarizzazione  $\lambda$  dei modelli ARX-L e ARX-L int. è fissato zona per zona e ora per ora attraverso una convalida incrociata di 10 gruppi. Per implementare SVM si utilizza una funzione kernel radiale con  $\epsilon$  e  $Y$  determinati da una convalida incrociata per 10 gruppi su una griglia di  $3 \times 7$  valori predeterminati. I parametri di regolazione di RF sono i seguenti: il numero di alberi  $N_t$  che formano la foresta è scelto sufficientemente grande, pari a  $N_T = 10000$ , e il numero di covariate  $F$  da selezionare in ciascun nodo è pari alla scelta tipica e pari a  $k/3$ . Infine, per quanto riguarda il modello GB il numero di alberi è sempre pari a 10000 e il parametro di *shrinkage* è pari a 0.1. In secondo luogo vengono implementati gli stessi modelli per la previsione dei consumi. Anche in questo caso viene sfruttata l'informazione passata data dai consumi al tempo  $t - 1, t - 2, \dots, t - 7$  indicata come  $y_1, y_2, \dots, y_7$ , e i seguenti regressori:

- il giorno dell'anno;
- il giorno della settimana;
- un calendario fittizio delle festività.

Come per la previsione dei prezzi nell'adattamento dei modelli RF e GB il giorno dell'anno è una variabile discreta con valori da 1 a 365. Per gli altri modelli invece vengono aggiunti i seguenti 32 regressori:  $\cos(\omega_j t)$ ,  $\sin(\omega_j t)$  con  $\omega_j = 2\pi j/365$  e  $j = 1, 2, \dots, 16$ . Anche l'informazione relativa al giorno della settimana è stata utilizzata in modo analogo all'analisi sui prezzi: nei modelli lineari sono state inserite delle variabili dummy,

mentre nei restanti modelli è stata utilizzata una variabile categoriale con sette classi. Infine, gli iperparametri sono stati fissati utilizzando i criteri descritti in precedenza.

## 3.2 Risultati

Nella seguente sezione vengono riportati i risultati dell'analisi effettuata attraverso le tecniche descritte per la previsione dei prezzi e della domanda. Le performance dei diversi modelli adattati ai dati sono confrontate tramite i criteri di valutazione descritti nel capitolo precedente. Vengono inoltre valutati i ruoli delle variabili esplicative presentando quelle che sono risultate più influenti nelle previsioni dei prezzi del mercato energetico.

### 3.2.1 Analisi dei prezzi

#### 3.2.1.1 Performance previsive dei modelli

Di seguito viene effettuato un confronto tra le prestazioni dei modelli utilizzati per prevedere i prezzi spot dell'energia sulla base degli indici di valutazione descritti nel Capitolo 2.

Le Tabelle 3.1 e 3.2 riportano i valori medi di RMSE ottenuti con i diversi modelli distinguendo tra ore di punta (14 ore) e ore non di punta (10 ore) per le due zone considerate. È interessante distinguere tra questi due tipi di orari perché le aziende adottano strategie diverse quando la domanda è alta o bassa (Grossi et al., 2022).

Ore di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	7.411	8.236	9.993	7.889	10.833	11.041
Centro-Nord	7.939	8.780	10.434	8.320	11.191	11.140

TABELLA 3.1: Valore medio dell'RMSE nelle ore di punta

Ore non di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	4.513	5.123	5.922	4.839	6.814	6.144
Centro-Nord	5.521	6.674	7.406	6.168	7.567	7.655

TABELLA 3.2: Valore medio dell'RMSE nelle ore non di punta

In Tabella 3.3 e Tabella 3.4 sono riportati i valori medi di RMSE in rapporto al prezzo medio orario distinguendo sempre tra ore di punta e ore non di punta per le due zone considerate.

Ore di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	0.124	0.138	0.168	0.133	0.182	0.185
Centro-Nord	0.133	0.147	0.175	0.139	0.188	0.186

TABELLA 3.3: Valore medio dell'RMSE rapportato con il prezzo medio orario nelle ore di punta

Ore non di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	0.092	0.103	0.120	0.098	0.139	0.125
Centro-Nord	0.111	0.132	0.146	0.124	0.152	0.154

TABELLA 3.4: Valore medio dell'RMSE rapportato con il prezzo medio orario nelle ore non di punta

Si osserva innanzitutto che i modelli hanno prestazioni migliori durante le ore non di punta. Questo avviene poiché quando la domanda è bassa la curva di offerta è piuttosto piatta e l'elasticità del prezzo è altrettanto bassa. Variazioni impreviste della domanda hanno quindi scarso effetto sul prezzo di equilibrio. Non si notano invece grandi differenze tra le performance predittive dei modelli per le due zone considerate: i valori di RMSE della zona Nord sono solo leggermente inferiori rispetto a quelli del Centro-Nord sia nelle ore di punta che nelle ore non di punta. Inoltre, è evidente come i modelli lineari risultano avere prestazioni migliori presentando un valore di RMSE inferiore. Le performance di RF e GB risultano decisamente peggiori rispetto a quelle degli altri modelli. Questo, in generale, può essere quindi dovuto al fatto che modelli più complessi



non riescono a catturare caratteristiche deterministiche dei prezzi.

Di seguito si considerano le performance predittive più dettagliatamente in ogni singola ora, continuando a distinguere tra ore di punta e ore non di punta.

In Tabella 3.5 e Tabella 3.6 è riportato per ogni ora il modello migliore, ovvero quello che presenta RMSE inferiore, e tra parentesi il numero di test di Diebold Mariano significativi. Questo numero varia tra 0 e 5, dove il valore 0 indica che l'RMSE del miglior modello non è significativamente inferiore all'RMSE degli altri 5 concorrenti. Il valore 5 indica invece che l'RMSE del miglior modello è sempre notevolmente inferiore.

Ore di punta		
Ora	Nord	Centro-Nord
8	ARX (5)	ARX (5)
9	ARX (4)	ARX (4)
10	ARX (4)	ARX (5)
11	ARX (5)	ARX (5)
12	ARX (5)	ARX (5)
13	ARX (5)	ARX (5)
14	ARX (4)	ARX (5)
15	ARX (5)	ARX (4)
16	ARX (5)	ARX (5)
17	ARX (5)	ARX (4)
18	ARX (5)	ARX (3)
19	ARX (4)	ARX (4)
20	ARX (4)	ARX (4)
21	SVM (4)	SVM (4)

TABELLA 3.5: Modello migliore in termini di RMSE inferiore e numero di Test di Diebold Mariano significativi (ore di punta)

Ore non di punta		
Ora	Nord	Centro-Nord
1	ARX (5)	ARX (5)
2	ARX (3)	ARX (5)
3	ARX (3)	ARX (5)
4	ARX (3)	ARX (5)
5	ARX (4)	ARX (5)
6	ARX (5)	ARX (5)
7	ARX (5)	ARX (5)
22	SVM (4)	SVM (5)
23	ARX (5)	ARX (5)
24	ARX (5)	ARX (5)

TABELLA 3.6: Modello migliore in termini di RMSE inferiore e numero di Test di Diebold Mariano significativi (ore non di punta)

Si può immediatamente notare come non vi siano evidenti differenze tra le ore di punta e le ore non di punta. Per la maggior parte delle ore il modello ARX presenta RMSE più basso in entrambe le zone. Quando questo modello risulta il migliore il numero dei test di Diebold Mariano significativi è quasi sempre pari a 4 e 5 e mai inferiore a 3. Questo indica che l'RMSE del modello autoregressivo lineare è quasi sempre significativamente inferiore agli RMSE dei modelli concorrenti, confermandone le migliori performance predittive. Solo nelle ore 21 e 22, in entrambe le zone, il modello migliore in termini di RMSE risulta SVM. Anche in questo caso il numero di test di Diebold Mariano significativi è pari a 4 o 5. RF e GB non risultano essere i modelli migliori per nessuna ora né al Nord né al Centro-Nord.

Le stesse considerazioni possono essere effettuate osservando l'indice MAE ottenuto dall'adattamento dei diversi modelli. I risultati basati su tale indice e i grafici di confronto delle previsioni ottenute con ogni modello sono riportati in Appendice.

### 3.2.1.2 Importanza delle variabili

Di seguito viene valutata l'importanza delle variabili esplicative presentando quelle che dall'analisi sono risultate più influenti nelle previsioni dei prezzi del mercato energetico. I ruoli dei singoli regressori sono analizzati attraverso le misure di importanza delle variabili (VIM) calcolate con le metodologie descritte nel capitolo precedente. Per valutare la rilevanza dei regressori dei modelli RF e GB è stato utilizzato il *Total Decrease*

in *Node Impurity* (TDNI), applicando la correzione proposta da Sandri & Zuccolotto, 2008, 2010. La rilevanza dei predittori negli altri modelli implementati, ARX, ARX-L e SVM, è stata invece valutata utilizzando un metodo *model-agnostic* basato su permutazioni delle variabili (Biecek & Burzykowski, 2021).

Per ognuna delle 24 ore viene quindi formulata una graduatoria delle covariate utilizzate sulla base del VIM. Per riassumere i risultati vengono prese in considerazione solo le covariate che occupano una delle prime cinque posizioni di tale graduatoria. Inoltre, al fine di discriminare la rilevanza della graduatoria, alle prime cinque posizioni vengono assegnati i valori riportati in Tabella 3.7.

Posizione 1	Posizione 2	Posizione 3	Posizione 4	Posizione 5
1	0.8	0.6	0.4	0.2

TABELLA 3.7: Valori assegnati alle prime cinque posizioni della graduatoria prodotta sulla base del VIM

Prendendo in esame i risultati ottenuti dall'adattamento del modello ARX per le due diverse zone, la Figura 3.3 mette in evidenza la rilevanza dei singoli regressori. Le barre rappresentano la percentuale ponderata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore. In primo luogo risulta rilevante l'informazione passata, in particolare i prezzi del giorno prima ( $y_1$ ). In secondo luogo viene preso in considerazione anche il giorno della settimana. Risultano particolarmente rilevanti i giorni sabato, domenica e lunedì e non giorni centrali della settimana. Inoltre, altre variabili informative sono la domanda prevista del giorno prima, i giorni di festa e in misura minore la generazione eolica prevista.

La Figura 3.4 descrive l'importanza delle variabili nel modello ARX-L. I risultati sono in accordo con quelli precedenti ottenuti per il modello ARX. Anche in questo caso risultano rilevanti l'informazione passata, in particolare i prezzi al lag 1 e al lag 7, e i giorni della settimana sabato, domenica e lunedì. La domanda prevista il giorno prima è un'altra variabile importante per entrambe le zone. Il calendario delle festività rientra tra i regressori più rilevanti solo per il Centro-Nord, comunque in bassa percentuale. Per entrambe le zone considerate la generazione di energia eolica non risulta mai tra le prime cinque posizioni nelle 24 ore.

Sono stati ottenuti dei risultati analoghi anche per il modello SVM come descritto in Figura 3.5. Assumono particolare rilevanza i prezzi al tempo  $t - 1$  e  $t - 7$ , il giorno della settimana, la domanda di energia prevista e in misura minore le festività. La generazione di energia eolica è una tra le variabili importanti solo per il Centro-Nord,

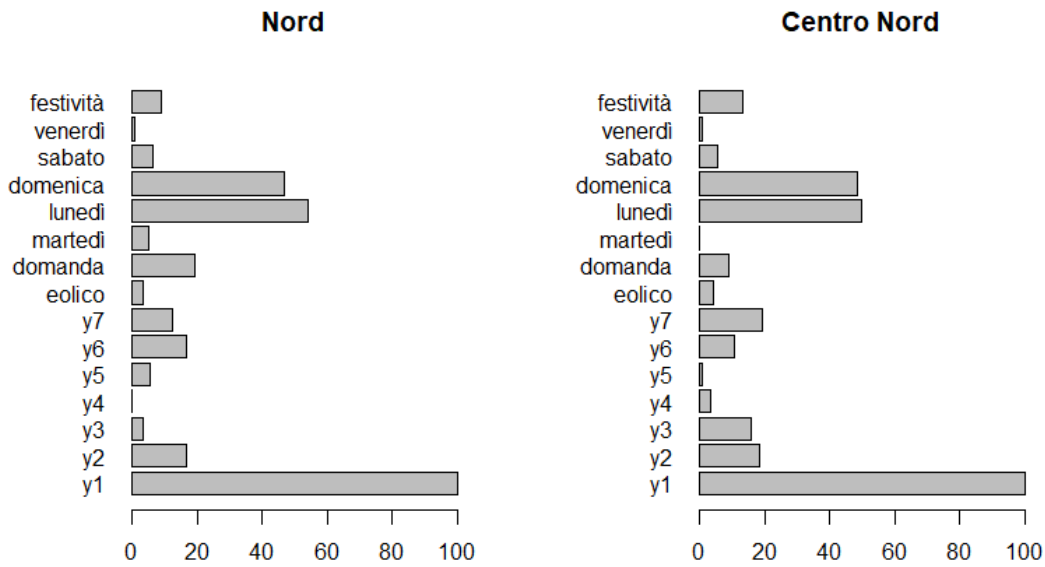


FIGURA 3.3: Misura di importanza delle variabili per il modello ARX. Percentuale pesata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore

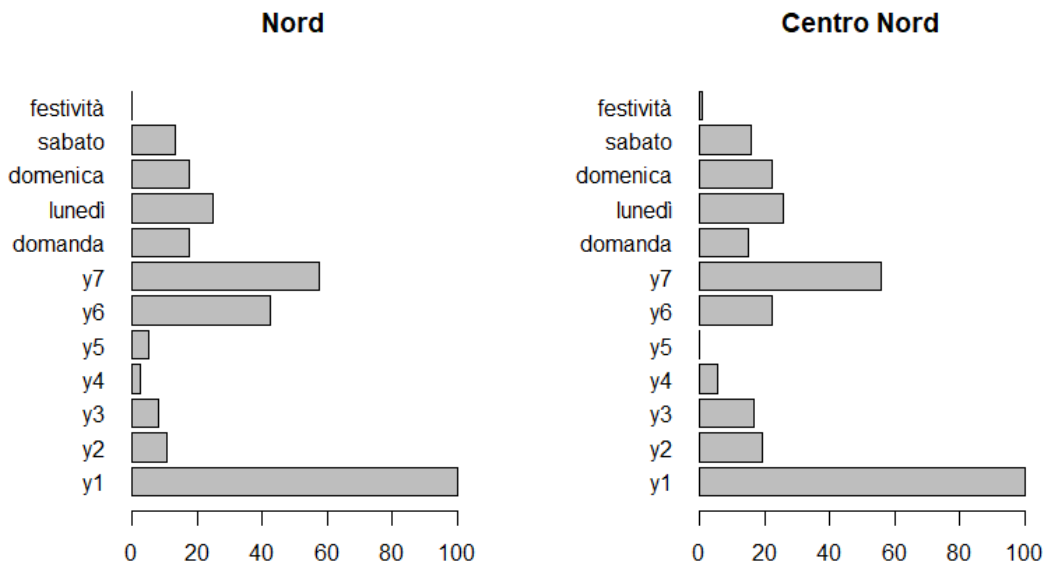


FIGURA 3.4: Misura di importanza delle variabili per il modello ARX-L. Percentuale pesata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore

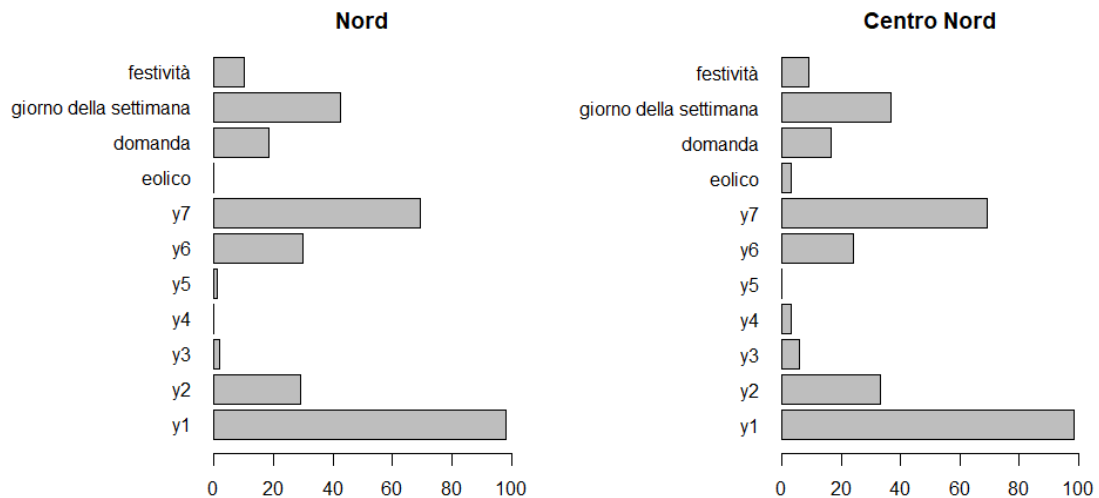


FIGURA 3.5: Misura di importanza delle variabili per il modello SVM. Percentuale pesata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore

presentando comunque una bassa percentuale del numero di volte in cui rientra tra le prime cinque posizioni.

L'importanza delle variabili nei modelli RF e GB si discosta invece leggermente dai risultati precedenti. Per quanto riguarda il modello RF si riportano i risultati in Figura 3.6. Non si osservano evidenti differenze tra le due zone considerate. I prezzi passati, in particolare quelli al lag 1 e al lag 7, rappresentano sempre i regressori di maggior rilievo. Tra le altre variabili considerate nell'analisi l'unica che rientra tra le prime cinque posizioni nelle 24 ore è il giorno della settimana.

Le variabili più rilevanti nel modello GB sono evidenziate in Figura 3.7. In questo caso si osservano invece differenze nei risultati ottenuti per le due zone in esame. Per quanto riguarda la zona Nord gli unici regressori considerati sono i prezzi ritardati, con particolare riferimento al lag 1 e lag 7, e il giorno della settimana. Si sono ottenuti quindi dei risultati analoghi a quelli del modello RF. Prendendo in considerazione invece la zona Centro-Nord risultano particolarmente influenti solo i prezzi ritardati fino al lag 5. Il prezzo al tempo  $t - 7$  non rientra nelle prime cinque posizioni per nessuna delle 24 ore, contrariamente a quanto visto per tutti gli altri modelli nelle diverse zone. Tra le altre variabili utilizzate le uniche che presentano una percentuale pesata maggiore di zero, anche se bassa, sono il giorno dell'anno e il calendario delle festività.

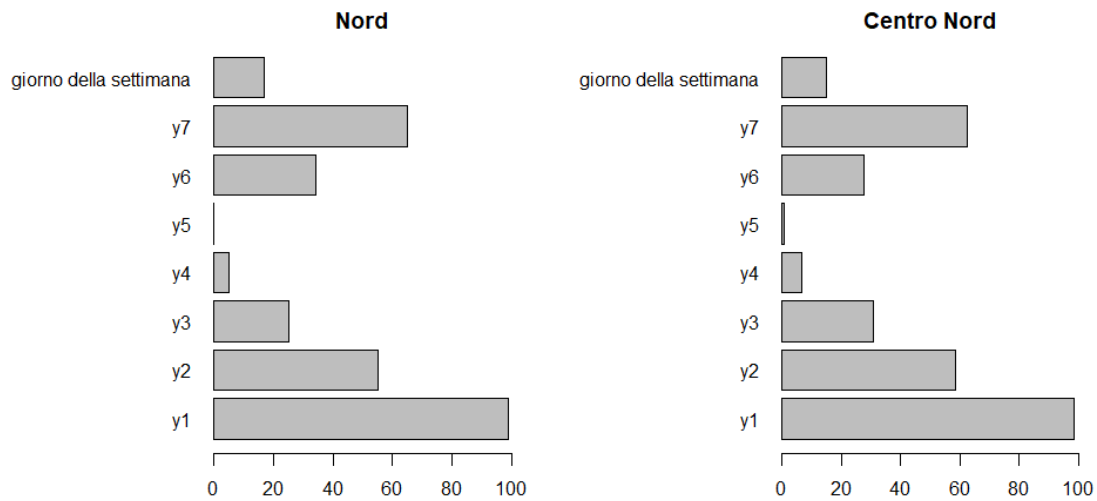


FIGURA 3.6: Misura di importanza delle variabili per il modello RF. Percentuale pesata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore

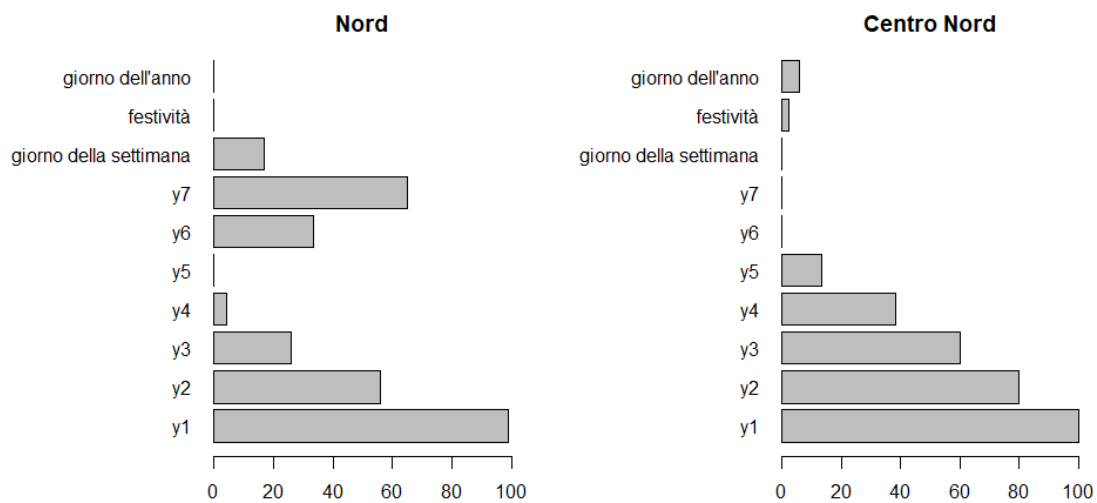


FIGURA 3.7: Misura di importanza delle variabili per il modello GB. Percentuale pesata del numero di volte in cui le variabili rientrano nelle prime cinque posizioni nelle 24 ore

### 3.2.2 Analisi dei consumi

#### 3.2.2.1 Performance previsive dei modelli

Di seguito vengono messe a confronto le prestazioni dei modelli utilizzati per prevedere i consumi di energia elettrica sulla base degli indici di valutazione descritti nel Capitolo 2. Le Tabelle 3.8 e 3.9 riportano i valori medi di RMSE ottenuti con i diversi modelli distinguendo tra ore di punta (14 ore) e ore non di punta (10 ore) per le due zone considerate.

Ore di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	4688.809	4313.462	4357.829	4347.114	4224.85	4409.780
Centro-Nord	814.607	649.248	657.770	643.191	661.85	666.704

TABELLA 3.8: Valore medio dell'RMSE nelle ore di punta

Ore non di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	4674.422	4315.482	4334.155	4371.358	4224.833	4411.506
Centro-Nord	815.619	657.514	664.366	658.016	673.791	684.580

TABELLA 3.9: Valore medio dell'RMSE nelle ore non di punta

In Tabella 3.10 e Tabella 3.11 sono riportati i valori medi di RMSE in rapporto al consumo medio orario distinguendo sempre tra ore di punta e ore non di punta per le due zone considerate.

Ore di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	0.227	0.209	0.211	0.211	0.205	0.214
Centro-Nord	0.208	0.166	0.168	0.165	0.169	0.171

TABELLA 3.10: Valore medio dell'RMSE rapportato con consumo medio orario nelle ore di punta.

Ore non di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	0.227	0.209	0.21	0.212	0.205	0.214
Centro-Nord	0.209	0.168	0.17	0.168	0.172	0.175

TABELLA 3.11: Valore medio dell'RMSE rapportato con consumo medio orario nelle ore non di punta

Dall'analisi dei valori ottenuti emerge come i modelli abbiano prestazioni migliori quando adattati ai dati del Centro-Nord, sia nelle ore di punta che nelle ore non di punta. A differenza di quanto ottenuto nell'analisi sui prezzi in questo caso i modelli non parametrici risultano migliori del modello lineare autoregressivo, presentando un valore di RMSE inferiore. Modelli più complessi, quindi, sembrano spiegare meglio la variabile risposta.

Di seguito si considerano le performance predittive più dettagliatamente in ogni singola ora, continuando a distinguere tra ore di punta e ore non di punta.

In Tabella 3.12 e Tabella 3.13 è riportato per ogni ora il modello migliore sulla base dell'indice RMSE e tra parentesi il numero di test di Diebold Mariano significativi, come riportato nell'analisi sui prezzi.



Ore di punta		
<b>Ora</b>	<b>Nord</b>	<b>Centro-Nord</b>
8	RF (3)	SVM (4)
9	RF (5)	SVM (5)
10	RF (4)	SVM (4)
11	RF (3)	SVM (4)
12	RF (5)	RF (4)
13	RF (2)	ARX-L (3)
14	RF (2)	SVM (3)
15	RF (5)	SVM (5)
16	RF (4)	SVM (5)
17	RF (3)	ARX-L int. (0)
18	RF (5)	RF (3)
19	RF (5)	ARX-L (3)
20	ARX-L (2)	ARX-L (4)
21	RF (3)	SVM (4)

TABELLA 3.12: Modello migliore in termini di RMSE inferiore e numero di Test di Diebold Mariano significativi (ore di punta)

Ore non di punta		
<b>Ora</b>	<b>Nord</b>	<b>Centro-Nord</b>
1	RF (5)	ARX-L (3)
2	RF (2)	ARX-L (4)
3	RF (4)	ARX-L (4)
4	ARX-L (4)	SVM (2)
5	RF (5)	ARX-L (2)
6	RF (5)	RF (2)
7	RF (2)	SVM (2)
22	RF (3)	SVM (4)
23	RF (2)	SVM (3)
24	RF (5)	ARX-L (3)

TABELLA 3.13: Modello migliore in termini di RMSE inferiore e numero di Test di Diebold Mariano significativi (ore non di punta)

Si osserva che non vi sono evidenti differenze tra le ore di punta e le ore non di punta, mentre si hanno risultati diversi tra le due zone considerate. Per quanto riguarda la zona Nord il modello RF presenta RMSE inferiore nella maggior parte delle ore, con almeno 2 test di Diebold Mariano significativi. Solo per la 4° e 20° ora il modello migliore è quello ARX-L con rispettivamente 4 e 2 test di Diebold Mariano significativi. Per quanto riguarda il Centro-Nord SVM e ARX-L sono i modelli migliori per quasi tutte le ore con un numero di test di Diebold Mariano significativi mai inferiore a 2. Per la 6° e 18° ora il modello con RMSE inferiore è RF con rispettivamente 2 e 3 test di Diebold Mariano significativi. Per la 17° ora il modello migliore secondo tale indice è invece ARX-L int. ma con 0 test di Diebold Mariano significativi. Le stesse considerazioni possono essere effettuate osservando l'indice MAE ottenuto dall'adattamento dei diversi modelli. I risultati basati su tale indice sono riportati in Appendice.

### 3.2.2.2 Riconciliazione delle previsioni

In questa sezione vengono presentati i risultati ottenuti dall'applicazione della tecnica di riconciliazione alle previsioni dei consumi energetici. In particolare sono state riconciliate le previsioni orarie con le previsioni a frequenza giornaliera utilizzando gli approcci di riconciliazione ottimale descritti nella sezione 2.4.3.

Per quanto riguarda la zona Nord le procedure di riconciliazione sono state applicate alle previsioni ottenute dall'adattamento del modello RF. Per la zona Centro-Nord invece sono state considerate le previsioni ottenute con il modello SVM. Queste scelte sono motivate dai risultati presentati nella sezione precedente.

Per entrambe le zone considerate sono state utilizzate le seguenti procedure di riconciliazione (a seconda della forma utilizzata per approssimare  $\Omega$ ):

- **t-ols** (Identità);
- **t-struct** (Strutturale);
- **t-wlsv** (*Series variance scaling*);
- **t-shr** (MinT-shr).

Di seguito vengono confrontate le previsioni di base e le previsioni riconciliate con frequenza oraria distinguendo tra le ore di punta e le ore non di punta come nel resto dell'elaborato.

L'accuratezza delle previsioni viene valutata tramite il rapporto tra i valori di RMSE calcolati sulla base delle previsioni riconciliate attraverso le diverse procedure e l'RMSE

calcolato a partire dalle previsioni di base:

$$\frac{RMSE_{ric}}{RMSE_{base}} \tag{3.1}$$

Le previsioni di base che non soddisfano nessun vincolo vengono quindi utilizzate come *benchmark*. A livello interpretativo se le previsioni fornite da una certa tecnica sono migliori di quelle di base il rapporto degli RMSE risulta essere minore di 1. La percentuale di guadagno in termini di accuratezza previsiva rispetto al *benchmark* è pari a

$$\left(1 - \frac{RMSE_{ric}}{RMSE_{base}}\right) \times 100$$

Nelle tabelle che seguono i dati in blu rappresentano i valori più bassi per ogni riga, cioè per ogni ora. Essi indicando quindi il metodo con cui è stato ottenuto il massimo guadagno rispetto alle previsioni di base. Per maggiore chiarezza viene riportato il valore 1 nella colonna riferita alle previsioni di base.

Le Tabelle 3.14 e 3.15 riportano i risultati ottenuti relativi alla zona Nord.

		Ore di punta			
Ora	base	ols	struc	wlsv	shr
8	1	0.9537035	0.9703938	0.9922664	<b>0.9486430</b>
9	1	0.9609270	0.9729282	0.9926114	<b>0.9604542</b>
10	1	0.9629125	0.9747434	0.9931648	<b>0.9619398</b>
11	1	0.9530422	0.9695253	0.9919744	<b>0.9491590</b>
12	1	0.9572781	0.9720655	0.9926085	<b>0.9541869</b>
13	1	0.9399631	0.9631124	0.9906013	<b>0.9277157</b>
14	1	0.9364978	0.9616176	0.9903202	<b>0.9199460</b>
15	1	0.9358525	0.9602296	0.9898173	<b>0.9246533</b>
16	1	0.9607806	0.9736738	0.9929313	<b>0.9592165</b>
17	1	<b>0.9781969</b>	0.9824759	0.9948652	0.9784718
18	1	0.9678267	0.9776048	0.9938660	<b>0.9674925</b>
19	1	0.9509388	0.9686556	0.9918197	<b>0.9455368</b>
20	1	0.9349857	0.9611337	0.9902658	<b>0.9150077</b>
21	1	0.9276967	0.9562023	0.9889496	<b>0.9099150</b>
media	1	0.9514716	0.968883	0.9918616	<b>0.9444527</b>

TABELLA 3.14: Rapporto tra RMSE delle procedure di riconciliazione e RMSE calcolato a partire dalle previsioni di base nella zona Nord (ore di punta)

Si osserva immediatamente che tutti i valori relativi alle previsioni riconciliate sono inferiori a 1. Questo, come detto, indica che le tecniche di riconciliazione apportano un miglioramento in termini di accuratezza delle previsioni. È inoltre evidente come

Ore non di punta					
Ora	base	ols	struc	wlsv	shr
1	1	0.9553294	0.9708569	0.9923003	<b>0.9519731</b>
2	1	0.9668346	0.9771731	0.9937845	<b>0.9662281</b>
3	1	<b>0.9687163</b>	0.9770452	0.9935496	0.9691148
4	1	0.9473351	0.9666636	0.9913496	<b>0.9405932</b>
5	1	0.9520477	0.9686983	0.9917314	<b>0.9485351</b>
6	1	0.9492195	0.9681092	0.9917603	<b>0.9417432</b>
7	1	0.9486368	0.9675749	0.9915977	<b>0.9417221</b>
22	1	0.9400692	0.9630581	0.9905692	<b>0.9284502</b>
23	1	<b>0.9784992</b>	0.9826659	0.9949141	0.9787569
24	1	0.9953430	0.9918813	0.9970754	<b>0.9918427</b>
media	1	0.9602031	0.9733727	0.9928632	<b>0.955896</b>

TABELLA 3.15: Rapporto tra RMSE delle procedure di riconciliazione e RMSE calcolato a partire dalle previsioni di base nella zona Nord (ore non di punta)

il metodo di riconciliazione ottimale migliore risultati essere t-shr a quasi tutte le ore. I valori medi riportati nell'ultima riga delle tabelle lo confermano. Nelle ore di punta è stato ottenuto un guadagno massimo del 9.01% (21° ora). Nelle ore non di punta il guadagno massimo è stato leggermente inferiore, pari al 7.15% (22° ora). In ogni caso non si evidenziano per questa zona importanti differenze tra le performance della tecnica di riconciliazione nelle ore di punta e nelle ore non di punta.

Per quanto riguarda il Centro-Nord analogamente le Tabelle 3.16 e 3.17 riportano il rapporto in (3.1) per ogni ora.

I risultati relativi alla zona Centro-Nord sono in accordo con quelli ottenuti per la zona Nord. I valori riportati risultano anche in questo caso tutti inferiori a 1 indicando un miglioramento dovuto alle tecniche di riconciliazione. In generale si sono ottenuti anche diversi valori inferiori a 0.90, il quale rappresenta invece il valore più basso ottenuto per la zona Nord. Il metodo con prestazioni migliori risulta essere t-shr anche per il Centro-Nord. Nelle ore di punta è stato ottenuto un guadagno massimo del 13.48% (8° ora). Nelle ore non di punta si hanno prestazioni addirittura migliori: è stato raggiunto un guadagno massimo pari al 16.1% (7° ora).

Le previsioni di base e quelle riconciliate sono state confrontate anche tramite l'indice MAE. I risultati ottenuti sono in linea con quelli appena presentati e sono riportati in forma estesa in Appendice.

Ore di punta					
Ora	base	ols	struc	wlsv	shr
8	1	0.9120081	0.9494381	0.9877336	<b>0.8651799</b>
9	1	0.9368446	0.9613284	0.9902172	<b>0.9233853</b>
10	1	0.9381374	0.9623440	0.9905068	<b>0.9236236</b>
11	1	0.9326211	0.9597395	0.9899703	<b>0.9119125</b>
12	1	0.9278808	0.9583252	0.9898411	<b>0.8911547</b>
13	1	0.9206653	0.9541723	0.9888372	<b>0.8808588</b>
14	1	0.9143226	0.9503967	0.9879055	<b>0.8735757</b>
15	1	0.9305460	0.9580805	0.9894944	<b>0.9123525</b>
16	1	0.9485241	0.9676535	0.9916803	<b>0.9410484</b>
17	1	0.9592060	0.9735029	0.9930457	<b>0.9558166</b>
18	1	0.9624885	0.9757725	0.9936586	<b>0.9589673</b>
19	1	0.9489228	0.9686026	0.9920284	<b>0.9383236</b>
20	1	0.9202289	0.9536110	0.9886519	<b>0.8842417</b>
21	1	0.9223112	0.9540025	0.9886169	<b>0.8956011</b>
media	1	0.9339077	0.9604978	0.9901563	<b>0.9111458</b>

TABELLA 3.16: Rapporto tra RMSE delle procedure di riconciliazione e RMSE calcolato a partire dalle previsioni di base nella zona Centro-Nord (ore di punta)

Ore non di punta					
Ora	base	ols	struc	wlsv	shr
1	1	0.9117772	0.9498668	0.9879264	<b>0.8555425</b>
2	1	0.9157845	0.9516426	0.9882708	<b>0.8699537</b>
3	1	0.9317296	0.9589174	0.9897216	<b>0.9129479</b>
4	1	0.9224084	0.9544360	0.9887822	<b>0.8921364</b>
5	1	0.9064841	0.9468806	0.9872148	<b>0.8474195</b>
6	1	0.9120829	0.9505307	0.9881644	<b>0.8457923</b>
7	1	0.9048650	0.9462709	0.9871185	<b>0.8390316</b>
22	1	0.9482910	0.9676374	0.9916959	<b>0.9402687</b>
23	1	<b>0.9795947</b>	0.9838315	0.9953105	0.9798983
24	1	0.9888562	0.9896011	0.9967842	<b>0.9883092</b>
media	1	0.9321874	0.9599615	0.9900989	<b>0.89713</b>

TABELLA 3.17: Rapporto tra RMSE delle procedure di riconciliazione e RMSE calcolato a partire dalle previsioni di base nella zona Centro-Nord (ore non di punta)



# Conclusioni

Nel presente elaborato sono state utilizzate diverse metodologie per la previsione dei prezzi e della domanda del mercato elettrico italiano. Il primo obiettivo dell'analisi è stato confrontare le performance previsive di modelli di *machine learning*, quali *Support Vector Machines* (SVM), *Random Forest* (RF) e *Gradient Boosting* (GB), con il modello autoregressivo lineare, tipicamente utilizzato nell'ambito delle previsioni energetiche. L'analisi, inoltre, mirava a valutare l'influenza dei regressori utilizzati per analizzare i prezzi spot dell'elettricità. Infine, scopo ultimo del lavoro di tesi era verificare l'efficienza di tecniche di riconciliazione ottimale delle previsioni applicandole ai risultati ottenuti attraverso l'analisi dei consumi energetici.

Nell'elaborato sono stati descritti il funzionamento e le principali caratteristiche del mercato energetico italiano. Successivamente sono stati presentati i diversi modelli adattati ai dati, le misure di importanza delle variabili prese in considerazione per l'analisi più approfondita di ciascun regressore, gli approcci di riconciliazione delle previsioni presenti in letteratura e gli indici presi in esame per la valutazione delle metodologie applicate. Infine sono stati riportati i risultati ottenuti dalle analisi.

Per quanto riguarda la previsione dei prezzi il modello autoregressivo lineare con variabili esogene si è rivelato essere quello migliore rispetto ai modelli concorrenti. I modelli di *machine learning*, in particolare RF e GB, hanno avuto performance previsive peggiori in ogni ora del giorno e in entrambe le zone d'Italia prese in considerazione.

In accordo con tale risultato l'analisi d'importanza delle variabili, in primo luogo, ha evidenziato la particolare rilevanza dell'informazione passata, ovvero i prezzi fino a sette giorni prima. In secondo luogo, la domanda di energia elettrica prevista e il giorno della settimana sono risultate variabili influenti. Si è rivelata, invece, di minore importanza la generazione eolica prevista un giorno prima.

Dall'analisi della domanda sono stati ottenuti risultati differenti.

I modelli di apprendimento automatico hanno avuto in questo caso prestazioni migliori rispetto al semplice modello autoregressivo. Prendendo in considerazione la zona Nord d'Italia, il modello RF è risultato il migliore in termini di accuratezza delle previsioni in

quasi tutte le 24 ore. Per quanto riguarda la zona Centro-Nord i risultati migliori sono stati ottenuti in particolare con il modello SVM per la maggior parte delle ore. Anche il modello ARX-L ha rivelato buone capacità previsive in alcune ore.

Sulla base dei risultati ottenuti nell'analisi della domanda, le procedure di riconciliazione ottimale sono state applicate alle previsioni di base ottenute dall'adattamento del modello RF per la zona Nord e del modello SVM per la zona Centro-Nord. Sono state riconciliate le previsioni orarie e le previsioni con frequenza giornaliera. L'efficacia di tali tecniche, riportata in letteratura, è stata confermata. È stato dimostrato, infatti, che riconciliando le previsioni di base si ha un guadagno in termini di accuratezza. La procedura con cui è stato ottenuto un maggior guadagno è risultata t-shr in entrambe le zone considerate.

Un ulteriore sviluppo del presente lavoro potrebbe essere quello di effettuare le analisi dei prezzi e della domanda tenendo conto di molte altre variabili che possono spiegarne la dinamica, come informazioni sul mercato infragiornaliero, fattori meteorologici e climatici o componenti socioeconomiche. Inoltre, dal momento che sono stati ottenuti dei risultati incoraggianti sfruttando le tecniche di riconciliazione ottimale, potrebbe essere interessante ampliare l'analisi. Si potrebbero, ad esempio, prendere in considerazione altri metodi, come quelli proposti da Hollyman et al., 2021 e Panagiotelis et al., 2023, e valutarne l'applicabilità a questo tipo di dati. Inoltre, potrebbero essere presi in esame anche altri livelli di aggregazione delle serie oltre a quello giornaliero.



# Appendice

## A Analisi dei prezzi: risultati basati sull'indice MAE

Di seguito sono riportati i risultati ottenuti dall'adattamento dei modelli per prevedere i prezzi spot dell'energia sulla base dell'indice MAE. Tali risultati sono in linea con quanto ottenuto sulla base dell'indice RMSE descritto nel Capitolo 3.

Le Tabelle A.1 e A.2 riportano i valori medi di MAE ottenuti con i diversi modelli distinguendo tra ore di punta (14 ore) e ore non di punta (10 ore) per le due zone considerate.

Ore di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	5.576	6.347	7.906	5.866	8.346	8.497
Centro-Nord	6.065	6.851	8.333	6.307	8.702	8.654

TABELLA A.1: Valore medio del MAE nelle ore di punta.

Ore non di punta						
Zona	ARX	ARX-L	ARX-L int.	SVM	RF	GB
Nord	3.405	3.947	4.682	3.654	5.225	4.763
Centro-Nord	4.144	5.235	5.919	4.699	5.867	5.917

TABELLA A.2: Valore medio del MAE nelle ore non di punta.

In Tabella A.3 e Tabella A.4 sono riportati i valori medi di MAE in rapporto al prezzo medio orario distinguendo sempre tra ore di punta e ore non di punta per le due zone considerate.

Ore di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	0.094	0.107	0.133	0.099	0.140	0.142
Centro-Nord	0.101	0.114	0.139	0.106	0.146	0.144

TABELLA A.3: Valore medio del MAE rapportato con il prezzo medio orario nelle ore di punta.

Ore non di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	0.069	0.079	0.094	0.074	0.107	0.097
Centro-Nord	0.083	0.103	0.117	0.094	0.118	0.119

TABELLA A.4: Valore medio del MAE rapportato con il prezzo medio orario nelle ore non di punta.

Di seguito si considerano le performance predittive più dettagliatamente in ogni singola ora, continuando a distinguere tra ore di punta e ore non di punta.

In Tabella A.5 e Tabella A.6 sono riportati per ogni ora il modello migliore in termini di MAE inferiore e tra parentesi il numero di test di Diebold Mariano significativi.

Ore di punta		
Ora	Nord	Centro-Nord
8	ARX (5)	ARX (5)
9	SVM (4)	ARX (4)
10	ARX (4)	ARX (5)
11	ARX (5)	ARX (5)
12	ARX (5)	ARX (5)
13	ARX (5)	ARX (5)
14	ARX (4)	ARX (5)
15	ARX (5)	ARX (4)
16	ARX (5)	ARX (5)
17	ARX (5)	ARX (4)
18	ARX (5)	ARX (3)
19	ARX (4)	ARX (4)
20	ARX (4)	ARX (4)
21	SVM (4)	SVM (4)

TABELLA A.5: Modello migliore in termini di MAE inferiore e numero di Test di Diebold Mariano significativi (ore di punta).

Ore non di punta		
Ora	Nord	Centro-Nord
1	ARX (5)	ARX (5)
2	ARX (3)	ARX (5)
3	ARX (3)	ARX (5)
4	SVM (3)	ARX (5)
5	ARX (4)	ARX (5)
6	ARX (5)	ARX (5)
7	ARX (5)	ARX (5)
22	SVM (4)	SVM (5)
23	ARX (5)	ARX (5)
24	ARX (5)	ARX (5)

TABELLA A.6: Modello migliore in termini di MAE inferiore e numero di Test di Diebold Mariano significativi (ore non di punta).

## B Grafici di confronto delle previsioni dei prezzi

I seguenti grafici confrontano le previsioni dei prezzi spot dell'energia. Vengono rappresentate le previsioni ottenute dall'adattamento di ogni modello e la serie osservata separatamente relative ai mesi di gennaio, aprile, luglio e novembre del 2020.

Di seguito sono riportati i grafici relativi alla zona Nord.

### Gennaio

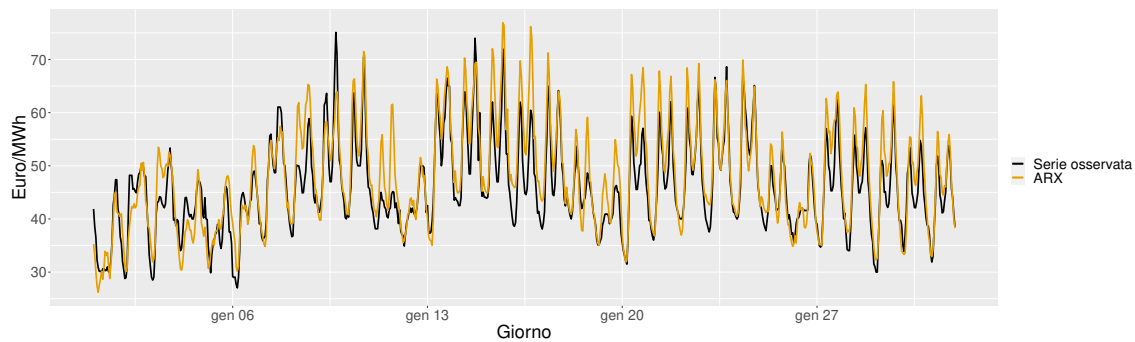


FIGURA B.1: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per gennaio 2020 nella zona Nord.

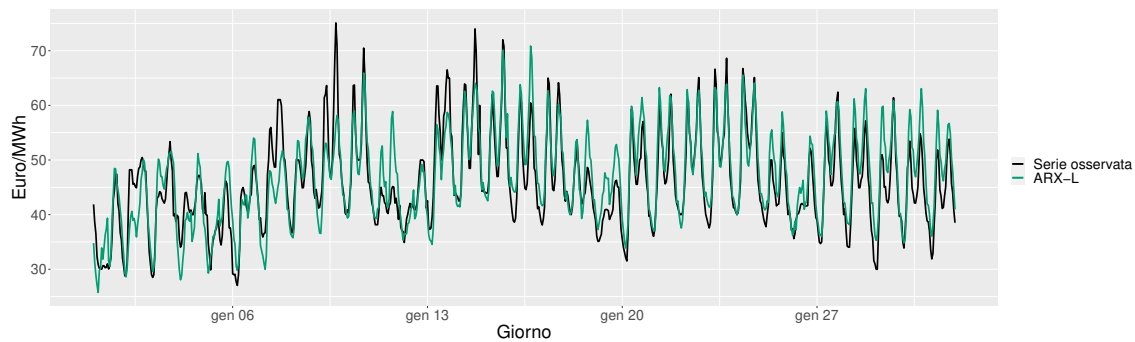


FIGURA B.2: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per gennaio 2020 nella zona Nord.

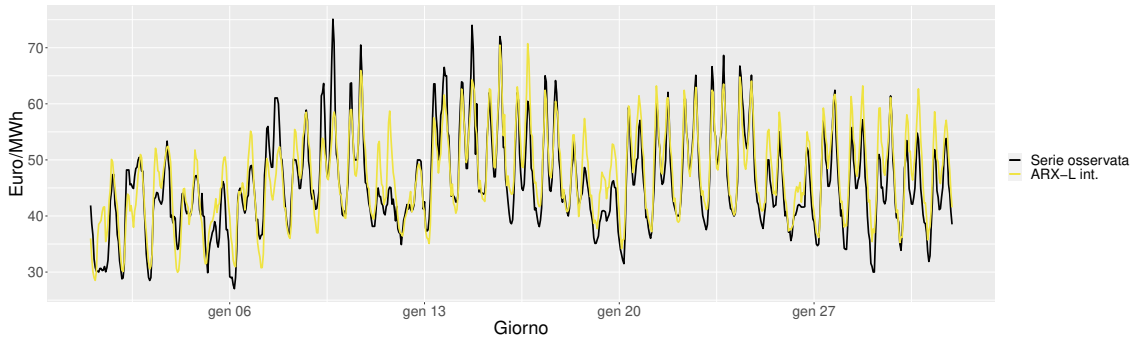


FIGURA B.3: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per gennaio 2020 nella zona Nord.

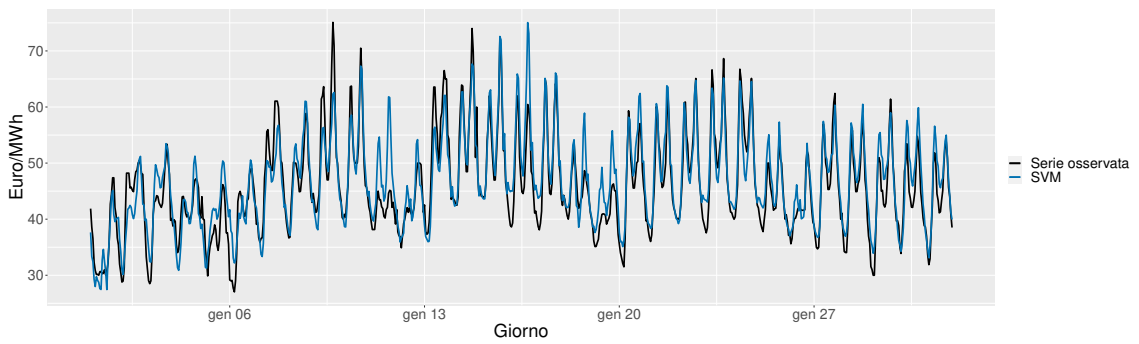


FIGURA B.4: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per gennaio 2020 nella zona Nord.

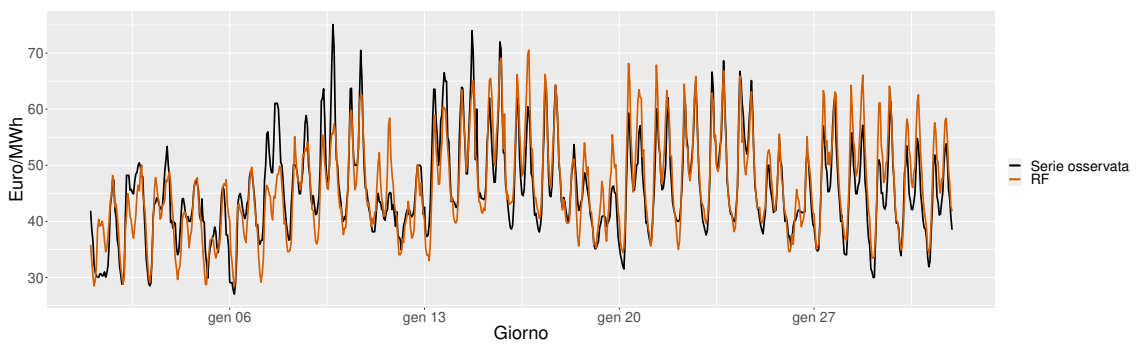


FIGURA B.5: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per gennaio 2020 nella zona Nord.

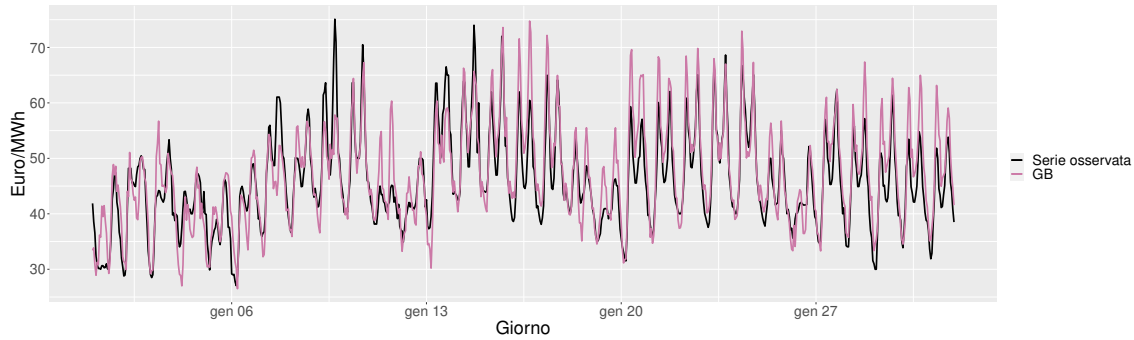


FIGURA B.6: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per gennaio 2020 nella zona Nord.

### Aprile

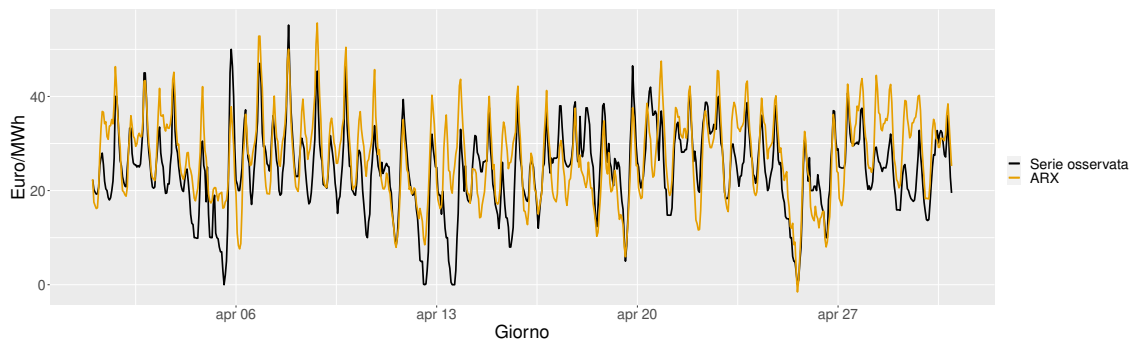


FIGURA B.7: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per aprile 2020 nella zona Nord.

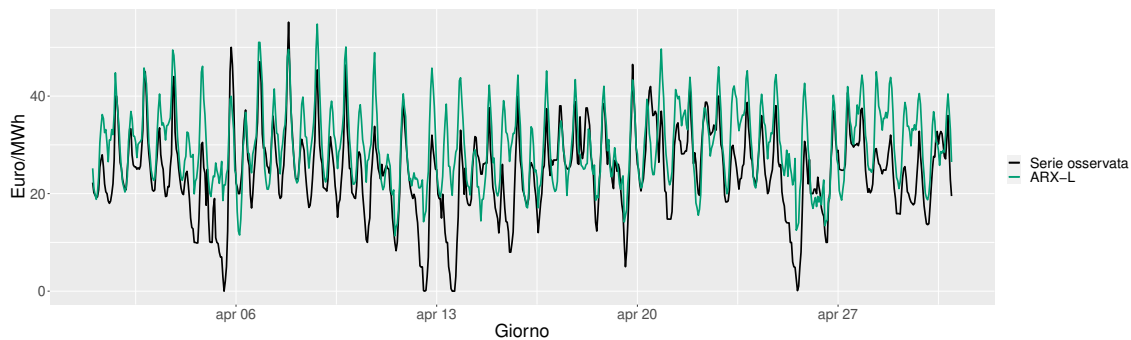


FIGURA B.8: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per aprile 2020 nella zona Nord.

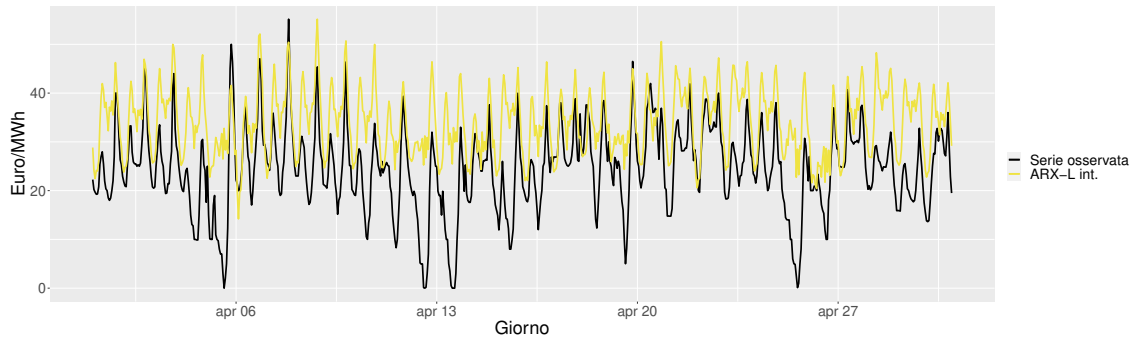


FIGURA B.9: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per aprile 2020 nella zona Nord.

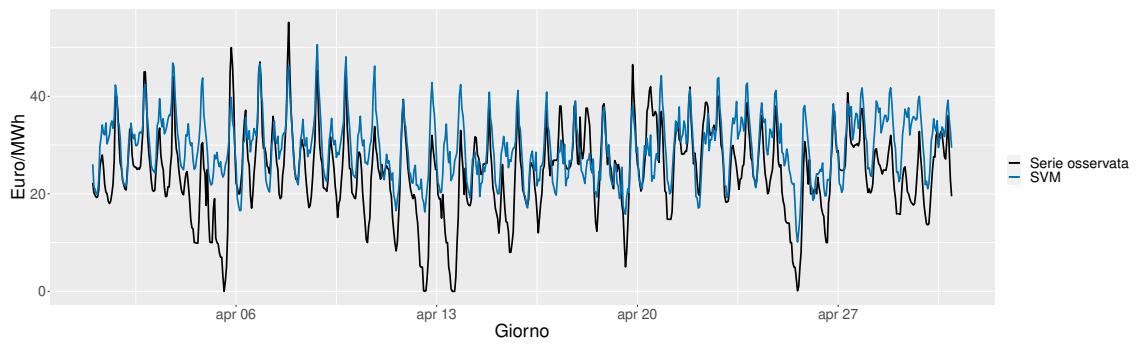


FIGURA B.10: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per aprile 2020 nella zona Nord.

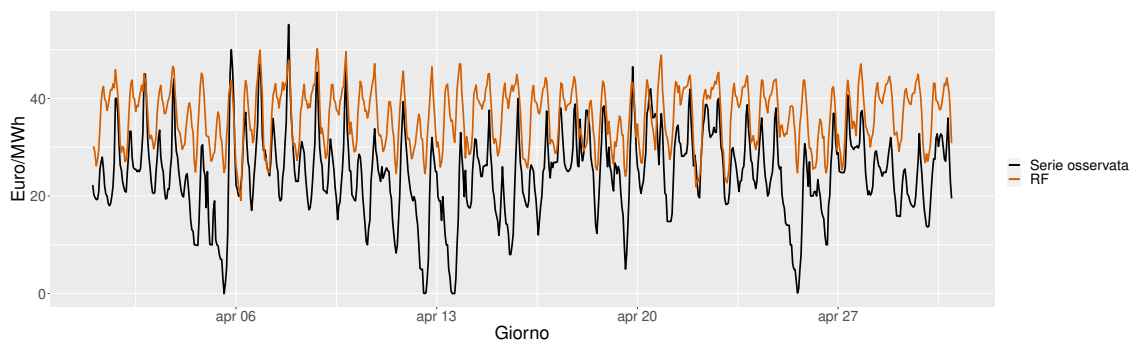


FIGURA B.11: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per aprile 2020 nella zona Nord.

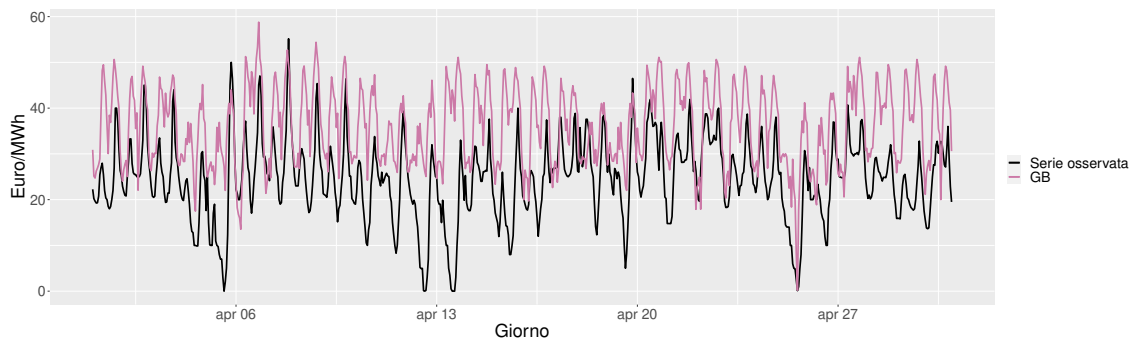


FIGURA B.12: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per aprile 2020 nella zona Nord.

## Luglio

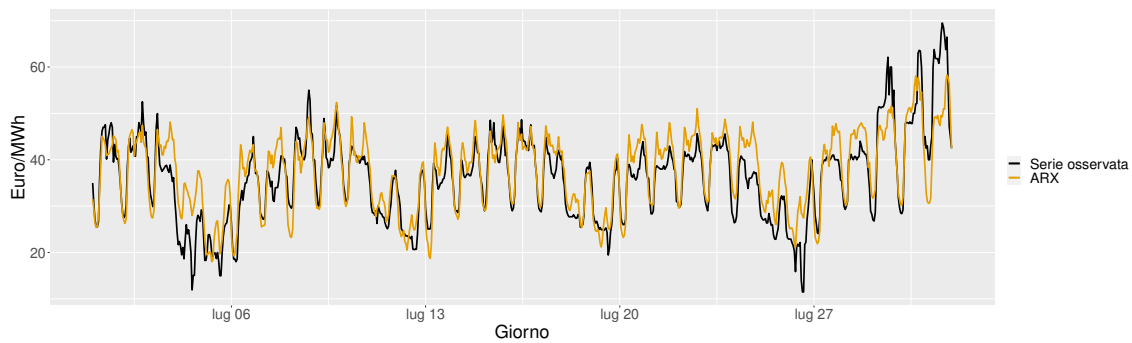


FIGURA B.13: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per luglio 2020 nella zona Nord.

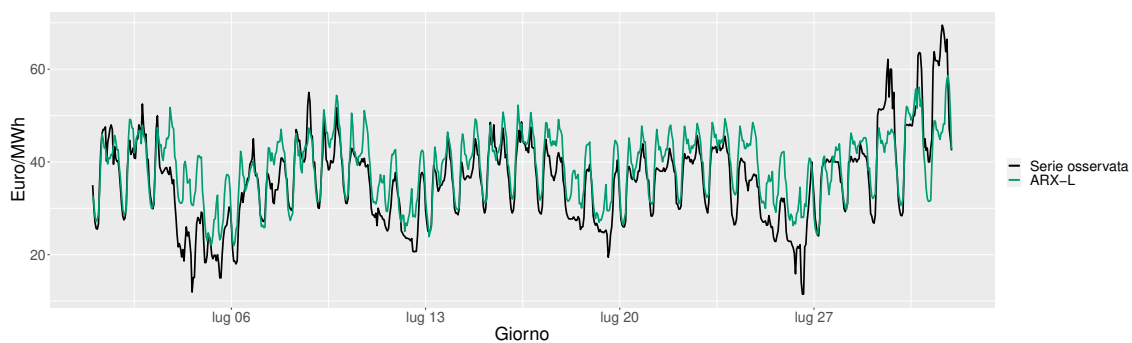


FIGURA B.14: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per luglio 2020 nella zona Nord.



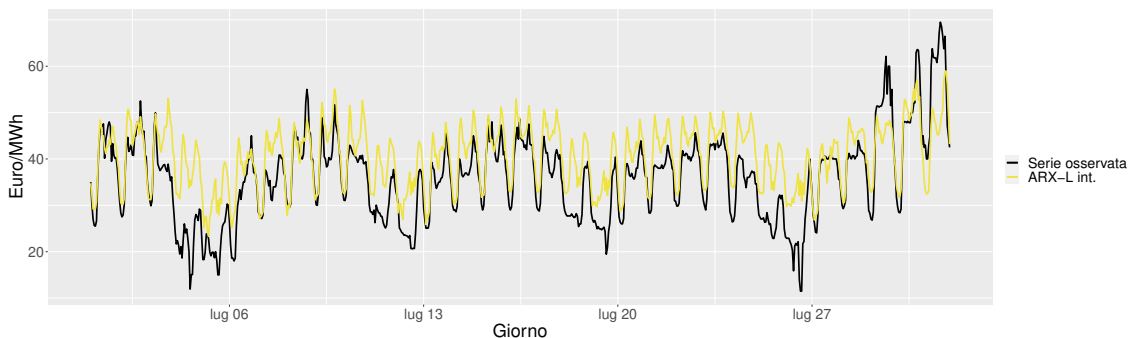


FIGURA B.15: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per luglio 2020 nella zona Nord.

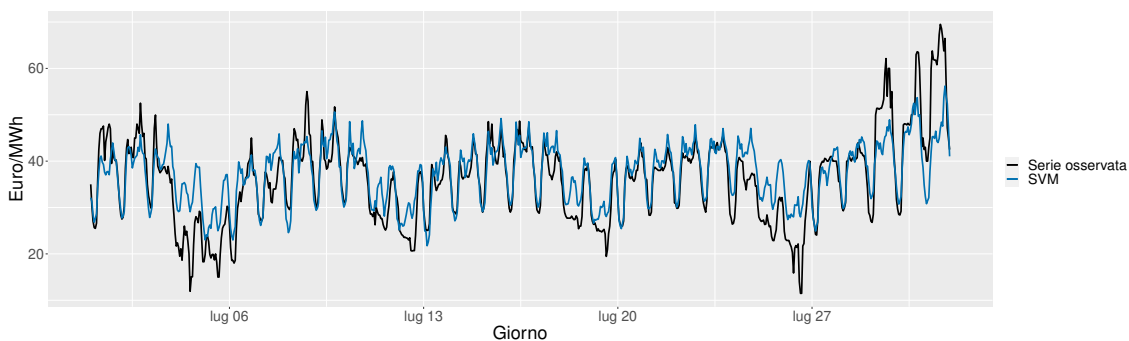


FIGURA B.16: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per luglio 2020 nella zona Nord.

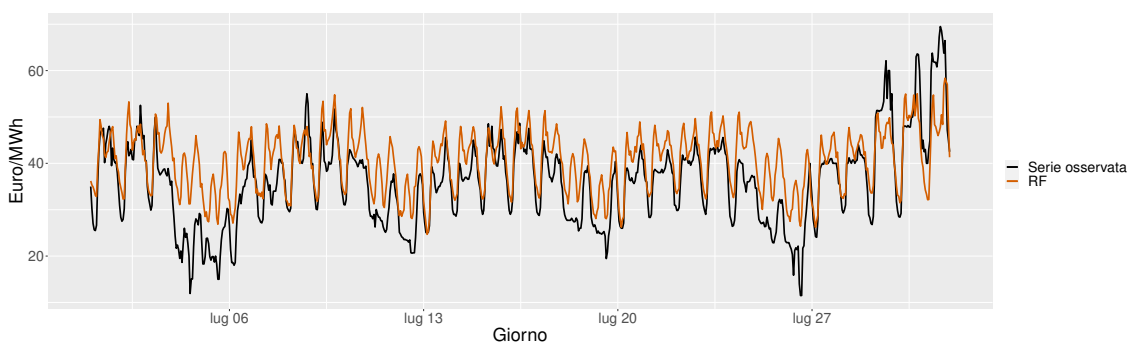


FIGURA B.17: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per luglio 2020 nella zona Nord.

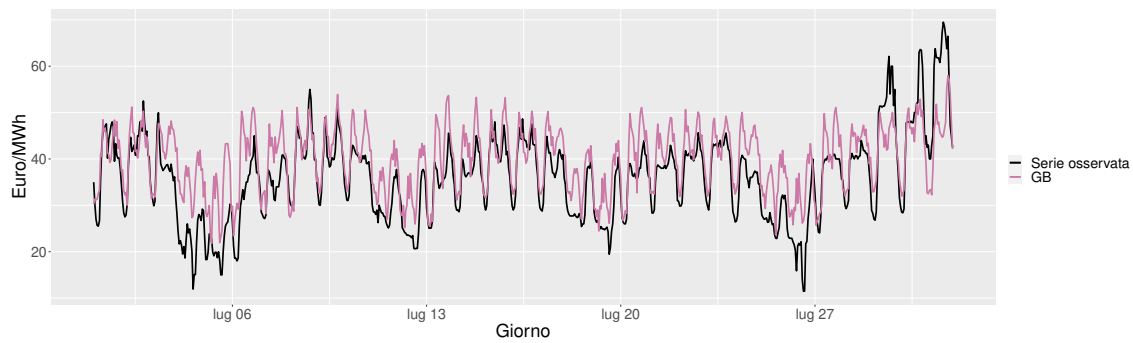


FIGURA B.18: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per luglio 2020 nella zona Nord.

### Novembre

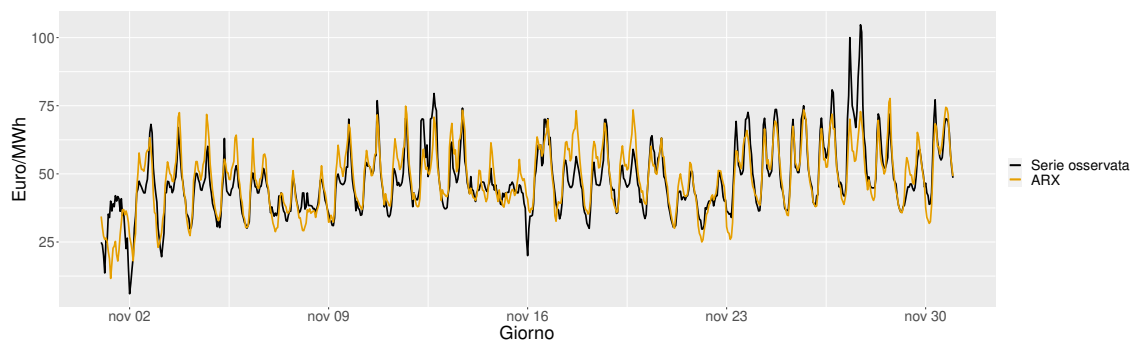


FIGURA B.19: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per novembre 2020 nella zona Nord.

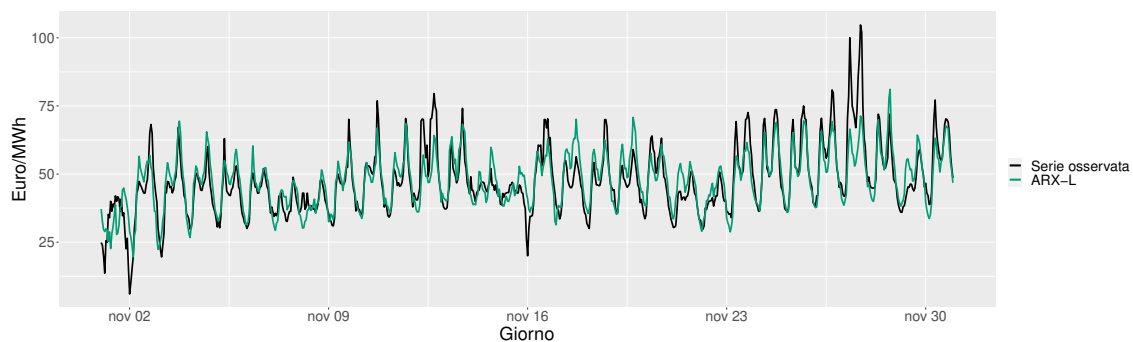


FIGURA B.20: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per novembre 2020 nella zona Nord.

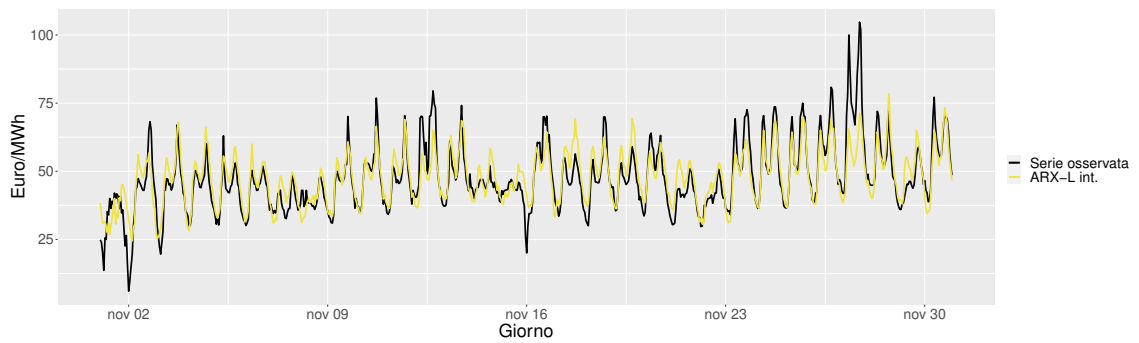


FIGURA B.21: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per novembre 2020 nella zona Nord.

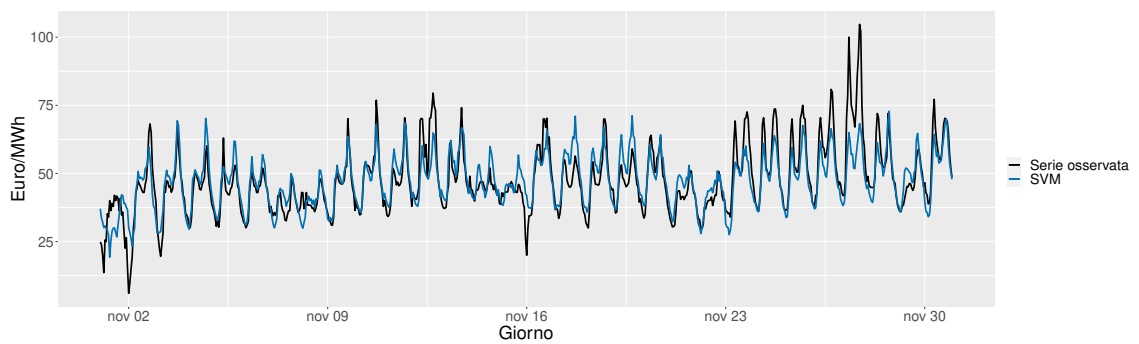


FIGURA B.22: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per novembre 2020 nella zona Nord.

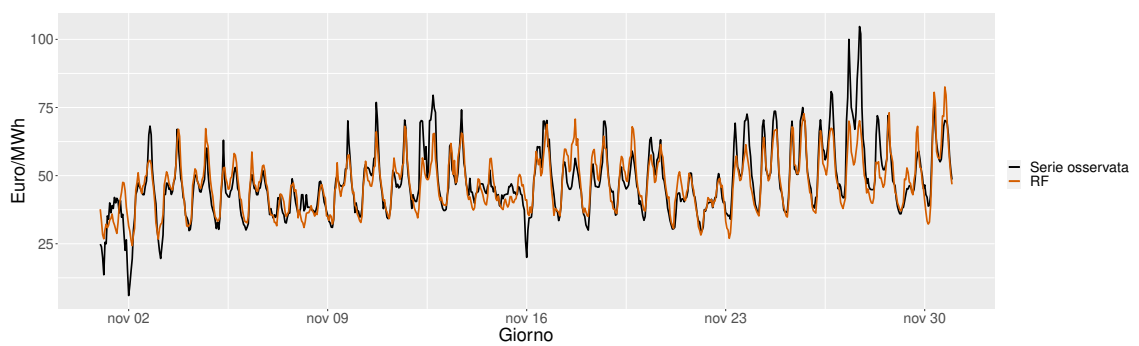


FIGURA B.23: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per novembre 2020 nella zona Nord.

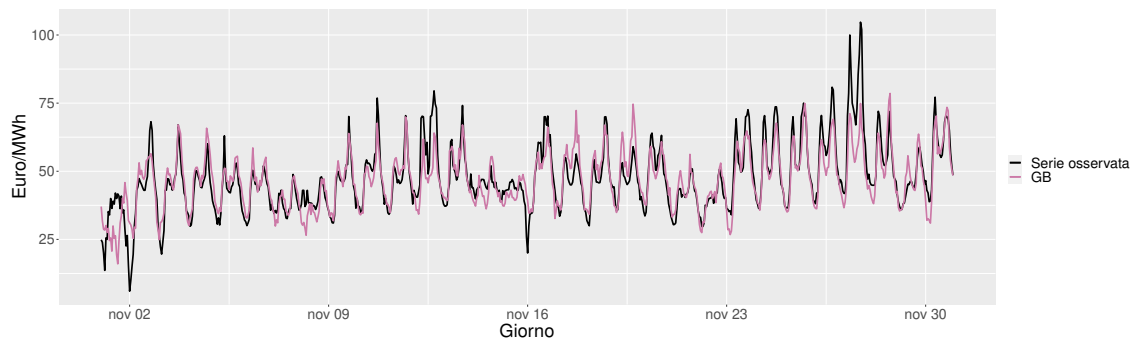


FIGURA B.24: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per novembre 2020 nella zona Nord.

Di seguito sono invece riportati i grafici relativi alla zona Centro-Nord nei mesi di gennaio, aprile, luglio e novembre del 2020.

### Gennaio

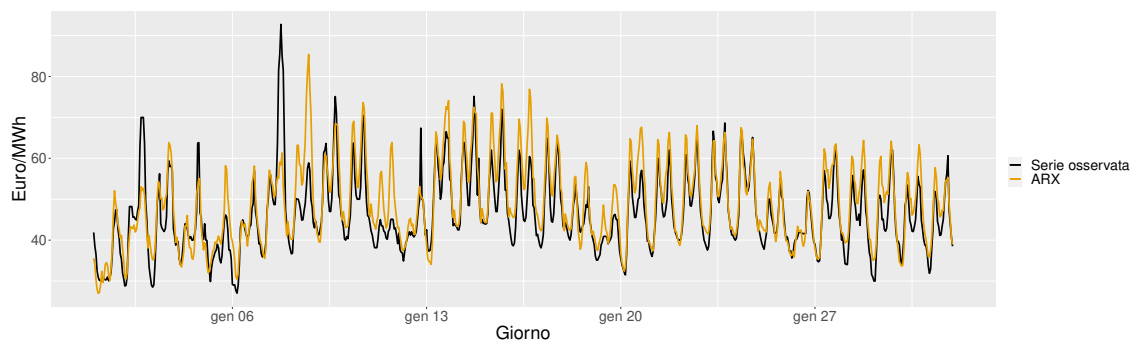


FIGURA B.25: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per gennaio 2020 nella zona Centro-Nord.

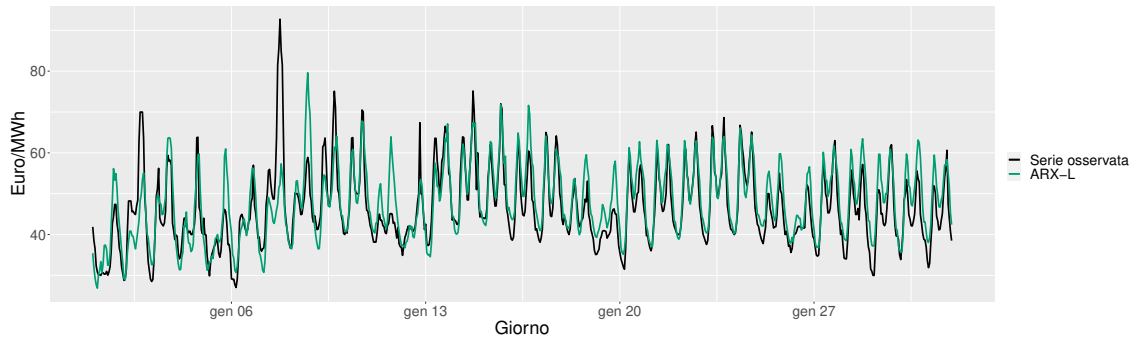


FIGURA B.26: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per gennaio 2020 nella zona Centro-Nord.

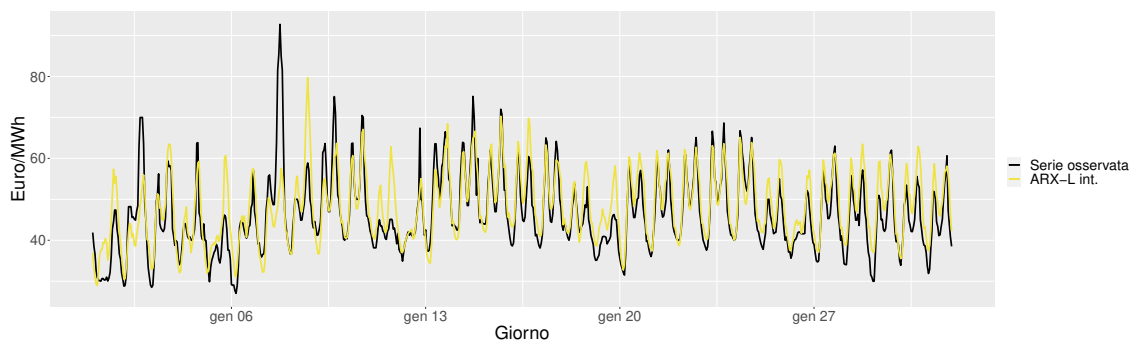


FIGURA B.27: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per gennaio 2020 nella zona Centro-Nord.

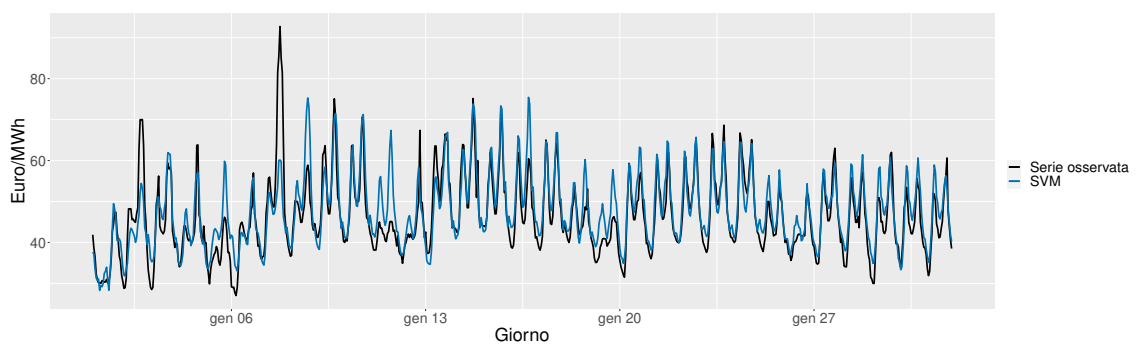


FIGURA B.28: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per gennaio 2020 nella zona Centro-Nord.

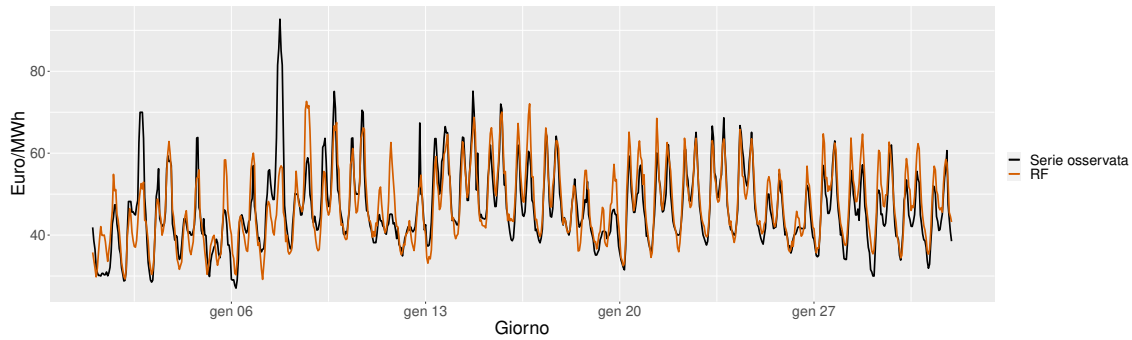


FIGURA B.29: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per gennaio 2020 nella zona Centro-Nord.

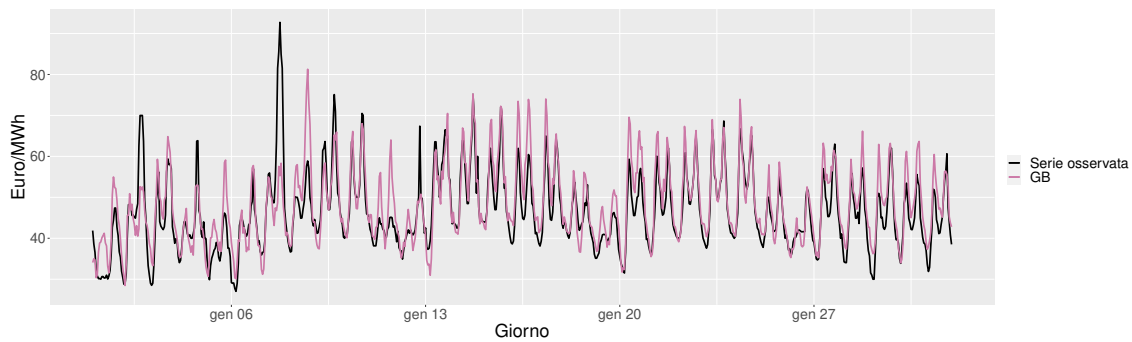


FIGURA B.30: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per gennaio 2020 nella zona Centro-Nord.

## Aprile

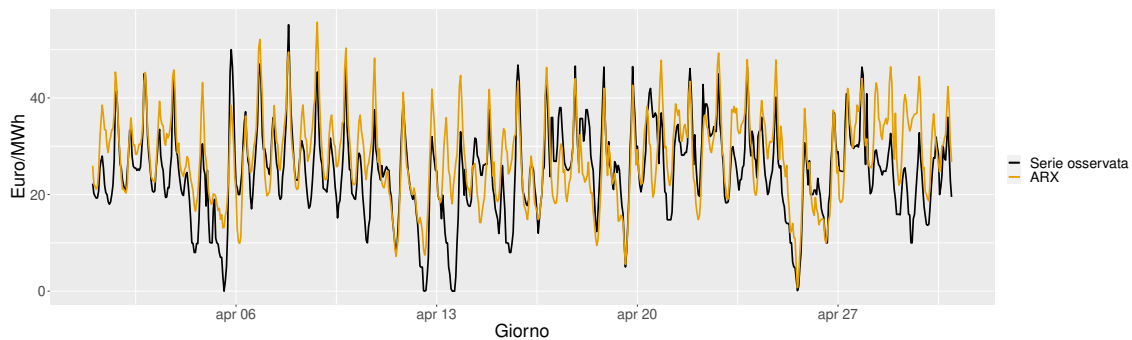


FIGURA B.31: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per aprile 2020 nella zona Centro-Nord.

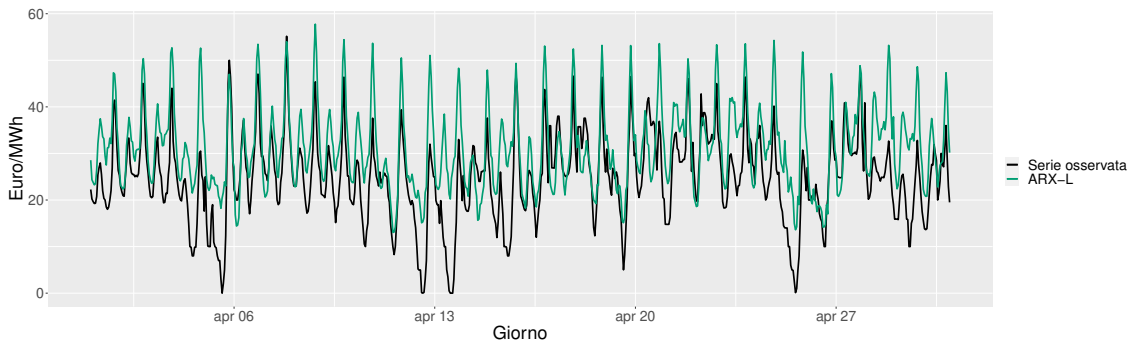


FIGURA B.32: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per aprile 2020 nella zona Centro-Nord.

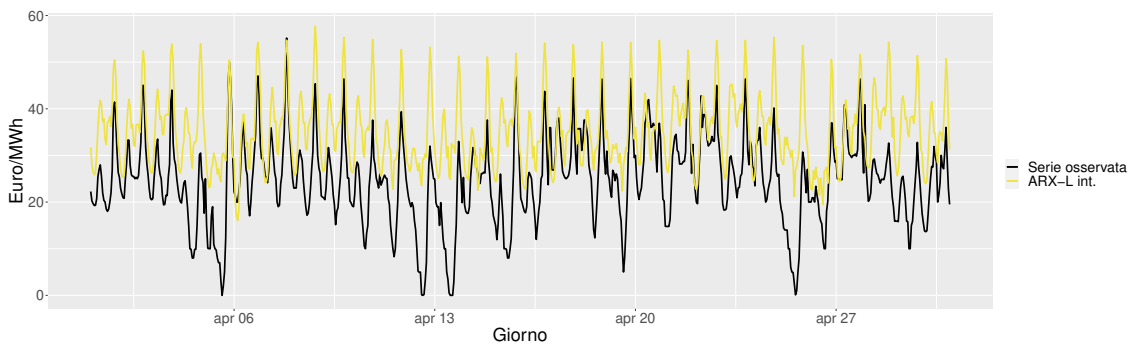


FIGURA B.33: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per aprile 2020 nella zona Centro-Nord.

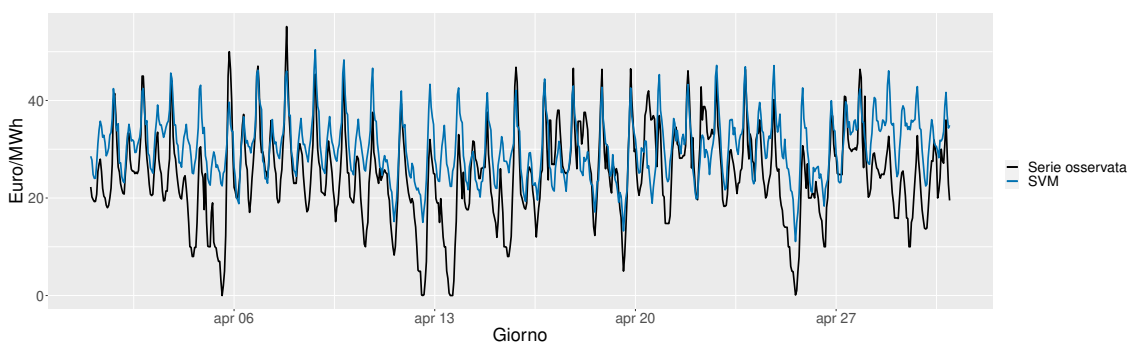


FIGURA B.34: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per aprile 2020 nella zona Centro-Nord.

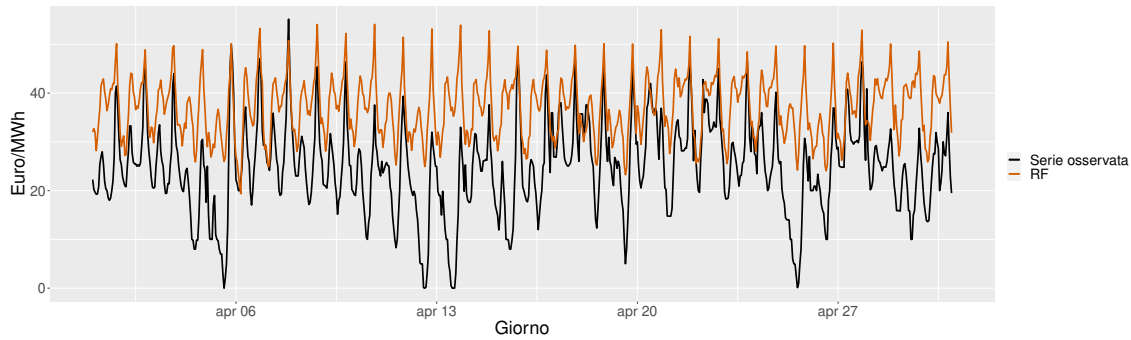


FIGURA B.35: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per aprile 2020 nella zona Centro-Nord.

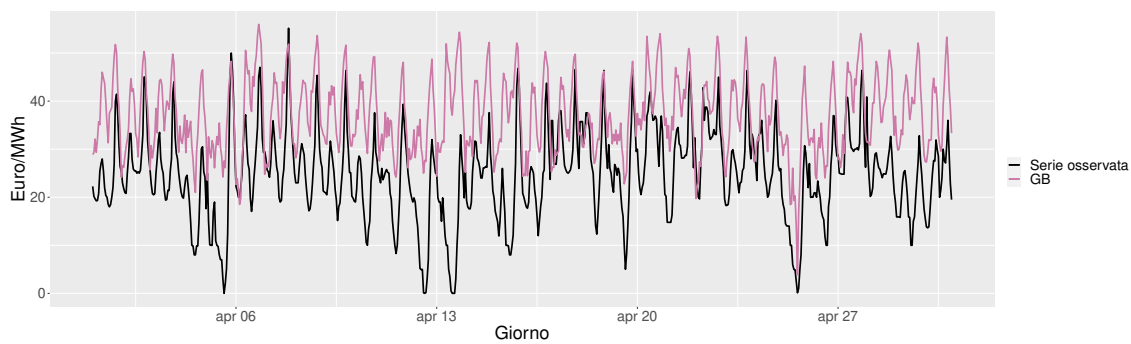


FIGURA B.36: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per aprile 2020 nella zona Centro-Nord.

## Luglio

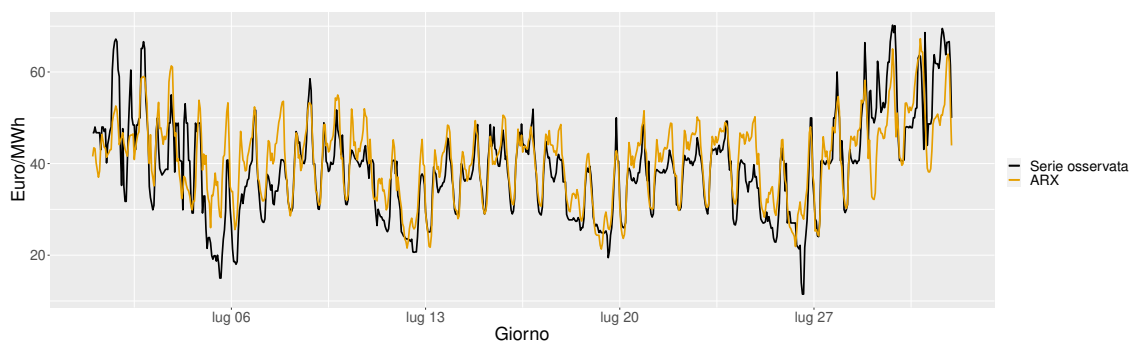


FIGURA B.37: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per luglio 2020 nella zona Centro-Nord.



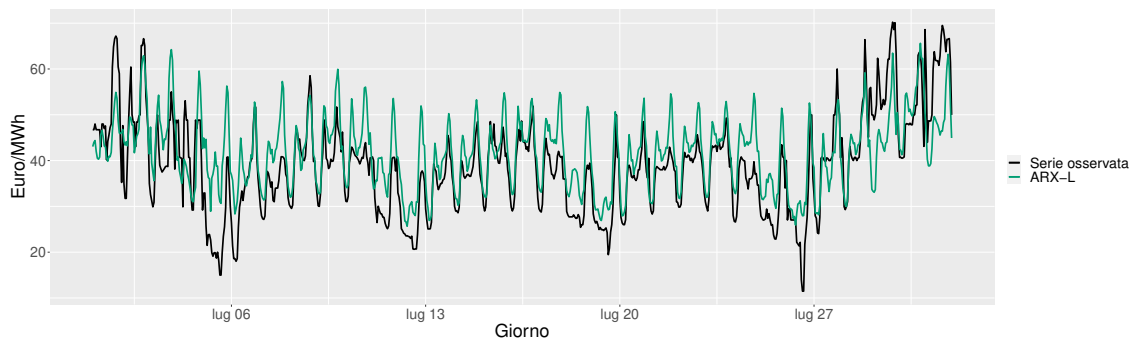


FIGURA B.38: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per luglio 2020 nella zona Centro-Nord.

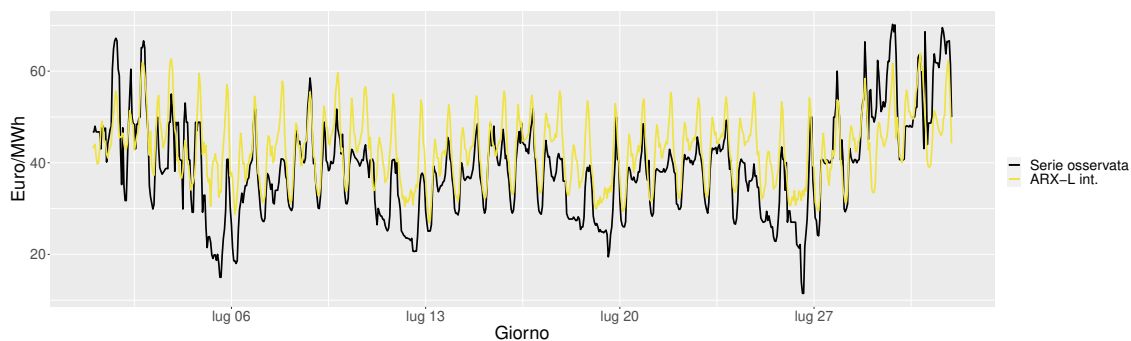


FIGURA B.39: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per luglio 2020 nella zona Centro-Nord.

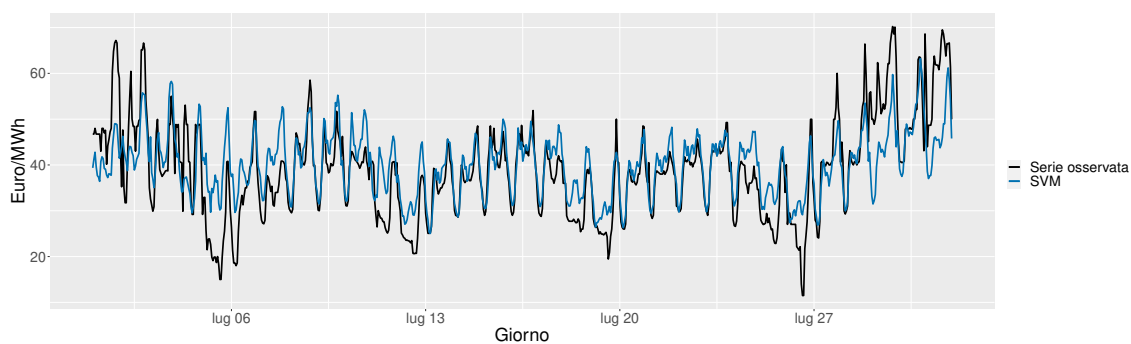


FIGURA B.40: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per luglio 2020 nella zona Centro-Nord.

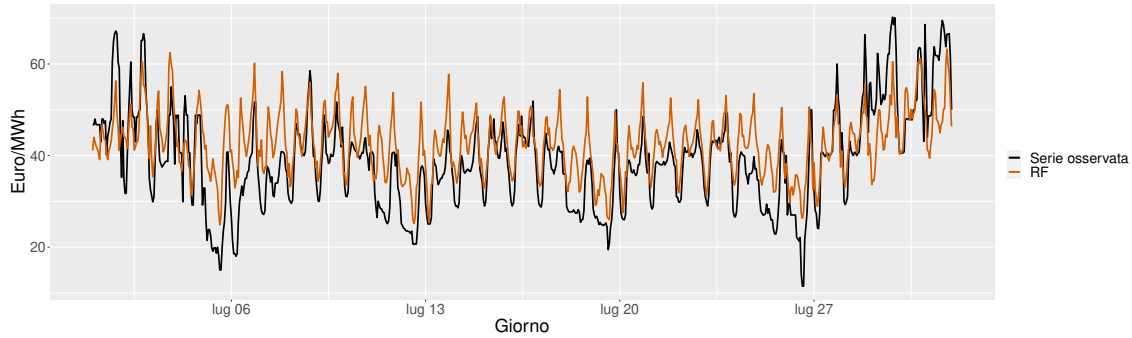


FIGURA B.41: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per luglio 2020 nella zona Centro-Nord.

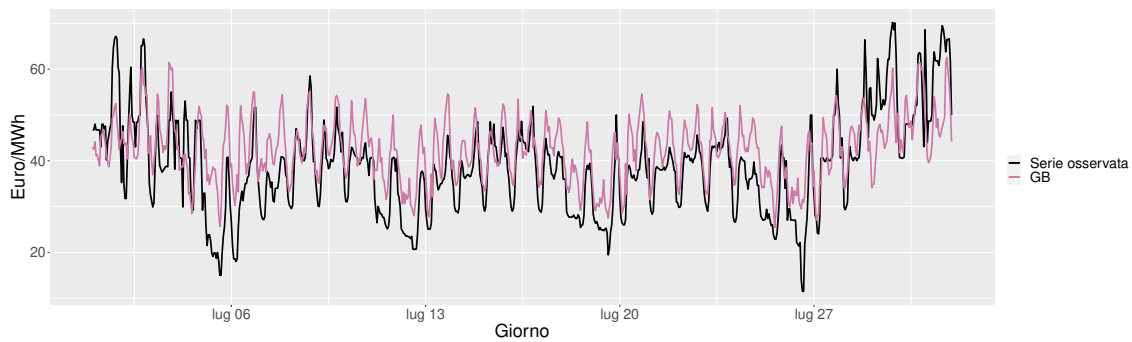


FIGURA B.42: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per luglio 2020 nella zona Centro-Nord.

## Novembre

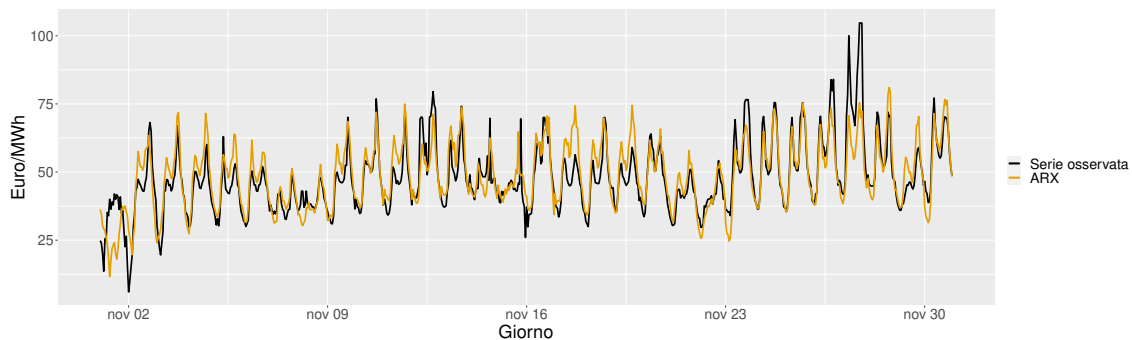


FIGURA B.43: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX per novembre 2020 nella zona Centro-Nord.

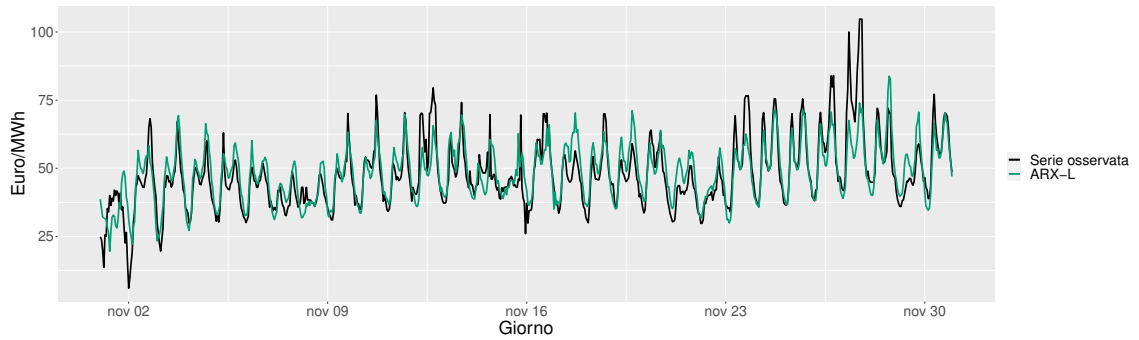


FIGURA B.44: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L per novembre 2020 nella zona Centro-Nord.

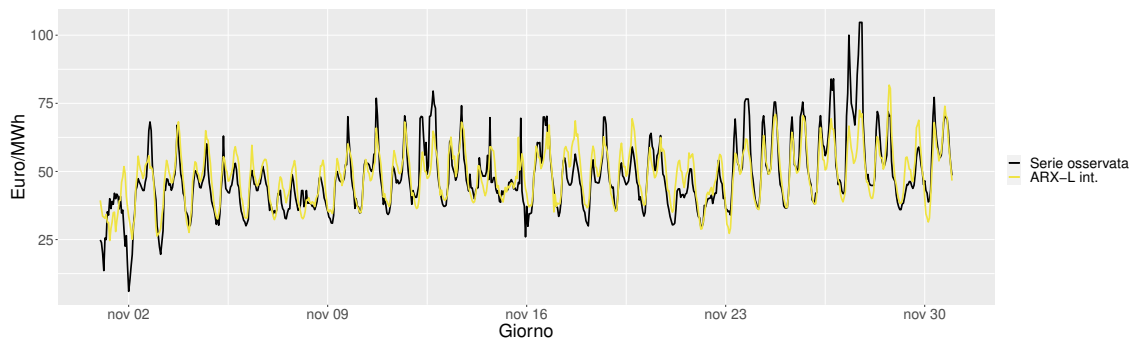


FIGURA B.45: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello ARX-L int. per novembre 2020 nella zona Centro-Nord.

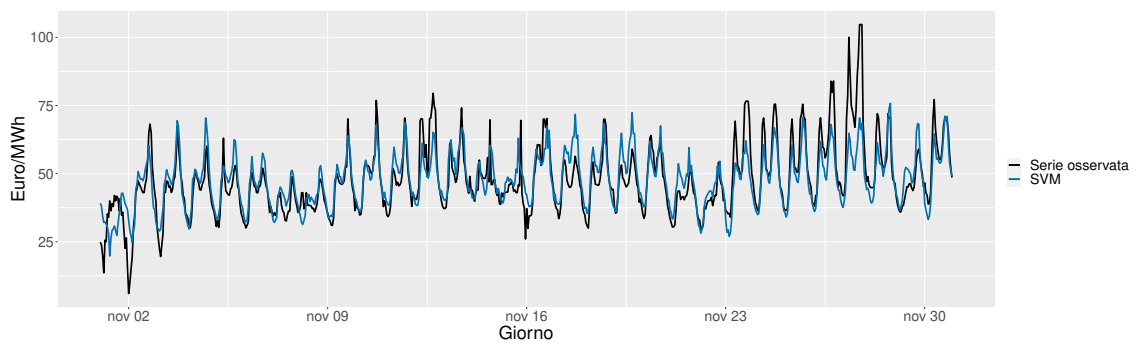


FIGURA B.46: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello SVM per novembre 2020 nella zona Centro-Nord.

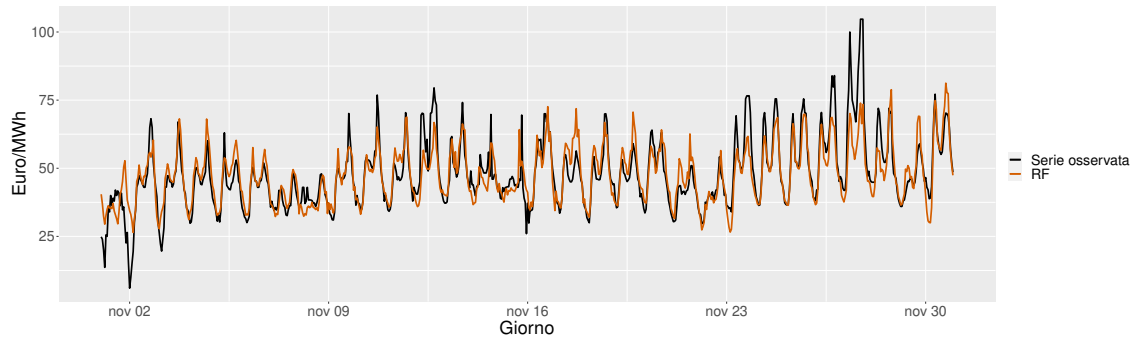


FIGURA B.47: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello RF per novembre 2020 nella zona Centro-Nord.

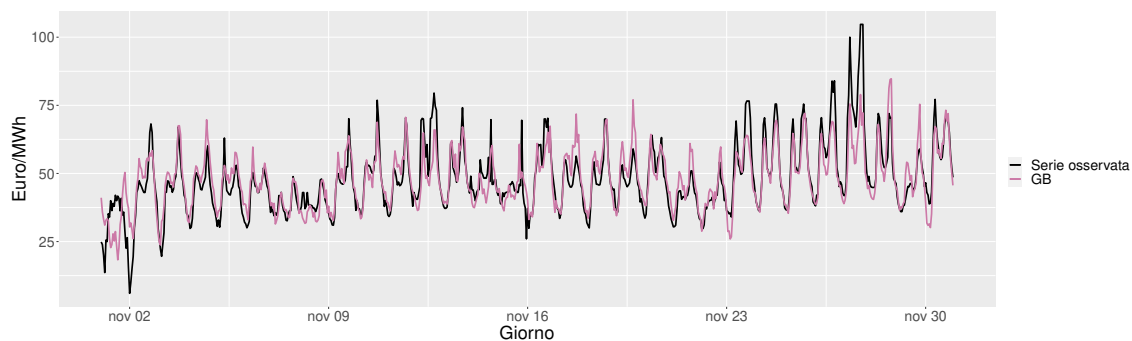


FIGURA B.48: Serie osservata dei prezzi spot dell'energia e previsioni ottenute con il modello GB per novembre 2020 nella zona Centro-Nord.

## C Analisi dei consumi: risultati basati sull'indice MAE

Di seguito sono riportati i risultati ottenuti dall'adattamento dei modelli per prevedere i consumi energetici sulla base dell'indice MAE. Tali risultati sono in linea con quanto ottenuto sulla base dell'indice RMSE descritto nel Capitolo 3.

Le Tabelle C.1 e C.2 riportano i valori medi di MAE ottenuti con i diversi modelli distinguendo tra ore di punta (14 ore) e ore non di punta (10 ore) per le due zone considerate.

Ore di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	3621.595	3302.411	3348.022	3278.272	3185.852	3342.421
Centro-Nord	641.118	496.112	507.751	483.517	500.261	501.838

TABELLA C.1: Valore medio del MAE nelle ore di punta

Ore non di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	3568.635	3313.460	3345.737	3303.024	3195.306	3363.731
Centro-Nord	635.701	501.939	511.105	495.308	510.156	515.093

TABELLA C.2: Valore medio del MAE nelle ore non di punta

In Tabella C.3 e Tabella C.4 sono riportati i valori medi di MAE in rapporto al consumo medio orario distinguendo sempre tra ore di punta e ore non di punta per le due zone considerate.

Ore di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	0.175	0.160	0.162	0.159	0.154	0.162
Centro-Nord	0.164	0.127	0.130	0.124	0.128	0.128

TABELLA C.3: Valore medio del MAE rapportato con consumo medio orario nelle ore di punta.

Ore non di punta						
<b>Zona</b>	<b>ARX</b>	<b>ARX-L</b>	<b>ARX-L int.</b>	<b>SVM</b>	<b>RF</b>	<b>GB</b>
Nord	0.173	0.161	0.162	0.160	0.155	0.163
Centro-Nord	0.163	0.128	0.131	0.127	0.131	0.132

TABELLA C.4: Valore medio del MAE rapportato con consumo medio orario nelle ore non di punta.

Di seguito si considerano le performance predittive più dettagliatamente in ogni singola ora, continuando a distinguere tra ore di punta e ore non di punta.

In Tabella C.5 e Tabella C.6 sono riportati per ogni ora il modello migliore in termini di MAE inferiore, e tra parentesi il numero di test di Diebold Mariano significativi.

<b>Ora</b>	<b>Ore di punta</b>	
	<b>Nord</b>	<b>Centro-Nord</b>
8	RF (3)	SVM (4)
9	RF (5)	SVM (5)
10	RF (4)	SVM (4)
11	RF (3)	SVM (4)
12	RF (5)	RF (4)
13	RF (2)	SVM (3)
14	RF (2)	SVM (3)
15	RF (5)	SVM (5)
16	RF (4)	SVM (5)
17	RF (3)	RF (2)
18	RF (5)	SVM (3)
19	RF (5)	SVM (3)
20	ARX-L (2)	SVM (5)
21	RF (3)	SVM (4)

TABELLA C.5: Modello migliore in termini di MAE inferiore e numero di Test di Diebold Mariano significativi (ore di punta)

Ore non di punta		
Ora	Nord	Centro-Nord
1	RF (5)	SVM (3)
2	RF (2)	SVM (4)
3	RF (4)	SVM (5)
4	RF (4)	SVM (2)
5	RF (5)	SVM (2)
6	RF (5)	ARX-L (2)
7	SVM (3)	SVM (2)
22	RF (3)	SVM (4)
23	RF (2)	SVM (3)
24	RF (5)	SVM (3)

TABELLA C.6: Modello migliore in termini di MAE inferiore e numero di Test di Diebold Mariano significativi (ore non di punta)

## D Riconciliazione delle previsioni: risultati basati sull'indice MAE

Di seguito vengono confrontate le previsioni di base e le previsioni riconciliate con procedure ottimali utilizzando l'indice MAE. Come nel Capitolo 3 l'accuratezza delle previsioni viene valutata tramite il rapporto tra i valori di MAE calcolati sulla base delle previsioni riconciliate attraverso le diverse procedure e l'indice MAE calcolato a partire dalle previsioni di base:

$$\frac{MAE_{ric}}{MAE_{base}}$$

Le previsioni di base che non soddisfano nessun vincolo vengono anche in questo caso utilizzate come *benchmark*. A livello interpretativo se le previsioni fornite da una certa tecnica sono migliori di quelle di base il rapporto degli indici MAE risulta essere minore di 1. La percentuale di guadagno in termini di accuratezza previsiva rispetto al *benchmark* è pari a

$$\left(1 - \frac{MAE_{ric}}{MAE_{base}}\right) \times 100$$

Nelle tabelle che seguono i dati in blu rappresentano i valori più bassi per ogni riga, cioè per ogni ora. Essi indicando quindi il metodo con cui è stato ottenuto il massimo guadagno rispetto alle previsioni di base. I valori riportati in rosso sono invece quelli

maggiori di 1 che segnalano un "peggioramento" dell'accuratezza delle previsioni riconciliate rispetto al *benchmark*. Per maggiore chiarezza viene riportato il valore 1 nella colonna riferita alle previsioni di base.

Le Tabelle D.1 e D.2 riportano i risultati ottenuti relativi alla zona Nord, mentre Tabelle D.3 e D.4 fanno riferimento alla zona Centro-Nord.

Ore di punta					
Ora	base	ols	struc	wlsv	shr
8	1	0.9439519	0.9655349	0.9900136	<b>0.9367847</b>
9	1	<b>0.9595630</b>	0.9650625	0.9907711	0.9598822
10	1	0.9691809	0.9778834	0.9934095	<b>0.9688425</b>
11	1	0.9449736	0.9632007	0.9904336	<b>0.9406311</b>
12	1	0.9539572	0.9692275	0.9913365	<b>0.9507521</b>
13	1	0.9127859	0.9483717	0.9874397	<b>0.8945521</b>
14	1	0.9319752	0.9575136	0.9889036	<b>0.9197501</b>
15	1	0.9382183	0.9610406	0.9891467	<b>0.9300004</b>
16	1	<b>0.9694072</b>	0.9738717	0.9921884	0.9702346
17	1	<b>0.9684944</b>	0.9784226	0.9939484	0.9701496
18	1	0.9674997	0.9759712	0.9931517	<b>0.9672259</b>
19	1	0.9443480	0.9643843	0.9908280	<b>0.9367975</b>
20	1	0.9246792	0.9543626	0.9885803	<b>0.9051742</b>
21	1	0.9173652	0.9463566	0.9865487	<b>0.9033230</b>
media	1	0.9461714	0.9643717	0.9904786	<b>0.9395786</b>

TABELLA D.1: Rapporto tra MAE delle procedure di riconciliazione e MAE calcolato a partire dalle previsioni di base nella zona Nord (ore di punta)

Ore non di punta					
Ora	base	ols	struc	wlsv	shr
1	1	0.9460228	0.9665178	0.9916230	<b>0.9397175</b>
2	1	0.9606916	0.9689483	0.9917935	<b>0.9604720</b>
3	1	<b>0.9526817</b>	0.9655963	0.9906921	0.9536726
4	1	0.9466763	0.9657934	0.9909525	<b>0.9424869</b>
5	1	0.9383997	0.9575397	0.9888817	<b>0.9327742</b>
6	1	0.9312022	0.9587241	0.9894243	<b>0.9210428</b>
7	1	0.9245055	0.9550331	0.9888203	<b>0.9109633</b>
22	1	0.9483283	0.9687602	0.9912423	<b>0.9365963</b>
23	1	0.9804478	0.9834415	0.9950149	<b>0.9798793</b>
24	1	<b>1.0054186</b>	<b>0.9958628</b>	0.9971707	0.9959778
media	1	0.9534374	0.9686217	0.9915615	<b>0.9473583</b>

TABELLA D.2: Rapporto tra MAE delle procedure di riconciliazione e MAE calcolato a partire dalle previsioni di base nella zona Nord (ore non di punta)



Ore di punta					
Ora	base	ols	struc	wlsv	shr
8	1	0.9089786	0.9485078	0.9877784	<b>0.8631444</b>
9	1	0.9193216	0.9525277	0.9883925	<b>0.9008234</b>
10	1	0.9102044	0.9479490	0.9870720	<b>0.8812906</b>
11	1	0.8910396	0.9378397	0.9850757	<b>0.8455131</b>
12	1	0.8866945	0.9384018	0.9855209	<b>0.8256956</b>
13	1	0.8985305	0.9403342	0.9853283	<b>0.8458189</b>
14	1	0.9014398	0.9432455	0.9864151	<b>0.8508629</b>
15	1	0.9209177	0.9536156	0.9890329	<b>0.8943398</b>
16	1	0.9300174	0.9584687	0.9891628	<b>0.9166535</b>
17	1	0.9240082	0.9543555	0.9886871	<b>0.9137269</b>
18	1	0.9252029	0.9538814	0.9886003	<b>0.9171405</b>
19	1	0.9244514	0.9536269	0.9884007	<b>0.9068676</b>
20	1	0.9011518	0.9440053	0.9861303	<b>0.8509360</b>
21	1	0.9075833	0.9466826	0.9867492	<b>0.8709998</b>
media	1	0.9106816	0.948103	0.9873105	<b>0.8774152</b>

TABELLA D.3: Rapporto tra MAE delle procedure di riconciliazione e MAE calcolato a partire dalle previsioni di base nella zona Centro-Nord (ore di punta)

Ore non di punta					
Ora	base	ols	struc	wlsv	shr
1	1	0.9172445	0.9482598	0.9869635	<b>0.8854051</b>
2	1	0.9102927	0.9497778	0.9877172	<b>0.8674821</b>
3	1	0.9163145	0.9507820	0.9881022	<b>0.8914784</b>
4	1	0.8955511	0.9410619	0.9860271	<b>0.8419420</b>
5	1	0.8724158	0.9296424	0.9836840	<b>0.7804327</b>
6	1	0.8861782	0.9372716	0.9851805	<b>0.8073195</b>
7	1	0.8937764	0.9397322	0.9853248	<b>0.8410742</b>
22	1	0.9475277	0.9667420	0.9913577	<b>0.9374106</b>
23	1	<b>0.9617185</b>	0.9738548	0.9927005	0.9643338
24	1	<b>0.9659048</b>	0.9765235	0.9938495	0.9703457
media	1	0.9166924	0.9513648	0.9880907	<b>0.8787224</b>

TABELLA D.4: Rapporto tra MAE delle procedure di riconciliazione e MAE calcolato a partire dalle previsioni di base nella zona Centro-Nord (ore non di punta)



# Bibliografia

- ATHANASOPOULOS, G., AHMED, R. A. & HYNDMAN, R. J. (2009). Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting* **25**, 146–166.
- ATHANASOPOULOS, G., GAMAKUMARA, P., PANAGIOTELIS, A., HYNDMAN, R. J. & AFFAN, M. (2020). Hierarchical forecasting. *Macroeconomic forecasting in the era of big data: Theory and practice* , 689–719.
- ATHANASOPOULOS, G., HYNDMAN, R. J., KOURENTZES, N. & PETROPOULOS, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research* **262**, 60–74.
- AZZALINI, A. & SCARPA, B. (2009). *Analisi dei dati e data mining*. Springer Science & Business Media.
- BATES, J. M. & GRANGER, C. W. (1969). The combination of forecasts. *Journal of the operational research society* **20**, 451–468.
- BIECEK, P. & BURZYKOWSKI, T. (2021). *Explanatory model analysis: Explore, explain and examine predictive models*. Chapman and Hall/CRC.
- BREIMAN, L. (2001). Random forests. *Machine learning* **45**, 5–32.
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R. & STONE, C. (1984). Cart. *Classification and regression trees* .
- CORTES, C. & VAPNIK, V. (1995). Support-vector networks. *Machine learning* **20**, 273–297.
- DI FONZO, T. & GIROLIMETTO, D. (2022). Forecast combination-based forecast reconciliation: Insights and extensions. *International Journal of Forecasting* .
- DIEBOLD, F. & MARIANO, R. (1995). Comparing predictive accuracy. *journal of business and economics statistics*, v. 13 .

- DRUCKER, H., BURGESS, C. J., KAUFMAN, L., SMOLA, A. & VAPNIK, V. (1996). Support vector regression machines. *Advances in neural information processing systems* **9**.
- DUNN, D. M., WILLIAMS, W. H. & DECHAINED, T. (1976). Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association* **71**, 68–71.
- FIANU, E. S., AHELEGBEY, D. F. & GROSSI, L. (2022). Modeling risk contagion in the italian zonal electricity market. *European Journal of Operational Research* **298**, 656–679.
- FRIEDMAN, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics* , 1189–1232.
- GHODDUSI, H., CREAMER, G. G. & RAFIZADEH, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics* **81**, 709–727.
- GIANFREDA, A. & GROSSI, L. (2012). Forecasting italian electricity zonal prices with exogenous variables. *Energy Economics* **34**, 2228–2239.
- GIROLIMETTO, D. (2020). Riconciliazione contemporanea, temporale e cross-temporale di previsioni di serie storiche .
- GNEITING, T. & KATZFUSS, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application* **1**, 125–151.
- GROSS, C. W. & SOHL, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of forecasting* **9**, 233–254.
- GROSSI, L., GOLIA, S. & PELAGATTI, M. (2022). Machine learning models and intraday market information for the prediction of italian electricity prices. *Forecasting* **5**, 81–101.
- GROSSI, L. & NAN, F. (2019). Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change* **141**, 305–318.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. & FRIEDMAN, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer.

- HASTIE, T., TIBSHIRANI, R., JAMES, G. & WITTEN, D. (2013). *An introduction to statistical learning*, vol. 112. Springer.
- HOLLYMAN, R., PETROPOULOS, F. & TIPPING, M. E. (2021). Understanding forecast reconciliation. *European Journal of Operational Research* **294**, 149–160.
- HYNDMAN, R. J., AHMED, R. A., ATHANASOPOULOS, G. & SHANG, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis* **55**, 2579–2589.
- HYNDMAN, R. J. & ATHANASOPOULOS, G. (2021). *Forecasting: principles and practice*. OTexts.
- HYNDMAN, R. J. & KOURENTZES, N. (2018). Thief: temporal hierarchical forecasting. *R package version 0.3*. URL <http://pkg.robjhyndman.com/thief> .
- KOURENTZES, N. & ATHANASOPOULOS, G. (2019). Cross-temporal coherent forecasts for australian tourism. *Annals of Tourism Research* **75**, 393–409.
- KUNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics* , 1217–1241.
- LAGO, J., DE RIDDER, F. & DE SCHUTTER, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy* **221**, 386–405.
- LEDOIT, O. & WOLF, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management* **30**, 110–119.
- LUCAS, A., PEGIOS, K., KOTSAKIS, E. & CLARKE, D. (2020). Price forecasting for the balancing energy market using machine-learning regression. *Energies* **13**, 5420.
- MARCJASZ, G., UNIEJEWSKI, B. & WERON, R. (2019). On the importance of the long-term seasonal component in day-ahead electricity price forecasting with narx neural networks. *International Journal of Forecasting* **35**, 1520–1532.
- NARDI, Y. & RINALDO, A. (2011). Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis* **102**, 528–549.
- NOWOTARSKI, J. & WERON, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* **81**, 1548–1568.

- NYSTRUP, P., LINDSTRÖM, E., PINSON, P. & MADSEN, H. (2020). Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research* **280**, 876–888.
- PANAGIOTELIS, A., ATHANASOPOULOS, G., GAMAKUMARA, P. & HYNDMAN, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting* **37**, 343–359.
- PANAGIOTELIS, A., GAMAKUMARA, P., ATHANASOPOULOS, G. & HYNDMAN, R. J. (2023). Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. *European Journal of Operational Research* **306**, 693–706.
- SANDRI, M. & ZUCCOLOTTO, P. (2008). A bias correction algorithm for the gini variable importance measure in classification trees. *Journal of Computational and Graphical Statistics* **17**, 611–628.
- SANDRI, M. & ZUCCOLOTTO, P. (2010). Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms. *Statistics and Computing* **20**, 393–407.
- SCHÄFER, J. & STRIMMER, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* **4**.
- SCHNÜRCH, S. & WAGNER, A. (2020). Electricity price forecasting with neural networks on epex order books. *Applied Mathematical Finance* **27**, 189–206.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288.
- TIMMERMANN, A. (2006). Forecast combinations. *Handbook of economic forecasting* **1**, 135–196.
- VAPNIK, V. & CHAPPELLE, O. (2000). Bounds on error expectation for support vector machines. *Neural computation* **12**, 2013–2036.
- WERON, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.
- WERON, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting* **30**, 1030–1081.

- 
- WICKRAMASURIYA, S. L., ATHANASOPOULOS, G. & HYNDMAN, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* **114**, 804–819.

