

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA TRIENNALE IN  
STATISTICA PER LE TECNOLOGIE E LE SCIENZE



RELAZIONE FINALE

## **I prezzi dell'energia elettrica nel Nord Italia: modelli di previsione**

Relatore Prof. Matteo Grigoletto  
Dipartimento di Scienze Statistiche

Laureanda Francesca Remigio  
Matricola 2003659

Anno Accademico 2022/2023



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 La previsione dei prezzi dell'energia</b>	<b>3</b>
1.1 La previsione nella letteratura passata . . . . .	3
1.2 La scelta del nord Italia . . . . .	4
1.3 Il mercato dell'energia . . . . .	5
1.3.1 Il Mercato elettrico a pronti . . . . .	6
1.3.1.1 Il Mercato del giorno prima . . . . .	6
1.3.1.2 Il Mercato infragiornaliero . . . . .	7
1.3.1.3 Il Mercato dei prodotti giornalieri . . . . .	8
1.3.1.4 Il Mercato per il Servizio di Dispacciamento . . . . .	8
1.3.2 Il Mercato elettrico a termine . . . . .	9
1.4 Il mercato dell'energia nell'Italia settentrionale . . . . .	9
<b>2 Un modello di previsione</b>	<b>11</b>
2.1 I modelli . . . . .	11
2.1.1 Modelli ARIMA . . . . .	11
2.1.2 Modelli GARCH . . . . .	12
2.1.3 Regressione LASSO . . . . .	13
2.2 Metodi di confronto . . . . .	13
2.2.1 Errore quadratico medio . . . . .	14
2.2.2 <i>Continuous Ranked Probability Score</i> . . . . .	14
2.2.3 Test di Diebold-Mariano . . . . .	15
2.2.4 <i>Model Confidence Set</i> . . . . .	16
2.3 I dati . . . . .	16
2.3.1 Le fonti . . . . .	16
2.3.2 Analisi esplorativa . . . . .	17
2.4 Specificazione del modello . . . . .	19
2.4.1 Modelli combinati . . . . .	20
2.4.2 Modelli ARFIMA . . . . .	22
2.4.3 Modelli GARCH . . . . .	23
2.4.4 LASSO . . . . .	24
2.5 Stima dei modelli . . . . .	24
2.6 Valutazione della performance previsiva . . . . .	25

---

<b>3</b>	<b>I risultati</b>	<b>27</b>
3.1	Performance previsiva . . . . .	27
3.2	Il modello stimato . . . . .	33
<b>4</b>	<b>Conclusioni</b>	<b>43</b>
	<b>Bibliografia</b>	<b>45</b>





# Introduzione

La previsione dei prezzi dell'energia è sempre stato un tema interessante e molto discusso a causa dei fattori che ne influenzano la volatilità. Sui prezzi dell'energia entrano in gioco infatti numerose componenti come l'imprevedibilità di domanda e offerta, l'impossibilità di immagazzinare grandi quantità di energia elettrica e infine la variabilità indotta dall'energia prodotta dalle fonti rinnovabili che, dipendendo dalle condizioni meteorologiche, non è prevedibile. Il mercato dell'energia è inoltre molto complesso e influenza particolarmente la difficoltà della previsione dei prezzi orari dell'energia.

Lo scopo di questo elaborato è di ripercorrere il recente studio di Billé et al. (2022) proponendo un approfondimento dei temi trattati e dei modelli di previsione utilizzati nello studio. Una buona previsione dei prezzi dell'energia elettrica è importante non solo per la previsione della quantità di energia da produrre in base alla domanda del mercato, ma anche per contrastare l'aumento dei costi dell'elettricità per il cliente finale e le speculazioni di mercato.

In particolare nel Capitolo 1 viene descritto il funzionamento del mercato dell'energia in Italia e viene motivata la scelta del nord Italia per lo studio; nel Capitolo 2 vengono descritti i modelli utilizzati per il modello finale di Billé et al. (2022) e sono approfonditi i metodi di confronto per valutare la capacità predittiva dei modelli proposti. Nei paragrafi successivi sono descritte le fonti dei dati utilizzati per stimare il modello, viene svolta una breve analisi esplorativa e sono specificati e stimati alcuni modelli di previsione che comprendono modelli combinati, modelli ARFIMA, modelli GARCH e infine la regressione LASSO. I modelli sono stimati sia in presenza di regressori esogeni che in assenza. In particolare i regressori considerati sono la domanda di energia prevista, l'energia prodotta dalle fonti rinnovabili, il prezzo dell'energia importata dall'estero, il prezzo del carbone e del gas naturale e il prezzo stimato per le emissioni di  $CO_2$ .

Nel Capitolo 3 vengono valutate la qualità e la precisione delle previsioni dei prezzi orari dell'energia elettrica con i metodi descritti nel Paragrafo 2.2. Emerge che l'utilizzo

di previsioni tempestive ottenute da fonti private migliora sostanzialmente la capacità predittiva rispetto all'utilizzo di previsioni ottenute da fonti gratuite. Sono inoltre riportati i risultati ottenuti per il modello stimato  $EX_4X$ , un modello combinato con la presenza di variabili esogene, che risulta il migliore modello in termini di previsione.

Nel Capitolo 4 sono riportate le conclusioni dello studio in cui si evidenzia l'importanza dell'utilizzo di previsioni provenienti da fonti private, che migliorano l'accuratezza delle previsioni e l'importanza di considerare nei modelli i regressori esogeni. Il modello che risulta migliore è il modello  $EX_4X$  per le previsioni puntuali mentre per le previsioni di densità risulta necessario considerare anche le specifiche del modello GARCH. Le variabili che risultano significative sono la domanda di energia prevista, l'energia proveniente dai settori fotovoltaico, eolico e idroelettrico e il prezzo del gas naturale.

# Capitolo 1

## La previsione dei prezzi dell'energia

### 1.1 La previsione nella letteratura passata

La previsione dei prezzi orari dell'elettricità è sempre stato un tema ampiamente discusso in letteratura a causa dei numerosi fattori che ne influenzano la volatilità. Questa previsione infatti risulta complicata proprio a causa delle componenti stocastiche che entrano in gioco, come ad esempio l'imperfetta prevedibilità di domanda e offerta di energia o la variabilità indotta dalle nuove normative del mercato energetico. Bisogna considerare anche l'impossibilità di immagazzinare grandi quantità di energia elettrica e, soprattutto negli ultimi anni, non si può ignorare la disponibilità di energia prodotta da fonti rinnovabili (*Renewable Energy Sources*, RES) che introduce ulteriore variabilità poiché dipende dalle condizioni meteorologiche. L'energia prodotta dalle RES, attualmente, non può essere trascurata in quanto più di un terzo dell'energia elettrica prodotta in Italia arriva da fonti rinnovabili. L'idroelettrico è il settore dominante, seguito dal solare fotovoltaico, le bioenergie, l'eolico e il geotermico. (Figura 1.1, IEA, 2023)

La presenza delle fonti rinnovabili ha prodotto un duplice effetto sul mercato dell'energia: da una parte ha condotto ad un sistema di produzione meno inquinante e meno costoso, d'altra parte, aumentando l'energia immessa sul mercato, l'equilibrio dei prezzi è diminuito.

In letteratura la previsione dei prezzi orari dell'elettricità è stata oggetto di numerosi studi e ha preso in esame diverse aree geografiche e mercati. Nello studio di Bunn et al. (2018) viene utilizzato un modello statistico a variabili latenti per prevedere il prezzo dell'elettricità in Austria, nel mercato dell'energia tedesco/austriaco. La previsione è data dalla densità della distribuzione *Skew-t*, i cui primi tre momenti vengono stimati

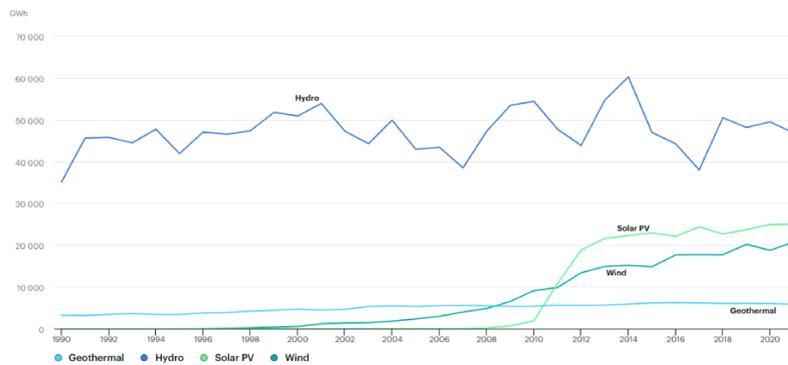


FIGURA 1.1: Produzione di energia elettrica da fonti rinnovabili, Italia 1990-2021

come combinazione lineare delle osservazioni passate e degli errori di previsione per l'energia proveniente dalle fonti rinnovabili. Lo studio ha mostrato come l'utilizzo di un predittore della funzione di densità ha aumentato la redditività del commercio dell'energia e ha ridotto il rischio finanziario rispetto all'utilizzo di modelli di regressione più tradizionali basati sul valore medio. Un'analisi simile viene presentata in uno studio più recente da Abramova & Bunn (2020).

Nello studio condotto in Italia di Lisi & Edoli (2018), viene utilizzato il segno di sbilanciamento, definito come il segno della forma algebrica dell'energia acquistata e venduta dal *Transmission System Operator* nazionale in tempo reale. L'analisi mostra che la probabilità che il segno di sbilanciamento sia positivo (o negativo) dipende dalle osservazioni passate ed è quindi in parte prevedibile.

Negli studi sopracitati viene presentata una valutazione economica del vantaggio derivante dall'uso del modello proposto. Altri studi meno recenti hanno affrontato il problema concentrandosi sull'aspetto economico e sui fattori di rischio che influenzano i prezzi infragionalieri dell'elettricità: nell'analisi proposta da Karakatsani & Bunn (2008) vengono esaminate le prestazioni di previsione del modello di regressione a cambiamento di regime, che consente di introdurre la discontinuità nei prezzi, e il modello di regressione con parametri variabili nel tempo, che consente una struttura variabile che si adatta ai cambiamenti delle normative e alla struttura del mercato. In conclusione lo studio, con i dati sul mercato britannico, ha portato a preferire il modello con parametri variabili il quale ha mostrato una capacità predittiva migliore.

## 1.2 La scelta del nord Italia

Nella maggior parte delle analisi vengono prese in esame diverse aree dell'Europa ma ci sono alcune caratteristiche evidenziate da Billé et al. (2022) che mostrano come la zona

settentrionale dell'Italia sia interessante per lo studio del mercato energetico.

Gli autori hanno osservato che spesso l'energia prodotta dal fotovoltaico in questa zona è sistematicamente sovrastimata nel mercato del giorno prima. Ciò richiede maggiori costi per ripristinare l'equilibrio nel mercato infragiornaliero. Per superare questa criticità alcuni paesi dell'Unione Europea, compresa l'Italia, hanno introdotto nel mercato dell'energia anche le unità RES.

L'Italia settentrionale inoltre:

- Essendo confinante con Francia, Svizzera, Austria e Slovenia ha la possibilità di importare energia elettrica ad un prezzo ridotto;
- La produzione di energia derivante dalle fonti rinnovabili è cresciuta, in particolare dal 2010 (Figura 1.1);
- La maggior parte dell'energia idroelettrica prodotta in Italia viene dalla zona delle Alpi;
- Circa la metà della domanda di energia del Paese viene dalla zona settentrionale;
- Tutte le centrali tradizionali, termoelettriche, idroelettriche e idrauliche agiscono in modo diverso sul mercato.

Le caratteristiche sopra evidenziate hanno portato gli autori dello studio di Billé et al. (2022) a scegliere dati provenienti dal nord Italia per presentare i modelli per la previsione del prezzo dell'energia elettrica che verranno analizzati in seguito.

### **1.3 Il mercato dell'energia**

In Italia al giorno d'oggi vi è il mercato libero dell'energia, nato il 16 marzo 1999 con l'emanazione del decreto legislativo n. 79 (d.lgs. n. 79/99), nell'ambito del processo di recepimento della direttiva comunitaria sulla creazione di un mercato interno dell'energia (96/92/CE). La liberalizzazione del mercato energetico ha aperto la possibilità alle aziende di proporre la propria offerta e ha permesso agli utenti di scegliere il proprio fornitore. La liberalizzazione del mercato ha inoltre coinvolto tutti gli aspetti legati alla produzione e alla compravendita dell'energia elettrica, comprese l'importazione e l'esportazione. Oggi infatti qualsiasi operatore è in grado di produrre e vendere energia elettrica mentre in precedenza ciò non era possibile.

Insieme al regime del libero mercato coesiste il regime vincolato: il consumatore, dovendo scegliere a quale operatore affidarsi, ha dalla sua parte il sistema del Servizio di

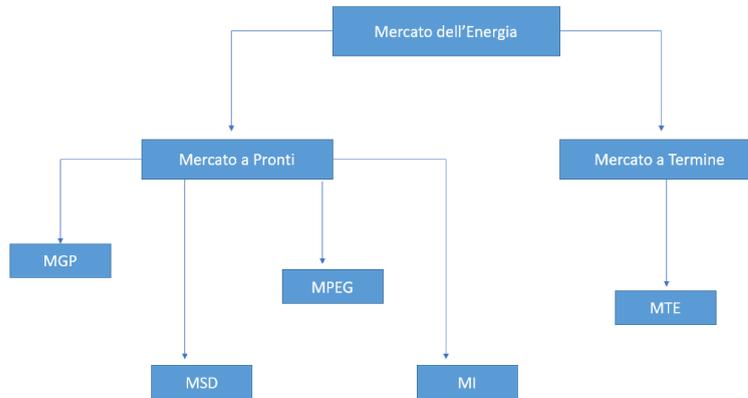


FIGURA 1.2: Il Mercato Elettrico

Maggior Tutela, nel quale interviene l'Autorità per l'Energia Elettrica e il Gas (AEEG), alla quale è affidato il compito di definire periodicamente il prezzo dell'energia.

Il mercato elettrico, come mostra la Figura 1.2 (GME (2009)) si articola in Mercato elettrico a pronti (MPE) e Mercato elettrico a termine (MTE), i quali a loro volta si dividono in altre sottocategorie che verranno analizzate in dettaglio nei paragrafi 1.3.1 e 1.3.2 (GME, 2009).

### 1.3.1 Il Mercato elettrico a pronti

Il Mercato elettrico a pronti si articola a sua volta in:

- Mercato del giorno prima (MGP), nel cui ambito si effettua la contrattazione dell'energia stabilendo i prezzi in base al rapporto tra domanda e offerta;
- Mercato infragiornaliero (MI) che consente agli operatori di modificare i programmi del MGP;
- Mercato dei prodotti giornalieri (MPEG) dove viene negoziata e consegnata l'energia;
- Il Mercato per il Servizio di Dispacciamento (MSD) che si articola a sua volta in fase di programmazione (MSD ex-ante) e mercato del bilanciamento (MD) ed entra in gioco per situazioni di congestioni o per quanto riguarda la riserva di elettricità.

#### 1.3.1.1 Il Mercato del giorno prima

Nel Mercato del giorno prima gli operatori presentano le offerte indicando il prezzo massimo e minimo al quale sono disposti ad acquistare o vendere energia elettrica.

L'apertura della seduta del MPG è fissata alle ore 8.00 del nono giorno antecedente il giorno di consegna. La chiusura, invece, scatta alle ore 12 del giorno che precede quello di consegna dell'energia.

Il MGP è un mercato d'asta e non un mercato a contrattazione continua: le offerte sono accettate dopo la chiusura della seduta di mercato, sulla base del merito economico e rispettano i limiti di transito tra le zone. Le offerte, riferite sia alle unità pompaggio che di consumo, vengono bilanciate al prezzo di equilibrio della zona virtuale a cui appartengono. Le zone virtuali sono delimitate dalla capacità di trasmissione dell'energia e in Italia sono: nord, centro-nord, centro-sud, meridione e Sicilia e Sardegna; le zone virtuali estere sono invece i punti di interconnessione tra l'Italia e i paesi confinanti (Austria, Francia, Svizzera e Slovenia). Il prezzo di equilibrio è definito, per ogni ora, dall'intersezione della curva di domanda e offerta, che è differente a seconda della zona. In Italia le offerte di acquisto accettate, sempre in base alla zona, vengono valorizzate al Prezzo Unico Nazionale (PUN), che è la media pesata dei prezzi nelle zone geografiche per le quantità di energia acquistata in tali zone.

### **1.3.1.2 Il Mercato infragiornaliero**

In questa fase gli operatori possono agire sulle offerte di acquisto e vendita effettuate nella fase precedente del MGP attraverso tre sessioni d'asta (MI-A) e una sessione continua di negoziazione (MI-XBID).

Le sessioni d'asta MI-A contengono tutta la capacità infragiornaliera di interconnessione delle diverse zone virtuali coinvolte nel mercato e si dividono in:

- MI-A1, che si svolge appena dopo la chiusura del MGP. Si apre alle ore 12:55 del giorno precedente il giorno di consegna e si chiude alle 15:00 dello stesso giorno;
- MI-A2, che si apre alle 12:55 del giorno precedente il giorno di consegna e si chiude alle 22:00 dello stesso giorno;
- MI-A3, che si apre alle 12:55 del giorno precedente il giorno di consegna e si chiude alle 10:00 del giorno di consegna. Gli esiti di questa fase sono comunicati alle 10:30 del giorno di consegna.

Le offerte, come nel mercato del MGP, vengono bilanciate al prezzo di equilibrio della zona alla quale appartengono e vengono valorizzate al Prezzo Unico Nazionale.

La sessione continua di negoziazione MI-XBID si divide in tre fasi nelle quali le offerte possono essere presentate per ogni unità e per portafoglio, per ciascuna fase il Gestore

dei Mercati Energetici GME (2009) organizza un book di negoziazioni diviso per zone virtuali e geografiche.

Ricapitolando nel Mercato infragiornaliero le fasi seguono l'ordine: 1) MI-A1, 2) MI-XBID fase I, 3) MI-A2, 4) MI-XBID fase II, 5) MI-A3, 6) MI-XBID fase III. Le sei fasi si susseguono in modo sequenziale e non sono sovrapposte.

### **1.3.1.3 Il Mercato dei prodotti giornalieri**

Sul Mercato dei prodotti giornalieri sono ammessi tutti gli operatori del mercato elettrico e le negoziazioni avvengono in modalità continua. I prodotti in questa fase del mercato sono negoziabili in due diversi modi:

- Differenziale unitario di prezzo che è l'espressione del differenziale, rispetto al PUN, al quale gli operatori sono disposti a vendere o acquistare;
- Prezzo unitario pieno che è il semplice valore unitario di scambio dell'energia elettrica.

Per entrambe le tipologie, il GME (2009) può quotare i profili di consegna che seguono:

- Baseload che permette di visualizzare i prezzi durante tutte le 24 ore della giornata;
- Peakload che si riferisce alle ore di punta di consumo.

Spetta al Gestore dei Mercati Energetici, la determinazione della posizione netta di ogni operatore e la sua registrazione in corrispondenti transazioni sulla PCE (Piattaforma Conti Energia) secondo quanto stabilito nella Disciplina del mercato elettrico.

### **1.3.1.4 Il Mercato per il Servizio di Dispacciamento**

Dal momento che l'energia elettrica non può essere immagazzinata, la quantità di energia prodotta deve essere pari a quella richiesta. Il cosiddetto dispacciamento, cioè la gestione dell'equilibrio dei flussi di energia sulla rete nazionale, è di pertinenza dell'azienda Terna s.p.a.. Nel caso in cui la domanda di energia dovesse discostarsi dai consumi effettivi, Terna provvede al cosiddetto bilanciamento, cioè a registrare la differenza tra il consumo programmato ed effettivo e a inviare disposizioni aggiornate alle centrali di generazione. Nel MSD le offerte sono accettate sulla base del merito economico e sono remunerate secondo la formula *pay-as-bid*, cioè al prezzo presentato.

Il Mercato per il Servizio di Dispacciamento si divide in due fasi:

- Fase di programmazione (MSD ex-ante) che si articola in sei sottofasi gestite dall'azienda Terna e dal GME;
- Mercato del Bilanciamento (MB) il quale è costituito da offerte continue a dettaglio orario.

### **1.3.2 Il Mercato elettrico a termine**

Il Mercato elettrico a termine è la sede dove sono negoziati i contratti energetici a termine con obbligo di consegna e ritiro. Le negoziazioni, in questo mercato, avvengono in modalità continua. Gli operatori possono partecipare liberamente specificando il numero, la tipologia, il periodo dei contratti e il prezzo al quale sono disposti a vendere o acquistare l'energia. Nel MTE sono ammessi sia contratti di tipo Baseload che di tipo Peakload e, alla fine di tutte le trattative, il Gestore dei Mercati Energetici registra sulla Piattaforma Conti Energia (PCE) la posizione netta in consegna sul Mercato elettrico a termine.

## **1.4 Il mercato dell'energia nell'Italia settentrionale**

La maggior parte della produzione industriale dell'Italia proviene dalla zona settentrionale, infatti, come evidenziato da Gianfreda et al. (2019), Gianfreda et al. (2016) e Shah & Lisi (2020), il consumo di elettricità del nord Italia (circa 175.396 GWh) corrisponde a più della metà del consumo nazionale annuale (circa 303.443 GWh). Il consumo di energia è consistentemente più elevato in questa zona rispetto all'intero consumo del Paese: la media del consumo per abitante è di 6326 kWh per abitante nel nord Italia, contro una media nazionale di 5024 kWh, dai dati di Terna (2018).

L'energia elettrica in questa zona inoltre proviene da diverse fonti: il 26% di essa deriva dall'energia idroelettrica, il 6% dall'energia solare e un ulteriore 8% deriva dalle biomasse. Il restante 60% dell'energia elettrica totale deriva da fonti tradizionali. Il fatto che il 40% di energia totale consumata nella zona settentrionale dell'Italia sia derivante da fonti rinnovabili, rende quest'area geografica interessante dal punto di vista statistico perché nella previsione del prezzo dell'energia in base ai dati del Mercato del giorno prima entrano in gioco delle ulteriori variabili che devono tener conto della previsione, un giorno in avanti, di energia che le fonti rinnovabili riescono a produrre.

Il nord Italia è inoltre direttamente confinante con quattro paesi esteri mentre le altre zone italiane hanno a disposizione solo collegamenti nazionali o un numero limitato di

collegamenti con l'estero. Dal 2011 e successivamente nel 2015, l'Italia ha stipulato degli accordi per l'importazione e l'esportazione dell'energia elettrica con i paesi confinanti, contribuendo così all'istituzione del market coupling in Europa, un meccanismo di coordinamento tra mercati elettrici organizzati in diversi Stati finalizzato alla gestione delle congestioni sulle reti di interconnessione (scambi transfrontalieri). Il market coupling ha istituito quindi un mercato unico dell'energia elettrica in Europa, consentendo così una maggiore facilità nella trasmissione di energia elettrica nelle varie reti e permettendo ai prezzi di allinearsi. Grazie alla vicinanza e alla presenza di reti di interconnessione tra il nord Italia e i paesi confinanti, è possibile importare energia elettrica a prezzi inferiori.

# Capitolo 2

## Un modello di previsione

### 2.1 I modelli

Il modello finale utilizzato per la previsione dei prezzi orari dell'energia elettrica da Billé et al. (2022) è un incrocio tra modelli autoregressivi e modelli di regressione: vengono presi in considerazione modelli ARIMA (paragrafo 2.1.1), modelli di tipo GARCH (paragrafo 2.1.2), regressione LASSO (paragrafo 2.1.3) e vengono introdotte delle variabili dummy per modellare la stagionalità.

#### 2.1.1 Modelli ARIMA

I modelli ARIMA (da *AutoRegressive Integrated Moving Average*), introdotti da Box e Jenkins (1976), fanno parte di una classe particolare di modelli per l'analisi delle serie storiche non stazionarie.

Un modello ARIMA( $p, d, q$ ) è una generalizzazione di un modello ARMA( $p, q$ ) per modellare processi non stazionari omogenei di grado  $d$  cioè processi che possono essere resi stazionari applicando  $d$  differenziazioni. In particolare in un modello ARIMA( $p, d, q$ ):

- $p$  rappresenta il numero di parametri autoregressivi;
- $q$  rappresenta il numero di parametri a media mobile;
- $d$  rappresenta l'ordine di differenziazione.

In generale sia  $\epsilon_t \sim WN(0, \sigma^2)$  e definito l'operatore ritardo come  $B^k Y_t = Y_{t-k}$ , il processo  $Y_t$  si dice ARIMA( $p, d, q$ ) se:

- $(1 - B)^d Y_t = X_t$

- $X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t - \sum_{i=1}^q \theta_i \epsilon_{t-i}$

Con  $X_t$  modello ARMA( $p, q$ ),  $\theta_i$  parametri a media mobile e  $\phi_i$  parametri autoregressivi.

Si possono generalizzare i modelli ARIMA( $p, d, q$ ) introducendo la classe dei modelli ARIMAX, *AutoRegressive Integrated Moving Average with eXogenous Inputs*, che introducono nel modello delle variabili extra (X) che ne migliorano la performance previsiva.

Un'ulteriore generalizzazione dei modelli ARIMA( $p, d, q$ ) è data dalla classe degli ARFIMA, acronimo di *AutoRegressive Fractionally Integrated Moving Average*, che permettono che l'ordine di differenziazione  $d$  sia non intero. Questi modelli sono utili per la modellazione di serie temporali a lunga memoria quindi risultano particolarmente appropriati per il problema della previsione dei prezzi dell'energia elettrica.

### 2.1.2 Modelli GARCH

Il modello GARCH è una generalizzazione del modello ARCH, da *AutoRegressive Conditional Heteroskedasticity*, cioè modello autoregressivo a eteroschedasticità condizionata. Introdotti da Engle (1982) i modelli ARCH sono particolarmente utilizzati in finanza e in econometria per l'analisi delle serie storiche. Essi hanno una componente erratica (a cui si può assegnare il compito di interpretare per esempio i rendimenti di un titolo) che non è autocorrelata e una varianza che non è costante nel tempo ed è autocorrelata. Secondo una formulazione classica quindi dato un processo  $\epsilon_t = \sigma_t z_t$  con  $z_t \sim N(0, 1)$  e  $\sigma_t^2 = \phi_0 + \sum_{i=1}^p \phi_i \epsilon_{t-i}^2$  si ha un modello ARCH( $p$ ), i parametri del quale possono essere stimati con i minimi quadrati. La varianza  $\sigma_t$ , in questo caso, segue un processo AR( $p$ ).

Se la varianza  $\sigma_t$  segue un processo ARMA( $p, q$ ) il modello diventa:

$$\sigma_t^2 = \phi_0 + \sum_{i=1}^p \phi_i \epsilon_{t-i}^2 + \sum_{i=1}^q \theta_i \sigma_{t-i}^2$$

Si tratta di un modello ARCH generalizzato: cioè un modello GARCH( $p, q$ ), con  $p$  parametri autoregressivi e  $q$  parametri a media mobile. Questa tipologia di modelli trova ampio impiego per i modelli sulle serie storiche finanziarie, che spesso presentano alta variabilità.

### 2.1.3 Regressione LASSO

La regressione LASSO (*Least Absolute Shrinkage and Selection Operator*), introdotta da Tibshirani (1996), è un metodo di analisi di regressione che esegue sia la regolarizzazione, per migliorare il modello in termini di previsione, sia una selezione delle variabili. La regressione LASSO è stata introdotta per i modelli di regressione lineare ma successivamente la sua applicazione si è estesa anche ad altri modelli come ad esempio i modelli lineari generalizzati.

I metodi di regolarizzazione, invece di controllare la complessità del modello ponendo un sottoinsieme dei coefficienti di regressione  $\beta_j$  a zero, come proposto nei metodi di *forward stepwise regression* e *backward stepwise regression*, contraggono (*shrink*) i coefficienti verso zero. Tra questi metodi, quelli maggiormente utilizzati, sono la regressione RIDGE, la LASSO e degli approcci ibridi come ad esempio la *rete elastica* o il *lasso raggruppato*. La regolarizzazione è particolarmente utile quando il numero di variabili presenti è molto elevato e, sebbene produca stime distorte dei coefficienti di regressione, riduce drasticamente la varianza delle stime ai minimi quadrati.

Il LASSO risolve il problema di minimo vincolato:

$$\min_{\beta} (y - X\beta)^T (y - X\beta)$$

soggetta al vincolo  $\sum_{j=1}^J |\beta_j| \leq s$  che rappresenta la penalizzazione. La soluzione non è esplicita ed è necessario ricorrere ad algoritmi numerici. Se  $s$  viene scelto sufficientemente piccolo, alcuni coefficienti di regressione vengono stimati pari a zero e si ottiene così una selezione continua del modello.

## 2.2 Metodi di confronto

Per valutare la capacità predittiva del modello, Billé et al. (2022) hanno utilizzato la radice dell'errore quadratico medio (RMSE, *Root Mean Square Error*), che è una misura dell'errore assoluto, la procedura del *Model Confidence Set* (MCS) che è un insieme di modelli costruito in modo tale da contenere il modello di previsione migliore dato un livello di confidenza (Hansen et al., 2011). Gli autori hanno utilizzato il test di Diebold e Mariano (DM) (Diebold & Mariano, 1995, West, 1996) per il confronto dei residui dei modelli considerati per la previsione e il CRPS (*Continuous Ranked Probability Score*) che generalizza l'errore medio assoluto rendendolo applicabile alle previsioni probabilistiche.

### 2.2.1 Errore quadratico medio

L'errore quadratico medio (MSE, *Mean Square Error*) è la deviazione standard dei residui di un modello statistico. I residui  $e_i$  sono definiti come la distanza tra i valori realmente osservati ( $y_i$ ) e i valori stimati dal modello ( $\hat{y}_i$ ):  $e_i = y_i - \hat{y}_i$ .

L'errore quadratico medio di uno stimatore  $\hat{\theta}$  rispetto a un parametro d'interesse  $\theta$  è definito come

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

L'errore quadratico medio dà una misura quantitativa per giudicare la qualità di uno stimatore prendendone in considerazione la varianza e la distorsione, infatti l'MSE può essere scomposto nella somma della varianza dello stimatore più il quadrato della distorsione (o bias):

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (Bias(\hat{\theta}, \theta))^2$$

dove la distorsione è definita come  $E(\hat{\theta} - \theta)$ .

La radice dell'errore quadratico medio (RMSE) fornisce un ulteriore indice statistico che corrisponde alla deviazione standard interna. L'RMSE può essere anche calcolato come deviazione standard dei residui indicando una misura assoluta di errore.

La formula per calcolare il valore dell'errore quadratico medio, dati  $y_1, \dots, y_n$  i valori osservati e  $\hat{y}_1, \dots, \hat{y}_n$  i valori stimati dal modello selezionato, è:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

### 2.2.2 *Continuous Ranked Probability Score*

Per valutare l'accuratezza delle previsioni probabilistiche, misure più semplici come l'errore medio assoluto o l'errore medio assoluto percentuale, non sono applicabili. In questi casi il sistema utilizzato è il *Continuous Ranked Probability Score* (CRPS), che è una generalizzazione dell'errore medio assoluto che lo rende applicabile alle previsioni probabilistiche. In generale il CRPS è una funzione che confronta il vero valore osservato con una sua previsione distributiva, proprietà che lo rende particolarmente applicabile in ambito del *machine learning* Bayesiano, i cui modelli generano previsioni distributive piuttosto che stime puntuali.

Il CRPS è stato applicato per esempio in ambito economico da Gneiting & Ranjan (2011) e Groen et al. (2013) e in uno studio sulla previsione dei prezzi dell'elettricità in Italia da Gianfreda et al. (2020).

Da una definizione classica (Hersbach, 2000) il CRPS è definito come la misura quadratica della distanza tra la funzione di ripartizione predetta dal modello e la funzione di ripartizione empirica delle osservazioni. Data  $Y$  una variabile casuale e sia  $F$  la funzione di ripartizione di  $Y$  definita come  $F(Y) = P(Y < x)$ , sia inoltre  $y$  il vettore delle osservazioni e  $F$  la funzione di ripartizione associata ad una previsione probabilistica empirica, il CRPS tra  $F$  e  $y$  è definito come:

$$CRPS(F, y) = \int_{\mathbb{R}} (F(x) - I(x - y))^2 dx$$

dove  $I$  è la funzione indicatrice. Come già espresso, il CRPS generalizza l'errore medio assoluto, a cui si riduce se la previsione è deterministica (Zamo & Naveau, 2018).

### 2.2.3 Test di Diebold-Mariano

Per confrontare le capacità predittive di due modelli possiamo valutare se, date le previsioni dei due modelli alternativi, e data una funzione di perdita, il valore atteso della funzione di perdita è uguale per le previsioni dei due modelli.

La funzione di perdita  $L(e_t)$ , con  $e_t = Y_t - \hat{Y}_t$  errore di previsione, specifica la perdita associata a un errore di previsione di ampiezza  $e_t$ . Tale funzione deve essere specificata in modo che  $L(0) = 0$  e  $L(e_t) > 0$  per  $|e_t| \neq 0$ .

L'ipotesi nulla del test di Diebold-Mariano è quindi l'uguaglianza della perdita attesa per i due modelli. Data una funzione di perdita  $l_{i,t} = L(Y_t, \hat{Y}_t^i)$  valutata per due modelli distinti,  $i = 1, 2$ , i due modelli hanno uguale performance predittiva se  $E(l_{1,t}) = E(l_{2,t})$ . La statistica test si basa dunque sul corrispondente campionario della quantità  $E(l_{1,t} - l_{2,t})$ .

Definito  $d_t = l_{1,t} - l_{2,t}$  si calcola quindi  $\bar{d}_t = \frac{1}{M} \sum_{t=1}^M d_t$ , dove  $M$  è il numero di previsioni effettuate con i due modelli. Sotto determinate condizioni, si verifica che  $\sqrt{M}(\bar{d}_t - \mu_d) \rightarrow N(0, \sigma_d^2)$ , dove la varianza può essere stimata in modo consistente. Di conseguenza, indicato con  $\mu_d$  il valore atteso del differenziale delle funzioni di perdita, l'ipotesi nulla può essere verificata con la statistica test

$$\frac{\bar{d}_t}{\sqrt{\frac{\sigma_d^2}{M}}} \rightarrow N(0, 1)$$

Il test di Diebold-Mariano viene utilizzato anche per i confronti multipli, con un'opportuna correzione sul livello di confidenza (ad esempio la correzione di Bonferroni).

### 2.2.4 *Model Confidence Set*

Dato un livello di confidenza il *Model Confidence Set* (MCS) contiene il modello di previsione migliore: è l'analogo di un intervallo di confidenza per un parametro, con la differenza che è riferito a un insieme di modelli. In maniera coerente dati poco informativi portano a un MCS con molti modelli, al contrario, se i dati sono molto informativi solo pochi modelli saranno contenuti nel MCS, pertanto l'MCS differisce dai criteri di selezione del modello esistenti che scelgono un unico modello indipendentemente dal contenuto informativo dei dati. Un altro vantaggio è che la procedura MCS consente di formulare affermazioni sulla significatività che sono valide nel senso tradizionale, dato un certo livello di confidenza.

## 2.3 I dati

Nel modello proposto da Billé et al. (2022) vengono prese in considerazione variabili esogene, cioè variabili il cui valore è noto o controllabile dalle autorità di politica economica, e variabili endogene, i valori delle quali sono spiegati dal modello. I dati utilizzati nel modello proposto provengono da più fonti, alcune delle quali private ed altre pubbliche.

Vengono presi in considerazione i prezzi dell'energia elettrica nel mercato del giorno prima nell'area settentrionale dell'Italia e le previsioni orarie della quantità di energia prodotta dalle fonti rinnovabili: energia solare, vento, biomasse ed energia idroelettrica. Vengono inoltre considerati i prezzi dell'energia elettrica che viene importata, i prezzi dei combustibili fossili come carbone e gas naturale e infine i costi per le emissioni di anidride carbonica.

### 2.3.1 Le fonti

I prezzi, espressi in €/MWh, sono stati raccolti dal sito web del Gestore dei Mercati Energetici (GME), mentre le previsioni della quantità di energia elettrica prodotta dalle fonti rinnovabili vengono raccolte dall'*European Network of Transmission System Operator for Electricity* (ENTSO-E) e dalla fonte privata *Refinitiv Thomson Reuters* (RTR) e l'unità di misura viene riscalata da MW a GW.

I dati provenienti da ENTSO-E vanno dall'anno 2015 all'anno 2019: queste informazioni sono pubblicate al più due ore dopo la chiusura delle contrattazioni del Mercato del giorno prima, mentre i dati di RTR sono disponibili solo a partire dal 2018 al 2019 e provengono da due fonti:

- Centro europeo per le previsioni meteorologiche a medio termine (ECMWF da *European Centre for Medium Range Weather Forecasts*);
- *Global Forecast System* (GFS) che è un sistema di previsione meteorologica numerica che contiene un modello informatico globale e un'analisi gestita dal *National Weather Service* degli Stati Uniti.

Entrambe le fonti utilizzano due tipi diversi di modelli meteorologici, uno deterministico senza componenti casuali detto *operational model* e uno probabilistico, che è meno preciso ma tiene conto delle variazioni delle condizioni meteorologiche detto *ensemble model*. In base alle pubblicazioni dei dati provenienti dai due modelli si utilizzano due diverse serie di previsione, entrambe contengono le ultime informazioni sul mercato con la differenza che una è detta *fast (F)*, in quanto include informazioni più recenti, l'altra serie, è detta *less fast (LF)*, perché contiene alcune informazioni meno recenti.

I dati per le informazioni sull'energia proveniente dalle fonti rinnovabili sono raccolti da ENTSO-E, ma non sono immediatamente disponibili in tempo reale e presentano quindi un ritardo.

Anche i dati sulle importazioni di energia e i prezzi di vendita nei paesi esteri sono presi da ENTSO-E: vengono raccolte sia informazioni sulla quantità di energia importata, sia sul prezzo orario delle importazioni per poi costruire una serie d'interesse che combina il prezzo medio orario dell'energia, calcolato come media dei prezzi orari nel mercato del giorno prima in Austria, Svizzera, Francia e Slovenia, pesato per la quantità importata.

Per tener conto dei costi dell'energia prodotta da fonti termoelettriche convenzionali, gli autori hanno preso in considerazione i prezzi del gas naturale dal TTF olandese (*Title Transfer Facility*), i prezzi dell'*ICE API2 Rotterdam Future* per il carbone e i dati provenienti da RTR per i prezzi sulle emissioni di anidride carbonica.

L'insieme di dati collezionati dagli autori contiene 35064 osservazioni per ogni variabile, da gennaio 2015 a dicembre 2019 fatta eccezione per i modelli che considerano i dati provenienti dalla fonte RTR, i quali sono disponibili solo a partire dal 2018.

### 2.3.2 Analisi esplorativa

I dati vengono utilizzati nel modello proposto nel seguito considerando le serie orarie e i dati mancanti o incompleti, perché non disponibili o non pubblicati, vengono sostituiti dall'informazione corrispondente osservata il giorno precedente alla stessa ora. Gli outlier vengono mantenuti nelle serie perché ritenuti rappresentativi di particolari caratteristiche del mercato dell'energia del nord Italia. Alle variabili non viene applicata

la trasformazione logaritmica per migliorare la distribuzione e stabilizzare la varianza in quanto la trasformata potrebbe non permettere alle dinamiche di volatilità del fenomeno di interesse di emergere.

Nella Tabella 2.1 sono riportate le statistiche di sintesi delle variabili considerate nel modello e nelle Figure 2.1 e 2.2 sono riportate le serie storiche di tutte le variabili esogene.

TABELLA 2.1: Statistiche di sintesi delle variabili calcolate sull'intero campione

	Min	Media	Max	Std. Dev.	Asimmetria	Curtosi
Price	1.000	52.345	206.120	16.364	1.107	3.426
Forecasted load	7.344	18.624	31.617	4.858	0.164	-1.107
Weighted import	0.000	43.075	249.340	15.551	0.911	2.883
Coal	4.280	6.961	9.840	1.598	0.143	-1.350
Natural gas	9.630	17.575	29.330	3.986	0.296	-0.367
CO2	0.440	1.330	3.316	0.877	0.829	-0.937
Forecasted solar	0.000	0.765	5.499	1.153	1.417	0.832
Forecasted wind	0.000	0.004	0.035	0.004	1.509	3.623
Hydro	0.550	3.910	10.510	2.029	0.348	-0.772
Biomass	0.044	0.128	0.237	0.036	0.818	-0.013
Waste	0.008	0.037	0.056	0.009	-0.532	0.058

I prezzi, l'energia solare, l'energia eolica e le importazioni mostrano asimmetria moderata e curtosi piuttosto alta; l'energia eolica, le biomasse e i rifiuti, sebbene nel nord Italia presentino valori piuttosto bassi, vengono ugualmente considerate nel modello per completezza: ciò consente di utilizzare il modello specificato nel seguito per altre zone di interesse, in cui queste variabili potrebbero fornire un contributo maggiore.

Nella Figura 2.3 sono riportate le serie dei prezzi dell'energia elettrica nel nord Italia del mercato del giorno prima alle ore 3, 9, 13, 15, 21 e 24. Si nota che i consumi e i prezzi dell'energia elettrica sono più alti nei giorni lavorativi e diminuiscono nel fine settimana. La stagionalità mensile è invece caratterizzata da livelli di consumo più alti nei mesi invernali di gennaio e febbraio e nei mesi estivi probabilmente dovuti all'utilizzo di sistemi di riscaldamento e raffreddamento rispettivamente. L'energia eolica e solare variano in base alle condizioni meteorologiche e a seconda delle ore.

Sono considerati il test di Jarque-Bera (JB) per verificare la normalità dei termini di errore (Jarque & Bera, 1987), i test per la stazionarietà di Dickey-Fuller (ADF) (Dickey & Fuller, 1979) e il test di Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Kwiatkowski et al., 1992).

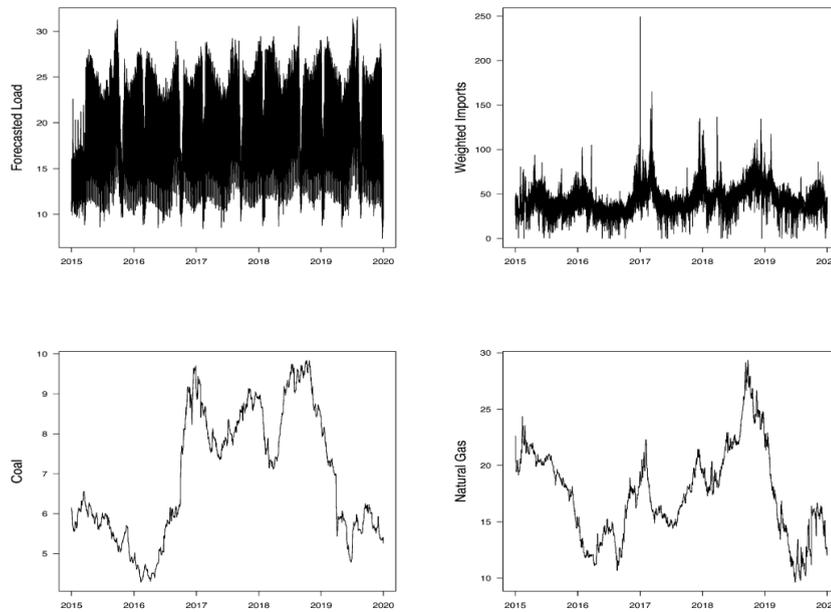


FIGURA 2.1: Serie storiche per le variabili *Forecasted Load*, *Weighted Import*, *Coal* e *Natural Gas*

Per il test JB non si ha normalità, le serie risultano stazionarie per il test di ADF e non stazionarie per il test KPSS.

## 2.4 Specificazione del modello

Il modello proposto nell'articolo di Billé et al. (2022) è composto da 58 tipologie di modelli divisi in cinque gruppi:

- Quattro modelli combinati (*expert models*) ( $EX_1$ ,  $EX_2$ ,  $EX_3$  e  $EX_4$ ), le loro rispettive estensioni per la presenza di variabili esogene ( $EX_1X$ ,  $EX_2X$ ,  $EX_3X$  e  $EX_4X$ ) e i modelli GARCH per tener conto della volatilità (Paragrafo 2.1.2);
- Modelli autoregressivi  $AR(p)$  e una loro estensione  $ARX(p)$  (Paragrafo 2.1.1);
- Modelli autoregressivi a media mobile  $ARMA(p,q)$  e una loro estensione  $ARMAX(p,q)$  (Paragrafo 2.1.1);
- Modelli autoregressivi integrati a media mobile  $ARIMA(p,d,q)$  con  $d$  non intero ( $ARFIMA(p,d,q)$ , estesi alla presenza di variabili esogene ( $ARFIMAX(p,d,q)$ ) (Paragrafo 2.1.1);
- Regressione LASSO (Paragrafo 2.1.3) ;

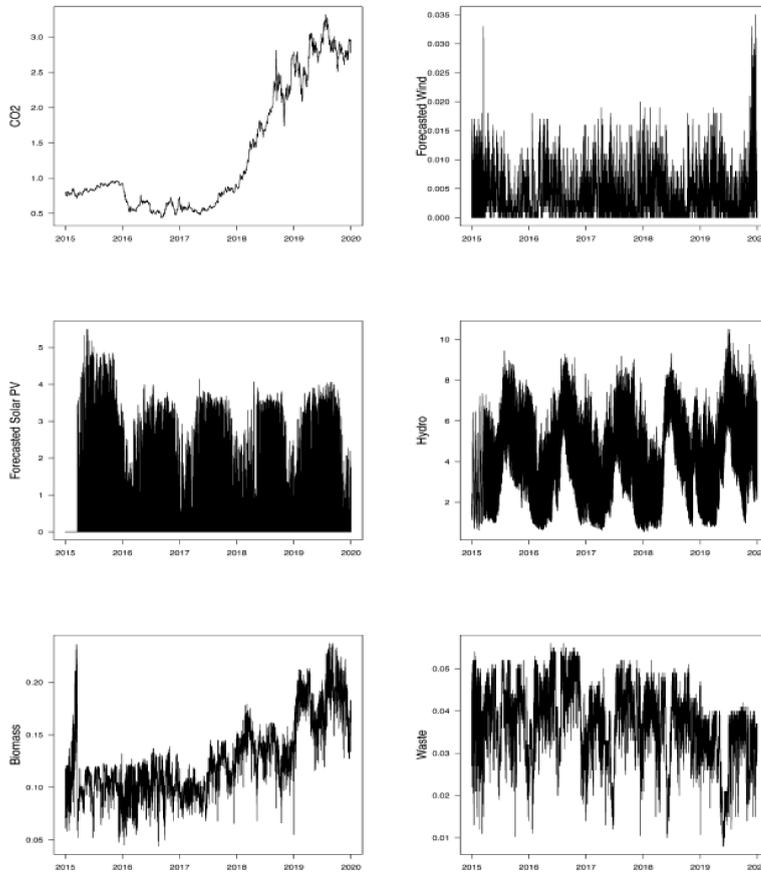


FIGURA 2.2: Serie storiche per le variabili *CO2*, *Forecasted Wind*, *Forecasted Solar PV*, *Hydro*, *Biomass* e *Waste*

### 2.4.1 Modelli combinati

La prima specificazione del modello ( $EX_1$ ) prende in considerazione i prezzi passati osservati uno, due e sette giorni prima e considera sei variabili dummy che indicano i giorni della settimana  $D_t^k$  dove  $D_t^1$  è pari a 1 per il lunedì e così via per  $k = 1, \dots, 6$ .

Sia  $y_t$  il prezzo orario dell'energia, il modello risulta:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \sum_{k=1}^6 \gamma_k D_t^k + \epsilon_t$$

Il modello viene esteso ( $EX_1X$ ) alla presenza dei regressori  $\mathbf{x}_t$  e  $\mathbf{z}_{t-1}$  che indicano rispettivamente al tempo  $t$  il carico previsto e l'energia solare ed eolica e al tempo  $t-1$  energia idroelettrica, biomasse, rifiuti e i prezzi delle importazioni, del gas naturale e della CO2.

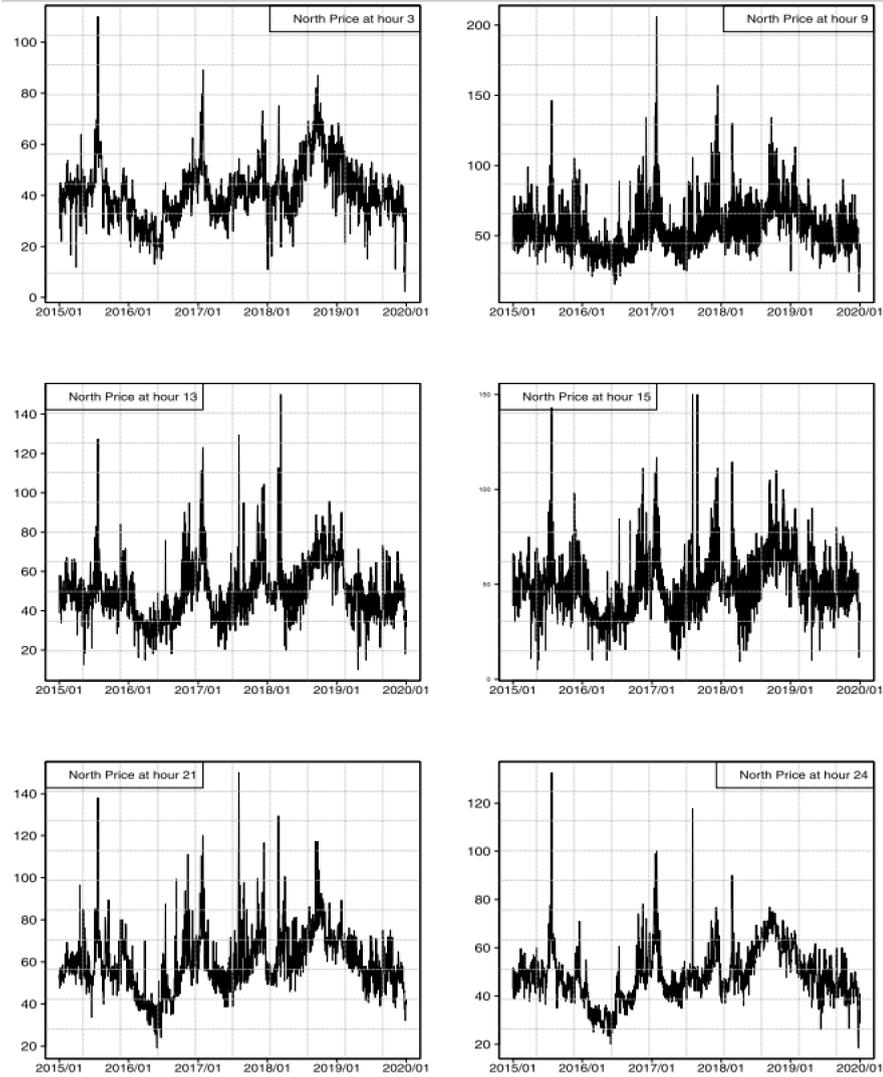


FIGURA 2.3: Prezzi dell'energia elettrica nel nord Italia del mercato del giorno prima alle ore 3, 9, 13, 15, 21 e 24

Risulta quindi:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \sum_{k=1}^6 \gamma_k D_t^k + \lambda' \mathbf{x}_t + \kappa' \mathbf{z}_{t-1} + \epsilon_t$$

Per la seconda specificazione del modello ( $EX_2$ ) vengono introdotti il prezzo massimo e minimo orari dell'energia osservati il giorno precedente, viene successivamente esteso, come visto per il modello  $EX_1$ , alla presenza delle variabili esogene. Il modello  $EX_2X$  diventa:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \beta_4 y_{min,t-1} + \beta_5 y_{max,t-1} + \sum_{k=1}^6 \gamma_k D_t^k + \lambda' \mathbf{x}_t + \kappa' \mathbf{z}_{t-1} + \epsilon_t$$

Il terzo modello  $EX_3$  include il prezzo orario all'ora 24 del giorno precedente:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \beta_4 y_{min,t-1} + \beta_4 5 y_{max,t-1} + \beta_6 y_{24,t-1} + \sum_{k=1}^6 \gamma_k D_t^k + \epsilon_t$$

analogamente si ottiene il modello  $EX_3X$  considerando i regressori.

L'ultimo modello  $EX_4$  prende in considerazione i prezzi svalutati, il modello così diventa:

$$y_t = \alpha_0 + \alpha_1 \bar{y}_t^\omega + \sum_{k=1}^8 \beta_k (y_{t-k} - \bar{y}_t^\omega) + \epsilon_t$$

dove  $\bar{y}_t^\omega$  è il valore medio del prezzo orario durante tutta la settimana e inoltre viene presa in considerazione la dipendenza del prezzo medio dai  $k = 8$  giorni passati.

Nel modello  $EX_4X$  vengono prese in considerazione le variabili dummy per i giorni della settimana, specificate come per il modello  $EX_1$  e i regressori  $\mathbf{x}_t$  e  $\mathbf{z}_{t-1}$ .

## 2.4.2 Modelli ARFIMA

Successivamente si passa ai modelli ARFIMA. Il primo modello che viene specificato è un modello AR(7) e una sua variante con il parametro  $p$  variabile. Formalmente il modello AR( $p$ ) è definito come:

$$y_t = \alpha + \sum_{k=1}^4 \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j + \sum_{r=1}^p \phi_r y_{t-r} + \epsilon_t$$

con  $D_t^k$  specificato diversamente dai modelli descritti nel Paragrafo 2.4.1,  $k = 1$  che indica il lunedì,  $k = 2$  il sabato,  $k = 3$  la domenica e  $k = 4$  che indica un giorno festivo che capita in un giorno feriale;  $M_t^j$  variabili dummy per i mesi da gennaio a novembre e  $\phi_r$  parametri autoregressivi con  $p = 1, \dots, 7$ . Il modello viene poi esteso analogamente alla presenza dei regressori  $\mathbf{x}_t$  e  $\mathbf{z}_{t-1}$  come per i modelli del Paragrafo 2.4.1 diventando un ARX( $p$ ) con  $p = 1, \dots, 7$ .

Viene proposto poi un modello ARMA( $p, q$ ) con  $p, q = 1, \dots, 7$  variabili o fissati, il modello formalmente è definito come:

$$y_t = \alpha + \sum_{k=1}^4 \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j + \sum_{r=1}^p \phi_r y_{t-r} + \sum_{s=1}^q \theta_s \epsilon_{t-s} + \epsilon_t$$

con le variabili  $D_t^k$ ,  $\gamma_t^j$  e  $\phi_r$  definite come nel modello precedente e  $\theta_s$  parametri a media mobile con  $q = 1, \dots, 7$ . I modelli quindi, estesi alla presenza dei regressori esogeni, possono essere ARMAX( $p, q$ ), ARMAX(1,7), ARMAX(7,1), ARMAX(7,7).

Infine per tener conto della memoria lunga, si considera un modello ARFIMA( $p, d, q$ ) definito come:

$$\Phi(L)(1 - L)^d(y_t - \mu_t) = \Theta(L)\epsilon_t$$

con  $\epsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma^2)$  e  $\mu_t = \mu + \sum_{k=1}^4 \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j$  dove  $D_t^k$  e  $\gamma_t^j$  sono definite come nel modello AR( $p$ ) specificato sopra e  $0 < d < 0.5$ . Questo modello viene poi confrontato con modelli in cui si considerano anche i regressori e, in particolare, vengono considerati i modelli ARFIMAX( $7, d, 7$ ) e ARFIMAX( $7, d, 0$ ) dove il parametro  $d$  varia da 0, 1 e 2.

### 2.4.3 Modelli GARCH

Per tener conto della volatilità variabile nel tempo, spesso osservata nei prezzi dell'energia, l'ultima classe di modelli considerata è la classe GARCH (Paragrafo 2.1.2). Questa classe di modelli è già stata utilizzata in precedenza per i modelli proposti negli studi di Bosco et al. (2007) i quali hanno utilizzato un modello ARMA-GARCH, e di Gianfreda & Grossi (2012), che hanno utilizzato un modello ARFIMA-GARCHX.

Nell'articolo di Billé et al. (2022) vengono quindi messi a confronto i modelli ARFIMA con le varianti dei modelli GARCH, con selezione automatica della lunghezza della componente autoregressiva e a media mobile. Sono quindi considerati il modello GARCH standard (SGARCH), il GARCH esponenziale (EGARCH), il GARCH in media (GARCH-M) e il GARCH a soglia (TGARCH), tutti con distribuzione normale.

Questa serie di modelli estende la serie presentata nei Paragrafi 2.4.2 e 2.4.1 con la differenza che la variabilità viene espressa senza perdita di generalità al giorno  $t$  come  $\sigma_t^2 = Var(\epsilon_t | \mathcal{F}_{t-1})$ .

In particolare il modello EGARCH (Nelson, 1991) permette alla varianza di rispondere in modo asimmetrico agli aumenti e in generale alle variazioni dei prezzi dell'elettricità, il modello TGARCH (Zakoian, 1994) permette di rendere conto delle asimmetrie nella volatilità dei prezzi mentre il modello GARCH-M (Kyritsis et al., 2017, Gianfreda & Scandolo, 2018) permette di considerare che i prezzi possano essere influenzati dalla variabilità passata e che la volatilità dei prezzi dell'elettricità è generalmente maggiore quando i prezzi sono più alti.

Le specifiche di questi modelli vengono estese alla presenza delle variabili esogene  $\mathbf{x}_t$  e  $\mathbf{z}_{t-1}$ .

### 2.4.4 LASSO

Alla fine viene considerata la regressione LASSO di altri modelli autoregressivi con 28 ritardi, in modo da tener conto del cambiamento del mercato dell'energia per le ultime quattro settimane: vengono proposti il modello  $AR(28)_{LASSO}$  e  $ARX(28)_{LASSO}$  estesi alla presenza dei regressori. Infine vengono anche stimati i modelli  $ARX(28) - GARCHX(1, 1) - M_{LASSO}$  per includere la volatilità.

## 2.5 Stima dei modelli

La procedura iterativa utilizzata per stimare i modelli proposti nel Paragrafo 2.4 permette un maggiore adattamento alle condizioni variabili dell'andamento dei prezzi sul mercato, all'aumento o alla diminuzione dell'energia prodotta dalle RES e più in generale all'andamento del mercato dell'energia. Per tutti i calcoli eseguiti viene utilizzato il software R ([www.r-project.org](http://www.r-project.org)).

Per tener conto della distorsione indotta dalla selezione iterativa, nelle previsioni del giorno prima, si confrontano i modelli iterativi con i modelli con ordini predeterminati.

La procedura iterativa per la stima dei modelli procede in due passi:

- Inizialmente vengono stimati gli ordini  $p$ ,  $q$  ed eventualmente l'ordine non intero  $d$  confrontando i modelli tramite il valore dell'AIC corretto (AICc);
- Successivamente vengono stimati i parametri dei modelli  $ARFIMAX(p, d, q)$  e  $ARFIMAX(p, q)$ -GARCH in presenza di regressori, si utilizzano gli ordini  $(\hat{p}, \hat{q})$  stimati al punto precedente mentre il parametro  $d$  viene stimato assieme agli altri parametri di interesse.

Per entrambi i punti si utilizza il pacchetto di R *nloptr* (Nelson, 2021) per l'ottimizzazione non lineare e infine, i parametri vengono stimati con la massima verosimiglianza.

I modelli con gli ordini fissati sono stimati con l'algoritmo di ottimizzazione di Broyden-Fletcher-Goldfarb-Shanno (BFGS) e in aggiunta viene effettuata una modifica per garantire la convergenza dell'algoritmo (SANN di Bélisle, 1992).

Per la LASSO si procede in modo che i regressori esogeni  $\mathbf{x}_t$  e  $\mathbf{z}_{t-1}$  si combinano con i 28 ritardi secondo un modello lineare. In questo modo si è in grado di selezionare un sottogruppo di regressori, ciascuno per ogni iterazione e per ogni ora. Per il compromesso tra varianza e distorsione viene utilizzato il criterio di convalida incrociata (*cross validation*, CV) che minimizza l'errore medio.

## 2.6 Valutazione della performance previsiva

Per la valutazione della performance previsiva dei modelli specificati nel Paragrafo 2.4 sono utilizzati una prima parte dei dati (Paragrafo 2.3) per stimare i parametri dei modelli, vengono poi calcolate le previsioni dei prezzi orari sulla base dei modelli proposti e infine le previsioni sono confrontate con la seconda parte dei dati osservati.

Nello studio di Billé et al. (2022) sono stati utilizzati i dati per i primi 730 giorni, da gennaio 2015 a dicembre 2016 per produrre le stime dei parametri dei modelli e i dati disponibili dal 2017 fino al 2019 per le previsioni. Sono state prodotte le previsioni per i tre anni successivi: dal 2017 al 2019 e confrontate con i dati disponibili di ENTSO-E. Nella pratica, per la previsione, viene utilizzata l'informazione al tempo  $t$  per produrre un insieme di 24 valori per i prezzi orari dell'energia previsti per il giorno successivo, quindi a  $t+1$ . Queste previsioni devono essere presentate prima della chiusura del mercato, cioè entro le 12:00 del giorno  $t$ .

Per calcolare le previsioni del prezzo orario dell'energia al tempo  $t+1$ , vengono usati i dati riferiti a quella specifica ora dei giorni precedenti. Si assume che gli operatori mandino entro le 12:00 le loro offerte: le offerte sono basate sulle previsioni dei prezzi al tempo  $t+1$  e le previsioni tengono conto delle variabili esogene osservate al tempo  $t-1$  e al tempo  $t$  stesso.

Per valutare le previsioni gli autori hanno considerato i metodi descritti nel Paragrafo 2.2, i cui risultati sono riportati nel Capitolo 3.



# Capitolo 3

## I risultati

### 3.1 Performance previsiva

Per valutare la qualità e la precisione delle previsioni dei prezzi orari dell'energia elettrica nel nord Italia gli autori Billé et al. (2022) hanno utilizzato i metodi descritti nel Paragrafo 2.2 del Capitolo 2. I valori dell'RMSE e i CRPS sono riportati nelle Tabelle 3.1 e 3.3 rispettivamente, i risultati sono calcolati alle ore 3, 9, 11, 13, 19 e 21 e nelle ultime due colonne sono riportate le medie dei due valori calcolate per tutte le 24 ore ( $Avg_{1-24}$ ) e per le ore di picco dalle 8 alle 20 ( $Avg_{8-20}$ ).

Osserviamo che l'inclusione dei regressori nel modello, in particolare nelle ore di punta, porta ad una diminuzione sia dell'RMSE che del CRPS. Considerando tutti i 58 modelli, si nota che il modello  $EX_4X$  è quello migliore per quanto riguarda la previsione puntuale con l'errore medio di circa 6 €/Mwh per le ore di picco e circa 7 €/MWh per le altre ore. Con l'inclusione della varianza variabile nel tempo, si nota che il modello non è significativamente migliore per quanto riguarda la previsione puntuale, tuttavia, ci si aspetta comunque un miglioramento in quanto il fenomeno di interesse è complesso e il fatto che la varianza non sia fissata per ogni istante di tempo aiuta a cogliere eventuali andamenti particolari dei prezzi dell'energia (Ziel et al., 2015). I modelli ARFIMA forniscono risultati simili e di poco superiori.

Inoltre è possibile notare che la precisione delle previsioni diminuisce durante le ore di incremento e decremento, in particolare nelle ore 9 e 19: questo fenomeno è dovuto al fatto che in queste fasi, per ristabilire l'equilibrio tra domanda e offerta, risulta necessario l'utilizzo di energia proveniente da fonti termoelettriche convenzionali. Durante le ore di punta le fonti rinnovabili, in particolare l'energia solare ed eolica producono energia elettrica a 0 €/MWh e detengono la priorità di dispacciamento. La produzione di

energia dalle RES è irregolare, ciò incide sugli errori di previsione, soprattutto nelle ore in cui la domanda è più alta, facendo aumentare i valori di RMSE e CRPS.

Il potere di prevedibilità durante le ore serali diminuisce poiché gli orizzonti di previsione sono più lunghi rispetto a quelli delle ore mattutine. Ciò è particolarmente influenzato dalle RES in quanto dipendono dalle previsioni delle condizioni meteorologiche la cui precisione diminuisce all'aumentare dell'orizzonte temporale di previsione.

La regressione LASSO non è necessaria per migliorare l'accuratezza in quanto non porta a sostanziali miglioramenti. Nessuno dei modelli infatti fornisce previsioni migliori del modello  $EX_4X$ , che è il modello preso a riferimento per il test DM, sotto l'ipotesi alternativa che qualsiasi altro modello produce previsioni migliori. Il modello  $EX_4X$  e le sue varianti GARCH sono sempre presenti nel *Model Confidence Set*, in tutte le ore considerate, e risultano sempre significativamente migliori in tutti i confronti a coppie (test DM).

Per le previsioni di densità, (Tabella 3.3) i valori dei CRPS mostrano che si ha un miglioramento sostanziale con l'introduzione dei modelli GARCH: la media dei CRPS nelle 24 ore di tutti gli altri i modelli è intorno a 0.3 mentre per i modelli GARCH si aggira tra 0.156 e 0.160. In generale risulta quindi fondamentale introdurre il termine di varianza variabile nel tempo per garantire una migliore performance previsiva.

Nelle Tabelle 3.5 e 3.6 sono riportati i valori di RMSE e CRPS dei modelli migliori,  $EX_4X$  e  $EX_4X - SGARCHX$ , in presenza di previsioni disponibili gratuitamente o di previsioni più tempestive ottenute da fonti private. Vengono confrontate le prestazioni di previsione dei modelli considerando le previsioni di ENSO-E con quelle ottenute per lo stesso modello considerando le previsioni da RTR. Inoltre si distinguono i modelli *Fast (F)* dai modelli *Less Fast (LF)*.

TABELLA 3.1: RMSE di tutti i modelli selezionati su alcune ore

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
EX <sub>1</sub>	6.468	12.501	10.822	9.834	11.836	8.911	10.973	9.150
EX <sub>2</sub>	6.437	12.162	10.533	9.636	11.748	8.809	10.760	9.016
EX <sub>3</sub>	5.984	11.960	10.309	9.411	11.804	8.756	10.616	8.824
EX <sub>4</sub>	5.204	9.469	8.284	7.998	8.365	6.643	8.516	7.074
EX <sub>1</sub> X	6.151	11.691	10.196	9.144	11.169	8.614	10.303	8.644
EX <sub>2</sub> X	6.113	11.563	10.062	9.062	11.175	8.636	10.237	8.602
EX <sub>3</sub> X	5.814	11.378	9.864	8.895	11.237	8.756	10.131	8.499
EX <sub>4</sub> X	4.860	8.691	7.708	7.318	8.106	6.466	7.871	6.615
EX <sub>4</sub> X – SGARCHX	4.861	8.834	7.712	7.274	8.284	6.449	7.919	6.628
EX <sub>4</sub> X – EGARCHX	4.912	8.790	7.805	7.230	8.207	6.473	7.907	6.626
EX <sub>4</sub> X – TGARCHX	4.894	8.835	7.791	7.286	8.401	6.595	7.992	6.689
EX <sub>4</sub> X – GARCHX – M	4.925	8.846	7.788	7.313	8.264	6.420	7.976	6.662
AR(7)	5.577	11.500	9.867	9.228	10.272	8.122	10.130	8.327
AR(p)	5.529	11.977	10.170	9.545	10.521	8.223	10.582	8.596
ARX(7)	5.489	10.706	9.162	8.521	9.932	7.837	9.412	7.847
ARX(p)	5.449	11.000	9.428	8.739	10.135	7.871	9.703	8.024
ARX(7)-SGARCHX	5.462	10.722	9.177	8.337	9.739	7.689	9.367	7.795
ARX(7)-EGARCHX	5.462	10.873	9.225	8.455	9.712	7.786	9.395	7.826
ARX(7)-TGARCHX	5.506	10.596	9.208	8.489	10.048	7.745	9.472	7.871
ARX(7)-GARCHX-M	5.470	10.692	9.184	8.348	9.804	7.689	9.405	7.843
ARX(p)-SGARCHX	5.436	10.924	9.404	8.591	10.088	7.777	9.736	8.034
ARX(p)-EGARCHX	5.446	11.074	9.536	8.618	10.021	7.899	9.720	8.030
ARX(p)-TGARCHX	5.498	11.001	9.484	8.979	10.309	7.837	9.932	8.164
ARX(p)-GARCHX-M	5.468	11.025	9.571	8.603	10.091	7.813	9.824	8.114
ARMA(7,7)	5.717	13.542	10.430	9.402	10.629	9.974	10.615	8.934
ARMA(1,7)	5.589	11.808	9.942	9.362	10.449	8.192	10.309	8.447
ARMA(7,1)	5.584	11.483	9.821	9.191	10.310	8.105	10.091	8.300
ARMA(p,q)	5.561	11.805	10.066	9.429	10.520	8.160	10.450	8.516
ARMAX(7,7)	5.533	10.418	9.197	8.493	9.950	7.685	9.339	7.824
ARMAX(1,7)	5.518	10.770	9.176	8.571	10.003	7.819	9.444	7.873
ARMAX(7,1)	5.494	10.710	9.136	8.498	9.968	7.832	9.394	7.836
ARMAX(p,q)	5.467	10.959	9.353	8.660	10.129	7.827	9.630	7.975
ARMAX(7,7)-SGARCHX	5.693	10.562	10.574	8.461	9.831	7.893	9.708	8.044
ARMAX(7,7)-EGARCHX	5.578	12.429	9.732	8.662	9.827	8.050	9.702	8.045
ARMAX(7,7)-TGARCHX	5.624	10.768	9.261	8.762	10.003	7.752	9.495	7.910
ARMAX(7,7)-GARCHX-M	5.689	10.746	9.228	8.530	9.959	7.719	9.519	8.135
ARMAX(p,q)-SGARCHX	5.453	10.867	9.314	8.649	10.060	7.737	9.625	7.959
ARMAX(p,q)-EGARCHX	5.456	10.875	9.376	8.522	10.138	7.779	9.626	8.009
ARMAX(p,q)-TGARCHX	5.507	10.959	9.248	8.638	10.303	7.822	9.781	8.060
ARMAX(p,q)-GARCHX-M	5.501	11.157	9.526	8.766	10.112	7.756	9.991	8.195

TABELLA 3.2: *Continua dalla pagina precedente*

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
ARFIMA(p,d,q)	5.572	11.261	9.827	9.267	10.054	8.223	10.063	8.289
ARFIMAX(p,d,q)	5.467	10.959	9.353	8.661	10.121	7.827	9.630	7.975
ARFIMAX(p,d,q)-SGARCHX	5.459	10.847	9.303	8.620	10.044	7.762	9.617	7.947
ARFIMAX(7,d,7)-SGARCHX	5.604	10.782	9.134	8.462	10.064	7.799	9.513	7.923
ARFIMAX(7,d,0)-SGARCHX	5.455	10.806	9.143	8.317	9.895	7.698	9.385	7.804
ARFIMAX(p,d,q)-EGARCHX	5.468	10.807	9.482	8.678	10.132	7.745	9.653	7.975
ARFIMAX(7,d,7)-EGARCHX	5.763	10.655	9.828	10.534	10.007	8.823	9.703	8.110
ARFIMAX(7,d,0)-EGARCHX	5.458	10.727	9.064	8.473	9.775	7.749	9.354	7.796
ARFIMAX(p,d,q)-TGARCHX	5.480	11.010	9.247	8.543	10.213	7.782	9.702	8.007
ARFIMAX(7,d,7)-TGARCHX	5.588	10.447	9.217	8.664	9.895	7.780	9.424	7.884
ARFIMAX(7,d,0)-TGARCHX	5.506	10.671	9.110	8.383	10.009	7.689	9.400	7.833
ARFIMAX(p,d,q)-GARCHX-M	5.480	10.999	9.424	8.671	10.178	7.839	10.068	8.262
ARFIMAX(7,d,7)-GARCHX-M	6.586	10.883	9.334	8.595	10.209	7.962	9.913	8.402
ARFIMAX(7,d,0)-GARCHX-M	5.480	10.782	9.084	8.289	10.021	7.714	9.436	7.852
AR(28) <sub>LASSO</sub>	6.415	12.395	10.802	9.960	11.462	9.007	10.914	9.117
ARX(28) <sub>LASSO</sub>	6.197	11.736	10.094	9.210	11.051	8.649	10.259	8.636
AR(28)-GARCH-M <sub>LASSO</sub>	6.394	12.614	10.845	10.045	11.859	9.171	11.143	9.275
ARX(28)-GARCHX-M <sub>LASSO</sub>	6.295	11.577	10.258	9.195	11.060	8.678	10.361	8.721

TABELLA 3.3: CRPS di tutti i modelli selezionati su alcune ore

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
EX <sub>1</sub>	0.255	0.332	0.317	0.298	0.323	0.313	0.317	0.298
EX <sub>2</sub>	0.254	0.332	0.318	0.299	0.323	0.313	0.317	0.298
EX <sub>3</sub>	0.253	0.333	0.318	0.300	0.324	0.313	0.318	0.298
EX <sub>4</sub>	0.252	0.323	0.309	0.291	0.314	0.309	0.309	0.292
EX <sub>1</sub> X	0.254	0.327	0.312	0.294	0.319	0.314	0.312	0.295
EX <sub>2</sub> X	0.253	0.327	0.312	0.294	0.319	0.314	0.312	0.295
EX <sub>3</sub> X	0.253	0.327	0.312	0.294	0.320	0.313	0.313	0.295
EX <sub>4</sub> X	0.251	0.320	0.307	0.289	0.313	0.308	0.307	0.290
EX <sub>4</sub> X - SGARCHX	0.125	0.199	0.174	0.156	0.178	0.166	0.178	0.158
EX <sub>4</sub> X - EGARCHX	0.125	0.198	0.173	0.155	0.180	0.164	0.178	0.158
EX <sub>4</sub> X - TGARCHX	0.124	0.199	0.174	0.155	0.178	0.167	0.179	0.158
EX <sub>4</sub> X - GARCHX - M	0.125	0.201	0.173	0.155	0.178	0.164	0.179	0.158
AR(7)	0.254	0.326	0.312	0.294	0.317	0.312	0.312	0.295
AR(p)	0.252	0.324	0.310	0.292	0.314	0.309	0.310	0.292

TABELLA 3.4: *Continua dalla pagina precedente*

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
ARX(7)	0.252	0.323	0.309	0.292	0.315	0.310	0.309	0.292
ARX(p)	0.252	0.324	0.310	0.292	0.314	0.310	0.310	0.293
ARX(7)-SGARCHX	0.124	0.201	0.176	0.158	0.177	0.163	0.179	0.158
ARX(7)-EGARCHX	0.124	0.205	0.179	0.159	0.177	0.166	0.181	0.159
ARX(7)-TGARCHX	0.124	0.200	0.175	0.154	0.179	0.162	0.178	0.158
ARX(7)-GARCHX-M	0.124	0.201	0.175	0.158	0.177	0.163	0.179	0.158
ARX(p)-SGARCHX	0.124	0.200	0.174	0.153	0.176	0.163	0.177	0.157
ARX(p)-EGARCHX	0.124	0.208	0.179	0.156	0.177	0.166	0.181	0.160
ARX(p)-TGARCHX	0.124	0.201	0.173	0.149	0.177	0.161	0.176	0.156
ARX(p)-GARCHX-M	0.124	0.201	0.175	0.154	0.176	0.163	0.178	0.158
ARMA(7,7)	0.264	0.341	0.326	0.308	0.328	0.324	0.326	0.307
ARMA(1,7)	0.264	0.338	0.324	0.307	0.328	0.322	0.324	0.306
ARMA(7,1)	0.264	0.338	0.324	0.307	0.328	0.322	0.324	0.306
ARMA(p,q)	0.262	0.337	0.323	0.305	0.327	0.320	0.323	0.304
ARMAX(7,7)	0.252	0.323	0.310	0.292	0.315	0.310	0.310	0.292
ARMAX(1,7)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARMAX(7,1)	0.252	0.323	0.309	0.292	0.315	0.310	0.309	0.292
ARMAX(p,q)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARMAX(7,7)-SGARCHX	0.125	0.202	0.179	0.157	0.180	0.165	0.180	0.159
ARMAX(7,7)-EGARCHX	0.126	0.203	0.183	0.159	0.179	0.166	0.182	0.160
ARMAX(7,7)-TGARCHX	0.125	0.201	0.178	0.156	0.180	0.163	0.180	0.159
ARMAX(7,7)-GARCHX-M	0.125	0.202	0.177	0.158	0.178	0.164	0.181	0.159
ARMAX(p,q)-SGARCHX	0.124	0.202	0.176	0.157	0.178	0.163	0.179	0.158
ARMAX(p,q)-EGARCHX	0.124	0.206	0.179	0.157	0.177	0.166	0.181	0.159
ARMAX(p,q)-TGARCHX	0.124	0.201	0.175	0.152	0.179	0.161	0.178	0.157
ARMAX(p,q)-GARCHX-M	0.124	0.202	0.176	0.158	0.177	0.163	0.179	0.158
ARFIMA(p,d,q)	0.263	0.339	0.325	0.307	0.328	0.321	0.325	0.306
ARFIMAX(p,d,q)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARFIMAX(p,d,q)-SGARCHX	0.124	0.201	0.175	0.157	0.178	0.164	0.178	0.158
ARFIMAX(7,d,7)-SGARCHX	0.125	0.202	0.178	0.158	0.179	0.164	0.181	0.160
ARFIMAX(7,d,0)-SGARCHX	0.125	0.201	0.176	0.158	0.177	0.163	0.179	0.158
ARFIMAX(p,d,q)-EGARCHX	0.125	0.206	0.179	0.158	0.177	0.166	0.181	0.160
ARFIMAX(7,d,7)-EGARCHX	0.126	0.204	0.182	0.161	0.179	0.167	0.183	0.161
ARFIMAX(7,d,0)-EGARCHX	0.125	0.205	0.179	0.159	0.178	0.167	0.181	0.160
ARFIMAX(p,d,q)-TGARCHX	0.124	0.200	0.176	0.152	0.179	0.162	0.178	0.157
ARFIMAX(7,d,7)-TGARCHX	0.125	0.200	0.178	0.157	0.179	0.164	0.180	0.159
ARFIMAX(7,d,0)-TGARCHX	0.124	0.200	0.176	0.154	0.179	0.162	0.178	0.158
ARFIMAX(p,d,q)-GARCHX-M	0.124	0.202	0.175	0.157	0.177	0.163	0.178	0.158
ARFIMAX(7,d,7)-GARCHX-M	0.126	0.203	0.177	0.158	0.178	0.165	0.180	0.160
ARFIMAX(7,d,0)-GARCHX-M	0.124	0.201	0.176	0.158	0.177	0.164	0.179	0.158
AR(28) <sub>LASSO</sub>	0.254	0.327	0.313	0.294	0.317	0.312	0.312	0.295
ARX(28) <sub>LASSO</sub>	0.256	0.327	0.311	0.294	0.318	0.313	0.312	0.295
AR(28)-GARCH-M <sub>LASSO</sub>	0.124	0.201	0.176	0.158	0.181	0.163	0.179	0.158
ARX(28)-GARCHX-M <sub>LASSO</sub>	0.125	0.200	0.173	0.153	0.179	0.166	0.178	0.158

TABELLA 3.5: RMSE dei modelli migliori con le previsioni di ENTSO-E e di ETR per modelli *Fast* e *Less Fast* per il 2019 per alcune ore

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
$EX_4X$	5.082	9.782	7.515	6.888	6.719	5.105	6.605	7.814
$EX_4X - F$	4.629	6.562	6.325	5.681	5.354	4.668	5.264	5.938
$EX_4X - LF$	4.624	6.553	6.329	5.694	5.357	4.673	5.266	5.941
$EX_4X - SGARCHX$	5.090	9.939	7.663	7.074	6.739	5.269	6.745	8.031
$EX_4X - SGARCHX - F$	4.592	6.573	6.273	5.525	5.400	4.645	5.302	5.932
$EX_4X - SGARCHX - LF$	4.580	6.593	6.307	5.597	5.362	4.647	5.303	5.937

TABELLA 3.6: CRPS dei modelli migliori con le previsioni di ENTSO-E e di ETR per modelli *Fast* e *Less Fast* per il 2019 per alcune ore

Modello	3	9	11	13	19	21	Avg <sub>1-24</sub>	Avg <sub>8-20</sub>
$EX_4X$	0.239	0.310	0.301	0.280	0.293	0.295	0.278	0.296
$EX_4X - F$	0.269	0.297	0.295	0.307	0.313	0.301	0.285	0.299
$EX_4X - LF$	0.269	0.297	0.295	0.307	0.313	0.301	0.285	0.299
$EX_4X - SGARCHX$	0.103	0.179	0.164	0.140	0.144	0.134	0.137	0.159
$EX_4X - SGARCHX - F$	0.098	0.128	0.113	0.117	0.141	0.121	0.112	0.125
$EX_4X - SGARCHX - LF$	0.099	0.126	0.112	0.117	0.144	0.122	0.112	0.125

I risultati mostrano chiaramente come l'utilizzo di previsioni professionali migliora sostanzialmente le previsioni sui prezzi dell'energia in particolare nelle ore di picco. Dopo le 21, con l'aumentare dell'orizzonte previsivo, l'uso di previsioni professionali non porta a vantaggi significativi.

Nella Tabella 3.5 ad eccezione dei modelli  $EX_4X$  e  $EX_4X - SGARCHX$  tutti i modelli riportati rientrano nel MCS; i modelli  $EX_4X - F$ ,  $EX_4X - LF$ ,  $EX_4X - SGARCHX - F$  e  $EX_4X - SGARCHX - LF$  risultano migliori rispetto al modello  $EX_4X$  preso come riferimento (dal test unilaterale DM) a un livello di significatività di 0.01%.

Nella Tabella 3.6, per il test unilaterale DM tutti i modelli risultano significativamente migliori del modello  $EX_4X$  preso a riferimento.

I valori riportati nelle tabelle mostrano l'importanza dell'utilizzo di previsioni calcolate professionalmente e evidenziano come considerare le ultime previsioni porti ad un potenziale miglioramento anche nel guadagno. Anche in questo caso, i modelli GARCH non migliorano sostanzialmente le previsioni puntuali ma migliorano le previsioni di densità.

## 3.2 Il modello stimato

Si considerano ora i coefficienti stimati (con i relativi intervalli di confidenza al 90%) del modello  $EX_4X$ , in particolare per le ore 3, 9, 15 e 21.

Il carico di energia previsto risulta significativo, in accordo con quanto evidenziato in letteratura, con un effetto positivo sul prezzo: il prezzo aumenta con l'aumentare del carico previsto (Figura 3.1). Nell'ora 3 si nota un andamento che è crescente fino alla fine del 2018 e decrescente dall'inizio del 2019. Per le ore 9 e 15 l'andamento è diverso: il carico influenza costantemente il prezzo fino al 2018 per poi diminuire progressivamente per tutta la durata del 2019: la diminuzione potrebbe dipendere dallo sviluppo del fotovoltaico che proprio negli ultimi anni ha subito delle considerevoli innovazioni ed è diventato particolarmente efficiente. Al contrario, per l'ora 21, l'andamento è decrescente da luglio 2017, probabilmente ciò è dovuto alla disponibilità di una quantità di energia convenzionale maggiore, sufficiente a coprire la domanda.

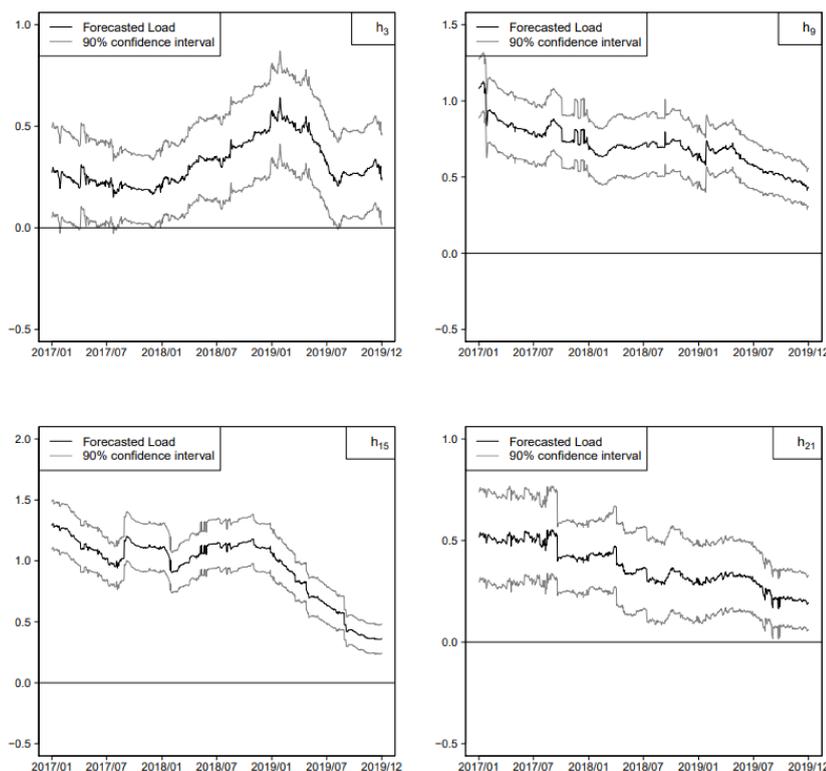


FIGURA 3.1: Coefficienti stimati per il carico previsto con i dati ENTSO-E per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

I coefficienti stimati per l'energia prodotta dal fotovoltaico risultano significativi alle ore 13 e 15 con un effetto negativo sul prezzo dell'energia: all'aumentare dell'energia

prodotta dal fotovoltaico diminuisce il prezzo medio nazionale dell'elettricità. Per l'ora 9 dall'ultima parte del 2017 e per l'ora 11 dalla metà del 2018, il coefficiente passa dall'essere significativo a non significativo. (Figura 3.2)

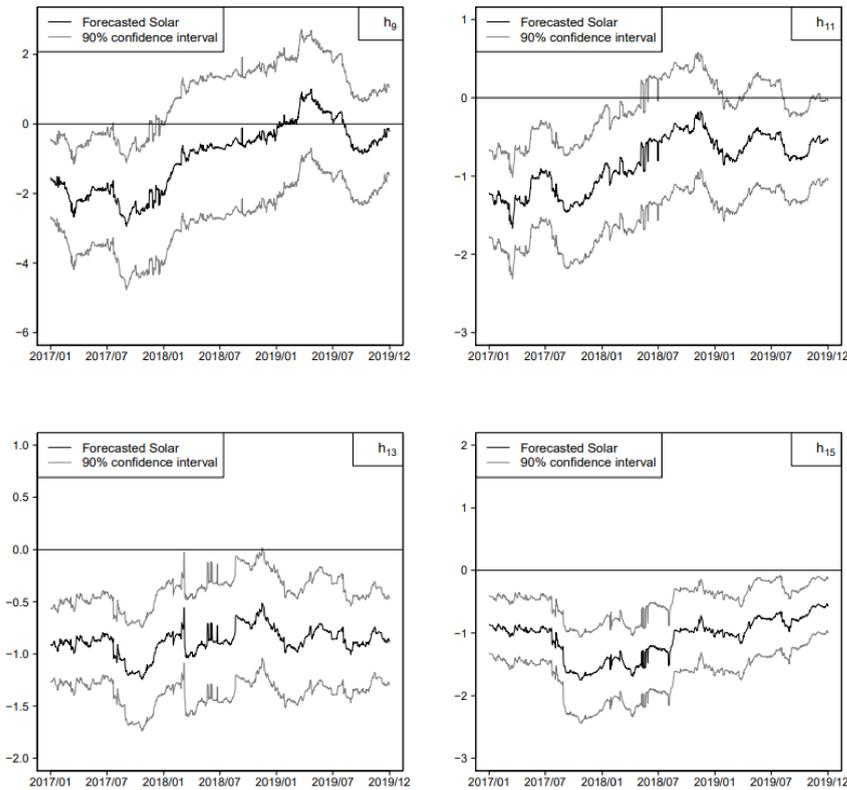


FIGURA 3.2: Coefficienti stimati per l'energia fotovoltaica con i dati ENTSO-E per il modello  $EX_4X$  alle ore 9, 11, 13, 15 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

Nonostante la produzione di energia eolica sia limitata nel nord Italia, i coefficienti stimati per le variabili che rappresentano l'energia eolica risultano significativi, con un effetto negativo sul prezzo. Fino all'ora 3 l'effetto dell'eolico risulta sempre significativo, non è mai significativo all'ora 15 mentre per le ore 9 e 21 risulta significativo fino a circa la metà del 2018 per poi diventare non significativo. (Figura 3.3)

Per quanto riguarda l'energia idroelettrica, (Figura 3.4) contrariamente a quanto ci si aspettasse, i coefficienti non risultano statisticamente significativi alle ore 9 e 21 ma risultano rilevanti (con effetto negativo) all'ora 3, probabilmente in quanto l'idroelettrico non entra in competizione con il fotovoltaico e, sebbene in misura minore, con l'eolico.

Per le importazioni di energia dall'estero, come mostra la Figura 3.5 i coefficienti non risultano significativi, pertanto i prezzi dell'energia all'estero non sembrano avere un'influenza sui prezzi dell'elettricità nel nord Italia. Guardando i grafici in Figura 3.6 e 3.7 si conclude allo stesso modo per le biomasse e i rifiuti.

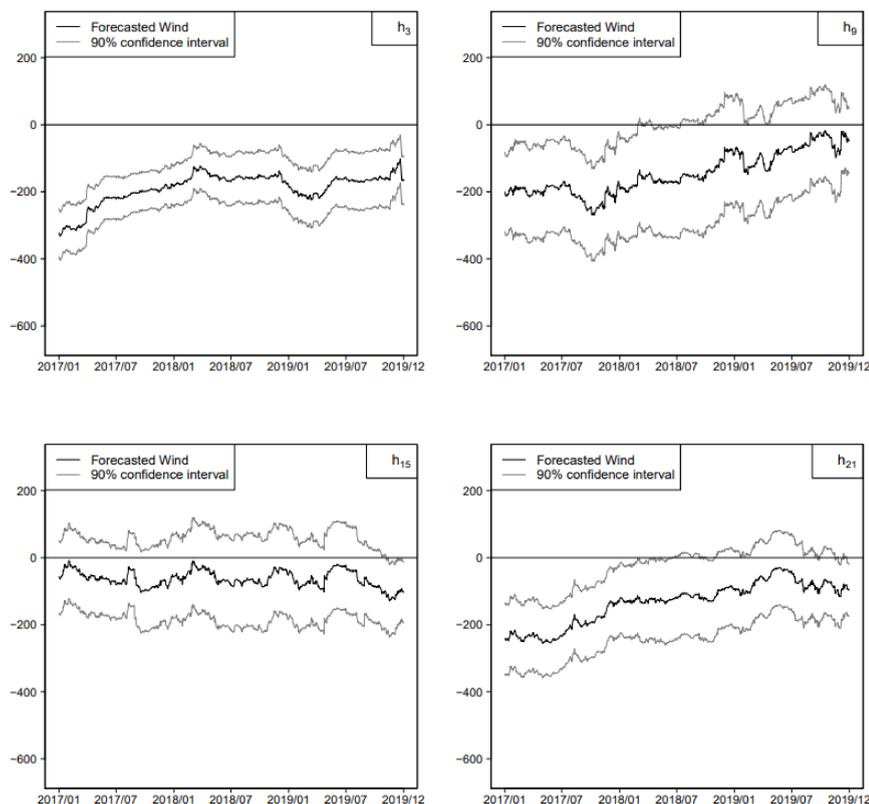


FIGURA 3.3: Coefficienti stimati per l'energia eolica con i dati ENTSO-E per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

Per quanto riguarda il carbone, il suo effetto risulta significativo all'ora 15 fino all'inizio del 2018 mentre, probabilmente a causa delle fonti rinnovabili, diventa via via non significativo (Figura 3.8).

La Figura 3.9 mostra l'andamento dei coefficienti per il gas naturale: i coefficienti risultano significativi all'ora 3 con un effetto positivo sul prezzo dell'elettricità.

Infine i prezzi delle emissioni di anidride carbonica non risultano mai rilevanti, come mostrano i grafici in Figura 3.10.

Bisogna tuttavia tenere conto che il modello  $EX_4X$  include la dipendenza dei prezzi dagli otto giorni precedenti (Paragrafo 2.4.1): ciò influisce sull'andamento dei coefficienti stimati per i regressori.

Il modello, in conclusione, prende in considerazione sia la serie storica passata dei prezzi, sia le informazioni sul contributo del carico previsto di energia, dell'energia solare, eolica e idroelettrica e del gas naturale, risultando un modello completo ed appropriato.

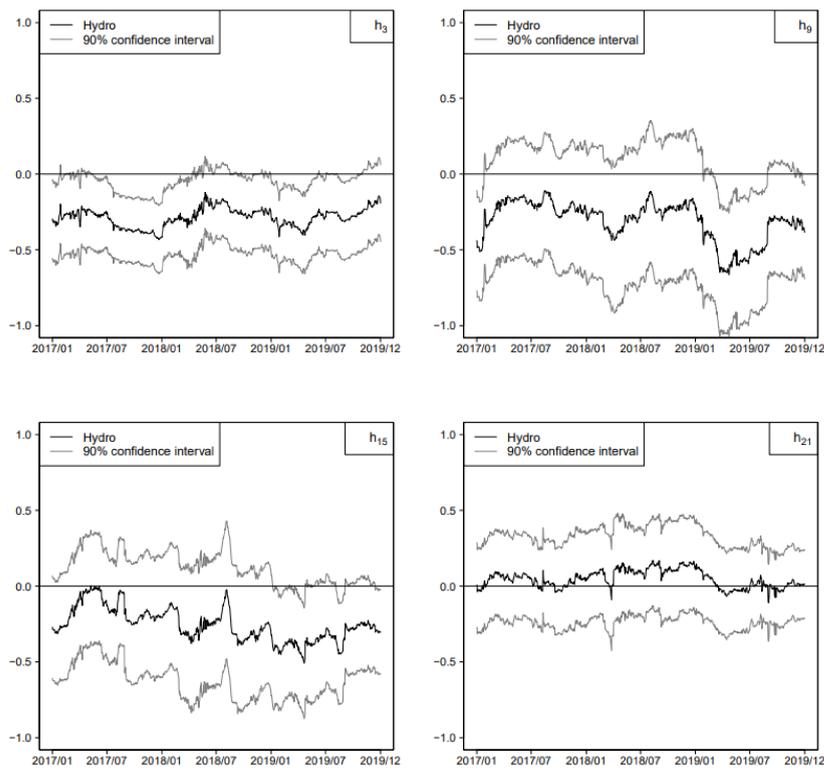


FIGURA 3.4: Coefficienti stimati per l'energia idroelettrica per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

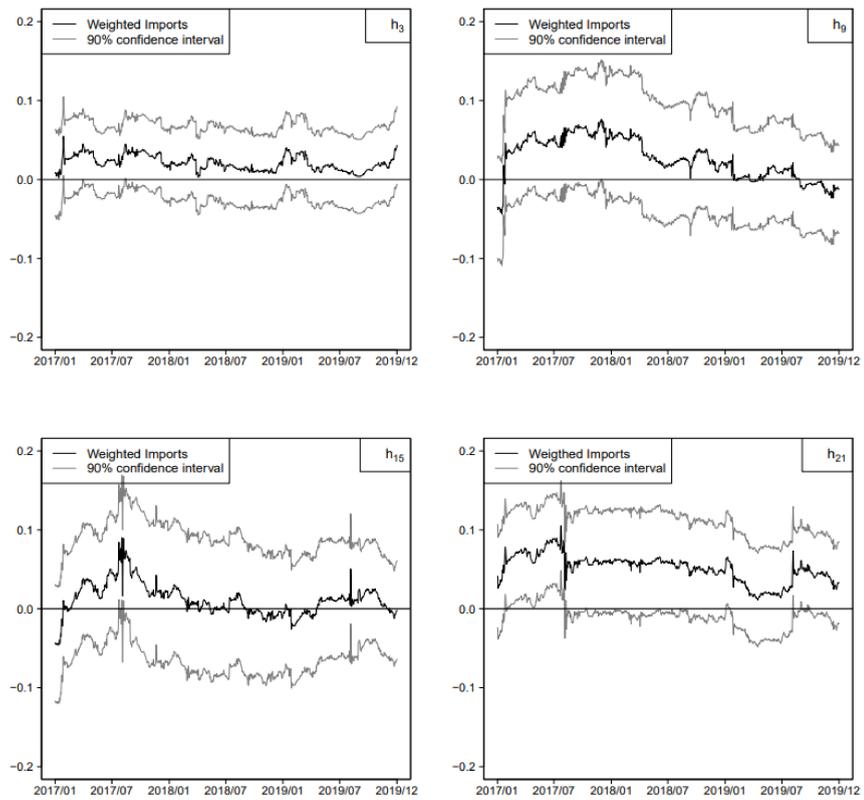


FIGURA 3.5: Coefficienti stimati per le importazioni per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

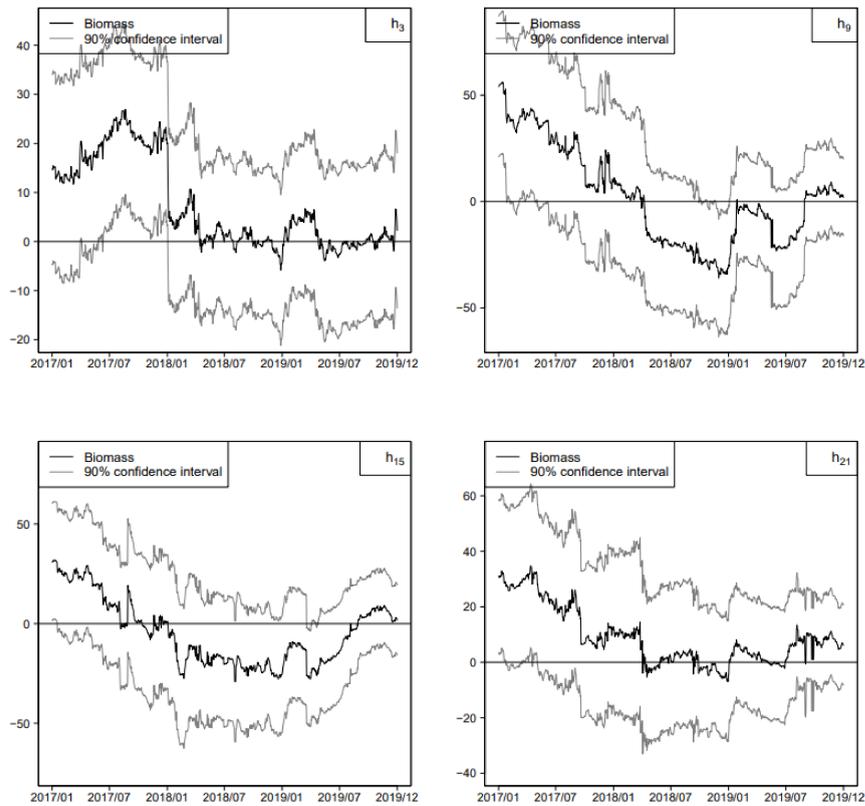


FIGURA 3.6: Coefficienti stimati per le biomasse per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

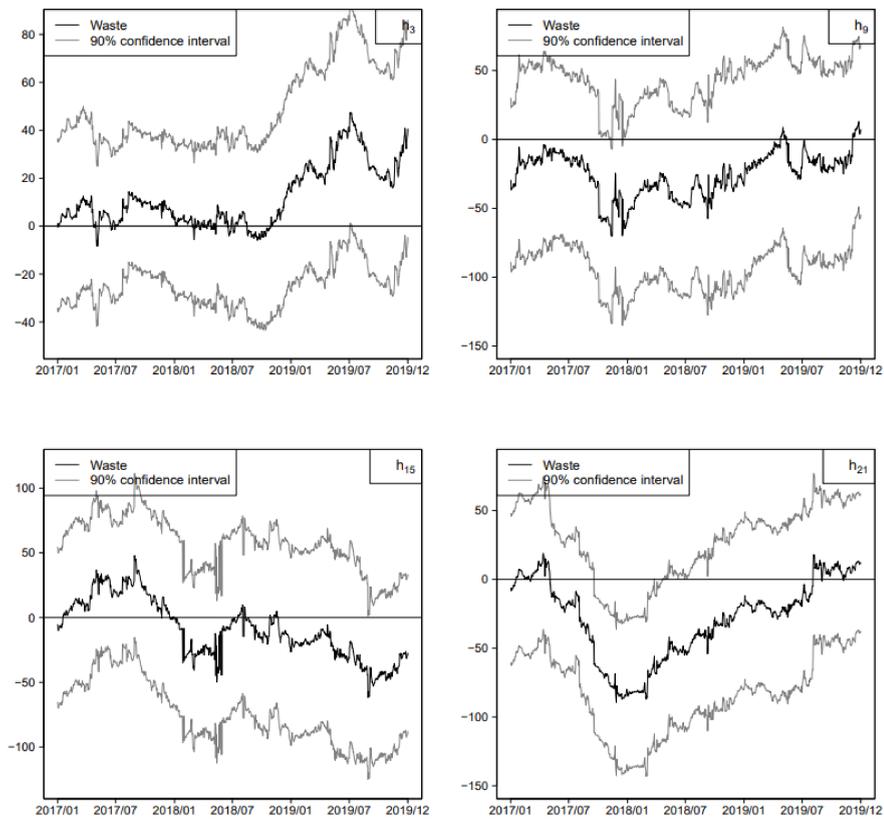


FIGURA 3.7: Coefficienti stimati per i rifiuti per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

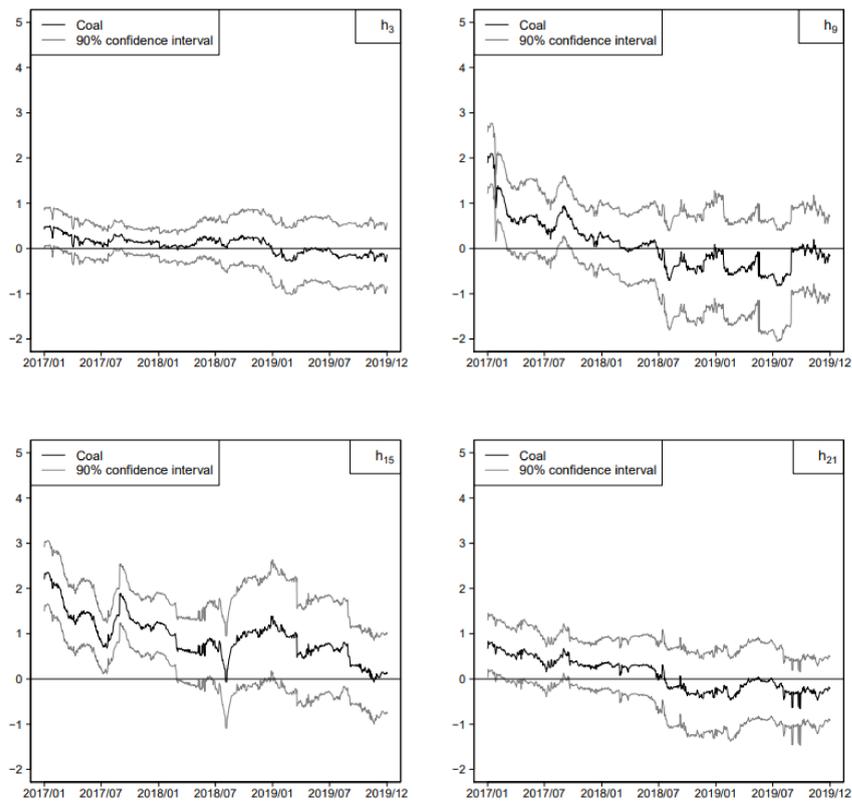


FIGURA 3.8: Coefficienti stimati per il carbone per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

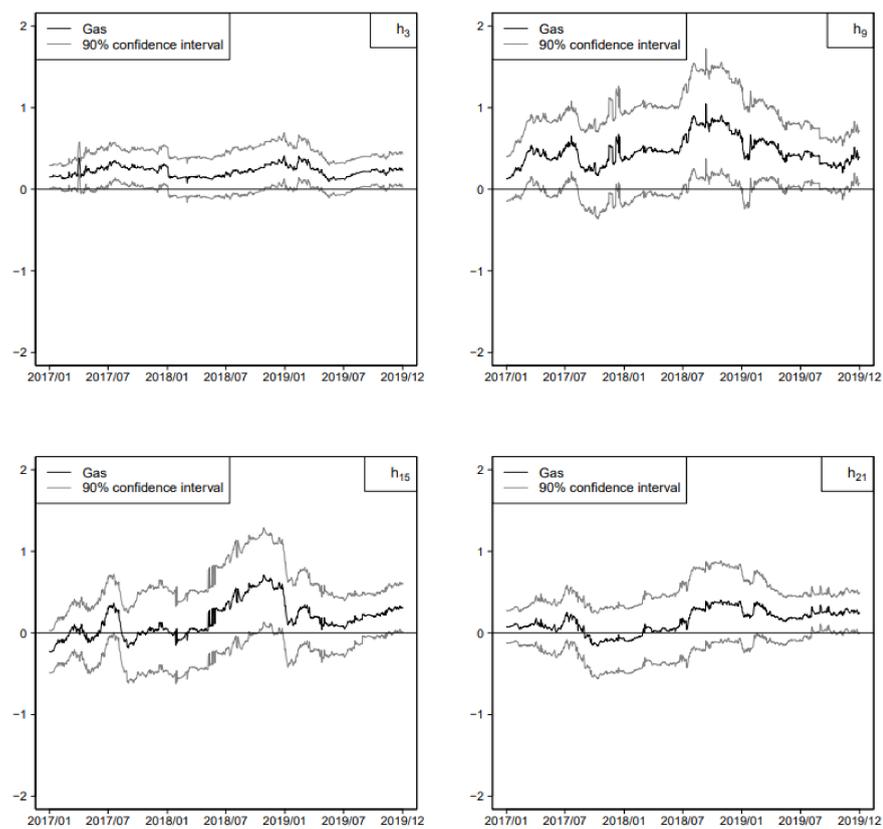


FIGURA 3.9: Coefficienti stimati per il gas naturale per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

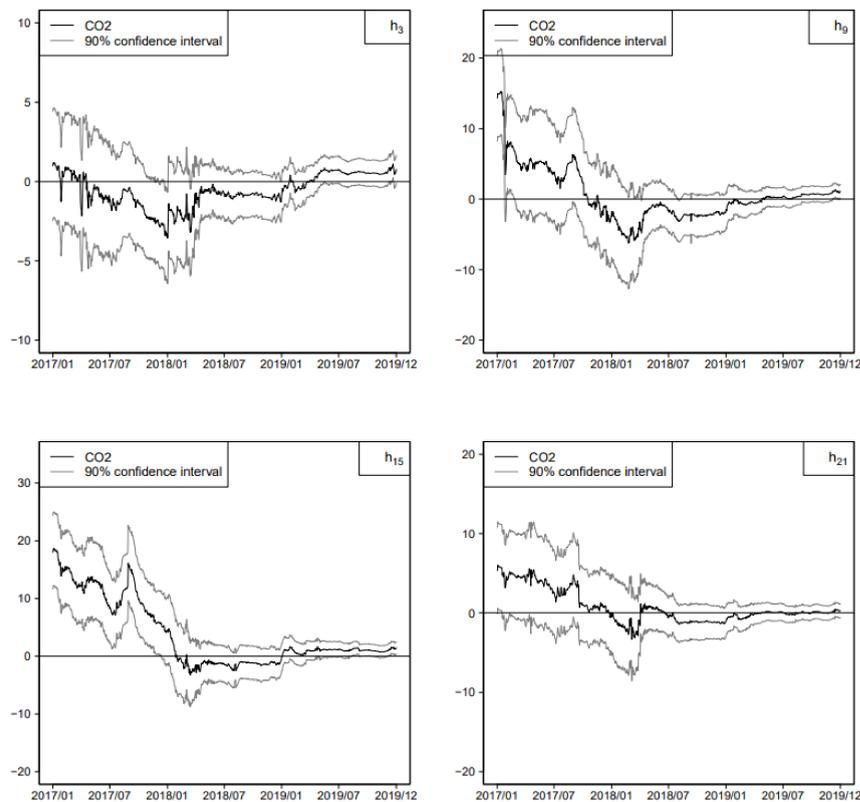


FIGURA 3.10: Coefficienti stimati per i prezzi dell'emissione di  $CO_2$  per il modello  $EX_4X$  alle ore 3, 9, 15, 21 con rispettivi intervalli di confidenza robusti al 90% riportati per il periodo fuori campione dal 01/01/2017 al 31/12/2019

# Capitolo 4

## Conclusioni

La previsione dei prezzi dell'elettricità è diventata una questione molto importante per quanto riguarda la pianificazione della produzione di energia, che è influenzata anche, come mostrato in precedenza, dalle condizioni meteorologiche che influiscono sulla quantità effettiva di energia prodotta dalle RES, che, come si può immaginare, ha un effetto sull'equilibrio tra domanda e offerta.

Nella trattazione, sono stati analizzati e confrontati diversi modelli di previsione: dai modelli combinati a modelli più complessi come gli ARFIMA con le specifiche dei modelli GARCH per il termine di varianza o ancora dei modelli ARIMA con parametri variabili e selezionati iterativamente durante il processo di stima.

In conclusione come visto nel Paragrafo 3.2 il modello che ha mostrato migliori capacità previsionive è il modello  $EX_4X$ , un modello combinato con la presenza delle variabili esogene. Il modello è risultato più efficiente in termini di previsione puntuale mentre, per la previsione di densità, il modello con le specifiche GARCH è risultato migliore. Le variabili esogene che sono risultate significative per il modello di previsione dei prezzi dell'energia nel nord Italia sono: la domanda di energia prevista, l'energia proveniente dai settori fotovoltaico, eolico e idroelettrico e anche il prezzo del gas naturale. Le altre variabili esogene, sebbene non siano risultate rilevanti nell'analisi, vengono comunque considerate affinché il modello possa essere esteso ad altre zone oltre che al nord Italia.

In conclusione:

- Le previsioni migliorano in presenza dei regressori;
- Per le previsioni puntuali è sufficiente considerare il modello  $EX_4X$  mentre per le previsioni di densità è necessario considerare le specifiche del modello GARCH;

- Si nota che l'utilizzo di dati professionali migliora consistentemente l'accuratezza delle previsioni del modello rispetto all' utilizzo di dati disponibili pubblicamente ai ricercatori.

Alla luce di ciò che è stato evidenziato nell'analisi, una buona previsione dei prezzi dell'energia elettrica attualmente sta diventando sempre più importante, non solo per la previsione della quantità di energia da produrre sulla base dell'energia disponibile e la domanda, ma anche per contrastare le speculazioni di mercato e l'aumento dei costi dell'elettricità per il cliente finale.

# Bibliografia

- ABRAMOVA, E. & BUNN, D. (2020). Forecasting the intra-day spread densities of electricity prices. *Energies* **13**, 687.
- BÉLISLE, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on d. *Journal of Applied Probability* **29**, 885–895.
- BILLÉ, A. G., GIANFREDA, A., DEL GROSSO, F. & RAVAZZOLO, F. (2022). Forecasting electricity prices with expert, linear, and nonlinear models. *International Journal of Forecasting* .
- BOSCO, B. P., PARISIO, L. P. & PELAGATTI, M. M. (2007). Deregulated wholesale electricity prices in italy: an empirical analysis. *International Advances in Economic Research* **13**, 415–432.
- BUNN, D. W., GIANFREDA, A. & KERMER, S. (2018). A trading-based evaluation of density forecasts in a real-time electricity market. *Energies* **11**, 2658.
- DICKEY, D. A. & FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association* **74**, 427–431.
- DIEBOLD, F. X. & MARIANO, R. S. (1995). Vcomparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, 253.
- ECMWF (1975). European centre for medium range weather forecasts. <https://www.ecmwf.int/en/about>.
- ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society* , 987–1007.
- ENTSO-E (2008). European network of transmission system operators. <https://www.entsoe.eu/about/>.

- GIANFREDA, A. & GROSSI, L. (2012). Forecasting italian electricity zonal prices with exogenous variables. *Energy Economics* **34**, 2228–2239.
- GIANFREDA, A., PARISIO, L. & PELAGATTI, M. (2016). The impact of res in the italian day-ahead and balancing markets. *The Energy Journal* **37**.
- GIANFREDA, A., PARISIO, L. & PELAGATTI, M. (2019). The res-induced switching effect across fossil fuels: An analysis of day-ahead and balancing prices. *The Energy Journal* **40**.
- GIANFREDA, A., RAVAZZOLO, F. & ROSSINI, L. (2020). Comparing the forecasting performances of linear models for electricity prices with high res penetration. *International Journal of Forecasting* **36**, 974–986.
- GIANFREDA, A. & SCANDOLO, G. (2018). Measuring model risk in the european energy exchange. *Handbook of Recent Advances in Commodity and Financial Modeling: Quantitative Methods in Banking, Finance, Insurance, Energy and Commodity Markets* , 89–110.
- GME (2009). Gestore dei mercati energetici. <https://www.mercatoelettrico.org/it/>.
- GNEITING, T. & RANJAN, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics* **29**, 411–422.
- GROEN, J. J., PAAP, R. & RAVAZZOLO, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics* **31**, 29–44.
- HANSEN, P. R., LUNDE, A. & NASON, J. M. (2011). The model confidence set. *Econometrica* **79**, 453–497.
- HERSBACH, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**, 559–570.
- IEA (2023). International energy agency. <https://www.iea.org/about>.
- JARQUE, C. M. & BERA, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique* , 163–172.

- KARAKATSANI, N. V. & BUNN, D. W. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting* **24**, 764–785.
- KWIATKOWSKI, D., PHILLIPS, P. C., SCHMIDT, P. & SHIN, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics* **54**, 159–178.
- KYRITSIS, E., ANDERSSON, J. & SERLETIS, A. (2017). Electricity prices, large-scale renewable integration, and policy implications. *Energy Policy* **101**, 550–560.
- LISI, F. & EDOLI, E. (2018). Analyzing and forecasting zonal imbalance signs in the italian electricity market. *The Energy Journal* **39**.
- Nelson (2021). Johnson. <http://github.com/stevengj/nlopt>.
- NELSON, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society* , 347–370.
- RTR (2018). Refinitiv thomson reuters. <https://www.thomsonreuters.com/en.html>.
- SHAH, I. & LISI, F. (2020). Forecasting of electricity price through a functional prediction of sale and purchase curves. *Journal of Forecasting* **39**, 242–259.
- Terna (2018). Statistical data on electricity in italy: Technical report. <https://www.terna.it/en/electric-system/statistical-data-forecast/statistical-publications>.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288.
- WEST, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society* , 1067–1084.
- ZAKOIAN, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control* **18**, 931–955.
- ZAMO, M. & NAVEAU, P. (2018). Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences* **50**, 209–234.
- ZIEL, F., STEINERT, R. & HUSMANN, S. (2015). Efficient modeling and forecasting of electricity spot prices. *Energy Economics* **47**, 98–111.

