

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE
**ANALISI STATISTICA DEL MERCATO ITALIANO
DELL'ELETTRICITÀ: MODELLI E PREVISIONI**

Relatore Prof. Luigi Grossi
Dipartimento di Scienze Statistiche

Laureando: Renato Tommasi
Matricola N° 1187257

Anno Accademico 2022/2023

Indice

Introduzione	2
1 Il Sistema Elettrico Nazionale e il Mercato Italiano dell'Energia Elettrica	5
1.1 Il sistema elettrico: caratteri generali e configurazione	5
1.1.1 Impianti di produzione	5
1.1.2 Trasmissione	9
1.1.3 Distribuzione	10
1.1.4 Carico (Load)	11
1.2 Breve storia del mercato elettrico in Italia e configurazione odierna	12
1.2.1 Il Mercato del Giorno Prima	14
1.2.2 Il Prezzo Unico Nazionale - PUN	16
2 Metodologia di Analisi	19
2.1 Modelli statistici per serie storiche di dati economici	20
2.2 Approccio univariato e multivariato	22
2.2.1 Modelli univariati selezionati	24
2.2.2 Modelli multivariati selezionati	28
2.3 Tecniche comparative adottate	31
3 Implementazione e Risultati Empirici	34

3.1	Data preprocessing	35
3.2	Modello naive	37
3.3	Modelli AR univariato e multivariato	38
3.3.1	Modello AR univariato	39
3.3.2	Modello AR multivariato	40
3.4	Modelli LASSO univariati e multivariati	41
3.4.1	LASSO univariati	42
3.4.2	LASSO multivariati	43
3.5	Confronto performance predittive	44
	Conclusioni	47
	Appendice	48
	Bibliografia	66

Introduzione

I mercati dell'energia elettrica oggi e ancor più in futuro saranno in primo piano nello scenario geopolitico mondiale. Le politiche energetiche dei paesi sono legate a doppio filo non solo con il problema del cambiamento climatico, ma anche con gli equilibri geopolitici tra le grandi potenze mondiali.

In questo contesto tumultuoso si inserisce una questione che non è solo di interesse puramente statistico, ma è anche uno strumento di fondamentale importanza per la stabilità del sistema e del mercato elettrico: la modellazione statistica delle fluttuazioni dei prezzi dell'energia elettrica con il fine di ottenere delle accurate previsioni short-term, in letteratura chiamata *short-term EPF (Electricity Price Forecasting)*.

In questa tesi si tenterà di aggiungere informazioni fondamentali per rispondere ad una domanda a cui non è ancora stata data una risposta definitiva, ovvero: qual è l'approccio migliore per modellare ed effettuare previsioni dei prezzi del mercato dell'energia elettrica? In particolare, è più adatto un approccio di tipo univariato, oppure multivariato? La tesi si basa sul lavoro di Ziel e Weron (2018), i quali hanno provato a dare una risposta a questa domanda, ma non sono giunti a conclusioni definitive, avendo ottenuto risposte contrastanti nei diversi dataset utilizzati per l'analisi. Tra le 12 serie storiche di prezzi relative a 12 diversi mercati dell'energia elettrica non figura però quella del mercato italiano.

L'obiettivo di questa tesi è quello di colmare questa lacuna e fornire dei risultati relativi al mercato italiano, utili, si spera, al miglioramento delle

tecniche di forecasting nell'ambito dei mercati elettrici.

Il **primo capitolo** contiene un'introduzione al sistema elettrico italiano, la sua configurazione odierna e le sue caratteristiche tecniche. Segue un breve riassunto della storia del mercato dell'energia elettrica in Italia e una spiegazione del meccanismo d'asta dal quale si ottiene il Prezzo Unico Nazionale (PUN) nel mercato del giorno prima (MGP). Il PUN è infatti la variabile di interesse che viene modellata e sulla quale si effettuano le previsioni.

Nel **secondo capitolo** si motivano le scelte dei modelli, dandone una giustificazione teorica. Si introducono i concetti relativi ai modelli naive, autoregressivi e autoregressivi con metodo di stima LASSO e si chiude con un'illustrazione dei metodi di valutazione e di comparazione selezionati.

Nel **terzo capitolo**, infine, vengono presentati e commentati i risultati ottenuti. In appendice, vengono illustrati i codici R utilizzati nella parte empirica.

Capitolo 1

Il Sistema Elettrico Nazionale e il Mercato Italiano dell'Energia Elettrica

1.1 Il sistema elettrico: caratteri generali e configurazione

Con “sistema elettrico” si intende l'insieme di tutti i componenti connessi ad un circuito elettrico. Naturalmente, quando si parla di sistema elettrico nazionale ci si riferisce ad un enorme circuito, o meglio, un'intricata ed estesa rete di circuiti interconnessi. In particolare, i principali elementi del sistema elettrico sono: gli impianti di produzione, le linee di trasmissione, le reti di distribuzione e infine il carico (load).

1.1.1 Impianti di produzione

Gli impianti di produzione sono le centrali elettriche, impianti industriali che producono energia elettrica mediante la trasformazione di una fonte di energia primaria. L'elettricità è infatti una fonte secondaria, in quanto non

si reperisce direttamente in natura ma va generata. La macchina dedicata alla conversione dell'energia è solitamente un generatore elettrico costituito da un alternatore mosso da una turbina (ad eccezione degli impianti fotovoltaici). L'altra componente fondamentale di una centrale è il trasformatore, cruciale per la trasmissione in alta o altissima tensione (AT o AAT). Le centrali elettriche vengono suddivise in base alla fonte energetica primaria che trasformano (o anche in base alla tecnologia che usano per farlo). Di seguito elencate le principali tipologie.

Centrali termoelettriche

Trasformano in energia elettrica l'energia chimica di una fonte primaria. La fonte primaria utilizzata è chiamata combustibile perché è tramite la combustione di quest'ultimo che si ottiene l'energia termica utile a far muovere la turbina che tramite l'alternatore genera energia elettrica. I combustibili possono essere carbon fossile, petrolio, gas naturale o i meno convenzionali biogas e biomassa (tuttavia la biomassa rappresenta per alcuni paesi una grossa porzione dell'approvvigionamento energetico ancora oggi, si veda ad esempio l'India¹). Sono caratterizzate da un basso costo di produzione da una buona agilità (generalmente possono essere spente e riaccese molto rapidamente) ma hanno il grosso difetto di emettere enormi quantità di gas-serra e di essere nella maggior parte dei casi molto inquinanti (oltre che spesso causa di gravissimi incidenti).

Centrali idroelettriche

Sono impianti che trasformano in energia elettrica l'energia cinetica di grandi masse d'acqua in movimento che aziona turbine e alternatori. Spesso comprendono la costruzione di una diga, opera ingegneristica colossale che permette di concentrare enormi masse d'acqua in prossimità della centrale,

¹<https://www.iea.org/reports/India>

conferendo quindi estrema agilità e rapidità di accensione e spegnimento, attributi utilissimi per la copertura di picchi improvvisi di domanda che possono verificarsi. La produzione è a emissioni zero e il fatto che siano uno dei pochissimi metodi di conservare (virtualmente) grandi quantità di energia (pompano l'acqua all'indietro e riportandola nei bacini di accumulo usando i surplus di energia elettrica prodotti da altre fonti non-programmabili come ad esempio le rinnovabili) le rende un *must-have*, con benefici che superano di gran lunga i dubbi che si possono avere sul loro impatto ambientale e sui (rarissimi ma devastanti) incidenti, come quello del Vajont del 9 ottobre 1963 o del bacino di Banqiao in Cina dell'agosto 1975 che fu causa di più di 85'000 vittime².

Centrali nucleari

In questi impianti l'energia termica che aziona le turbine è generata dalla fissione nucleare dell'uranio-235 e del plutonio-239 (contenuti in un combustibile composto principalmente da uranio-238) che avviene nei reattori nucleari. È una tecnologia di produzione molto poco agile: i tempi di accensione e spegnimento sono molto lunghi. Le centrali nucleari vengono infatti utilizzate principalmente per coprire la domanda fissa (base-load) e svolgono egregiamente questo compito dato che godono di grande stabilità di produzione. In Italia e in pochi altri paesi sono state dismesse in seguito agli incidenti di Chernobyl del 1986 e Fukushima del 2011 (che però ha curiosamente causato una singola vittima³, anche se non è ancora chiaro se anche quella sarebbe da asserire all'incidente della centrale o al terrificante terremoto che ha scatenato l'incidente stesso), tuttavia hanno il grande vantaggio di garantire una produzione sostanzialmente a emissioni zero e priva di impatto ambientale,

²Il numero di vittime citato è stato reperito al sito <http://www.sjsu.edu> tuttavia esiste un rapporto ufficiale del dipartimento idrologico della provincia di Henan, desecretato solo nel 2005, che fa salire il numero di vittime a 171'000

³<https://www.unscear.org>

tant'è che sembra sia una tecnologia imprescindibile in un futuro sostenibile del settore elettrico mondiale⁴.

Centrali solari

Centrali solari: ce ne sono di due tipi, fotovoltaiche e termodinamiche. Le prime sono impianti composti da moduli fotovoltaici che convertono direttamente la luce solare in energia elettrica sfruttando l'effetto fotovoltaico, le seconde invece sfruttano il calore generato dalla radiazione solare per scaldare un mezzo che permette di azionare delle turbine. Ci sono due approcci: o tramite pannelli solari, che al loro interno contengono il fluido da scaldare che viene poi fatto fluire fino alla turbina, oppure tramite un complesso sistema di specchi che convoglia la radiazione solare in un punto in cui è presente il fluido da scaldare che evaporando attiva le turbine. Queste tecnologie oltre ad essere puramente rinnovabili e quindi parte fondamentale di praticamente ogni scenario futuro sostenibile proposto (di particolare interesse quelli proposti dal *Joint Research Centre - JRC*), sono sicuramente tra le meno inquinanti anche se il loro impatto ambientale è ancora dibattuto (richiedono infatti larghissime superfici di terreno e alcuni dei materiali che le compongono, le terre rare nello specifico, si portano dietro non poche questioni etiche ed ecologiche tutt'ora aperte⁵) e le loro caratteristiche le rendono complicate da gestire dato che la loro produzione è molto suscettibile alle condizioni atmosferiche: vengono infatti definite non-programmabili.

Centrali eoliche

In queste centrali è l'energia cinetica del vento a mettere in azione turbine e alternatori per generare energia elettrica. Questa tecnologia permette una produzione molto pulita e rinnovabile (anch'essa quasi sempre presente negli

⁴<https://publications.jrc.ec.europa.eu>

⁵<https://hbr.org>

scenari futuri sostenibili proposti da svariati enti di ricerca tra cui il JRC⁶) ma si porta dietro gli stessi difetti del solare: produzione non-programmabile data l'alta suscettibilità alle condizioni climatiche e dubbio impatto ambientale (richiedono aree vastissime e non è ancora chiaro quanto il loro impatto sulla fauna sia sostenibile⁷).

In Italia in particolare, nel 2021 la produzione era suddivisa come segue: 65,6% termoelettrico, 16,4% idroelettrico, 8,7% fotovoltaico, 7,2% fotovoltaico e il restante 2% circa da altre fonti⁸. Per quanto riguarda il fuel-mix del 2021, cioè le percentuali delle diverse fonti primarie utilizzate sul totale di energia prodotta è il seguente: 48,1% gas naturale, 42,3% fonti rinnovabili, 5,1% carbone, 0,9% prodotti petroliferi e 3,6% altre fonti⁹.

1.1.2 Trasmissione

Le linee di trasmissione sono lunghi cavi in alta e altissima tensione (AT dai 36kV ai 150kV e AAT dai 220kV ai 380kV) che trasportano l'energia elettrica dalle centrali alle reti di distribuzione. La rete di trasmissione è composta principalmente da linee elettriche aeree in alluminio, ma non mancano anche sistemi più complessi come i cavi interrati o addirittura lunghissimi cavi sottomarini. Fondamentali sono, inoltre, le stazioni elettriche, ovverosia impianti dove fisicamente si interconnettono le diverse linee di trasmissione e servono non solo per la trasformazione nei diversi livelli di tensione (voltaggio) ma anche per eseguire le manovre di apertura, chiusura e sezionamento delle linee, processi fondamentali per la gestione dei flussi di energia e per evitare disservizi. È infatti fondamentale mantenere un perfetto bilancio tra la quantità di energia consumata e prodotta: i principi della fisica stessi im-

⁶<https://www.cleanenergyministerial.org>

⁷Rapporto IPCC, 2011

⁸<https://www.terna.it/pubblicazioni-statistiche>

⁹<https://www.gse.it>

pongono che in un circuito non ci siano mai sbilanci tra le due, pena gravi incidenti o comunque disservizi su larga scala. Ogni rete di trasmissione sincronizzata deve quindi necessariamente avere un ente di controllo, chiamato *System Operator* (SO) che si occupi di provvedere alla gestione dei flussi per mantenere il perfetto equilibrio della rete e che si occupi anche non solo della sua manutenzione ma anche della sua espansione e del suo perfezionamento. In Europa, i System Operators delle diverse reti nazionali-internazionali sono raggruppati nel *European Network of Transmission System Operators for Electricity* (ENTSO-E). In Italia, il Transmission System Operator è Terna SpA.

1.1.3 Distribuzione

La rete di distribuzione è una fitta e intricata serie di cavi elettrici a media tensione (tra gli 1 e i 35 kV) e a bassa tensione (tra i 50 e i 1000 V) che portano l'energia elettrica ai diversi utenti finali. È collegata alla rete di trasmissione nelle stazioni di trasformazione e nelle cabine primarie, che abbassano gradualmente il voltaggio e infine, nelle cabine secondarie, lo portano al livello necessario all'utenza (in Europa 220-240V). La rete in media tensione è chiamata distribuzione primaria mentre quella in bassa tensione è chiamata distribuzione secondaria. Una prima differenza sostanziale con la rete di trasmissione è che spesso si vengono a creare svariate reti di distribuzione sconnesse tra di loro: l'idea è cercare di portare le cabine secondarie il più vicino possibile al consumatore finale (per evitare perdite di energia dovute a linee troppo lunghe) ma anche cercare di raggruppare più utenze possibili per collegarle alla stessa cabina (costruendone quindi meno possibili, dato che sono molto costose e impedirebbero altrimenti economie di scala). Una seconda e importante differenza è quella legata alla frequenza e all'importanza dei disservizi: essendo spesso sconnesse tra loro, la gravità degli incidenti è spesso molto limitata e contenuta, d'altro canto però, essendoci molte più connessioni, il tasso di incidenti o disservizi è più alto. L'attività

di gestione, manutenzione e miglioramento della rete di distribuzione è compito dei *Distribution System Operators* (DSOs). In Europa si contano più di 2500 DSOs; in Italia nel 2021 erano 123, secondo quanto riporta l'Anagrafica dell'ARERA relativamente al numero degli iscritti al 2021 ¹⁰.

La distribuzione dell'elettricità è anche il segmento della catena del valore dove ci si aspetta che avvenga il maggior contributo all'innovazione energetica a basse emissioni di CO_2 . Lo rileva il report del JRC "*Distribution System Operator Observatory 2020*"¹¹ dedicato al ruolo dei DSOs, che si è evoluto con l'obiettivo di un sistema elettrico più sostenibile, rendendo queste società e le infrastrutture di rete che gestiscono uno strumento chiave per raggiungere l'impegno dell'UE verso la neutralità climatica entro il 2050.

La società di distribuzione di maggior importanza in Italia è E-Distribuzione, che distribuisce energia elettrica a oltre il 90% dei consumatori sul suolo nazionale. Si tratta di un'azienda del marchio Enel: inizialmente chiamata Enel-Distribuzione, ha cambiato nome in E-Distribuzione solo nel 2016.

1.1.4 Carico (Load)

Il carico, tecnicamente parlando, sarebbe un qualsiasi punto di un circuito in cui la corrente viene utilizzata per compiere un'azione (in altre parole, un punto in cui vi è un prelievo di energia dal sistema), tuttavia è pratica comune utilizzare questo termine per riferirsi alla quantità di energia prelevata dal sistema per essere utilizzata. Va specificato quindi che con "carico" ci si riferisce spesso al consumo totale di energia in una data circoscrizione spazio-temporale. In Europa, il *load* per il settore domestico è suddiviso nelle seguenti fasce di consumo annuo:

1. *Very small*: consumo annuo inferiore a 1000 kWh;
2. *Small*: consumo annuo compreso tra 1000 e 2500 kWh;

¹⁰<https://www.arera.it>

¹¹<https://ses.jrc.ec.europa.eu/distribution-system-operators-observatory>

3. *Medium*: consumo annuo compreso tra 2500 e 5000 kWh;
4. *Large*: consumo annuo compreso tra 5000 e 15'000 kWh;
5. *Very Large*: consumo annuo superiore a 15'000 kWh.

La classificazione per il settore industriale invece ha cinque fasce che vanno complessivamente dai 20 ai più di 150'000 MWh di consumi annui.

Ma con che tipo di accordi economici avviene lo scambio di energia elettrica tra le diverse parti del sistema elettrico?

1.2 Breve storia del mercato elettrico in Italia e configurazione odierna

Nel 1883 a Milano in via Santa Radegonda a due passi dal Duomo, in uno stabile precedentemente usato come teatro, veniva costruito il primo impianto per la produzione di energia elettrica in Europa continentale, il secondo al Mondo¹². Questo impianto termoelettrico (oggi non più esistente) era costruito e gestito da Edison. Per i primi anni dello sviluppo di questo tipo di tecnologia, la configurazione normativa ed economica fu lasciata relativamente libera: le prime società erano nate infatti da iniziative di imprenditori e/o da consorzi, talvolta, di origine pubblica. Come iniziative private si possono citare sicuramente: SADE (Società Idroelettrica di Elettricità) fondata nel 1905 a Venezia e CEdiS (Consorzio Elettrico di Storo) fondato a Storo (TN) nel 1903. Il 6 dicembre 1962 entrò in vigore la legge che istituì ENEL, acronimo di Ente *Nazionale Energia Elettrica*. Questo ente aveva lo scopo di nazionalizzare tutti gli impianti di produzione, trasmissione e distribuzione dell'energia elettrica in Italia e di "portare l'elettricità in tutte le case degli italiani". Tutte le attività vengono quindi sostanzialmente portate in capo allo stato e la situazione non cambierà di molto per circa 30 anni. Il

¹²<http://www.storiadimilano.it/citta/milanotecnica/elettricit/radegonda0.htm>

16 marzo 1999 viene infine emanato il Decreto Legislativo n.79 (noto come *Decreto Bersani*), come attuazione della direttiva 96/92/CE recante norme comuni per il mercato interno dell'energia elettrica. Questo decreto pone le basi per la riforma del mercato elettrico in Italia e determina la suddivisione dell'Enel in varie società distinte in base alle attività svolte, in particolare: produzione, trasmissione, distribuzione e vendita.

Per quanto riguarda la trasmissione, il processo di liberalizzazione poneva l'esigenza di un System Operator indipendente dall'influenza dei diversi operatori del settore e dai principali attori del mercato dell'energia elettrica. Il decreto legislativo individuò la soluzione nella costituzione del Gestore della Rete di Trasmissione Nazionale (GRTN), società per azioni cui proprietà veniva conferita al ministero dell'Economia. Tuttavia, il 1° Novembre 2005 le attività del GTRN sono state trasferite a Terna S.p.A. e il GRTN ha cambiato nome in GSE: *Gestore dei Servizi Energetici*, che ha oggi come compito principale quello di incentivare e promuovere la produzione di energia elettrica da fonti rinnovabili¹³.

L'aspetto più importante del Decreto Legislativo n.79 del 1999 è però la creazione del mercato elettrico italiano (*Italian Power Exchange - IPEX*), la sede delle transazioni aventi oggetto la compravendita all'ingrosso di energia elettrica, entrato ufficialmente in funzione (aprendo alla possibilità di partecipare alla domanda) il 1° Gennaio 2005. Secondo il Decreto, la competizione in questo mercato doveva fondarsi sull'*obiettività*, sulla *neutralità* e sulla *trasparenza*. Con l'obiettivo di garantire queste esigenze, vennero istituiti il *Gestore dei Mercati Energetici* (GME) e, come organo di controllo del precedente, il *Gestore dei Servizi Energetici* (GSE).

Entrando nel dettaglio della conformazione del mercato elettrico italiano, si trova la prima distinzione, quella tra il Mercato a Pronti (MPE, in inglese *Spot Electricity Market*) e il Mercato a Termine (MTE, in inglese *Forward Electricity Market*).

¹³<https://www.sorgenia.it/guida-energia/grtn>

Il Mercato a Termine è la sede delle negoziazioni di contratti a termine con obbligo di consegna e ritiro riguardanti lo scambio di energia elettrica. Le negoziazioni si svolgono continuamente: gli operatori presentano le loro proposte indicando la tipologia dei contratti, il periodo di consegna e il prezzo a cui sono disposti ad acquistare e/o vendere. Al mercato a termine possono partecipare tutti gli operatori del mercato elettrico con il GME che agisce come controparte centrale.

Il Mercato a pronti (MPE) si riferisce invece alla borsa elettrica e ogni operatore ha come controparte il mercato stesso. È suddiviso in quattro sezioni: Mercato del Giorno Prima (MGP), Mercato Infragiornaliero (MI), Mercato dei Prodotti Giornalieri (MPEG) e Mercato per i Servizi di Dispacciamento (MSD).

Il motivo per cui esistono queste diverse sezioni risiede proprio nella natura del bene che viene scambiato: l'energia elettrica. Il problema principale è che l'energia elettrica è un bene che non può essere conservato in grandi quantità sistematicamente ma soprattutto non si può decidere quanto ne viene "preso" e siccome il sistema elettrico va sempre mantenuto in equilibrio, pena gravi disservizi e ingenti danni economici, la quantità di energia in più che viene estratta deve anche essere immediatamente immessa da qualcun altro e se ne dovesse venire estratta meno di quanto accordato, qualcuno dovrà per forza produrne di meno.

1.2.1 Il Mercato del Giorno Prima

Nel Mercato del Giorno Prima (MGP), che comprende la maggior parte delle transazioni di compravendita di energia elettrica, viene fissata una programmazione di produzione e consumi per le 24 ore del giorno successivo (da cui il nome) e tutti gli operatori elettrici della filiera di generazione e trading all'ingrosso di energia elettrica possono partecipare su base facoltativa. Il mercato si svolge in un'unica sessione in asta implicita per offerte di acquisto

e vendita relative al giorno successivo. Apre alle 8:00 di nove giorni prima del giorno di consegna e si chiude alle 9:00 del giorno precedente la consegna.

Domanda e Offerta

Generalmente l'offerta è rappresentata dai produttori, cioè le società che gestiscono gli impianti di produzione di energia elettrica. Un'offerta di vendita nell'asta del MGP altro non è che una coppia di numeri: una quantità (in MWh) e un prezzo unitario (€/MWh). Essi indicano la disponibilità del produttore a vendere una quantità di energia non superiore alla quantità espressa ad un prezzo unitario non inferiore a quello espresso. Per quanto riguarda la domanda il discorso è molto simile: ogni acquirente si rende disponibile ad acquistare una quantità non superiore a quella espressa nell'offerta di acquisto (indicata sempre in MWh) ad un prezzo unitario non superiore a quello espresso nell'offerta d'acquisto (in €/MWh). A questo punto entra in gioco il GME che si occupa di accettare le offerte di vendita e acquisto e procede col calcolare le curve di offerta e domanda aggregate, utilizzando un criterio di ordinamento non decrescente per le offerte di vendita e non crescente per le offerte di acquisto. Infine, si calcola l'intersezione delle due curve per ottenere la quantità totale e il prezzo di equilibrio del mercato (PM). Si stabilisce quindi una lista degli impianti che hanno effettuato un'offerta ad un prezzo inferiore o uguale a quello di mercato e per ognuno di essi vengono stabilite le relative quantità da produrre per ogni ora del giorno successivo. Stesso processo per gli operatori che hanno acquistato energia, ma al contrario: vengono accettate le offerte solo di coloro che hanno offerto un prezzo unitario maggiore o uguale al prezzo di mercato (PM)¹⁴. Una volta definita la programmazione può essere ancora leggermente aggiustata nel mercato infragiornaliero (MI) che apre alle 10:30 del giorno prima della consegna, in cui gli operatori possono modificare le posizioni stabilite nel MGP. Tuttavia, durante il giorno successivo, complici situazioni imprevedibili come tempera-

¹⁴<https://www.docenti.unina.it>

ture atmosferiche estreme o semplici naturali oscillazioni, il load (carico) e gli impianti di produzione spesso non rispettano la programmazione stabilita nel MGP. Ecco che serve quindi un mercato aggiuntivo, il Mercato per i Servizi di Dispacciamento (MSD) che si occupa di permettere transazioni con il System Operator (Terna S.p.A.) che deve assicurare il corretto funzionamento del sistema in *real-time*.

1.2.2 Il Prezzo Unico Nazionale - PUN

Il sistema elettrico in Italia è caratterizzato da una conformazione particolarmente prona a congestioni tra zone. Una zona è una porzione della rete elettrica che ha dei limiti di trasmissione da e verso altre zone geografiche. Una congestione invece è un fenomeno che si verifica quando a una particolare linea di trasmissione è richiesto di trasportare una quantità di energia maggiore di quella che può tecnicamente reggere. Generalizzando il concetto, una congestione è un fenomeno che si verifica quando tra due diverse zone di un sistema elettrico sarebbe necessario un passaggio di energia che non può avvenire per limiti tecnici delle linee di trasmissione. In Italia, fino alla fine del 2020, erano presenti 7 zone estere virtuali, 6 zone fisiche, Nord (NORD), Centro Nord (CNOR), Centro Sud (CSUD), Sud (SUD), Sardegna (SARD) e Sicilia (SIC) e 5 poli di produzione limitata, illustrate nell'immagine 1.1¹⁵:

È fondamentale tenere conto dei limiti fisici della rete nella stipulazione dei programmi del MGP, infatti molto spesso la programmazione che emergerebbe dall'intersezione delle curve di domanda e offerta nazionali sarebbe proprio tecnicamente impossibile da mettere in atto. Quando si verifica questa situazione, si procede a calcolare e intersecare le curve di domanda e offerta aggregate relative però alle offerte di ogni singola zona necessaria al-

¹⁵http://www.costanteinvernizzi.it/Site_1/Welcome.html. Dal gennaio 2021, le zone geografiche sono diventate 7 (la Calabria è stata divisa dal SUD ed è diventata una zona a sé stante), mentre i poli di produzione limitata sono stati soppressi. L'Umbria è passata dal centro-nord al centro-sud.



Figura 1.1: Suddivisione territoriale del mercato zonale italiano fino al dicembre 2020.

la prevenzione delle congestioni, stabilendo quindi diversi prezzi zonali. Per chiarire: se nella prima programmazione si calcola che si verificherebbe una congestione tra due zone, ad esempio CNOR e CSUD, si procederebbe a stimare due prezzi zonali, costituiti dall'intersezione delle curve di domanda e offerta aggregate delle due zone virtuali che si sono venute creare, ossia il gruppo di zone "sopra" a CNOR e quelle "sotto" a CSUD.

Essendo la produzione principalmente concentrata nel Nord del paese, la situazione più frequente è lo split in due o tre zone. Un mercato unico emerge invece meno frequentemente. Molto raramente emergono più di quattro zone¹⁶.

Tuttavia, per non creare squilibri sul fronte della domanda, nel mercato italiano le offerte di acquisto nel mercato del giorno prima vengono valutate al *Prezzo Unico Nazionale* (PUN), che altro non è che la media pesata dei prezzi zonali emersi nel MGP, con i consumi zonali come pesi. Essendo quindi

¹⁶Beltrami et al. (2021)

il prezzo di acquisto di ogni acquirente all'ingrosso di energia elettrica del mercato del giorno prima, il PUN è un ottimo indicatore del costo dell'energia elettrica ed è quindi di primario interesse la sua analisi sia dal punto di vista economico sia, come in questa tesi, dal punto di vista statistico.

Capitolo 2

Metodologia di Analisi

Lo studio di una qualsiasi realtà economica si avvale di una sempre più vasta gamma di tecniche statistiche fondamentali per la descrizione, l'interpretazione e la previsione della dinamica temporale di tutti quei fenomeni che possono essere rappresentati sotto forma di serie storica. Si pensi ad esempio alle serie mensili degli indici della produzione industriale, alle variazioni giornaliere dei titoli nel mercato azionario, ai dati sulle vendite settimanali di un'azienda, i dati trimestrali sull'occupazione oppure, naturalmente, il Prezzo Unico Nazionale (PUN) orario dell'energia elettrica. Una serie storica, dunque, altro non è che una serie di misure ripetute nel tempo (rilevate ad intervalli regolari) di uno stesso fenomeno di interesse oppure, ancora meglio, *“una successione di dati numerici nella quale ogni dato è associato ad un particolare istante od intervallo nel tempo”*¹.

Il PUN è quindi un ottimo esempio di serie storica: è una serie di dati numerici riferiti ad ogni singola ora di ogni singolo giorno dell'anno. Si presta quindi perfettamente all'analisi statistica.

¹Vianelli, 1983, p.28

2.1 Modelli statistici per serie storiche di dati economici

L'analisi statistica delle serie storiche generalmente ha svariati obiettivi, ma i principali sono sempre il tentare di chiarire il meccanismo causale che l'ha generata, dare una descrizione parsimoniosa delle sue caratteristiche (secondo il *Rasoio di Occam*) e anche prevedere l'evoluzione del fenomeno di interesse cui la serie si riferisce.

Una vasta gamma di approcci e modelli diversi sono stati sviluppati nel corso degli anni per rispondere a queste necessità. L'approccio moderno all'analisi delle serie storiche impone il trattamento della componente casuale u_t del processo generatore:

$$Y_t = f(t) + u_t$$

Come un processo casuale a componenti correlate, cioè che $Cov(u_r, u_s) \neq 0$ per qualche $r \neq s$. Si tratta quindi la serie Y_t come un processo stocastico ovvero un fenomeno che evolve seguendo leggi probabilistiche. Si può dire, semplificando, che un processo stocastico è una collezione di variabili casuali indicizzate nel tempo, con una distribuzione congiunta caratterizzata dalla funzione media:

$$\mu_t = E(Y_t)$$

dalla funzione di varianza:

$$\sigma_t^2 = Var(Y_t) = E[(Y_t - \mu_t)^2]$$

e dalla funzione di autocovarianza:

$$\gamma_{t_1, t_2} = E[(Y_{t_1} - \mu_{t_1})(Y_{t_2} - \mu_{t_2})]$$

è inoltre molto utile l'utilizzo della versione normalizzata dell'autocovarianza, ovvero la funzione di autocorrelazione (ACF, *AutoCorrelation Function*):

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sigma_{t_1}^2 \cdot \sigma_{t_2}^2}$$

che permette di quantificare la correlazione tra un valore della serie e i diversi valori passati della serie stessa. Utilissima per verificare la correlazione tra due valori della serie al netto delle correlazioni che hanno con gli altri valori della serie stessa, è la funzione di autocorrelazione parziale (PACF, *Partial AutoCorrelation Function*):

$$P_k = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

Stazionarietà e invertibilità

Per poter modellare efficacemente un processo stocastico è necessario però che sia stazionario, in particolare, è sufficiente che sia stazionario in senso debole (o del secondo ordine), ossia che la sua media sia costante nel tempo e che la sua funzione di autocovarianza dipenda esclusivamente dal ritardo, cioè dalla distanza nel tempo delle due osservazioni per la quale si sta calcolando:

$$E(Y_t) = \mu \quad \text{Cov}(Y_t, Y_{t+k}) = \gamma_k$$

Segue quindi che anche la varianza del processo è assunta finita e costante nel tempo (per $k = 0$ la funzione di autocovarianza coincide con la varianza).

È necessario però anche che il processo sia invertibile, ovvero che data una funzione lineare $h(\cdot)$ e un processo *White Noise* ε_t (una variabile casuale a media nulla e varianza finita), sia possibile scrivere, per ogni t :

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t$$

Questa condizione permette di identificare univocamente un processo stocastico, essendo che, data una qualsiasi funzione di autocovarianza esiste uno e un solo processo stocastico stazionario e invertibile.

È per questi motivi che, qualsiasi sia la serie storica analizzata, il primo passo da fare è sempre e comunque quello di trattare la serie per agevolare l'analisi. Primo fra tutti è il processo di standardizzazione, ovvero trasformare i dati in modo che abbiano media circa nulla e varianza circa unitaria.

Un'altra trasformazione molto spesso utilizzata è quella utile a *stabilizzare la varianza*. Molto spesso infatti le serie storiche, in particolar modo le serie storiche dei prezzi dei mercati elettrici, presentano dati anomali, osservazioni estremamente più grandi (o più piccole) del loro tipico livello: queste osservazioni vengono chiamate *spikes*. Inoltre spesso soffrono di periodi di estrema volatilità in cui le *spikes* sono decisamente più frequenti, aumentando quindi la varianza della serie stessa, salvo poi ritornare in periodi più calmi in cui la varianza torna sui livelli tipici. Come detto sopra, questa caratteristica combacia con quella di un processo non stazionario: è pertanto necessario prendere contromisure per evitare che la serie su cui si adattino i modelli presenti questa dipendenza temporale della varianza (o della media). Nel capitolo successivo verranno illustrate le scelte fatte per standardizzare la serie e stabilizzarne la varianza.

2.2 Approccio univariato e multivariato

Come esposto nel primo capitolo, in particolare nel paragrafo 1.2.1, i prezzi $P_{d,h}$ dell'energia elettrica delle $h = 1, \dots, 24$ ore del giorno d vengono stabiliti in un'unica asta (tenuta il giorno prima della consegna) e quindi pubblicati contemporaneamente. Questo fatto suggerisce un *framework* di tipo multivariato, dove la serie storica dei prezzi viene trattata come una serie multidimensionale i cui valori osservati sono realizzazioni del vettore casuale $\mathbf{P}_d = (P_{d,1}, \dots, P_{d,24})'$. Inoltre, i profili delle serie dei prezzi divisi per ora mostrano importanti distinzioni tra di loro, riflettendo in modo significativo la variazione, ora per ora durante ogni giornata, della domanda di energia elettrica, i costi di produzione e gestione degli impianti, le limitazioni delle infrastrutture dal punto di vista operativo e in ultima istanza anche delle strategie d'asta messe in atto dagli operatori del mercato (Gianfreda et al., 2016; Karakatsani e Bunn, 2008; Shahidehpour et al., 2002; Ziel, Weron 2018). D'altra parte però è altrettanto ragionevole considerare

la serie storica dei prezzi come un'unica serie univariata $P_t = P_{24d+h}$ ad “alta frequenza”, adatta quindi ad un approccio analitico univariato. Questo approccio affonda le sue radici nella tradizione dell'analisi di serie storiche nei mercati finanziari e nei mercati delle materie prime (Ziel, Weron 2018) ed è ancora oggi molto popolare, soprattutto nell'ambito ingegneristico delle previsioni sui prezzi elettrici affrontati con metodologie tecnologicamente molto avanzate, principalmente reti neurali (si guardi Aggrawal et al., 2009 per una rassegna). Entrambi gli approcci sono stati proposti ampiamente nella letteratura dell'EPF (*Electricity Price Forecasting*), tra i tanti, Cuaresma et al. (2004), Misiorek et al. (2006), Zhou et al. (2006), Garcia-Martos et al. (2007), Karakatsani e Bunn (2008), Gianfreda e Grossi (2012), Lisi e Nan (2014), Alonso et al. (2016), Gaillard et al. (2016), Hagfors et al. (2016), Maciejowska et al. (2016), Nowotarski e Weron (2016), Uniejewski et al. (2016) e Ziel (2016a), Grossi e Nan (2019), suggeriscono l'utilizzo di 24 (48 o più) modelli stimati indipendentemente, uno per ogni periodo del load (per il mercato italiano sono 24, in altri mercati i prezzi sono stabiliti ogni 30 o anche ogni 15 minuti). Per quanto riguarda gli studi in cui vengono proposti modelli per serie storiche di tipo univariato abbiamo Nogales et al. (2002), Contreras et al. (2003), Conejo et al. (2005), Zareipour et al. (2006), Paraschiv et al. (2015) e Ziel et al. (2015a) e studi in cui invece vengono utilizzate reti neurali (sempre con un approccio di tipo univariato) sono, tra i tanti, Rodriguez e Anders (2004), Amjady (2006), Pao (2007), Amjady et al. (2010), Abedinia et al. (2015), Kim (2015), Dudek (2016), Keles et al. (2016) e Rafiei et al. (2017). È quindi evidente che ancora non è chiaro quale approccio sia più adatto nell'EPF. In questa tesi verranno utilizzati entrambi e applicati ai prezzi dell'IPEX (Italian Power Exchange), uno dei pochi mercati dell'energia elettrica europei non trattati nello studio comparativo di Ziel e Weron (2018) che si proponeva di contribuire alla soluzione della questione. Di seguito sono elencati e spiegati i diversi modelli utilizzati, divisi nei due approcci di cui sopra.

2.2.1 Modelli univariati selezionati

Nell'approccio univariato, la serie storica di prezzi P_t viene modellata come:

$$P_t = f(P_{t-1}, P_{t-2}, \dots) + \varepsilon_t$$

dove $f(\cdot)$ è una qualsiasi funzione delle variabili esplicative (in questa tesi saranno esclusivamente i valori passati della serie stessa o loro rielaborazioni) e ε_t termine stocastico, con l'assunzione implicita (o esplicita) sulla sua distribuzione:

$$\varepsilon_t \sim N(0, \sigma^2)$$

La previsione viene computata per le 24 ore successive, corrispondenti quindi ad una previsione da uno a *24-step ahead*. In questa tesi si utilizzerà una tecnica ricorsiva (o iterata), dove, dopo aver effettuato la previsione *one-step ahead*:

$$\hat{P}_t = f(P_{t-1}, P_{t-2}, P_{t-3}, \dots)$$

la previsione *two-step ahead* \hat{P}_{t+1} è data da:

$$\hat{P}_{t+1} = f(\hat{P}_t, P_{t-1}, P_{t-2}, P_{t-3}, \dots)$$

e così via.

Il difetto di questa procedura è che permette una propagazione degli errori di previsione commessi ad ogni step previsivo, errori che si accumulano e producono stime spesso eccessivamente imprecise. Tuttavia ha il pregio, a differenza delle alternative possibili, di utilizzare tutta l'informazione disponibile.

Di seguito verranno presentati i modelli selezionati e verranno introdotte le loro caratteristiche principali.

Modello naive

Il principale *benchmark* adottato in questa tesi è il modello chiamato nel seguito semplicemente **naive**. Appartiene alle tecniche di tipo *similar-day*.

Funziona così: il valore predetto per l' h -esima ora dei Martedì, Mercoledì, Giovedì e Venerdì è posto uguale al valore osservato nell' h -esima ora del giorno precedente, mentre l' h -esima ora dei Lunedì, Sabati e Domeniche è posta uguale al valore osservato all' h -ora del corrispondente giorno della settimana prima.

È definito come:

$$P_{d,h} = \begin{cases} P_{d-7,h} + \varepsilon_{d,h} & \text{se } \text{DoW}_{d,h}^k = 1 \text{ per } k = 1, 6, 7; \\ P_{d-1,h} + \varepsilon_{d,h} & \text{altrimenti.} \end{cases}$$

Dove $\text{DoW}_{d,h}^k$ è una variabile *dummy* che vale 1 nel k -esimo giorno della settimana (il Lunedì ha $k = 1$, Martedì $k = 2$, ecc.) e 0 altrimenti con $k = 1, \dots, 7$. Questa tecnica è stata introdotta e utilizzata nell'ambito dell'EPF per la prima volta da Nogales et al. (2002) e, secondo questo stesso studio e quello di Conejo et al. (2005), il metodo produce previsioni migliori di molti modelli più complicati. Essendo dunque un ottimo strumento di controllo della performance previsiva dei modelli selezionati, è stato adottato anche in questa tesi. Va specificato che questo è l'unico modello tra quelli selezionati che viene adattato direttamente sulla serie di prezzi $P_{d,h}$ non trasformati. Tutti i modelli presentati da qui in avanti, infatti, sono stati adattati sulla serie di dati standardizzati e trasformati, che verranno indicati generalmente con Y_t .

Modello AR(p)

Il primo modello leggermente più complesso che è stato selezionato è un classico modello AR (*AutoRegressive*) univariato. I modelli AR sono un elemento della classe di modelli denominati ARIMA (*AutoRegressive Integrated Moving Average*). Nello specifico, i modelli AR nascono dall'idea che nelle serie storiche di dati, la variabile casuale che genera il processo sia funzione lineare delle sue realizzazioni passate con l'aggiunta di un termine casuale tipicamente *White Noise* (Rumore Bianco), cioè una variabile casuale a media

nulla e varianza finita. In particolare, ogni AR è definito dal suo ordine p , che corrisponde al numero di valori precedenti (detti “ritardi”) che influenzano la realizzazione futura. In particolare, il modello utilizzato in questa tesi è definito come segue:

$$Y_t = \bar{Y}_{HoW,t} + \phi_0 + \sum_{k=1}^p \phi_k (Y_{t-k} - \bar{Y}_{HoW,t-k}) + \varepsilon_t$$

dove ϕ_k sono i parametri autoregressivi e p l’ordine selezionato del processo AR. Inoltre, con \bar{Y}_{HoW} si indica la media dei valori della serie trasformata condizionati all’ora della settimana. \bar{Y}_{HoW} assumerà quindi uno solo tra 168 valori, ovvero le medie delle 168 serie orario-settimanali (una per ognuna delle 168 ore della settimana). Con $HoW_{d,h}^k$ o HoW_t^k si intende inoltre una variabile dummy che assume valore 1 in corrispondenza della k -esima ora settimanale e 0 altrimenti.

I parametri di questo modello vengono stimati risolvendo le equazioni di *Yule-Walker* e viene determinato l’ordine p ottimale utilizzando il *criterio di Akaike* (AIC, *Akaike Information Criterion*), scegliendolo ogni volta da un $p_{max} = 196$. In Ziel et al. (2015a) il modello AR è stato un ottimo benchmark per i prezzi del mercato EPEX.DE+AT (acronimo per il mercato elettrico di Germania e Austria), performando molto meglio di svariati modelli più sofisticati. Per questo motivo, oltre che per la sua presenza fissa negli studi comparativi sui metodi di EPF, è stato selezionato in questa tesi. Verrà in seguito indicato con AR.

Modelli univariati stimati con LASSO

Come notato in svariati lavori precedenti, tra i tanti si citano Amjady e Keynia (2009b), González et al. (2015), Karakatsani e Bunn (2008), Keles et al. (2016), Maciejowska (2014), Voronin e Partanen (2013), Weron (2014) e Gianfreda e Grossi (2012), un aspetto cruciale nella definizione di modelli adeguati per l’EPF è effettuare una scelta appropriata delle variabili esplica-

tive da considerare. Molto spesso la scelta viene effettuata arbitrariamente, tendenzialmente basandosi su considerazioni di carattere qualitativo sulle dinamiche della formazione del prezzo dell'energia elettrica oppure basandosi sull'analisi di alcuni indicatori di tipo statistico, spesso però non per forza robusti statisticamente.

Un'alternativa si presenta però con il metodo di stima LASSO o lasso (*Least Absolute Shrinkage and Selection Operator*) appartenente ai metodi di *shrinkage* (detti anche di regolarizzazione), che, in generale, sono metodi che stimano il modello completo (quello formulato con tutte le variabili esplicative a disposizione nel dataset) utilizzando un algoritmo che riduce (in modulo) verso lo zero il valore dei coefficienti delle variabili meno influenti, dando quindi più peso a quelle più importanti (James et al., 2013). Il metodo LASSO, introdotto da Tibshirani (1996), tuttavia, ha una particolarità che lo rende di estremo interesse per la modellazione dei prezzi del mercato dell'energia elettrica: non si limita infatti ad avvicinare allo zero i parametri delle variabili meno importanti, molto spesso infatti le azzerava proprio, effettuando, de facto, una selezione vera e propria delle migliori variabili esplicative del modello. La formulazione generale del modello, costruito ispirandosi al modello fARX (*full AutoRegressive with eXogenous variables*) di Uniejewski et al. (2016) e alla formulazione proposta da Ziel, Weron (2018), è la seguente:

$$Y_t = \sum_{k=1}^{168} \phi_{0,k} HoW_t^k + \sum_{k=1}^{196} \phi_{1,k} Y_{t-k} + \sum_{k=1}^{168} \phi_{2,k} HoW_t^k Y_{t-1} + \sum_{k=1}^{168} \phi_{3,k} HoW_t^k Y_{t-24} + \phi_{4,1} Y_{t-24,min} + \phi_{4,2} Y_{t-24,max} + \varepsilon_t$$

dove il primo termine si riferisce agli effetti sul valore atteso dell'ora della settimana, il secondo termine rappresenta la parte autoregressiva (composta p_{max} termini), il terzo e il quarto sono gli effetti periodici, ovvero l'interazione tra la dummy HoW e i prezzi dell'ora precedente e di 24 ore prima.

Gli ultimi due termini sono gli effetti non lineari, in quanto considerano i valori massimo e minimo del giorno precedente (massimo e minimo non sono funzioni lineari). Questo modello verrà indicato con **lasso_{p,nl}**. In questa tesi sono stati adattati anche versioni semplificate, in particolare, una versione senza effetti non lineari:

$$Y_t = \sum_{k=1}^{168} \phi_{0,k} HoW_t^k + \sum_{k=1}^{p_{max}} \phi_{1,k} Y_{t-k} + \sum_{k=1}^{168} \phi_{2,k} HoW_t^k Y_{t-1} + \sum_{k=1}^{168} \phi_{3,k} HoW_t^k Y_{t-24} + \varepsilon_t$$

che verrà d'ora in avanti chiamata **lasso_p** e, infine, una versione ancora più semplificata, che non comprende né effetti non lineari, né effetti periodici:

$$Y_t = \sum_{k=1}^{168} \phi_{0,k} HoW_t^k + \sum_{k=1}^{p_{max}} \phi_{1,k} Y_{t-k} + \varepsilon_t$$

d'ora in avanti chiamato semplicemente **lasso**.

I risultati eccezionali ottenuti in Ziel, Weron (2018) e Golia, Grossi e Pelagatti (2022) dai modelli stimati con metodo lasso supportano la scelta di includerli come modello principale di questa tesi.

2.2.2 Modelli multivariati selezionati

Per quanto riguarda l'approccio multivariato, la struttura più utilizzata per la serie storica 24-dimensionale dei prezzi è un semplice insieme di 24 modelli univariati, atti a modellare ognuna delle 24 serie storiche che nascono dal condizionamento alle 24 ore del giorno:

$$\begin{cases} P_{d,1} = f_1(P_{d-1,1}, P_{d-2,1}, P_{d-3,1}, \dots) + \varepsilon_{d,1} \\ \vdots \\ P_{d,24} = f_1(P_{d-1,24}, P_{d-2,24}, P_{d-3,24}, \dots) + \varepsilon_{d,24} \end{cases}$$

con $\varepsilon_{d,h}$ termine casuale con assunzione di normalità:

$$\varepsilon_{d,h} \sim N(0, \sigma_{d,h}^2)$$

e $f_h(\cdot)$ funzione dei valori precedenti della serie o loro rielaborazioni. In questo caso le previsioni $\hat{P}_{d,h}$, $h = 1, \dots, 24$ sono tutte *one-step ahead*.

Il grande vantaggio di questo tipo di modelli è che sono di facile implementazione e inoltre richiedono un numero ridotto di parametri da stimare per ognuno dei 24 modelli, rendendoli più leggeri dal punto di vista computazionale. Tuttavia, hanno come grosso difetto quello di trascurare le relazioni di dipendenza che potrebbero sussistere tra le serie di prezzi delle diverse ore: non fanno infatti alcuna assunzione sulla distribuzione congiunta degli $\varepsilon_{d,h}$ per le diverse ore. Viene quindi modellata solo la distribuzione marginale del vettore casuale 24-dimensionale $\mathbf{P}_d = (P_{d,1}, \dots, P_{d,24})'$. Esistono delle alternative che si propongono di considerare e modellare anche le strutture di interdipendenza tra le diverse serie orarie, ma non verranno considerate in questa tesi perché troppo dispendiose dal punto di vista computazionale o perché poco performanti rispetto ai modelli presentati.

Di seguito, dunque, verranno elencati e descritti nel dettaglio i diversi modelli di tipo multivariato selezionati.

Modello AR multivariato

Il primo modello multivariato selezionato è semplicemente la versione multivariata del modello AR(p) univariato esposto in precedenza. L'idea è quella di costruire 24 modelli AR(p_h), uno per ogni ora del giorno $h = 1, \dots, 24$ e adattarlo alla relativa serie (trasformata) dell'ora h del giorno. Per ognuno dei 24 modelli, come nel caso univariato, l'ordine p_h del modello viene selezionato in base all'AIC da un massimo di $p_{max} = 8$ e i parametri $\phi_{k,h}$ vengono stimati con le equazioni di Yule-Walker (Hyndman e Athanasopoulos, 2013; Ziel e Steinert, 2016). La formulazione generale del modello è la seguente, con h fissato:

$$Y_{d,h} = \bar{Y}_{HoW,d,h} + \phi_{0,h} + \sum_{k=1}^{p_h} \phi_{k,h} (Y_{dk,h} \bar{Y}_{HoW,d-k,h}) + \varepsilon_{d,h}$$

dove, come nell'AR univariato, $\bar{Y}_{HoW_{d,h}}$ è la media dei valori relativi ad una delle 168 ore della settimana e $\varepsilon_{d,h} \sim N(0, \sigma_h^2)$. Come la sua controparte univariata, anche questo modello è molto popolare nella letteratura di EPF e per questo motivo è stato selezionato per questa tesi, sebbene sia generalmente meno accurato delle formulazioni AR univariate (Ziel e Weron, 2018). Verrà indicato in seguito con **24AR**.

Modelli multivariati stimati con LASSO

Esattamente come l'AR multivariato (denominato **24AR**), anche il modello di tipo AR multivariato qui di seguito presentato consiste nella specificazione di un set di 24 modelli riferiti ad ognuna delle 24 ore del giorno, stimati indipendentemente l'uno dall'altro, utilizzando però l'algoritmo LASSO per la stima dei parametri. Il modello selezionato è una versione ridotta di quello formulato da Ziel e Weron (2018):

$$Y_{d,h} = \sum_{j=1}^7 \phi_{0,j} DoW_{d,h}^j + \sum_{i=1}^{24} \sum_{k=1}^8 \phi_{k,i} Y_{d-k,i} + \sum_{k=1}^8 \phi_{k,min} Y_{d-k,min} + \sum_{k=1}^8 \phi_{k,max} Y_{d-k,max} + \varepsilon_t$$

dove il primo termine si riferisce all'effetto della dummy DoW (*Day of the Week*, giorno della settimana) di $Y_{d,h}$, il secondo termine è la parte auto-regressiva e gli ultimi due sono gli effetti non lineari. Si farà riferimento a questa formulazione con il nome di **24lasso_{nl}**.

È stata considerata anche la seguente semplificazione:

$$Y_{d,h} = \sum_{j=1}^7 \phi_{0,j} DoW_{d,h}^j + \sum_{i=1}^{24} \sum_{k=1}^8 \phi_{k,i} Y_{d-k,i} + \varepsilon_t$$

ovvero un modello simile a quello precedente, ma senza gli effetti non lineari. Ci si riferirà a questo modello con **24lasso**.

Questo tipo di modelli, insieme ai lasso univariati, sono stati i migliori in assoluto nello studio comparativo di Ziel e Weron (2018). Per questo motivo sono stati inseriti in questa tesi.

2.3 Tecniche comparative adottate

Per valutare l'accuratezza delle previsioni di un qualsiasi modello, la strada più semplice, più ovvia e anche la più seguita nella letteratura è quella di concentrarsi sul modulo degli errori di previsioni, definiti semplicemente come:

$$AE_{d,h} = |P_{d,h} - \hat{P}_{d,h}| = |\hat{\varepsilon}_{d,h}|$$

Dove AE sta per *Absolute Error*, $P_{d,h}$ il prezzo osservato per l'ora h del giorno d e $\hat{P}_{d,h}$ la corrispondente previsione fornita dai modelli. Serve però un modo per riassumere in un unico valore l'informazione fornita dalla serie storica degli errori assoluti di previsione. Si adotta a questo proposito il MAE (*Mean Absolute Error*), ovvero la media campionaria della serie degli errori assoluti:

$$MAE = \frac{1}{24 \cdot D} \sum_{d=1}^D \sum_{h=1}^{24} AE_{d,h}$$

Dove D indica il numero totali di giorni per cui è stata effettuata la previsione dei prezzi orari. Spesso utile per serie storiche di dati in cui non si presentano valori prossimi allo zero, è anche il MAPE (*Mean Absolute Percentage Error*) che altro non è che la media degli errori percentuali di previsione:

$$MAPE = \frac{1}{24 \cdot D} \sum_{d=1}^D \sum_{h=1}^{24} \frac{AE_{d,h}}{P_{d,h}} \cdot 100\%$$

Un'altra strada consiste nel considerare la serie dei quadrati degli errori e riassumerne poi le caratteristiche con la relativa media, presa poi sotto radice per tornare sulla scala della risposta (quindi i prezzi). Questo indice, chiamato RMSE (*Root Mean Square Error*) è definito come:

$$RMSE = \sqrt{\frac{1}{24 \cdot D} \sum_{d=1}^D \sum_{h=1}^{24} (P_{d,h} - \hat{P}_{d,h})^2}$$

e la sua versione percentuale, il RMSPE (*Root Mean Square Percentage Error*):

$$RMSPE = \sqrt{\frac{1}{24 \cdot D} \sum_{d=1}^D \sum_{h=1}^{24} \left(\frac{P_{d,h} - \hat{P}_{d,h}}{P_{d,h}} \right)^2}$$

Un ultimo passo nell'analisi delle performance è il *test di Diebold-Mariano*, il quale permette una comparazione quantitativa e statisticamente valida tra le previsioni ottenute da due diversi modelli. In particolare, il *Diebold-Mariano test (DM test)* è la seguente quantità pivotale:

$$DM = \bar{d} / \sqrt{\sum_{k=-M}^M \hat{\gamma}_d(k) / T} \sim N(0, 1)$$

con T numero totale di previsioni effettuate, $M = T/3$ e \bar{d} è la media campionaria della serie storica delle differenze tra gli output di una stessa qualsiasi funzione di perdita calcolata sugli errori di previsioni di due modelli diversi, ovvero la media della serie $d_t = L(\hat{\varepsilon}_{1,t}) - L(\hat{\varepsilon}_{2,t})$ dove $L(\cdot)$ è la *loss function*. Inoltre, $\hat{\gamma}_d(k)$ è la stima della funzione di autocovarianza della serie delle differenze d_t calcolata al ritardo k . L'ipotesi nulla H_0 di non significativa differenza di performance predittiva tra i due modelli si rifiuta al livello di significatività α se $|DM| < z_{\alpha/2}$ con $z_{\alpha/2}$ quantile della normale standard di ordine $\alpha/2$.

In particolare, si è utilizzata in questa tesi la correzione per campioni finiti del DM-test di *Harvey-Leybourne-Newbold* (1997):

$$HLN - DM = \sqrt{1 + \frac{1-2h+h(h-1)}{T}} \cdot DM \sim t_{T-1}$$

dove T è il numero totale di previsioni effettuate e h l'orizzonte previsivo di ogni passo. È il default della funzione `dm.test(.)` della libreria *forecast* del software R.

Calcolando gli indici sopracitati per ogni modello, si potrà dunque ottenere una valutazione quantitativa dell'accuratezza delle previsioni effettuate.

Capitolo 3

Implementazione e Risultati Empirici

L'analisi è stata effettuata utilizzando i dati ufficiali del PUN orario, rilasciati direttamente dal GME e scaricabili liberamente dal sito del GME¹. Per procedere con l'analisi dati si è scelto di utilizzare una finestra di calibrazione di quattro anni di dati (dal 01/01/2015 al 31/12/2018), 35064 dati orari che considerano anche il 29/02 dell'anno bisestile 2016, con la quale si sono stimati i parametri dei modelli e poi una finestra di un anno intero come *out-of-sample* test (8760 dati orari dalla prima ora del 01/01/2019 all'ultima ora del 31/12/2019). Fanno eccezione i modelli **naive**, **AR** e **24AR** per le quali sono state utilizzate procedure leggermente differenti. Ciò che non cambia è la finestra previsiva di 8760 dati del 2019 che è stata mantenuta comune per tutti i modelli in modo da poterne comparare efficacemente i risultati.

¹<https://www.mercatoelettrico.org>

3.1 Data preprocessing

Dopo aver scaricato i dati dal sito ufficiale del GME e averli uniti in un unico file .xlsx in modo da contenere tutti i dati dal 01/01/2015 al 31/12/2019, si procede con l'analisi. Seguendo i passaggi svolti da Ziel, Weron (2018), il primo passo nella trattazione empirica dei dati è quello di standardizzare la serie originale, che come si vede in figura 3.1 ha media diversa da zero e varianza certamente diversa da uno.

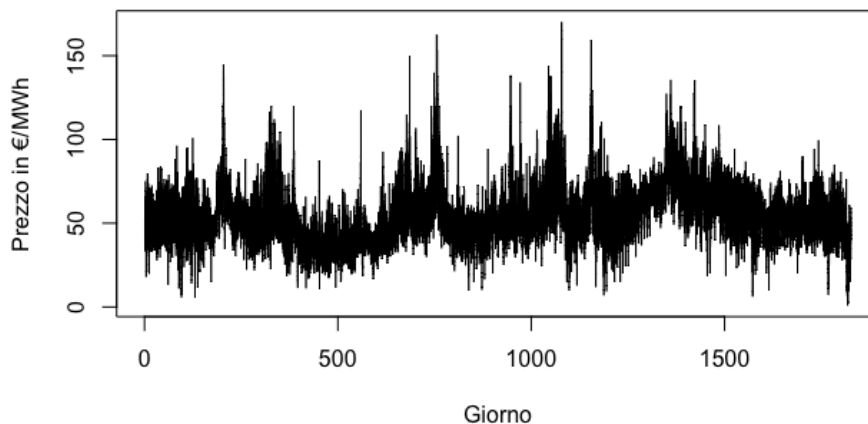


Figura 3.1: Serie originale del PUN orario dell'energia elettrica nel mercato italiano dal giorno 01/01/2015 al giorno 31/12/2019 in €/MWh

Il metodo proposto dagli autori dell'articolo, chiamato "*Normalizzazione (Median, MAD)*", consiste nel togliere da ogni dato la sua mediana e dividere il risultato per la *deviazione assoluta dalla mediana* cioè il MAD (*Median Absolute Deviation*) aggiustato dal fattore $1/z_{0.75} \approx 1.4826$ con $z_{0.75}$ terzo quartile della normale standard, per ottenere una consistenza asintotica alla deviazione standard (Nel software R questa correzione è l'impostazione di default del comando `mad(.)`). La serie standardizzata è dunque la seguente:

$$X_t = [P_t - \text{median}(P_t)] / \text{mad}(P_t)$$

dove con P_t si intende la serie originale dei prezzi in €/MWh. A questo punto si procede con la stazionarizzazione della serie. La media già non sembra dipendere dal tempo, mentre la varianza è quasi certamente dipendente dal tempo, si veda di seguito il grafico 3.2 relativo alla serie standardizzata.

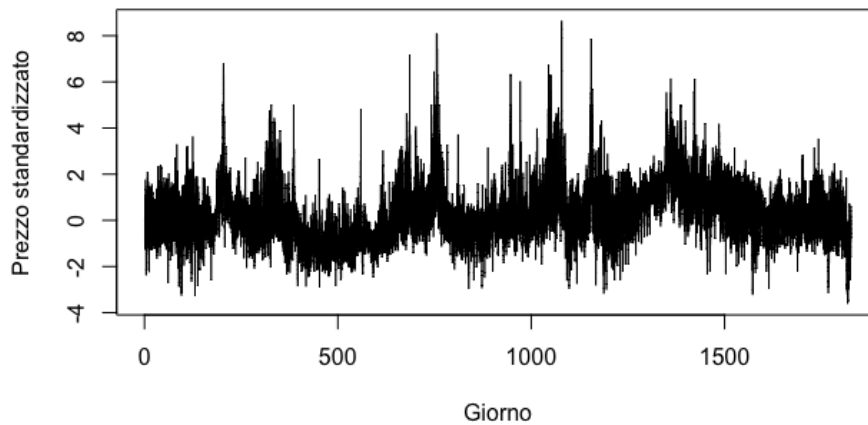


Figura 3.2: Serie standardizzata del PUN orario dell'energia elettrica nel mercato italiano dal giorno 01/01/2015 al giorno 31/12/2019

Seguendo ancora il suggerimento di Ziel e Weron (2018), si effettua la trasformazione con seno iperbolico inverso per stabilizzare la varianza:

$$Y_t = \text{arcsinh}(X_t) = \ln(X_t + \sqrt{X_t^2 + 1})$$

In appendice A (3.5) i comandi utilizzati sul software R per trasformare la serie.

Si riporta nella figura 3.3 il grafico della serie trasformata per mostrare l'effetto della trasformazione sulla varianza della serie.

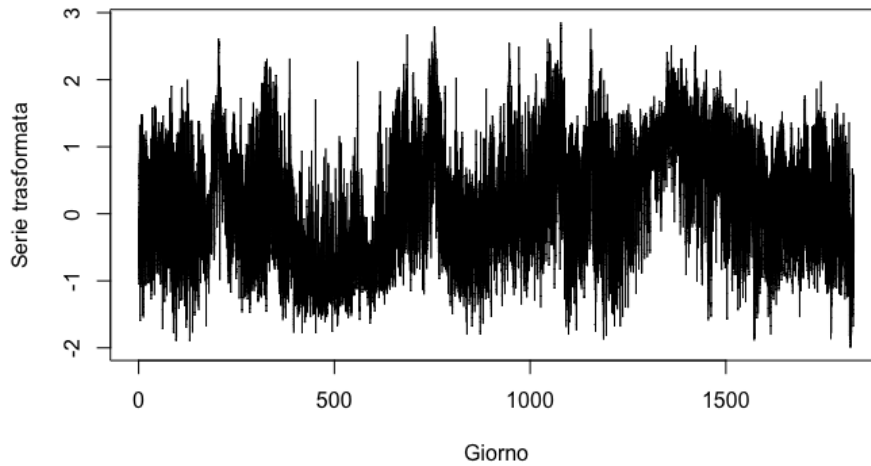


Figura 3.3: Serie trasformata Y_t , dal giorno 01/01/2015 al giorno 31/12/2019

Si conferma la correttezza delle assunzioni distributive, mostrando, nell'immagine 3.4, l'istogramma della serie come approssimazione della sua distribuzione.

Dopo aver ottenuto la serie trasformata, indicata con Y_t o $Y_{d,h}$, si passa all'adattamento dei modelli.

3.2 Modello naive

Questo è stato l'unico modello in cui si è lavorato esclusivamente con la serie di prezzi originale non trasformata $P_{d,h}$. Utilizzando il software R, si effettuano direttamente le previsioni per tutta la finestra temporale scelta come orizzonte previsivo (ovvero i 8760 dati orari relativi all'intero 2019), ponendo come previsione per ogni dato orario dei Martedì, Mercoledì, Giovedì e Venerdì il dato della corrispondente ora osservato il giorno precedente,

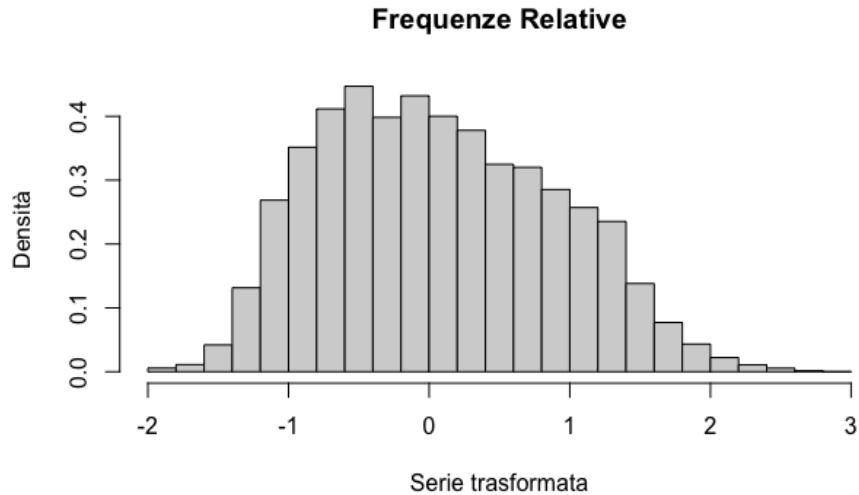


Figura 3.4: Istogramma delle frequenze relative di Y_t

mentre per i Lunedì, Sabati e Domeniche si utilizza come previsione il dato dell'ora corrispondente relativa al corrispondente giorno della settimana.

In appendice B (3.5) i comandi utilizzati per ottenere il vettore con gli 8760 valori predetti. In figura 3.5 si mostrano le previsioni per i primi 3 giorni (72 dati orari) del 2019 confrontate coi valori effettivamente osservati.

3.3 Modelli AR univariato e multivariato

Per questi due modelli il procedimento seguito è stato quello denominato *rolling window*. Nello specifico, si procede selezionando una finestra di dati per stimare i parametri del modello, effettuare una previsione di orizzonte k e poi spostare la finestra di calibrazione di k passi avanti, in modo da mantenere il numero complessivo di osservazioni utilizzate per stimare il modello, ma sfruttando anche i k nuovi valori osservati (ovvero i dati reali delle corrispondenti k previsioni effettuate al passo precedente) ed effettuare quindi

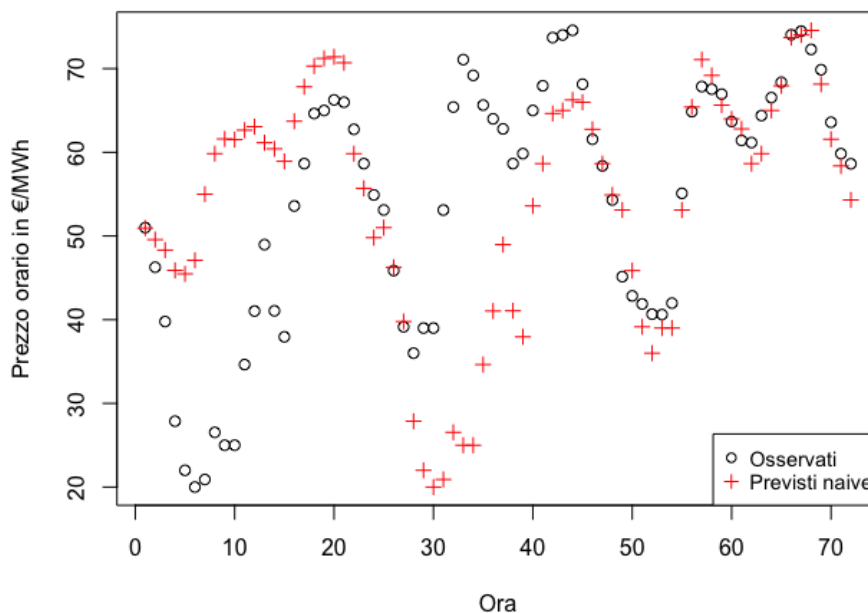


Figura 3.5: Confronto tra valori osservati e predetti dal modello **naive** per i primi 3 giorni del 2019

una nuova previsione di orizzonte k e così via. Nei paragrafi successivi viene illustrata la procedura declinata nei due diversi approcci.

3.3.1 Modello AR univariato

Per l'AR univariato si è utilizzata una finestra di calibrazione di 4 anni (35064 dati) e come orizzonte previsivo si è scelto $k = 24$. La giustificazione di questa scelta risiede nel fatto che i prezzi orari dell'energia elettrica vengono pubblicati contemporaneamente, quindi non avrebbe alcun senso utilizzati il dato osservato, ad esempio, nella prima ora del giorno per effettuare la previsione sulla seconda ora del giorno.

Una volta effettuate le 24 previsioni relative alle 24 ore del giorno successivo, la finestra di calibrazione scorre in avanti di 24 osservazioni e si procede con una nuova stima del modello, non considerando più i 24 dati del 01/01/2015 e considerando invece i 24 dati osservati del 01/01/2019, mantenendo quindi fisso il numero totale di osservazioni utilizzate per la stima dei parametri (cioè 35064 osservazioni). Si procede quindi con il *forecast* dei 24 dati relativi alle 24 ore del 02/01/2019 e si fa scorrere nuovamente la finestra in avanti di 24 osservazioni, togliendo dalla finestra di calibrazione quelle relative al 02/01/2015 e inserendovi quelle osservate il 02/01/2019. Iterando questo processo si arriva infine ad ottenere le 8760 previsioni del 2019 (effettuate 24 alla volta). Ricordandosi che il modello viene adattato sui dati differenziati per la media dell'ora della settimana corrispondente, per ottenere il vettore delle previsioni sulla scala della serie trasformata serve infine aggiungere la media rispettiva ad ogni previsione ottenuta per la serie differenziata.

In appendice C (3.5) i comandi utilizzati per l'implementazione nel software R.

3.3.2 Modello AR multivariato

Per questo modello invece, avendo le 24 serie orarie, per ottenere le previsioni relative al giorno successivo (nel primo caso, il 01/01/2019) è sufficiente un orizzonte di previsione $k = 1$ per ognuno dei 24 modelli stimati. Analogamente all'AR univariato, si fa scorrere la finestra di calibrazione questa volta però di un solo passo avanti, dato che l'orizzonte previsivo è 1 per ogni modello. Si stimano quindi i nuovi 24 modelli sulla nuova finestra di calibrazione e si ottengono le 24 previsioni ad un passo (una da ognuno dei 24 modelli) relative alle 24 ore del giorno 02/01/2019 e via così.

Come per l'AR univariato il passaggio finale consiste nell'aggiungere alle 8760 previsioni ottenute le relative medie dell'ora della settimana così da avere le previsioni sulla scala della serie trasformata.

In appendice C (3.5) i comandi utilizzati per l'implementazione in R.

In figura 3.6 si mostrano le previsioni dei modelli AR univariato e multivariato per i primi 3 giorni (72 dati orari) del 2019 confrontate coi valori effettivamente osservati.

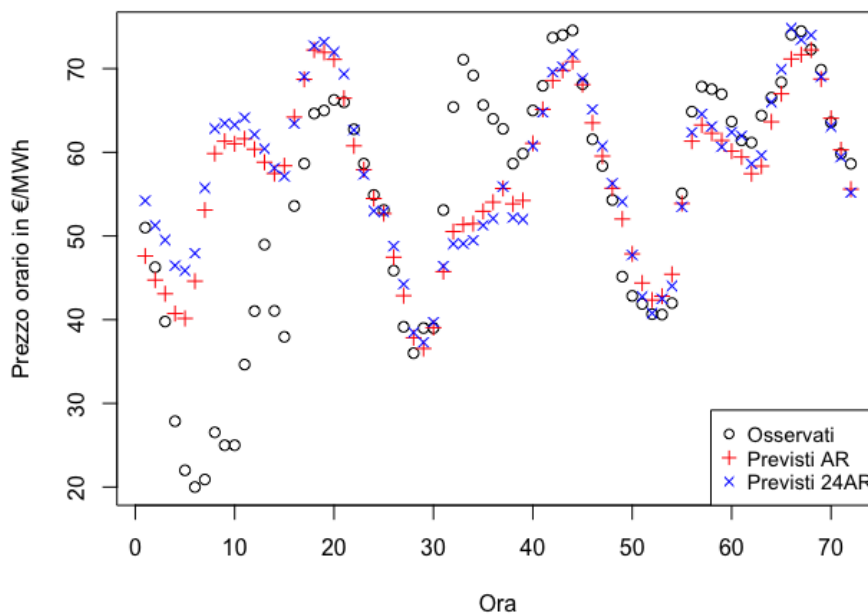


Figura 3.6: Confronto tra valori osservati e predetti dai modelli **AR** e **24AR** per i primi 3 giorni del 2019

3.4 Modelli LASSO univariati e multivariati

Per tutti e 5 i modelli stimati con LASSO, è stata selezionata una finestra di calibrazione di quattro anni di dati (01/01/2015-31/12/2018) per stimare i parametri del modello, ottenendo quindi la stima dei parametri dei 5 modelli scelti, utilizzandoli poi per ottenere le previsioni di orizzonte massimo 24

sulla finestra di test, inserendo come esplicative anche i dati osservati, via via aggiunti ad ogni previsione. In appendice D (3.5) vengono mostrati gli script utilizzati per ottenere le stime dei parametri dei modelli di tipo LASSO univariati, mentre in appendice E (3.5) sono mostrati gli script utilizzati per stimare i modelli di tipo LASSO multivariati.

3.4.1 LASSO univariati

In figura 3.7 si mostrano le previsioni ottenute dai tre modelli implementati per i primi 3 giorni (72 dati orari) del 2019 confrontate coi valori effettivamente osservati.

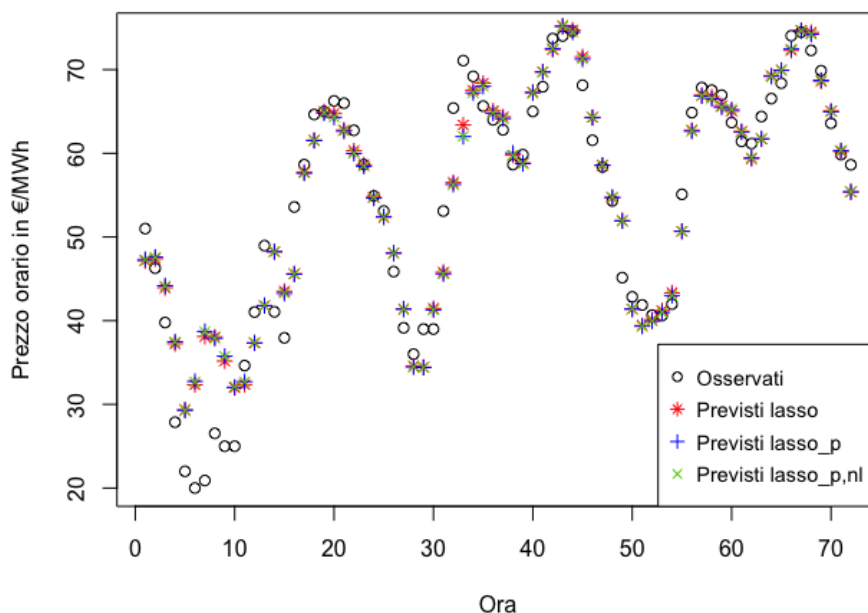


Figura 3.7: Confronto tra valori osservati e predetti dai modelli *lasso*, *lasso_p* e *lasso_{p,nl}* per i primi 3 giorni del 2019

Come si può vedere già in prima battuta, le previsioni puntuali fornite dai tre modelli univariati stimati con LASSO sono estremamente accurate.

3.4.2 LASSO multivariati

In figura 3.8 si mostrano le previsioni ottenute dai due modelli implementati per i primi 3 giorni (72 dati orari) del 2019 confrontate coi valori effettivamente osservati.

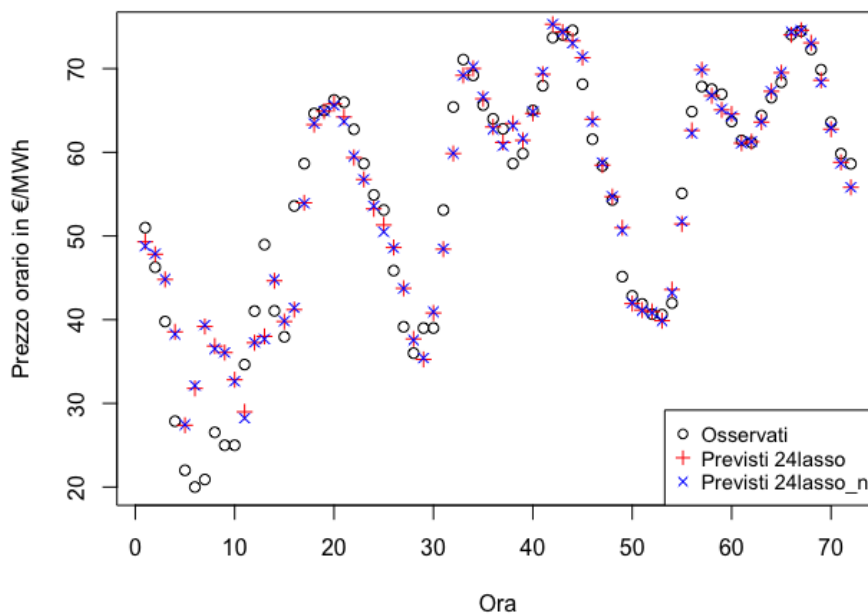


Figura 3.8: Confronto tra valori osservati e predetti dai modelli **24lasso** e **24lasso_{nl}** per i primi 3 giorni del 2019

Come anche nella loro versione univariata, i modelli LASSO multivariati sembrano, già in prima battuta, aver performato molto bene.

3.5 Confronto performance predittive

L'ultimo passaggio dell'analisi consiste nell'effettuare la valutazione della bontà delle previsioni per ognuno dei modelli di cui sopra e compararli tra di loro per cercare quale modello ha dato i risultati migliori. Si veda in figura 3.9 una prima comparazione tra le previsioni dei modelli principali considerati, effettuata mostrando in grafico le previsioni per le prime 48 ore del 2019 (2 giorni).

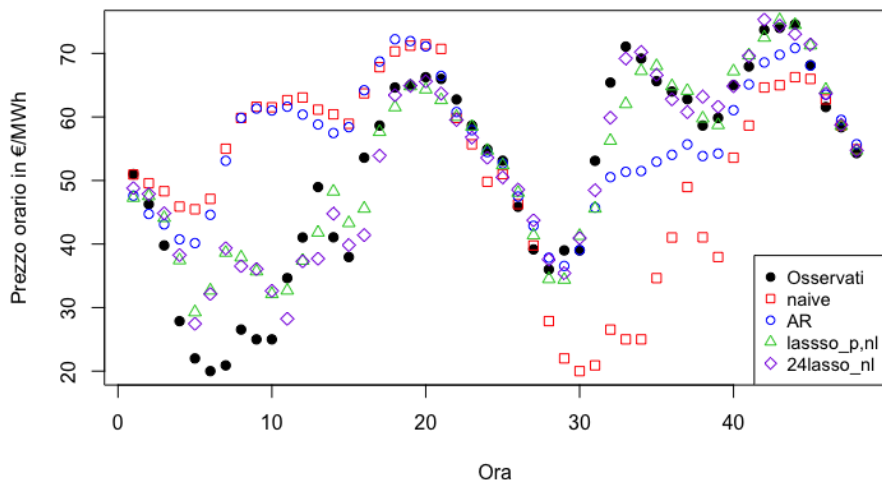


Figura 3.9: Confronto tra valori osservati e predetti dai modelli **naive**, **AR**, **lasso_{p,nl}** e **24lasso_{nl}** per i primi 2 giorni del 2019

Per prima cosa sono stati calcolati tutti gli indici illustrati in precedenza: MAE e RMSE in €/MWh e MAPE e RMSPE rispettive versioni percentuali. In appendice F (3.5) viene illustrato il codice utilizzato.

Nella seguente tabella sono riportati i risultati del calcolo degli indici selezionati (MAE e RMSE in €/MWh, MAPE e RMSPE corrispondenti versioni percentuali) sulle previsioni di tutti i modelli.

	MAE	MAPE(%)	RMSE	RMSPE(%)
naive	5.81	13.83	8.02	55.27
AR	4.14	9.08	5.73	21.54
24AR	4.61	10.77	6.26	34.22
lasso	2.10	4.52	2.92	11.82
lasso_p	2.09	4.46	2.91	11.50
lasso_{p,nl}	2.09	4.47	2.91	11.52
24lasso	2.16	4.57	3.00	11.21
24lasso_{nl}	2.16	4.57	3.01	10.98

Tabella 3.1: Indici di performance predittiva dei modelli considerati.

Si procede quindi con il calcolo del Diebold-Mariano test, per dare una valutazione statisticamente significativa della comparazione dei modelli. In appendice F (3.5) vengono illustrati il codice utilizzato e i risultati ottenuti.

Nella figura 3.10 si mostrano i risultati del DM-test incrociato calcolato su tutte le coppie possibili di modelli. Il test è stato effettuato con ipotesi bilaterale ed essendo stato effettuato due volte per ogni coppia di modelli, in un caso la statistica test viene positiva e nell'altro invece negativa, ma rimanendo uguale in modulo. Quindi se la statistica test ha un valore in modulo maggiore del quantile 0.975 di una distribuzione t -student con 8759 gradi di libertà, significa che la differenza nella performance previsiva dei due modelli è significativa. In particolare, il modello in colonna è considerato come il "primo" dei due nella differenza che permette la costruzione del test, dunque per valori positivi e "grandi" del test si ha che il modello in colonna ha performato meglio di quello in riga. Naturalmente vale il contrario per valori negativi della statistica test. I risultati delle statistiche test ottenute sono tutti significativi ad un livello di significatività del 5%.

Dalla tabella 3.1 e dall'immagine 3.10 si può vedere che in generale i modelli stimati con LASSO hanno performance ampiamente migliori della

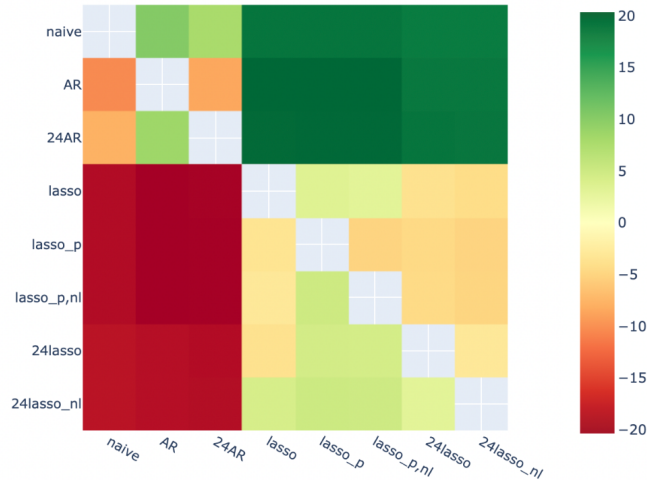


Figura 3.10: Heat map della tabella coi risultati del DM-test. Come esempio, il modello **naive** è stato il peggiore (la relativa colonna è priva di riquadri verdi) mentre la colonna del **lasso_p** ha sostanzialmente ogni riquadro tendente al verde ed è per tanto il migliore.

concorrenza: con un MAE medio di 2.12€/MWh e un MAPE medio di 4.5%, una precisione più che doppia rispetto ai modelli alternativi. Li premia anche il RMSPE, che tende a dare più peso ad errori più grandi (penalizza di più le previsioni molto diverse dal valore osservato), con un valore di 11.4%. Analizzando i modelli LASSO si nota che, sebbene i due modelli multivariati siano stati più "vicini" al valore osservato nei loro errori, hanno comunque sbagliato "più frequentemente" della controparte univariata (hanno RMSPE minore, ma MAPE maggiore) anche se è difficile dare un valore statistico a queste considerazioni. Secondo il test di Diebold e Mariano, il modello che ha performato meglio è senza dubbio il **lasso_p**, tuttavia è difficile dire se questo risultato sia dovuto al tipo di approccio selezionato oppure piuttosto alle esplicative selezionate. Guardando gli indici calcolati e il risultato del DM-test dei modelli **lasso** e **24lasso** che hanno identiche esplicative, ma

seguono un differente approccio, si può vedere che il modello univariato è stato significativamente più accurato nelle previsioni puntuali. Tornando infine alla domanda da cui muove l'intento questa tesi, quale approccio prevede meglio nel breve periodo?

Il suggerimento finale, dato che entrambi gli approcci hanno ottenuto ottimi risultati, è quello di utilizzare una procedura di *shrinkage* di tipo LASSO per la stima dei parametri e di considerare versioni miste dei due approcci per massimizzare i punti di forza e sopperire ai punti deboli: come concludono anche Ziel e Weron (2018), può essere sufficiente anche una semplice media tra le due previsioni.

Conclusioni

In questo turbolento periodo storico, gli squilibri geopolitici a livello internazionale si riflettono sul mercato dell'energia elettrica sotto forma di prezzi in media molto più alti e al contempo anche più volatili. In questo contesto l'*Electricity Price Forecasting* (EPF) assume un ruolo centrale non solo come supporto per gli attori del mercato dell'energia elettrica, ma anche come strumento fondamentale per il corretto funzionamento del mercato e del sistema elettrico. L'obiettivo di questa tesi è quello di fornire ulteriori informazioni utili, si spera, al miglioramento delle tecniche di modellazione e previsione per quanto riguarda il mercato elettrico. In particolare si vuole stabilire se nel *short-term* EPF è più adatto un approccio di tipo di multivariato o univariato. Analizzando le performance previsive ottenute dai diversi modelli risulta evidente la superiorità dei modelli stimati con LASSO che, con un MAPE medio di 4.5%, si impongono tra i migliori modelli in questo ambito. I risultati ottenuti confermano e supportano le conclusioni tratte da Ziel e Weron (2018), ovvero che i modelli stimati con procedura LASSO sono certamente più performanti. Al tempo stesso, però, è molto difficile stabilire con certezza se vi sia una differenza statisticamente significativa tra l'approccio univariato e l'approccio multivariato. La risposta alla domanda a cui questa tesi cerca di rispondere sembra essere che il miglior approccio all'EPF sia misto: anche una semplice media tra i valori predetti dai modelli univariati e multivariati può produrre risultati migliori, a patto che la procedura di stima coinvolga un algoritmo di *shrinkage* come il LASSO.

Appendice

Appendice A: trasformazione della serie

```
library("readxl")
# si seleziona il file .xlsx
dati_15.19 <- data.frame(read_excel(file.choose()))
dati_15.19_TS <- ts(dati_15.19[,3], start=c(1, 1),
                    frequency = 24)

# STANDARDIZZAZIONE #
Yab<-(dati_15.19_TS -
      median(dati_15.19_TS))/mad(dati_15.19_TS)

# TRASFORMAZIONE con SENO IPERBOLICO INVERSO #
Yt_official <- log(Yab+sqrt((Yab^2)+1))
```

Appendice B: modello naive

```
# L'idea è tagliare il dataset in modo che parta dalla
# prima ora del 29/12/18 perchè è il dato più vecchio
# utile alla previsione. I dati del 29-30-31/12/2018
# servono per predire i valori di sabato domenica e
# lunedì 5-6-7/01/2019. I dati orari del 31/12/2018
# servono anche per predire quelli
```

```

# di Martedì 01/01/2019, poi si continua così

# Si inizia creando il vettore che identifichi il
# giorno della settimana di ogni ora del dataset
# Si parte dall'ora 1 di sabato 29/12/2018
# 1 sarà sabato, 2 domenica, 3 lunedì ecc.

DayOfTheSample <- c()
for (i in 1:368){
  DayOfTheSample <- append(DayOfTheSample,
                           c(rep(i,24)))}

# Si accoppia il vettore dei giorni al vettore dei
# prezzi orari partendo dal dato 34993, relativo
# all'ora 1 del 29/12/2018
Pt.perNaive <- cbind(dati_15.19_TS[34993:43824],
                    DayOfTheSample)

### MATRICE CON PREVISIONI ORARIE PER OGNI ORA
# DI OGNI GIORNO DELLA SETTIMANA ###

prezzi.predetti.naive <- matrix(NA,
                               nrow = 1272, ncol = 7)
# 1272 righe perchè ci sono 53 martedì nel 2019
# 53*24==1272. Le righe dalla 1249 compresa in poi
# sono NA per le colonne degli altri DoW
colnames(prezzi.predetti.naive)<-c("Mar", "Mer", "Gio",
                                   "Ven", "Sab",
                                   "Dom", "Lun")

```

```

# Ora con cicli for si riempiono le colonne dei giorni
# con le relative previsioni orarie

# Martedì: è speciale perchè ci sono 53 martedì nel 2019
# (di tutti gli altri giorni ce ne sono 52)
for (i in 0:52){
  prezzi.predetti.naive[(1+24*i):(24+24*i),1] <-
  Pt.perNaive[DayOfTheSample==3+7*i,1]}
# riempie 24 righe alla volta della colonna del
# martedì con i valori osservati nel giorno precedente

# Mercoledì
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),2] <-
  Pt.perNaive[DayOfTheSample==4+7*i,1]}

# Giovedì
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),3] <-
  Pt.perNaive[DayOfTheSample==5+7*i,1]}

# Venerdì
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),4] <-
  Pt.perNaive[DayOfTheSample==6+7*i,1]}

# da qui in poi invece si prendono i valori del giorno
# corrispondente della settimana prima:

```

```

# Sabato
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),5] <-
  Pt.perNaive[DayOfTheSample==1+7*i,1]}

# Domenica
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),6] <-
  Pt.perNaive[DayOfTheSample==2+7*i,1]}

# Lunedì
for (i in 0:51){
  prezzi.predetti.naive[(1+24*i):(24+24*i),7] <-
  Pt.perNaive[DayOfTheSample==3+7*i,1]}

# Si dispongono ora le previsioni in un vettore

previsioni.naive <- rep(NA, 8760)
for (settimana in 0:51){
  for (i in 0:6){
    previsioni.naive[(1+24*i+168*settimana):
                      (24+24*i+168*settimana)] <-
    prezzi.predetti.naive[(1+24*settimana):
                          (24+24*settimana),i+1]}}

# si aggiungono infine i 24 valori dell'ultimo giorno
previsioni.naive[8737:8760]<-
prezzi.predetti.naive[1249:1272,1]

```

Appendice C: modelli AR e 24AR

```
#####  
## MODELLO AR ##  
#####  
previsioni.AR <- rep(NA, 8760)  
  
for (i in 0:364){  
  matrice.Yt.HoW<-matrix(c(Yt_official[(1+24*i):  
                                (35064+24*i)] ,  
                            rep(NA, 48)), nrow = 209 ,  
                            ncol = 168, byrow = T)  
  Ymedio_HoW <- colMeans(matrice.Yt.HoW, na.rm = T)  
  #SERIE DIFFERENZIATA  
  seriexAR <- Yt_official[(1+24*i):(35064+24*i)] -  
    c(rep(Ymedio_HoW, 208), Ymedio_HoW[1:120])  
  
  fitAR <- ar(seriexAR , aic = TRUE, order.max = 196)  
  # si effettua la previsione e si riaggiunge  
  # la media HoW  
  previsioni.AR[(1+24*i):(24+24*i)] <-  
    predict(fitAR , n.ahead = 24, se.fit = F)[1:24] +  
    Ymedio_HoW[121:144]}  
  
#####  
## MODELLO 24AR ##  
#####  
previsioni.24AR<-c()  
for (i in 0:364){  
  # matrice con in colonna i dati delle singole ore  
  # della settimana (HoW) 209 righe perchè in 4 anni
```

```

# ci sono 208 settimane piene più 5 giorni
# (di cui uno è il giorno di anno bisestile)
matrice.Ydh.HoW<-matrix(c(Yt_official[(1+24*i):
                                (35064+24*i)]),
                        rep(NA, 48)), nrow = 209,
                        ncol = 168, byrow = T)
Ymedio_HoW <- colMeans(matrice.Ydh.HoW, na.rm = T)

# si differenzia la serie trasformata togliendo ad
# ogni dato la relativa media tra le 168 calcolate
serie.24AR <- Yt_official[(1+24*i):(35064+24*i)] -
  c(rep(Ymedio_HoW, 208), Ymedio_HoW[1:120])

# si crea la matrice con le 24 serie storiche relative
# ad ogni ora del giorno (una in ogni colonna)
matrice.Ydh.24AR <- matrix(serie.24AR, nrow = 1461,
                          ncol = 24, byrow = T)

# Stima dei 24 modelli AR, appendendo ad un vettore
# le 24 previsioni (una per ogni modello stimato)
previsioni.SenzaMedia24AR <- rep(NA, 24)
for (ora in 1:24){
  fit.24AR <- ar(matrice.Ydh.24AR[,ora], aic = TRUE,
                order.max = 8)
  previsioni.SenzaMedia24AR[ora]<-
    predict(fit.24AR, n.ahead = 1, se.fit = F)[1]}
# aggiungendo la media HoW relativa ,
# si ottengono le previsioni effettive
previsioni.24AR <-
  append(previsioni.24AR,

```



```
(previsioni.SenzaMedia24AR +
  Ymedio_HoW[121:144]))}
```

Appendice D: modelli LASSO univariati

```
# i primi 196 valori dei prezzi non hanno le esplicative
# necessarie. Il primo valore con tutte le esplicative
# è il 197-esimo
Yt.lasso.train <- Yt_official[-c(1:196,35065:43824)]
```

```
### MATRICE DUMMY HoW ###
# si utilizza la funzione make.dummy()
# di sse_060921.lib.txt
source(file.choose())
```

```
HoW.completo <- make.dummy(43824, freq = 168,
                          start = 73)
```

```
HoW <- HoW.completo[-(1:196),]
# il primo dato ha 1 nella 101-esima colonna perchè
# il primo dato che ha tutte le esplicative è la
# quinta ora del venerdì
```

```
### TERMINI AUTOREGRESSIVI ###
# 196 colonne coi prezzi da 1 a 196 precedenti
prezzi.shiftati <- matrix(NA, nrow = 43628, ncol=196)
for (i in 1:196){
  prezzi.shiftati[,i]<-Yt_official[(197-i):(43824-i)]}
#####
### MODELLO SEMPLICE ###
#####
```

```

# (senza effetti periodici e non-lineari)

### MATRICE DI ESPLICATIVE ###
X.train.semplice <-
  as.matrix(cbind(HoW, prezzi.shiftati))[1:34868,]
X.test.semplice <-
  as.matrix(cbind(HoW, prezzi.shiftati))[34869:43628,]

### STIMA con TRAINING SET per stima parametri ###
library(glmnet)
lasso.semplice <- cv.glmnet(x = X.train.semplice,
                           y = Yt.lasso.train)

### PREVISIONI con TEST SET ###
previsioni.lasso <- rep(NA, 8760)
previsioni.lasso[1:8760] <-
  as.matrix(predict(lasso.semplice,
                   newx = X.test.semplice)[,1])

### EFFETTI PERIODICI ###
# ora prima (dato precedente) e ora corrispondente del
# giorno prima (24h prima, 24 dati prima) sono termini
# di interazione, per ottenerli in forma matriciale si
# moltiplica per la matrice di dummy
interazione_HoW.Yt_meno1 <-
  HoW*Yt_official[-c(1:195, 43824)]
interazione_HoW.Yt_meno24 <-
  HoW*Yt_official[-c(1:172, 43801:43824)]

#####
### MODELLO CON EFFETTI PERIODICI ###

```

```

#####

### MATRICE DI ESPLICATIVE DEL MODELLO ###
X.train.p <-
  as.matrix(cbind(HoW, prezzi.shiftati ,
                  interazione_HoW.Yt_meno1,
                  interazione_HoW.Yt_meno24))[1:34868,]
X.test.p <-
  as.matrix(cbind(HoW, prezzi.shiftati ,
                  interazione_HoW.Yt_meno1,
                  interazione_HoW.Yt_meno24))[34869:43628,]

### STIMA con TRAINING SET per stima parametri ###
lasso.p <- cv.glmnet(x = X.train.p,
                    y = Yt.lasso.train)

### PREVISIONI con TEST SET ###
previsioni.lasso.p <- rep(NA, 8760)
previsioni.lasso.p[1:8760] <-
  as.matrix(predict(lasso.p, newx = X.test.p)[,1])

### MASSIMI e MINIMI giornalieri per
# EFFETTI NON-LINEARI ###
# Matrice con in colonna le 24 serie orarie
Ydh.24 <- matrix(data = Yt_official , nrow = 1826,
                 ncol = 24, byrow = T)

massimi.giornalieri.ogniOra <- c()
minimi.giornalieri.ogniOra <- c()
for (i in 1:1826){
  massimi.giornalieri.ogniOra <- append(

```

```

    massimi.giornalieri.ogniOra ,
    rep(max(Ydh.24[i,]),24))
minimi.giornalieri.ogniOra <- append(
    minimi.giornalieri.ogniOra ,
    rep(min(Ydh.24[i,]),24))}

# Si tolgono i primi e gli ultimi dati
max.shiftati <-
    massimi.giornalieri.ogniOra[-c(1:172, 43801:43824)]
min.shiftati <-
    minimi.giornalieri.ogniOra[-c(1:172, 43801:43824)]

#####
### MODELLO CON EFFETTI PERIODICI E NON-LINEARI ###
#####

### MATRICE DI ESPLICATIVE ###
X.train.nl.p <-
    as.matrix(cbind(HoW, prezzi.shiftati ,
                    interazione_HoW.Yt_meno1,
                    interazione_HoW.Yt_meno24,
                    max.shiftati , min.shiftati))[1:34868,]
X.test.nl.p <-
    as.matrix(cbind(HoW, prezzi.shiftati ,
                    interazione_HoW.Yt_meno1,
                    interazione_HoW.Yt_meno24,
                    max.shiftati , min.shiftati))[34869:43628,]

### STIMA con TRAINING SET per stima parametri ###
lasso.nl.p <- cv.glmnet(x = X.train.nl.p,

```

```
y = Yt.lasso.train)
```

```
### PREVISIONI con TEST SET ###  
previsioni.lasso.nl.p <- rep(NA, 8760)  
previsioni.lasso.nl.p[1:8760] <-  
  as.matrix(predict(lasso.nl.p, newx = X.test.nl.p)[,1])
```

Appendice E: modelli LASSO multivariati

```
# matrice con le 24 serie storiche di ogni ora del  
# giorno messe in colonna  
Ydh.24lasso <- matrix(data = Yt_official, nrow = 1826,  
                      ncol = 24, byrow = T)  
  
### matrice con in colonna le 24 "Y" dei 24 modelli ###  
Ydh.24lasso.train <- Ydh.24lasso[-c(1:8, 1462:1826),]  
  
### MATRICE DUMMY DoW ###  
# si utilizza la funzione make.dummy()  
# di sse_060921.lib.txt  
source(file.choose())  
  
# 01/01/2015 era Giovedì quindi lo start è a 4, Lunedì=1  
DoW.completo <- make.dummy(1826, freq = 7, start = 4)  
DoW <- DoW.completo[-(1:8),]  
  
# qui il dato utile più vecchio è il 192-esimo, cioè  
# l'ora 24-esima dell'8-avo giorno addietro il primo  
# giorno che ha tutte le esplicative è il 9/1/2015
```

```

#####
### 24LASSO SEMPLICE ###
#####
# senza non-linear effects
Xdh.train.parziale.semplice <- as.matrix(DoW)[1:1453,]
Xdh.test.parziale.semplice <- as.matrix(DoW)[1454:1818,]

matrice.prev.24lasso.semplice <-
  matrix(NA, nrow = 365, ncol = 24)

for (ora in 0:23){
  matrice.ritardi <- matrix(NA, nrow=1818, ncol=192)
  for (n in 0:191){
    for (k in 0:1817){
      matrice.ritardi[k+1,n+1] <-
        Yt_official[192 - n + ora + 24*k]}
    #Costruzione matrice esplicative:
    Xdh.train.semplice <-cbind(matrice.ritardi[1:1453,],
                               Xdh.train.parziale.semplice)
    Xdh.test.semplice <-cbind(matrice.ritardi[1454:1818,],
                              Xdh.test.parziale.semplice)
    # stima modello e previsione
    lasso24.semplice <-
      cv.glmnet(x = Xdh.train.semplice ,
               y = Ydh.24lasso.train[,ora+1])
    matrice.prev.24lasso.semplice[,ora+1] <-
      predict(lasso24.semplice, newx = Xdh.test.semplice)}

previsioni.24lasso <- rep(NA, 8760)
for (k in 0:364){

```

```

previsioni.24lasso[(1+24*k):(24+24*k)] <-
  as.matrix(matrice.prev.24lasso.semplice[k+1,1:24])}

### MASSIMI e MINIMI giornalieri
# per EFFETTI NON-LINEARI ###
massimi.giornalieri <- rep(NA, 1826)
minimi.giornalieri <- rep(NA, 1826)
for (i in 1:1826){
  massimi.giornalieri[i] <- max(Ydh.24lasso[i,])
  minimi.giornalieri[i] <- min(Ydh.24lasso[i,])}
# Si costruiscono ora le due matrici che entreranno
# nella matrice delle esplicative:

# 1818 è la lunghezza di ogni serie oraria
# tagliata (priva dei primi 8 dati)
massimi.ritardati <- matrix(NA, nrow = 1818, ncol = 8)
minimi.ritardati <- matrix(NA, nrow = 1818, ncol = 8)
for (j in 0:7){
  massimi.ritardati[,j+1] <-
    massimi.giornalieri[(8-j):(1825-j)]
  minimi.ritardati[,j+1] <-
    minimi.giornalieri[(8-j):(1825-j)]}

#####
### 24 MODELLI CON EFFETTI NON-LINEARI ###
#####

## MATRICE ESPLICATIVE parziale ##
Xdh.train.parziale <-
  as.matrix(cbind(DoW, massimi.ritardati,
                 minimi.ritardati))[1:1453,]

```

```

Xdh.test.parziale <-
  as.matrix(cbind(DoW, massimi.ritardati ,
                  minimi.ritardati))[1454:1818,]
matrice.prev.24lasso.nl <-
  matrix(NA, nrow = 365, ncol = 24)

### STIMA DEI 24 MODELLI ###
library(glmnet)
for (ora in 0:23){
  matrice.ritardi <- matrix(NA, nrow=1818, ncol=192)
  for (n in 0:191){
    for (k in 0:1817){
      matrice.ritardi[k+1,n+1] <-
        Yt_official[192 - n + ora + 24*k]}
  # Costruzione matrice esplicative complete:
  Xdh.train <- cbind(matrice.ritardi[1:1453,],
                    Xdh.train.parziale)
  Xdh.test <- cbind(matrice.ritardi[1454:1818,],
                   Xdh.test.parziale)
  # Stima modello e previsione
  lasso24.nl <-
    cv.glmnet(x = Xdh.train ,
              y = Ydh.24lasso.train[,ora+1])
  matrice.prev.24lasso.nl[,ora+1] <-
    predict(lasso24.nl, newx = Xdh.test)}

previsioni.24lasso.nl <- rep(NA, 8760)
for (k in 0:364){
  previsioni.24lasso.nl[(1+24*k):(24+24*k)] <-
    as.matrix(matrice.prev.24lasso.nl[k+1,1:24])}

```


Appendice F: performance predittive

```
# Dopo aver rinominato le previsioni di ogni modello
# si procede a ritrasformare nella scala dei Prezzi/MWh
a <- median(dati_15.19_TS)
b <- mad(dati_15.19_TS)

P.predetti <- b*sinh(cbind(AR, AR24, lasso , lasso.p,
                        lasso.p.nl , lasso24 ,
                        lasso24.nl)) + a

# aggiunti anche i naive si ha la matrice completa
# con le previsioni di tutti i modelli
P.predetti <- cbind(naive , P.predetti)

#### CALCOLO INDICI ####
# Si utilizza la funzione for.perf() la cui
# formulazione è riportata:

for.perf<-function(serie , previsioni)
{
y.orig<-serie
previ<-previsioni
hor<-length(previsioni)
n.obs<-length(serie)
# valori osservati sull'orizzonte previsivo:
#  $y_{(T-h+1)}, \dots, y_{(T)}$ 
oss.out<-y.orig[(n.obs-hor+1):n.obs]
err.p<-oss.out-previ # errori di previsione
EM<-mean(err.p) # Errore Medio
```

```

VE<-var(err.p)          # Varianza errori di previsione
EQM<-mean(err.p^2)     # Mean Squared Error
# Scomposizione di EQM
ES<-EM^2               # Errore Sistemtico
# Standard deviation previsioni
sd.previ<-sqrt(var(previ)*((hor-1)/hor))
# Standard deviation valori osservati
sd.oss<-sqrt(var(oss.out)*((hor-1)/hor))
# Errore in varianza
EV<-(sd.previ-sd.oss)^2
# Errore in covarianza
EC<-2*(1-corr(previ,oss.out))*sd.previ*sd.oss
q.ES<-ES/EQM
q.EV<-EV/EQM
q.EC<-EC/EQM
EAM<-mean(abs(err.p))
EQM<-mean((err.p)^2)
RMSE<-sqrt(EQM)      # RMSE
# Errori percentuali
p.err<-err.p/oss.out*100
RMSPE<-sqrt(mean(p.err^2))
MAPE<-mean(abs(p.err))
# Calcoli per indice di Theil
p.naive<-y.orig[(n.obs-hor):(n.obs-1)]
e.naive<-oss.out-p.naive
EQM.naive<-mean(e.naive^2)
U.theil<-sqrt(EQM)/sqrt(EQM.naive)
output<-list(EAM=EAM,RMSE=RMSE,q.ES=q.ES,q.EV=q.EV,
             q.EC=q.EC,RMSPE=RMSPE,MAPE=MAPE,
             U.theil=U.theil)}

```

```

# Si carica la funzione for.perf()

source(file.choose())

P.originali1519 <- c(dati_15.19[,3])
matrice.indici <- matrix(NA, nrow=8, ncol=8)
for (i in 1:8){
  matrice.indici[,i] <-
    unlist(for.perf(P.originali1519 , P.predetti[,i]))}

P.osservati <- c(dati_15.19[35065:43824,3])
AE <- abs(P.predetti-P.osservati)

#####
### HLN-DM TEST ###
#####

library(forecast)

StatisticaTest.DM <- matrix(NA, ncol = 8, nrow = 8)
p.val.DM <- matrix(NA, ncol = 8, nrow = 8)
for (i in 1:8){
  for (j in 1:8){
    if (i!=j){
      dmtest <- dm.test(AE[,i], AE[,j], h=24, power = 1)
      StatisticaTest.DM[i,j] <- dmtestprova$statistic
      p.val.DM[i,j] <- dmtest$p.value}}}}

# Quindi se le entrate di StatisticaTest.DM sono

```

*# positive e "grandi" (maggiori del quantile 0.975
della distribuzione t-student con 8759 gradi di
libertà) allora il modello in colonna è meglio
del modello in riga.*

Bibliografia

Abedinia, O., Amjady, N., Shafie-Khah, M., Catalao, J., 2015. Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Convers. Manag.* 105, 642–654.

Aggarwal, S., Saini, L., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. *Int. J. Electr. Power Energy Syst.* 31, 13–22.

Alonso, A.M., Bastos, G., García-Martos, C., 2016. Electricity price forecasting by averaging dynamic factor models. *Energies* 9 (8), 600.

Amjady, N., 2006. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Trans. Power Syst.* 21, 887–996.

Amjady, N., Keynia, F., 2009b. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Trans. Power Syst.* 24 (1), 306–318.

Amjady, N., Daraeepour, A., Keynia, F., 2010. Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. *IET Gener. Transm. Distrib.* 4 (3), 432–444.

Beltrami F., Fontini F., Grossi L., 2021. The value of carbon emission reduction induced by Renewable Energy Sources in the Italian power market. *Ecological Economics*, 189, 107149

Conejo, A.J., Contreras, J., Espínola, R., Plazas, M.A., 2005. Forecasting electricity prices for a day-ahead pool-based electric energy market. *Int. J. Forecast.* 21, 435–462.

Contreras, J., Espínola, R., Nogales, F., Conejo, A., 2003. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* 18 (3), 1014–1020.

Cretì A., Fontini F., 2019. *Economics of Electricity Markets, Competition and Rules*. Cambridge University Press.

Cuaresma, J.C., Hlouskova, J., Kossmeier, S., Obersteiner, M., 2004. Forecasting electricity spot-prices using linear univariate time-series models. *Appl. Energy* 77 (1), 87–106.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *J. Bus. Econ. Stat.* 13, 253–263.

Dudek, G., 2016. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *Int. J. Forecast.* 32, 1057–1060.

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.

Gaillard, P., Goude, Y., Nedellec, R., 2016. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *Int. J. Forecast.* 32 (3), 1038–1050.

Garcia-Martos, C., Rodriguez, J., Sanchez, M., 2007. Mixed models for short-run forecasting of electricity prices: application for the Spanish market. *IEEE Trans. Power Syst.* 22, 544–551.

Gianfreda, A., Grossi, L., 2012. Forecasting Italian electricity zonal prices with exogenous variables. *Energy Econ.* 34, 2228–2239.

Gianfreda, A., Parisio, L., Pelagatti, M., 2016. The RES-induced switching effect across fossil fuels: an analysis of the Italian day-ahead and balancing prices and their connected costs. RG Working Paper. <https://doi.org/10.13140/RG.2.2.30195>. 12320.

Golia S., Grossi L., Pelagatti M., 2022. Machine Learning Models and Intra-Daily Market Information for the Prediction of Italian Electricity Prices. *Forecasting*, 2023, 5, 81–101.

González, C., Mira-McWilliams, J., Juárez, I., 2015. Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, bagging and random forests. *IET Gener. Transm. Distrib.* 9 (11), 1120–1128.

Grossi L., Nan F., 2019 . Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change* 141 (2019) 305–318.

Hagfors, L., Kamperud, H., Paraschiv, F., Prokopczuk, M., Sator, A., Westgaard, S., 2016. Prediction of extreme price occurrences in the German day-ahead electricity market. *Quant. Finan.* 16, 1929–1948.

Harvey D., Leybourne S., Newbold P., 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, vol. 13, issue 2, 281-291.

Hyndman, R., Athanasopoulos, G., 2013. *Forecasting: Principles and practice*. Online at <http://otexts.org/fpp/>.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning with Applications in R*. Springer, New York.

Karakatsani, N., Bunn, D., 2008. Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. *Int. J. Forecast.* 24, 764–785.

Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Appl. Energy* 162, 218–230.

Kim, M., 2015. Short-term price forecasting of Nordic power market by combination Levenberg-Marquardt and Cuckoo search algorithms. *IET Gener. Transm. Distrib.* 9 (13), 1553–1563.

Lisi, F., Nan, F., 2014. Component estimation for electricity prices: procedures and comparisons. *Energy Econ.* 44, 143–159.

Lisi F., Di Fonzo T., 2005. *Serie Storiche Economiche*. Carocci, Roma

Maciejowska, K., 2014. Fundamental and speculative shocks, what drives electricity prices? IEEE Conference Proceedings - EEM14. <https://doi.org/10.1109/EEM.2014.6861289>.

Maciejowska, K., Nowotarski, J., Weron, R., 2016. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *Int. J. Forecast.* 32 (3), 957–965.

Misiorek, A., Trück, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. *Stud. Nonlinear Dyn. Econ.* 10 (3).Article 2.

Nogales, F.J., Contreras, J., Conejo, A.J., Espinola, R., 2002. Forecasting next-day electricity prices by time series models. *IEEE Trans. Power Syst.* 17, 342–348.

Nowotarski, J., Weron, R., 2016. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. *Energy Econ.* 57, 228–235.

Pao, H.T., 2007. Forecasting electricity market pricing using artificial neural networks. *Energy Convers. Manag.* 48, 907–912.

Paraschiv, F., Fleten, S.-E., Schürle, M., 2015. A spot-forward model for electricity prices with regime shifts. *Energy Econ.* 47, 142–153.

Rafiei, M., Niknam, T., Khooban, M.-H., 2017. Probabilistic forecasting of hourly electricity price by generalization of ELM for usage in improved wavelet neural network. *IEEE Trans. Ind. Inf.* 13, 71–79.

Rodriguez, C.P., Anders, G.J., 2004. Energy price forecasting in the Ontario competitive power system market. *IEEE Trans. Power Syst.* 19 (1), 366–374.

Shahidehpour, M., Yamin, H., Li, Z., 2002. *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Wiley.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288.

Uniejewski, B., Nowotarski, J., Weron, R., 2016. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* 9 (8), 621.

Vianelli, S., 1983. *L'analisi delle serie temporali nello sviluppo storico e metodologico della statistica*, Milano

Voronin, S., Partanen, J., 2013. Price forecasting in the day-ahead energy market by an iterative method with separate normal price and price spike frameworks. *Energies* 6, 5897–5920.

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: a statistical approach*. Chichester: Wiley.

Weron, R., 2014. Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int. J. Forecast.* 30, 1030–1081.

Zareipour, H., Canizares, C.A., Bhattacharya, K., Thomson, J., 2006. Application of public-domain market information to forecast Ontario's wholesale electricity prices. *IEEE Trans. Power Syst.* 21 (4), 1707–1717.

Zhou, M., Yan, Z., Ni, Y., Li, G., Nie, Y., 2006. Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach. *IEE Proc.: Gener. Transm. Distrib.* 153 (2), 187–195.

Ziel, F., 2016a. Forecasting electricity spot prices using LASSO: on capturing the autoregressive intraday structure. *IEEE Trans. Power Syst.* 31, 4977–4987.

Ziel, F., Steinert, R., 2016. Electricity price forecasting using sale and purchase curves: the X-model. *Energy Econ.* 59, 435–454.

Ziel, F., Steinert, R., Husmann, S., 2015a. Efficient modeling and forecasting of electricity spot prices. *Energy Econ.* 47, 89–111.

Ziel F., Weron R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* 70, 396–420.

Sitografia

ARERA (Autorità di Regolazione per Energia Reti e Ambiente, Attività dei distributori, 2022, <https://www.arera.it/it/dati/eem71.htm>

Costante M. Invernizzi, http://www.costanteinvernizzi.it/Site_1/Welcome.html

GME (Gestore Mercati Energetici), <https://www.mercatoelettrico.org/It/Tools/Accessodati.aspx?ReturnUrl=%2fit%2fdownload%2fDatiStorici.aspx>

GSE (Gestore Servizi Energetici), Fuel Mix, Determinazione del Mix Energetico per gli Anni 2020-2021, 15 Settembre 2022, <https://www.gse.it/servizi-per-te/news/fuel-mix-determinazione-del-mix-energetico-per-gli-anni-2020-2021>

Harvard Business Review, The Dark Side of Solar Power, Atalay Atasu - Serasu Duran - Luk N. Van Wassenhove, 18 Giugno 2021, <https://hbr.org/2021/06/the-dark-side-of-solar-power>

Iea, Energy in India today, 2021, <https://www.iea.org/reports/india-energy-outlook-2021/energy-in-india-today>

IPCC-WG3 Erneuerbare Energien, SRREN Report, 2011, <https://www.ipcc-wg3.de/srren-report/>

Joint Research Centre (European Commission) – IRENA (International Renewable Energy Agency), Benchmarking Scenario Comparisons Key Indicators for the Clean Energy Transition, 2021. <https://www.cleanenergyministerial.org/content/uploads/2022/03/irena-jrc-benchmarking-scenario-comparisons-2021.pdf>

JRC (European Commission), Smart Electricity System and Interoperability, Distribution System Operators Observatory, 2020, <https://ses.jrc.ec.europa.eu/distribution-system-operators-observatory>

JRC Publications Repository (European Commission), The Role of Nuclear Energy in a Low-Carbon Energy Future, 2013/01/08, <https://publications.jrc.ec.europa.eu/repository/handle/JRC72017>

San José state university economics department, The Catastrophic Dam Failures in China in August 1975, Thayer Watkins, 26 Aprile 2002, <https://web.archive.org/web/20020426022951/http://www.sjsu.edu/faculty/watkins/aug1975.htm>

Sorgenia, GRTN (Getstore della Rete di Trasmissione Nazionale), <https://www.sorgenia.it/guida-energia/grtn>

Storia di Milano, La Centrale elettrica di via Santa Radegonda, Gian Luca Lapini, 30 Gennaio 2004, <http://www.storiadimilano.it/citta/milanotecnica/elettricita/radegonda0.htm>

TERNA, Produzione di Energia Elettrica per Fonte, 2020, <https://app.powerbi.com/view?r=eyJrIjoiZGUzY2U3NTgtZWZmZS00MjM2LTgwZGIhYTgwZGI1YmQ0MmY1IiwidCI6ImVjY2Q3MzRlLTcwMjItNDcwOS1hYmE1LWE1ZGQ3NzkyOWUyNyIsImMiOiJh9&pageName=ReportSection>

UniNa, Il mercato dell'energia elettrica, <https://www.docenti.unina.it/webdocenti-be/allegati/materiale-didattico/483882>

United Nations Scientific Committee, Sources, Effects And Risks Of Ionizing Radiation UNSCEAR 2013 Report, 2014, https://www.unscear.org/docs/reports/2013/13-85418_Report_2013_Annex_A.pdf