

Università degli studi di padova

Dipartimento di scienze statistiche
Corso di laurea triennale in statistica
per l'economia e l'impresa



Relazione finale

**Applicazione di modelli statistici per la
previsione dei prezzi dell'energia elettrica nei
principali mercati elettrici Europei**

Relatore: **Prof. Luigi Grossi**

Laureando: **Vittorio Valle**

Matricola: **1229911**

Anno accademico 2022/2023

Indice

Introduzione	3
1 Sistema elettrico	5
1.1 Rete elettrica	5
1.1.1 Produzione	6
1.1.2 Trasmissione	8
1.1.3 Distribuzione	9
1.2 Mercato energetico	10
1.3 Borsa elettrica	11
1.4 Mercati dell'energia elettrica da un punto di vista temporale .	12
1.5 Mercato del giorno prima	13
2 Metodologie	17
2.1 Notazione	17
2.2 Processi stocastici	17
2.2.1 Stazionarietà	18
2.3 Processi ARMA	19
2.3.1 Processi MA e AR	19
2.3.2 Processi ARMA	22
2.4 Processi ARIMA	23
2.5 Processi SARIMA	24
2.6 Modelli per serie storiche con variabili esogene: ARMAX . . .	25
2.7 Procedura di identificazione e validazione del modello	26
2.8 Valutazione delle previsioni e confronto tra modelli	28
3 Presentazione dei dati	31
4 Analisi dati	34
4.1 Relazioni con i predittori	37
4.2 Trattamento dei picchi	40
4.3 Adattamento dei modelli e previsioni Out of sample	42

4.4	Previsioni Rolling windows	45
4.5	Previsioni Rolling windows caso Germania	48
5	Conclusione	52

Introduzione

Negli ultimi decenni con la deregolamentazione e l'introduzione della concorrenza è emersa una nuova sfida per gli operatori del sistema energetico, l'estrema volatilità dei prezzi ha reso necessario uno studio approfondito sulla previsione dei prezzi dell'energia che risulta ad oggi fondamentale per il processo decisionale e lo sviluppo della strategia di un'azienda energetica e nella gestione stessa del mercato. Le soluzioni proposte possono essere classificate in termini della lunghezza dell'orizzonte di previsione, è consuetudine parlare di previsione dei prezzi a breve termine, medio termine e lungo termine (anche se non vi è un chiaro riferimento a quali devono essere le soglie discriminatorie). L'obiettivo principale delle previsioni a lungo termine consiste nella pianificazione degli investimenti (costruzioni di nuove centrali, implementare strategie di approvvigionamento ecc.), per quanto riguarda le previsioni di medio termine sono generalmente preferite per i calcoli di bilanci, la gestione del rischio e la pianificazione operativa. Una parte importanti degli studi fatti in materia di previsione dei prezzi si concentra nel breve termine, nel mercato a pronti dell'energia le imprese sono tenute a fare un'offerta in termini di prezzo e quantità, gli ordini di acquisto sono accettati in base ad un criterio di merito (prima quelli con i prezzi più bassi e poi in ordine crescente) fino a quando la domanda totale risulta soddisfatta (Weron, 2006), qualora una centrale elettrica fosse in grado di prevedere i prezzi orari potrà regolare il proprio programma di produzione massimizzando i suoi profitti, in virtù di questo le tecniche di previsione a breve termine (tipicamente orarie) risultano di primaria importanza nelle operazioni di mercato quotidiane. La formazione stessa del prezzo orario avviene all'interno della borsa elettrica in particolare nel mercato del giorno prima secondo le regole espresse dal gestore dei mercati dell'energia (GME) per le quali il prezzo nasce attraverso un meccanismo di asta doppia che permette, in base alle offerte fatte dalle società di produzione e dalle società di distribuzione, di creare la curva di domanda e di offerta il cui punto d'incontro stabilisce il prezzo e la quantità di equilibrio. La struttura temporale del prezzo dunque dipenderà da vari fattori come fattori temporali legati al periodo dell'anno, al giorno della

settimana e all'ora del giorno costruendo una struttura di correlazione molto importante da considerare in ottica di previsione del prezzo, dipenderà anche da fattori esogeni come variabili metereologiche e disponibilità di fonti rinnovabili ugualmente importanti da inserire nei modelli previsionali. Una grande varietà di metodologie sono state sperimentate in quest'ambito, dai metodi basati su un approccio statistico ai metodi basati sull'intelligenza artificiale, nel presente elaborato verranno trattati esclusivamente alcuni modelli basati su un approccio statistico i quali, utilizzando una combinazione dei valori prezzi passati e fattori esogeni, riescono a prevedere il prezzo dell'energia. Questi modelli risultano particolarmente interessanti perché un'interpretazione fisica può essere collegata ai loro componenti permettendo agli operatori del sistema di capirne il comportamento (Weron, 2006). L'obbiettivo del presente elaborato è quello di prevedere efficacemente i prezzi dell'energia attraverso modelli statistici nei principali mercati Europei. Il presente elaborato è composto come segue: il capitolo 1 descrive il funzionamento del sistema elettrico, dalla rete elettrica ai mercati che lo compongono, con un particolare riferimento al mercato del giorno prima. Il capitolo 2 tratta le metodologie statistiche utilizzate nella previsione dei prezzi dell'energia, in più sono forniti i principali indicatori per valutare in primo luogo la bontà di adattamento del modello tenendo in considerazione il principio di parsimonia e successivamente per valutare la bontà previsionale del modello. Il capitolo 3 contiene un approfondimento alla fonte di dati utilizzata e un approfondimento riguardante le principali difficoltà da affrontare per possedere dati di qualità nel settore energetico. Il capitolo 4 infine riporta le analisi sui modelli che sono stati applicati rispondendo agli obbiettivi d'indagine.

Capitolo 1

Sistema elettrico

1.1 Rete elettrica

Da molto tempo ormai l'energia elettrica è di fatto un bene economico cioè un bene immateriale dotato di prezzo volto a soddisfare i bisogni di chi lo utilizza, se da un lato questa sembra una semplice considerazione dall'altro lato il sistema implementato per permettere l'utilizzo di tale prodotto non lo è affatto. L'energia elettrica, infatti, perché possa essere utilizzata necessita innanzitutto di un sistema composto da infrastrutture e da figure societarie che le gestiscono con il fine ultimo di produrre (attraverso una trasformazione di fonti di energia primaria) e trasportare l'energia elettrica fino alle abitazioni di tutte le famiglie presenti in un determinato territorio. Per permettere che vi sia uno scambio economico oltre ad una filiera di trasporto risulta necessaria l'implementazione di una struttura economica composta da una sequenza di mercati che tengano conto dell'organizzazione della rete elettrica e delle caratteristiche intrinseche dell'energia. Tutto questo dà origine al sistema elettrico che per essere compreso risulta necessario analizzare in primo luogo la struttura base della rete elettrica.

La rete elettrica viene costruita e adattata in considerazione del fatto che

1. L'energia elettrica non può essere stoccata né in grandi quantità né nel lungo periodo, non che sia fisicamente impossibile da fare ma il costo risulta essere troppo elevato e quindi sconveniente, questo limite impone di fatto che l'energia generata dev'essere immediatamente immessa in rete;
2. La quantità di energia presente all'interno della rete dev'essere sempre costante questo per evitare squilibri di sistema come i blackout, in considerazione del fatto che nel sistema elettrico sono presenti gli agenti

quali Consumatori finali e produttori di energia il limite in questione definisce che la quantità di energia prodotta e immessa in rete dev'essere costantemente uguale all'energia utilizzata dai consumatori finali.

La rete elettrica è il sistema implementato al fine di riuscire, dalla produzione dell'energia, a trasportare l'energia elettrica fino ai consumatori finali che possono essere abitazioni oppure piccole, medie o grandi imprese, a differenza di altri prodotti l'energia non è semplice da trasportare e per tale ragione necessita di un sistema organizzato in 3 diverse fasi le quali sono: Produzione, Trasmissione e Distribuzione.

1.1.1 Produzione

L'energia elettrica è una fonte di energia secondaria cioè non si trova pronta all'uso in natura ma dev'essere generata a partire da fonti di energia primarie come il sole, il vento, i combustibili ecc. Le centrali elettriche, dette anche Impianti, hanno proprio il compito di utilizzare tali fonti di energia primaria con lo scopo di trasformarle in energia elettrica, le centrali elettriche possono essere classificate in base alla tipologia di fonte primaria utilizzata, alcuni esempi sono le centrali Termoelettriche (la cui fonte primaria è l'energia chimica del combustibile), centrali Idroelettriche (utilizzano l'energia cinetica di grandi masse d'acqua), centrali Geotermiche (utilizzano il calore del suolo), centrali Eoliche (utilizzano il vento), centrali Fotovoltaiche (utilizzano come l'irradiamento solare) ecc.

Risulta possibile fare un'ulteriore classificazione sempre per quanto riguarda la fonte primaria utilizzata ma, in questo caso, distinguendola in fonte rinnovabile o fonte non rinnovabile per estensione è possibile classificare le centrali nelle centrali che utilizzano fonti rinnovabili o centrali che utilizzano fonti non rinnovabili (queste ultime vengono nominate centrali termiche), in particolare una fonte primaria viene considerata rinnovabile quando il suo tasso di riproduzione è istantaneo o avviene su una scala temporale paragonabile al suo tasso di esaurimento (Cretì & Fontini, 2019), in parole povere una fonte è rinnovabile quando è sempre disponibile in natura oppure se la sua velocità di riproduzione coincide con la velocità con cui viene usata dalle centrali per produrre energia, al fine di ottenere una visione più chiara è possibile considerare l'energia proveniente dalle radiazioni solari in questo caso il sole è a tutti gli effetti una fonte di energia rinnovabile in quanto il sole emana istantaneamente radiazioni che possono essere catturate dall'uomo, ma anche la biomassa può essere considerata una fonte di energia rinnovabile a seconda del tasso di riproduzione della stessa rispetto al suo tasso di esaurimento ad esempio l'abbattimento di legno proveniente da foreste a rotazione bre-

ve può essere considerata una forma rinnovabile di produzione dell'energia, mentre l'abbattimento di foreste senza un criterio di rotazione non è affatto una forma rinnovabile di produzione dell'energia, un altro esempio che vale la pena citare sono i rifiuti urbani ovvero essendo tutti d'accordo che la vita sulla terra implica che l'essere umano produca rifiuti questi potrebbero essere considerati una fonte di energia rinnovabile.

Le centrali elettriche dopo aver prodotto l'energia la dovranno subito immettere nella rete attraverso un trasformatore che avrà il compito di alzare la tensione (viene alzata la tensione con lo scopo di ridurre al minimo la dispersione dell'energia che corrisponde a $P = R \cdot I^2$, dove P è la potenza, R la resistenza e I la corrente, diminuendo I e tenendo presente che la potenza è $P = V \cdot I$ per mantenere lo stesso livello di P è necessario aumentare V ovvero la tensione) e da qui in avanti il compito delle centrali elettriche è terminato e subentra la fase della trasmissione.

La fase di produzione si compone quindi di un insieme di produttori che operano in un mercato di tipo concorrenziale, il mercato concorrenziale si caratterizza per il fatto che tutti i produttori producono e vendono ad un unico prezzo ovvero il prezzo di mercato che si stabilisce come equilibrio tra domanda e offerta, il mercato è inoltre a libero accesso cioè, al limite del rispetto di alcune normative ambientali e di sicurezza, chiunque potrà entrare in questo mercato aprendo una centrale elettrica.

La produzione in Italia avviene ancora in gran parte sfruttando fonti non rinnovabili Figura 1.1 (gas naturale, carbone e petrolio), anche se è in continuo aumento lo sviluppo delle fonti rinnovabili: energia geotermica, idroelettrica, solare ed eolica.

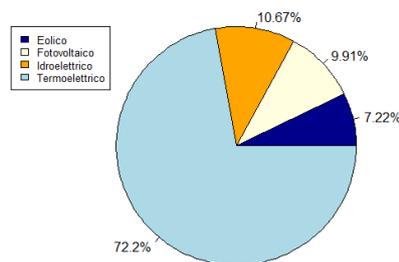


Figura 1.1: Quote di energia elettrica prodotta in Italia suddivise in base alla tipologia di fonte primaria utilizzata (Fonte: TERNA)

1.1.2 Trasmissione

La fase di trasmissione dell'energia elettrica ha il compito di trasportare l'energia dal momento in cui viene prodotta fino alle cabine primarie (facenti parte della fase di distribuzione), in particolare la rete di trasmissione Italiana è composta da 431 stazioni di trasformazione e smistamento che collegano tra loro linee elettriche allo stesso livello di tensione e ripartiscono tra loro l'energia elettrica e 74442 km di linee elettriche che consentono il trasporto dell'energia in lunga distanza ad alta tensione per minimizzare la dissipazione.

La figura adibita alla gestione della fase di trasmissione è il *Transmission System Operator* (TSO), in Italia è TERNA, la quale provvede, oltre ad occuparsi dei servizi di manutenzione della rete di trasmissione e dell'ampliamento della stessa, al servizio di dispacciamento che consiste nella gestione dei flussi della rete, considerato il fatto che l'energia presente all'interno della rete dev'essere sempre la stessa e non è stoccabile in grandi quantità il TSO avrà il compito di coordinare le centrali elettriche in modo da mantenere il sistema funzionante, questo lavoro verrà fatto tenendo in considerazione le contrattazioni che vengono svolte prima della consegna fisica dell'energia. Nei principali mercati europei la fase di trasmissione viene gestita da un unico soggetto in regime di monopolio naturale, esiste un monopolio naturale quando una singola impresa può produrre un prodotto a prezzi più bassi rispetto al caso in cui nello stesso mercato vi partecipino più imprese, questo avviene semplicemente perché nelle linee elettriche i costi fissi sono molto elevati ma i costi marginali sono praticamente nulli infatti una volta che la linea viene costruita il costo marginale di servire una famiglia supplementare con quella stessa linea è praticamente nullo pertanto maggiore è il numero di famiglie che possono essere servite con un singolo filo, minori saranno i costi poiché il costo fisso viene ripartito sul numero dei clienti.

Ad onor del vero è possibile fare delle distinzioni più precise per quanto riguarda la figura del TSO, è stato detto che il TSO si occupa del servizio di dispacciamento e dei servizi di manutenzione delle reti e dell'ampliamento delle stesse in modo da garantire un corretto funzionamento dell'intero sistema, in realtà il TSO è il soggetto che è sia proprietario della rete sia colui che si occupa di dispacciare l'energia ma è possibile (non nel caso italiano) che vi sia un proprietario della rete separato da colui che gestisce i flussi di energia e in questo caso si chiamerà *System Operator* (SO), il proprietario invece avrà il compito di mantenere e ampliare la rete (Cretì & Fontini, 2019).

1.1.3 Distribuzione

La fase di distribuzione è la fase finale della rete elettrica cioè corrisponde all'ultimo tragitto che l'energia deve fare per arrivare fino all'utente finale, in particolare la rete di distribuzione è anch'essa costituita da un insieme molto fitto di linee elettriche che partono dalle 'cabine primarie' che sono i nodi di interconnessione tra la rete di trasmissione e la rete di distribuzione in particolare queste cabine segnano il confine tra le due reti, le cabine primarie sono un impianto elettrico che ha la funzione di trasformare l'energia in ingresso ad alta tensione in energia a media tensione (In Italia sono presenti circa 2000 cabine primarie). Dalle cabine primarie vi sono due possibilità:

1. Tenendo presente che ora l'energia è in media tensione vi possono essere alcuni utenti finali come grandi industrie o centri commerciali che utilizzano direttamente energia in media tensione e quindi la linea elettrica arriverà direttamente nella sede di questi utenti finali;
2. D'altro canto vi sono utenti finali come le singole abitazioni i cui elettrodomestici o strumenti di vario genere non sono in grado di gestire una media tensione, perciò la linea elettrica non andrà direttamente nelle loro case ma passerà attraverso una 'cabina secondaria' che avrà anch'essa il compito di abbassare la tensione (da media a bassa tensione), dopodiché dalla cabina secondaria la linea elettrica arriverà a casa dell'utente finale.

In questa fase non c'è un unico soggetto in regime di monopolio come nel caso della fase di trasmissione ma vi sono dei soggetti chiamati *Distribution system operator* (DSO) che sono proprietari di un tratto di rete, oltre quindi al compito di garantire il servizio di distribuzione, vale a dire garantire la qualità delle reti attraverso servizi di manutenzione dovrà anche occuparsi della vendita al dettaglio dell'energia, la vendita al dettaglio è l'attività di vendita di elettricità agli utenti finali a tal fine però un prerequisito è che l'elettricità utilizzata sia in qualche modo misurata questo compito spetta ai contatori, la vendita al dettaglio comunque non è sempre svolta dal DSO stesso ma dai dettaglianti che acquistano energia dai DSO rendendo competitivo il mercato e quindi facendo diventare i carichi a tutti gli effetti dei clienti e vengono trattati come tali, utilizzando tutti i possibili strumenti di marketing quali il prezzo, le relazioni con i clienti, garantendo servizi di efficienza energetica e così via.

1.2 Mercato energetico

All'interno di questa sezione verrà descritta la struttura di mercato del sistema elettrico comune ai principali stati Europei, quindi verrà descritto il modo in cui interagiscono i partecipanti del sistema elettrico attraverso i mercati di cui esso è composto, verrà quindi fatta una panoramica dei mercati presenti all'interno del sistema con un principale riferimento alla borsa elettrica.

Il modello di mercato descritto in questa sezione non è l'unico modello applicabile, in altri stati infatti vi possono essere sistemi gestionali molto diversi ad esempio modelli in cui la gestione del sistema elettrico spetta ad un'unica società, in Italia tra l'altro il modello che abbiamo attualmente non corrisponde a quello esistente prima degli anni Novanta, in quel periodo ENEL gestiva tutto il sistema dalla fase di produzione alla fase di distribuzione compresa la vendita a dettaglio, a partire dal 1991 però si è proceduto verso la cosiddetta Liberalizzazione del mercato che ci ha portati al modello di gestione del sistema elettrico attualmente è in vigore.

L'attuale struttura gestionale che organizza i principali mercati Europei prevede un modello completamente liberalizzato nel senso che tutti i compiti svolti dagli agenti del sistema sono disaggregati cioè attribuiti a figure/società distinte l'una dall'altra. Ricordando che l'energia ha delle caratteristiche intrinseche che non le permettono di essere trasportata come viene trasportato un comune prodotto materiale, l'energia necessita infatti di un insieme di servizi quali il dispacciamento, servizi di trasmissione (che includono la manutenzione della rete, l'ampliamento della stessa ecc.) e servizi di distribuzione (che includono la vendita al dettaglio), questi servizi ovviamente hanno un costo che viene remunerato dal carico ma che viene tenuto separato dal costo dell'energia elettrica in particolare non sarà né il TSO né il DSO a stabilire il prezzo dell'energia elettrica ma saranno loro a stabilire il costo che l'utente finale dovrà sostenere per avergli trasportato l'energia dalla produzione fino alla sua abitazione.

Il prezzo dell'energia elettrica si formerà invece all'interno della 'borsa elettrica' ovvero il luogo in cui avviene l'incontro tra domanda e offerta per la compravendita dell'energia elettrica all'ingrosso, questo risulta essere un mercato centralizzato nel senso che vi è un organo che organizza le regole e le caratteristiche di coloro che vi partecipano, quest'organo viene chiamato 'Gestore del mercato dell'energia' (GME). Successivamente alla determinazione delle regole di scambio dell'energia e di formazione del prezzo (che saranno spiegate nel capitolo successivo) il Gestore del mercato dovrà garantire un continuo e organizzato svolgimento delle contrattazioni, successivamente alla stipulazione dei contratti il GME avrà il compito di comunicarli al TSO che, alla luce delle contrattazioni avvenute, sarà in grado di fornire il servizio di

dispacciamento garantendo così un corretto scambio di energia.

1.3 Borsa elettrica

La borsa elettrica, nota anche come mercato elettrico o mercato dell'energia, è un ambiente di scambio elettronico in cui vengono negoziati e commercializzati contratti per l'acquisto e la vendita di energia elettrica. Essa costituisce il luogo centrale dove produttori, distributori, rivenditori e grandi consumatori si incontrano per scambiarsi l'energia elettrica. Ricordando i limiti dell'energia elettrica questi portano inevitabilmente a far sì che le contrattazioni all'interno della borsa elettrica si svolgono prima della consegna fisica prevista ed è necessario che si svolgano in un arco temporale sempre più vicino al tempo dell'immissione in rete dell'energia. In aggiunta a ciò nella borsa elettrica vi sono anche mercati in cui il servizio che viene scambiato non è l'energia bensì tutti quei servizi ausiliari alla trasmissione di tale energia, alla luce di questo i mercati presenti nella borsa elettrica possono essere classificati in base a 2 variabili:

1. Il tipo di bene o servizio che viene scambiato ovvero l'energia, i servizi ancillari e i servizi di capacità;
2. La distanza temporale tra il momento dello scambio contrattuale e il momento della consegna fisica dell'energia elettrica.

per quanto concerne i servizi ancillari e i servizi di capacità sono quei servizi che consentono al TSO di poter svolgere le sue funzioni di dispacciamento, in particolare in relazione al fatto che il TSO formula un programma di dispacciamento prima della consegna fisica dell'energia, basato sulle contrattazioni avvenute precedentemente che tengono conto della curva di domanda, non è detto sia perfetto, durante la consegna fisica dell'energia infatti ci potrebbero essere delle fluttuazioni dei carichi o variazioni dal lato della produzione, in questi casi il TSO farà affidamento ai servizi ancillari (come i servizi di bilanciamento) che aiutano a far fronte a questi problemi in tempo reale, ovviamente anche questi servizi hanno dei mercati all'interno della borsa elettrica e le contrattazioni tra il TSO e i produttori che intendono offrire tali servizi avvengono anch'esse prima della consegna fisica dell'energia, in alcuni casi vi sono anche contratti annuali.

1.4 Mercati dell'energia elettrica da un punto di vista temporale

In questa sezione verranno analizzati i mercati, il cui bene scambiato è l'energia elettrica, da un punto di vista temporale, prima di entrare nella fattispecie tecnica dei mercati risulta necessario fare una precisazione, l'energia scorre nella rete in un tempo continuo ma per quanto riguarda i mercati ciò non può avvenire nel senso che le contrattazioni devono essere avviate a un punto temporale specifico, concluse in un altro punto del tempo e riferirsi ad un intervallo temporale definito, preso in considerazione questo aspetto i principali mercati Europei suddividono il piano temporale in fasce orarie e gli scambi di energia elettrica vengono effettuati in relazione ad uno specifico slot orario e ovviamente prima di quel determinato slot orario.

Con riferimento ad una determinata fascia oraria è possibile osservare la sequenza temporale dei mercati che inizia con il 'mercato del giorno prima' il cui nome risulta essere già evocativo del fatto che le contrattazioni in questo mercato avvengono esattamente il giorno precedente l'immissione in rete dell'energia elettrica, in Europa il sopraccitato mercato viene tradizionalmente aperto il mattino e chiuso all'incirca verso le 12:00, il mercato del giorno prima è il luogo in cui domanda e offerta si incontrano formando il prezzo orario dell'energia.

Al termine del mercato del giorno prima verrà aperto il 'mercato infragiornaliero' la cui chiusura, denominata 'chiusura del gate', avverrà un'ora prima della fascia oraria presa a riferimento che coincide con l'immissione in rete dell'energia. Questo mercato non ha lo scopo di definire il prezzo dell'energia ma si prefigge come obiettivo quello di permettere ai produttori che hanno partecipato al mercato del giorno prima, ed è stata accettata la loro offerta, di aggiustare le loro posizioni, potrebbe accadere infatti che un impianto, dopo aver fatto un'offerta sul mercato del giorno prima e dopo che quell'offerta è stata accettata dal meccanismo di mercato, scopra di non poter consegnare tutta quell'energia, questo può succedere perché l'offerta (in termini di quantità) proposta dal suddetto impianto è stata fatta il giorno prima quindi non nel tempo reale e teniamo presente che potrebbero esserci problemi strutturali (o costi molto alti da parte della centrale) qual ora quell'energia non venisse consegnata, qui entra in gioco il mercato infragiornaliero che permette a questo produttore di fare delle contrattazioni ad esempio con un'altra centrale che fornirà l'energia elettrica che manca al suo posto.

Oltre al mercato infragiornaliero e il mercato del giorno prima vi possono essere i contratti bilaterali che, nel contesto della borsa elettrica, sono accordi o contratti tra due parti indipendenti per l'acquisto o la vendita di energia

elettrica a un prezzo e in una quantità predeterminati. Questi contratti sono negoziati direttamente tra i partecipanti al mercato elettrico, come produttori, distributori o grandi consumatori, senza coinvolgere la borsa elettrica come intermediario, un tipico esempio possono essere le grandi aziende con grandi unità produttive quali Amazon, Nike, Google ecc. queste società negoziano accordi di medio -lungo termine direttamente con i fornitori.

1.5 Mercato del giorno prima

Vista l'importanza del mercato del giorno prima nella generazione del prezzo dell'energia e nella formazione dei programmi di immissione in rete questa sezione tratterà più nel dettaglio le regole che insistono nel mercato. Il mercato del giorno prima è organizzato secondo la forma della cosiddetta 'asta doppia' ovvero è un'asta dove entrambe le parti offrono dei prezzi in cambio delle quantità di energia in questo modo è possibile costruire delle curve di domanda e di offerta. In merito alla curva di domanda essa si forma attraverso le società di distribuzione che agiscono come intermediari tra i consumatori e il mercato dell'energia elettrica, i DSO quindi parteciperanno al mercato facendo delle offerte le quali saranno poi posizionate in modo decrescente, in base al prezzo, formando la curva di domanda. La curva di domanda risulta una curva piuttosto anelastica ovvero una curva quasi verticale per la quale un aumento di prezzo comporta una leggera diminuzione delle quantità questo fenomeno è dettato dal fatto che, indipendentemente dal prezzo, una certa quantità di energia viene sempre utilizzata dai carichi, si pensi ai macchinari ospedalieri o industriali, o più semplicemente ad apparecchi elettronici come il frigorifero o la televisione, in virtù di questo fenomeno la curva di domanda avrà una forma piuttosto verticale che potrà essere spostata più o meno verso destra mantenendone comunque la sua struttura. La curva di offerta invece è formata in base alle offerte fatte dalle società produttive, i produttori infatti dovranno proporre all'interno del mercato il prezzo e la quantità di energia che sono disposti a produrre in quella determinata fascia oraria, tipicamente il prezzo in corrispondenza del quale un impianto accetta di produrre energia è il prezzo che copre i suoi costi marginali che sono di fatto i costi del combustibile (per quanto riguarda i combustibili fossili il fatto di bruciare ad esempio un metro cubo di gas produce anidride carbonica e in Europa abbiamo un regolamento che obbliga a pagare per il danno ambientale che producono aumentando il costo marginale), la curva di offerta sarà dunque una curva a gradini dove ogni gradino corrisponderà ad un produttore, l'altezza del gradino darà come informazione il prezzo dell'energia per il quale la centrale è disposta a produrre ovvero il costo di un unità di combustibile e

la lunghezza del gradino invece darà come informazione la quantità massima di energia che il produttore è disposto a generare. Ordinando in modo crescente questi gradini si ottiene la curva di offerta di mercato, ovviamente nel momento in cui i produttori manifestano la loro disponibilità a vendere una determinata quantità di energia ad un determinato prezzo e la loro offerta viene accettata essa dà luogo ad un impegno vincolante ad immettere in rete la quantità concordata.

Tenendo in considerazione la curva di domanda e la curva di offerta relative ad una determinata fascia oraria del giorno successivo il punto di equilibrio sarà il punto di intersezione tra le due curve e proprio in quel punto si formerà il prezzo dell'energia elettrica il quale viene chiamato *System Marginal Price* (SMP) Figura 1.2. Il meccanismo di prezzo vigente sulla borsa elettrica italiana e sui principali paesi Europei prevede che per ciascuna fascia oraria venga costruita la curva di domanda e di offerta di mercato, venga identificato il prezzo di equilibrio (SMP) e la quantità di equilibrio dopodiché l'intera produzione di energia elettrica relativa a quella fascia oraria viene remunerata al prezzo di equilibrio, in questo modo tutte le centrali che hanno costi marginali più alti del SMP non produrranno energia in quanto la loro offerta non viene accettata, la centrale che produce energia in corrispondenza del prezzo di equilibrio avrà dei costi marginali uguali al prezzo di equilibrio perciò al massimo potrà sostenere i suoi costi variabili mentre tutte le centrali che hanno costi marginali più bassi del prezzo di equilibrio riceveranno un prezzo maggiore rispetto al loro costo marginale generando un margine di contribuzione unitario con il quale copriranno i costi fissi (Weron, 2006).

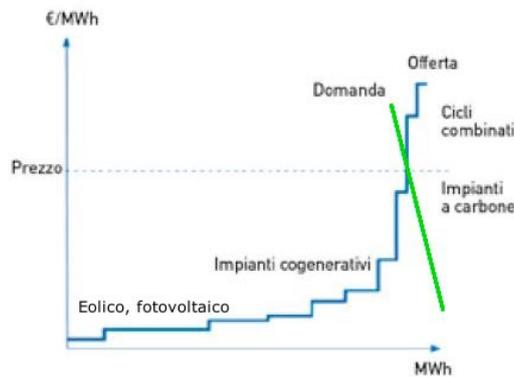


Figura 1.2: Meccanismo di formazione del prezzo nel mercato del giorno prima attraverso l'incrocio tra la curva di domanda e la curva di offerta (Fonte: ingDemurtas)

In questo caso è possibile pensare che le centrali con costi marginali abbastanza alti non produrranno mai, in alcuni casi è vero infatti gli impianti ad olio combustibile sono ormai stati dismessi per gli elevati costi marginali, in altri casi invece non è così cioè dal momento che il mercato del giorno prima si riferisce ad una determinata fascia oraria del giorno successivo la curva di domanda tenderà a cambiare nelle varie fasce orarie, tenderà a spostarsi verso destra nelle ore in cui le industrie avviano la loro produzione mentre tenderà a spostarsi verso sinistra nelle ore notturne quindi in diverse situazioni abbiamo diverse curve di domanda e più la domanda sarà spostata verso destra (a parità di curva di offerta) più verranno coinvolte nella produzione le centrali con costi marginali più alti.

Un aspetto significativo tiene in considerazione le centrali che utilizzano fonti di energia rinnovabile per produrre energia, come spiegato in precedenza, la curva di offerta viene costruita predisponendo le offerte dei produttori in termini di prezzo crescente (che coincide con il costo marginale del combustibile) per poi fare una selezione in base ad un criterio di merito (Creti & Fontini, 2019), tenendo in considerazione il fatto che il costo delle fonti rinnovabili è nullo i produttori che utilizzano fonti rinnovabili offrono sul mercato una certa quantità ad un prezzo praticamente nullo, in funzione delle regole di borsa, inserendo le energie rinnovabili, la curva di offerta subirà uno spostamento verso destra abbassando il prezzo dell'energia. Un limite delle fonti rinnovabili però è la loro variabilità, in altri termini la quantità di energia producibile con un impianto fotovoltaico o eolico dipende dalla presenza di sole o di vento quindi se un impianto a gas può, al netto di guasti tecnici, produrre la stessa quantità di energia ad ogni ora del giorno questo non è possibile per gli impianti ad energia rinnovabile, questo fenomeno ha un risvolto negativo anche sui prezzi. Le centrali elettriche, infatti, possono essere descritte mediante i loro costi i quali vengono suddivisi in costi variabili e costi fissi, alcune centrali hanno alti costi fissi e bassi costi variabili (centrali fotovoltaiche o eoliche) altre hanno bassi costi fissi e alti costi variabili (centrali a gas) altre ancora si trovano nella media (centrali a carbone), la presenza di centrali ad energia rinnovabile ha un risvolto negativo sulle centrali con costi variabili e fissi nella media poiché diminuendo il prezzo e la loro produzione non sono più in grado, come prima, di sostenere i propri costi fissi rendendole meno competitive, se quindi queste centrali sparissero dal mercato nel momento in cui le centrali rinnovabili non sono più in grado di produrre energia entrano in esercizio le centrali con costi marginali più alti alzando il prezzo, in conclusione la forte presenza di rinnovabili nel sistema potrebbe anche far scendere in media il prezzo dell'energia ma aumentandone la variabilità.

Quando non c'è congestione che, nel contesto dell'energia elettrica, si riferisce a una situazione in cui la capacità di trasmissione disponibile nella rete elettrica non è sufficiente per soddisfare la domanda di energia o per garantire un flusso efficiente dell'energia tra le diverse zone o regioni, SMP è l'unico prezzo per l'intero sistema (Weron, 2006), questo non è così nel caso in cui ci siano problemi di congestione per i quali il sistema elettrico di un paese viene suddiviso in zone. Il mercato Italiano dell'energia elettrica è un chiaro esempio di 'mercato zonale' (Grossi et.al, 2022) le quali zone sono Nord (NOR), Centro-Nord (CNOR), Centro-Sud (CSOU), Sud (SOU), Sicilia (SIC), Sardegna (SAR) e Calabria, in Italia i partecipanti al mercato fanno offerte per ciascuna zona e fascia oraria nel mercato del giorno prima, e i prezzi di equilibrio sono stabiliti in base a queste offerte specifiche. Il prezzo complessivo dell'energia elettrica viene poi determinato attraverso una media ponderata dei prezzi nelle diverse zone questo è noto come il "Prezzo Medio Ponderato Zonale" (PMPZ) che è calcolato considerando i prezzi di equilibrio nelle diverse zone e pesandoli in base alla quantità di energia scambiata in ciascuna zona.

Capitolo 2

Metodologie

2.1 Notazione

Nella presente sezione vengono introdotte le notazioni che saranno utilizzate nelle sezioni successive: si considera Y il processo stocastico generatore della serie storica y_1, \dots, y_n dove n corrisponde al numero di elementi della serie, mentre con Y_t si considera la variabile casuale, facente parte del processo stocastico, al tempo t .

con il simbolo \mathcal{E} viene indicato il processo stocastico 'White noise' ovvero un processo a media nulla $E(\mathcal{E}_t) = 0 \forall t$, con varianza pari a $Var(\mathcal{E}_t) = \sigma_\epsilon^2 \forall t$ e a componenti incorrelate $Cov(\mathcal{E}_t, \mathcal{E}_{t-s}) = 0 \forall t, s$ con $s \neq 0$, tutto questo può essere riassunto con la notazione $\mathcal{E}_t \sim \mathcal{WN}(0, \sigma_\epsilon^2)$.

Viene indicato con B l'operatore ritardo (*Backflip operator*) che, se applicato ad una variabile del processo stocastico, permette di considerare il suo ritardo cioè $BY_t = Y_{t-1}$, più in generale $B^d Y_t = Y_{t-d}$.

Nell'ambiente dei processi stocastici si fa riferimento specialmente alla funzione di autocovarianza e alla funzione di autocorrelazione, queste vengono indicate rispettivamente con il simbolo γ e ρ , inoltre è importante definire il 'lag' ovvero lo scostamento temporale tra due variabili casuali quindi se consideriamo la variabile casuale al tempo t e una variabile casuale al tempo s il lag sarà $k = |t - s|$.

2.2 Processi stocastici

Nella presente sezione verrà introdotto uno strumento fondamentale nell'approccio statistico all'analisi di dati temporali ovvero i processi stocastici a componenti correlate, i processi stocastici sono un insieme infinito di variabili casuali indicizzate da un indice temporale, legate tra loro da una struttura

di correlazione e di cui la serie storica è realizzazione di una parte finita del processo.

Un processo stocastico è definito da uno ‘spazio di probabilità’ ovvero da una terna (Ω, F, P) dove Ω corrisponde allo spazio degli eventi elementari formato da tutti i possibili esiti del processo, com’è noto nel caso di una variabile casuale lo spazio degli eventi elementari coincide con il supporto della stessa mentre in un processo stocastico lo spazio delle probabilità è rappresentato dall’insieme di tutte le possibili traiettorie o realizzazioni che il processo può assumere nel tempo. F invece corrisponde alla σ - algebra che rappresenta l’insieme di tutti gli eventi che possono essere misurati o osservati lungo le traiettorie del processo nel tempo. Infine vi è P ovvero la misura di probabilità sugli eventi contenuti in Ω .

Pertanto ogni $\omega \in \Omega$ si può vedere come una possibile realizzazione del processo stocastico. Un processo aleatorio si presenta quindi come una funzione a due variabili (ω, t) a valori in \mathbb{R} ovvero:

$$Y(\omega, t) : \Omega \times T \rightarrow \mathbb{R} \quad (2.1)$$

fissando $\omega = \omega_0$ si ottiene la funzione $Y(\omega_0, t)$ che rappresenta una realizzazione del processo al variare di $t \in T$.

2.2.1 Stazionarietà

È possibile caratterizzare un processo stocastico mediante i suoi momenti in particolare momento primo, momento secondo e autocovarianza, nello specifico è possibile dire che il momento primo del processo stocastico risulta essere $E(Y_t) = \mu_t$ ovvero la traiettoria che passa per le medie delle variabili casuali del processo, è possibile definire la varianza come $Var(Y_t) = \sigma_t^2$ e infine l’autocovarianza come $Cov(Y_t, Y_{t-s}) = \gamma_{t,s}$.

Alla luce di questa caratterizzazione è possibile definire la stazionarietà in senso debole, intuitivamente un processo stocastico è stazionario in senso debole quando il suo comportamento risulta invariante nel tempo, l’assunzione di stazionarietà in senso debole coinvolge i momenti del processo stocastico in particolare si assume che il momento primo e momento secondo non dipendono dal tempo e la covarianza, che esprime il legame lineare esistente tra due variabili, non dipende dal tempo ma dal numero di istanti temporali che separa le due variabili. Quindi un processo è detto stazionario in senso debole se soddisfa le seguenti condizioni:

1. $E(Y_t) = \mu$
2. $Var(Y_t) = \sigma^2$
3. $Cov(Y_t, Y_{t-k}) = \gamma_k$

2.3 Processi ARMA

In questa sezione saranno introdotti una classe di processi stocastici che giocano un ruolo chiave nella modellizzazione di una serie storica, i processi ARMA(p, q), questi processi si caratterizzano per il fatto che mettono assieme i processi autoregressivi e i processi a media mobile, prima quindi di trattare il caso generale dei processi ARMA(p, q) verranno trattati i due casi specifici. Nella presente sezione saranno esaminati solamente processi stocastici debolmente stazionari.

2.3.1 Processi MA e AR

Viene richiamata nella presente sezione la notazione utilizzata per indicare un processo a media 0, varianza σ_ϵ^2 e componenti incorrelate ovvero $\mathcal{E}_t \sim \mathcal{WN}(0, \sigma_\epsilon^2)$, viene inoltre richiamato l'operatore ritardo B tale che $B^d Y_t = Y_{t-d}$.

Risulta possibile dare la definizione di un processo a Media mobile di ordine q , infatti un processo è detto a Media mobile di ordine q , la cui notazione è MA(q), se può essere scritto nella forma:

$$Y_t = \mu + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1} + \dots + \theta_q \mathcal{E}_{t-q} \quad (2.2)$$

Con l'ausilio dell'operatore ritardo è possibile riscrivere il modello come segue:

$$Y_t = \mu + \theta(B) \mathcal{E}_t \quad (2.3)$$

dove $\theta(B)$ è il polinomio caratteristico di grado q dato da

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (2.4)$$

Il momento primo di un processo stocastico MA(q) risulta essere:

$$\begin{aligned} E(Y_t) &= E\left(\mu + \sum_{i=1}^q \theta_i \mathcal{E}_{t-i}\right) \\ &= \mu + \sum_{i=1}^q \theta_i E(\mathcal{E}_{t-i}) \\ &= \mu \end{aligned} \quad (2.5)$$

La varianza di un processo stocastico MA(q) risulta essere:

$$\begin{aligned}
Var(Y_t) &= Var\left(\mu + \sum_{i=1}^q \theta_i \mathcal{E}_{t-i}\right) \\
&= Var\left(\sum_{i=1}^q \theta_i \mathcal{E}_{t-i}\right) \\
&= \sum_{i=1}^q Var(\theta_i \mathcal{E}_{t-i}) \\
&= \sigma_\epsilon^2 \left(1 + \sum_{i=1}^q \theta_i^2\right)
\end{aligned} \tag{2.6}$$

La covarianza di un processo stocastico MA(q) risulta essere:

$$\begin{aligned}
Cov(Y_t, Y_{t-k}) &= Cov(\mu + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1} + \dots + \theta_q \mathcal{E}_{t-q}, \mu + \mathcal{E}_{t-k} + \theta_1 \mathcal{E}_{t-1-k} + \dots + \theta_q \mathcal{E}_{t-q-k}) \\
&= \begin{cases} 0 & \text{se } k > q \\ \sigma_\epsilon^2 (\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) & \text{se } k = 1, \dots, q \\ \sigma_\epsilon^2 (1 + \sum_{i=1}^q \theta_i^2) & \text{se } k = 0 \end{cases}
\end{aligned} \tag{2.7}$$

Data questa struttura della covarianza è possibile ricavare la funzione di autocorrelazione di un processo stocastico MA(q):

$$\rho_k = \begin{cases} 0 & \text{se } k > q \\ \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \sum_{i=1}^q \theta_i^2} & \text{se } k = 1, \dots, q \\ 1 & \text{se } k = 0 \end{cases} \tag{2.8}$$

Come si può notare dalla funzione di autocorrelazione un processo a media mobile presenta un legame lineare con le variabili casuali che si trovano ad un numero di istanti temporali al massimo di $\pm q$ e diminuisce esponenzialmente all'aumentare del ritardo, il legame lineare si interrompe nel momento in cui il ritardo risulta superiore a q .

Si può proseguire ulteriormente dando la definizione di un processo di tipo Autoregressivo di ordine p , infatti un processo è detto Autoregressivo di ordine p , la cui notazione è AR(p), se può essere scritto nella forma:

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \mathcal{E}_t \tag{2.9}$$

Con l'ausilio dell'operatore ritardo è possibile riscrivere il modello come segue:

$$\phi(B)Y_t = \mu + \mathcal{E}_t \tag{2.10}$$

dove $\phi(B)$ è il polinomio caratteristico di grado p dato da

$$\phi(B) = 1 + \phi_1 B + \dots + \phi_p B^p \quad (2.11)$$

Per analizzare le caratteristiche di un processo autoregressivo consideriamo il caso in cui $p = 1$, perciò il momento primo di un processo stocastico AR(1) risulta essere:

$$\begin{aligned} E(Y_t) &= E(\mu + \phi_1 Y_{t-1} + \mathcal{E}_t) \\ &= \mu + \phi_1 E(Y_{t-1}) + E(\mathcal{E}_t) \\ &= \mu + \phi_1 E(Y_{t-1}) \end{aligned}$$

Recuperando l'assunzione di stazionarietà tale che la media del processo stocastico non dipende da t allora $E(Y_t) = E(Y_{t-1})$ quindi:

$$\begin{aligned} E(Y_t) - \phi_1 E(Y_{t-1}) &= \mu \\ E(Y_t) \cdot (1 - \phi_1) &= \mu \\ E(Y_t) &= \frac{\mu}{1 - \phi_1} \end{aligned} \quad (2.12)$$

La varianza di un processo stocastico AR(1) risulta essere:

$$\begin{aligned} Var(Y_t) &= Var(\mu + \phi_1 Y_{t-1} + \mathcal{E}_t) \\ &= \phi_1^2 Var(Y_{t-1}) + Var(\mathcal{E}_t) \\ &= \phi_1^2 Var(Y_{t-1}) + \sigma_\epsilon^2 \end{aligned}$$

Anche in questo caso è possibile recuperare l'assunzione di stazionarietà tale che la varianza del processo stocastico non dipende da t allora $Var(Y_t) = Var(Y_{t-1})$ quindi:

$$\begin{aligned} Var(Y_t) - \phi_1^2 Var(Y_t) &= \sigma_\epsilon^2 \\ Var(Y_t) \cdot (1 - \phi_1^2) &= \sigma_\epsilon^2 \\ Var(Y_t) &= \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \end{aligned} \quad (2.13)$$

La covarianza di un processo stocastico AR(1) considerando $\mu = 0$ risulta essere:

$$\begin{aligned} Cov(Y_t, Y_{t-k}) &= Cov(\phi_1 Y_{t-1} + \mathcal{E}_t, \phi_1 Y_{t-k-1} + \mathcal{E}_{t-k}) \\ &= \begin{cases} \frac{\sigma_\epsilon^2}{1 - \phi_1^2} & \text{se } k = 0 \\ \phi_1^k \gamma_0 & \text{se } k \neq 0 \end{cases} \end{aligned} \quad (2.14)$$

Data questa struttura della covarianza è possibile ricavare la funzione di autocorrelazione di un processo stocastico AR(1) con $\mu = 0$:

$$\rho_k = \begin{cases} 1 & \text{se } k = 0 \\ \phi_1^k & \text{se } k \neq 0 \end{cases} \quad (2.15)$$

Come si può notare dalla funzione di autocorrelazione, tenendo in considerazione il fatto che $|\phi_1| < 1$, ogni variabile casuale facente parte del processo possiede un legame lineare con ogni variabile casuale, la correlazione diminuisce esponenzialmente all'aumentare di k e risulta pari a 0 solamente per $k \rightarrow \infty$.

2.3.2 Processi ARMA

Un processo ARMA(p, q) è un processo ottenuto unendo in un unico modello una componente Autoregressiva di ordine p e una componente a Media mobile di ordine q , quindi un processo è detto autoregressivo a media mobile, la cui notazione è ARMA(p, q), se può essere scritto nella forma:

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1} + \dots + \theta_q \mathcal{E}_{t-q} \quad (2.16)$$

Con l'ausilio dell'operatore ritardo è possibile riscrivere il modello come segue:

$$\phi(B)Y_t = \mu + \theta(B)\mathcal{E}_t \quad (2.17)$$

dove $\phi(B)$ e $\theta(B)$ sono i polinomi caratteristici di grado rispettivamente p e q e sono così definiti:

$$\begin{aligned} \phi(B) &= 1 + \phi_1 B + \dots + \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \end{aligned} \quad (2.18)$$

Per analizzare le caratteristiche di un processo autoregressivo consideriamo il caso in cui $p = 1$ e $q = 1$, perciò il momento primo di un processo stocastico ARMA(1, 1) risulta essere:

$$\begin{aligned} E(Y_t) &= E(\mu + \phi_1 Y_{t-1} + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1}) \\ &= \mu + \phi_1 E(Y_{t-1}) + E(\mathcal{E}_t) + \theta_1 E(\mathcal{E}_{t-1}) \\ &= \mu + \phi_1 E(Y_{t-1}) \\ &= \frac{\mu}{1 - \phi_1} \end{aligned} \quad (2.19)$$

La varianza di un processo ARMA(1, 1) risulta essere:

$$\begin{aligned} Var(Y_t) &= Var(\mu + \phi_1 Y_{t-1} + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1}) \\ &= \phi_1^2 Var(Y_{t-1}) + Var(\mathcal{E}_t) + \theta_1^2 Var(\mathcal{E}_{t-1}) + 2\phi_1 \theta_1 Cov(Y_{t-1}, \mathcal{E}_{t-1}) \end{aligned}$$

Per procedere con il calcolo della varianza è necessario sviluppare il termine di covarianza:

$$\begin{aligned} Cov(Y_{t-1}, \mathcal{E}_{t-1}) &= Cov(\mu + \phi_1 Y_{t-2} + \mathcal{E}_{t-1} + \theta_1 \mathcal{E}_{t-2}, \mathcal{E}_{t-1}) \\ &= Cov(\mathcal{E}_{t-1}, \mathcal{E}_{t-1}) \\ &= \sigma_\epsilon^2 \end{aligned}$$

Risulta possibile procedere con il calcolo della varianza:

$$\begin{aligned} Var(Y_t) &= \phi_1^2 Var(Y_{t-1}) + \sigma_\epsilon^2 + \theta_1^2 \sigma_\epsilon^2 + 2\phi_1 \theta_1 \sigma_\epsilon^2 \\ &= \frac{\sigma_\epsilon^2 (1 + 2\phi_1 \theta_1 \sigma_\epsilon^2 + \theta_1^2)}{1 - \phi_1^2} \end{aligned} \quad (2.20)$$

La covarianza di un processo stocastico ARMA(1, 1) considerando $\mu = 0$ risulta essere:

$$\begin{aligned} Cov(Y_t, Y_{t-k}) &= Cov(\mu + \phi_1 Y_{t-1} + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1}, \mu + \phi_1 Y_{t-1-k} + \mathcal{E}_{t-k} + \theta_1 \mathcal{E}_{t-1-k}) \\ &= \begin{cases} \frac{\sigma_\epsilon^2 (1 + 2\phi_1 \theta_1 \sigma_\epsilon^2 + \theta_1^2)}{1 - \phi_1^2} & \text{se } k = 0 \\ \sigma_\epsilon^2 \left[\frac{(\phi_1 - \theta_1)(1 - \theta_1 \phi_1)}{1 - \phi_1^2} \right] & \text{se } k = 1 \\ \phi_1^{k-1} \gamma_1 & \text{se } k \geq 2 \end{cases} \end{aligned} \quad (2.21)$$

Data questa struttura della covarianza è possibile ricavare la funzione di autocorrelazione di un processo stocastico ARMA(1, 1) con $\mu = 0$:

$$\rho_k = \begin{cases} 1 & \text{se } k = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \theta_1 \phi_1)}{1 - \phi_1^2} & \text{se } k = 1 \\ \phi_1^{k-1} \rho_1 & \text{se } k \geq 2 \end{cases} \quad (2.22)$$

Dalla funzione di autocorrelazione è possibile notare che il legame tra 2 variabili casuali, facenti parte del processo astocastico ARMA(1, 1), che si trovano ad una distanza di $q = 1$ è governato sia dalla componente autoregressiva sia dalla componente a media mobile, successivamente per $k > q$ l'autocorrelazione è governata unicamente dalla componente autoregressiva.

2.4 Processi ARIMA

I processi stocastici ARMA tengono in considerazione l'assunzione di stazionarietà, le serie storiche reali però non sempre possono essere considerate come una realizzazione di un processo ARMA(p, q) stazionario, è possibile infatti che non siano stazionarie in media il che comporta l'impossibilità di

assumere una media costante al variare del tempo. Per trattare queste situazioni è possibile fare affidamento ad un processo autoregressivo-integrato a media mobile, la cui notazione è ARIMA(p, d, q), che permette di rendere stazionario un modello attraverso un numero d di differenze prime (d corrisponde all'ordine del modello) in particolare se la media del processo stocastico segue un andamento polinomiale del tempo applicando un numero di differenze prime pari all'ordine del polinomio è possibile ottenere un processo stocastico stazionario sul quale applicare poi un modello ARMA(p, q), per capire questo concetto supponiamo che la media del processo cresca linearmente con il tempo:

$$Y_t = \beta_0 + \beta_1 t + \mathcal{E}_t \quad (2.23)$$

Applicando ora una differenza prima si ottiene:

$$\begin{aligned} (1 - B)Y_t &= Y_t - Y_{t-1} \\ &= \beta_0 + \beta_1 t + \mathcal{E}_t - \beta_0 - \beta_1(t-1) - \mathcal{E}_{t-1} \\ &= \beta_0 + \beta_1 t + \mathcal{E}_t - \beta_0 - \beta_1 t + \beta_1 - \mathcal{E}_{t-1} \\ &= \beta_1 + \mathcal{E}_t - \mathcal{E}_{t-1} \end{aligned} \quad (2.24)$$

Com'è possibile notare applicando una differenza prima ad un processo stocastico con una media che varia nel tempo come un polinomio di grado 1 si ottiene un nuovo processo stocastico stazionario in media.

Quindi un processo è definito autoregressivo-integrato a media mobile, la cui notazione è ARIMA(p, d, q), se può essere scritto nella forma:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\mathcal{E}_t \quad (2.25)$$

dove $\phi(B)$ e $\theta(B)$ sono i polinomi caratteristici di grado rispettivamente p e q e sono così definiti:

$$\begin{aligned} \phi(B) &= 1 + \phi_1 B + \dots + \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \end{aligned} \quad (2.26)$$

2.5 Processi SARIMA

Finora si sono tenuti in considerazione modelli di generazione dei dati, racchiusi nell'insieme dei modelli ARIMA(p, d, q), la cui correlazione teorica tendeva a diminuire all'aumentare dello sfasamento temporale tra le due variabili prese in considerazione, questo però non è sempre così infatti non viene tenuto conto di una componente essenziale delle serie temporali ovvero della

stagionalità, è possibile infatti che una variabile abbia una correlazione significativa con un'altra variabile che si trova allo stesso istante temporale ma in un periodo differente, per capirci se consideriamo una serie mensile il cui modello generatore dei dati si assume sia di tipo MA(1) la variabile casuale che si trova al mese di luglio risulta essere correlata con le variabili del mese di giugno e di agosto non tenendo in considerazione che quella variabile ha un legame anche con la variabile del mese di luglio dell'anno precedente, i processi stocastici che includono anche una correlazione tra i periodi sono i modelli SARIMA(p, d, q)(P, D, Q)_s che possono essere definiti come segue:

Un processo è detto autoregressivo-integrato a media mobile stagionale, la cui notazione è SARIMA(p, d, q)(P, D, Q)_s, se può essere scritto nella forma:

$$\Phi(B^s)\phi(B)(1 - B^s)^D(1 - B)^d Y_t = \Theta(B^s)\theta(B)\mathcal{E}_t \quad (2.27)$$

dove $\Phi(B^s)$ e $\Theta(B^s)$ sono i polinomi caratteristici di grado rispettivamente P e Q e sono così definiti:

$$\begin{aligned} \Phi(B^s) &= 1 + \Phi_1 B^s + \dots + \Phi_P B^{s-P} \\ \Theta(B^s) &= 1 + \Theta_1 B^s + \dots + \Theta_Q B^{s-Q} \end{aligned} \quad (2.28)$$

2.6 Modelli per serie storiche con variabili esogene: ARMAX

I modelli ARIMA e SARIMA sono processi stocastici che legano la singola variabile casuale con il proprio passato ma che non utilizzano esplicitamente informazioni contenute in altre serie storiche, in molti casi, tuttavia, una variabile non è collegata esclusivamente al suo passato ma può essere anche influenzata dai valori di altre serie temporali (Weron, 2006). I prezzi dell'energia elettrica rientrano proprio in questo caso, infatti il prezzo formatosi nel mercato del giorno prima è generalmente regolato da vari fattori esogeni come il volume di energia rinnovabile offerta. Per catturare queste relazioni è possibile utilizzare modelli che sono generalizzazioni dei modelli esposti nelle sezioni precedenti, cioè la generalizzazione dei modelli AR, ARMA, ARIMA e SARIMA sono ARX, ARMAX, ARIMAX e SARIMAX. In questo modello il valore corrente della serie è espresso linearmente in funzione dei suoi valori passati, in termini dei valori precedenti del rumore e, inoltre, in termini dei valori di una o più variabili esogene, un modello è quindi definito a media mobile autoregressivo con variabili esogene x_1, \dots, x_k , la cui notazione è

ARMAX(p, q, r_1, \dots, r_k), se può essere scritto nella forma:

$$\phi(B)Y_t = \theta(B)\mathcal{E}_t + \sum_{i=1}^k \beta^i(B)x_t^i \quad (2.29)$$

dove r_i è l'ordine dei fattori esogeni e $\beta^i(B) = \beta_0^i + \beta_1^i(B) + \dots + \beta_{r_i}^i B^{r_i}$, se $r_i = 1, \forall i$ allora si ha la combinazione lineare:

$$\sum_{i=1}^k \beta^i(B)x_t^i = \beta_0 + \beta_1 x_{t,1} + \dots + \beta_k x_{t,k}$$

2.7 Procedura di identificazione e validazione del modello

La modellazione di una serie temporale è una procedura iterativa che avviene con una serie di passaggi ben definiti (Weron, 2006), il primo passo consiste nello specificare la struttura appropriata e gli ordini del modello, l'intero processo viene guidato dal principio di parsimonia per il quale considerando due modelli che descrivono adeguatamente i dati il modello migliore è quello con meno parametri. L'identificazione può essere effettuata guardando i grafici della funzione di autocorrelazione parziale e globale (ACF e PACF) o attraverso una procedura iterativa automatizzata che consiste nell'adattare molte diverse strutture possibili del modello e nell'utilizzare un criterio statistico di bontà di adattamento per selezionare il modello migliore, i criteri di informazione permettono di tenere in considerazione una penalizzazione che cresce in funzione del numero di parametri utilizzati, tra i più noti vi sono il Criterio di Akaike corretto (AICC) e il Criterio Bayesiano (BIC) descritti come segue:

$$AICC = -2 \log(\mathcal{L}) + \frac{2dn}{n-d-1} \quad (2.30)$$

$$BIC = -2 \log(\mathcal{L}) + d \log(n) \quad (2.31)$$

dove n corrisponde alla lunghezza della serie storica, d corrisponde al numero di parametri del modello e \mathcal{L} corrisponde alla funzione di verosomiglianza. Alla luce della formulazione di questi criteri il modello migliore sarà quello con il valore inferiore.

Il secondo passo sarà stimare i parametri del modello scelto al punto precedente, considerando il caso più generale di un modello ARMAX i parametri da stimare saranno i parametri della componente autoregressiva ϕ_1, \dots, ϕ_p , i parametri della componente a media mobile $\theta_1, \dots, \theta_q$, i parametri associati alle variabili esogene β_0, \dots, β_k e infine il termine di varianza σ^2 , il metodo che viene utilizzato è il metodo di massima verosomiglianza per il quale si

assume che i dati y_1, \dots, y_t siano osservazioni di una serie temporale gaussiana, in questo modo è possibile individuare la funzione di verosomiglianza e massimizzarla in corrispondenza dei parametri, si noti che le stime dei parametri associati alle variabili esogene β_0, \dots, β_k con il metodo della massima verosomiglianza corrispondono alle stime con i minimi quadrati (OLS).

Il terzo e ultimo passo consiste nell'eseguire una serie di test sui residui per valutare se le assunzioni fatte sulla struttura del modello possono essere considerate valide, nella presente sezione verranno mostrati i test che sono più comunemente usati per validare il modello in termini di bontà di adattamento pur sapendo che un'altra serie di test, sempre con l'obiettivo di validare il modello, vengono eseguiti sulle previsioni.

- **ACF/PACF Test:** Assumendo la Normalità delle osservazioni risulta possibile fare un test sulla significatività delle correlazioni tra le variabili di fatto il test prevede che se il modello risulta appropriato almeno il 95% dei valori della correlazione dei residui risultano compresi tra $\pm 1.96/\sqrt{n}$ dove n è la lunghezza della serie.
- **Test di Ljung - Box:** Invece di guardare se il 95% delle autocorrelazioni dei residui risulta contenuta all'interno delle bande è possibile eseguire la statistica test seguente:

$$Q = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j} \quad (2.32)$$

La distribuzione della statistica test può essere approssimata come una $\chi^2(h)$ quindi un grande valore di Q suggerisce che le autocorrelazioni sono troppo grandi perché i dati siano considerati realizzazione del processo stocastico assunto al punto iniziale

- **Test R^2 :** Con riferimento ai modelli ARMAX è possibile anche considerare un test che permette di capire la varianza catturata (spiegata) dai predittori, la sua formulazione è la seguente:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{Y})^2}{\sum_{t=1}^n (y_t - \bar{Y})^2} \quad (2.33)$$

dove $\hat{y}_t = \hat{\beta}_0 + \beta_1 \hat{x}_{t,1} + \dots + \beta_k \hat{x}_{t,k}$ e $\bar{Y} = \sum_{t=1}^n y_t$.

2.8 Valutazione delle previsioni e confronto tra modelli

Nella presente sezione vengono descritti i criteri di valutazione utilizzati per confrontare i modelli in termini di accuratezza delle previsioni, i modelli in questione vengono già considerati validi in termini di bontà di adattamento cioè i test della sezione precedente applicati ai seguenti modelli sono quantomeno considerati accettabili. Viene definita con il simbolo \hat{y}_t la previsione della serie storica al tempo t e con il simbolo $e_t = y_t - \hat{y}_t$ viene indicato l'errore di previsione corrispondente, si utilizza il simbolo T per indicare in numero di previsioni. Ovviamente per svolgere questo tipo di test è necessario considerare un insieme di istanti temporali in cui sono presenti dei dati osservati e sui quali verranno fatte le previsioni. Ci sono più tecniche di previsione che possono essere prese in considerazione, tra le più popolari vi sono:

1. **Previsioni Out - of sample:** per questa tipologia di previsioni il modello viene adattato su un insieme di dati che rimarrà costante per ogni previsione, successivamente il modello sarà utilizzato per effettuare previsioni su nuovi istanti temporali che cambieranno in base all'orizzonte temporale
2. **Previsioni Rolling - windows:** Per questo tipo di previsioni invece ciò che rimane fisso per ogni previsione è l'orizzonte temporale mentre invece per ogni previsione cambierà il bacino di dati in cui verrà adattato il modello (mantenendo uguale la numerosità dei dati) e di conseguenza cambierà anche l'istante base da cui si parte per fare la previsione.

Indipendentemente dalla tecnica utilizzata è possibile valutare la bontà di previsione di un determinato modello attraverso l'**Errore quadratico medio**:

$$MSE = \frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2 \quad (2.34)$$

Questo indicatore risulta molto importante perchè si può scomporre addittivamente, $MSE = ES + EV + EC$, dove:

- Errore sistematico: $ES = (\frac{1}{T} \sum_{t=1}^T e_t)^2$ questo indice permette di cogliere se è presente un errore sistematico nelle previsioni;
- Errore in varianza: $EV = (\sigma_{\hat{y}} - \sigma_y)^2$ dove con il simbolo $\sigma_{\hat{y}}$ viene indicata la deviazione standard della serie delle previsioni mentre con

il simbolo σ_y viene indicata la deviazione standard della serie dei valori osservati, questo indicatore permette di capire quanta componente di variabilità non spiegata incide nel totale dell'errore quadratico medio;

- Errore in covarianza: $EC = 2(1 - r_{\hat{y}y})\sigma_{\hat{y}}\sigma_y$ dove $r_{\hat{y}y}$ è il coefficiente di correlazione tra i valori previsti dal modello e i valori osservati, questo indice non ha una particolare interpretazione.

Oltre all'errore quadratico medio è spesso utilizzata la radice dell'errore quadratico medio:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad (2.35)$$

La performance previsiva può essere anche valutata attraverso l'**Errore assoluto medio**:

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

Altri indicatori utilizzabili per valutare le performance previsive del modello fanno uso non dell'errore di previsione ma dell'errore di previsione percentuale $p_t = \frac{e_t}{y_t}$ in questo modo risulta possibile calcolare l'**Errore quadratico medio percentuale**:

$$MSPE = \frac{1}{T} \sum_{t=1}^T p_t^2 \quad (2.37)$$

e l'**Errore assoluto medio percentuale**:

$$MAPE = \frac{1}{T} \sum_{t=1}^T |p_t| \quad (2.38)$$

Tutti questi indicatori risultano informativi nel momento in cui due o più modelli vengono confrontati tra loro, qualora infatti un indicatore venisse misurato in un modello non da modo di capire se il modello prevede 'bene' o 'male' ma misurando lo stesso indicatore in due modelli diversi è possibile capire quale dei due risulta, per l'indicatore, migliore ovvero quello con il valore più basso. Vi è però un indice che permette di essere informativo senza dover per forza essere confrontato ed è l'**Indice U di tail** che tiene in considerazione un modello benchmark che produce delle 'previsioni naive' ovvero delle previsioni che non possono essere considerate delle buone previsioni infatti vengono calcolate come $\hat{y}_t = y_{t-1}$ cioè le previsioni per il modello

benchmark corrispondono al valore osservato all'istante precedente, l'indice u di tail quindi è definito come segue:

$$\begin{aligned}
 U &= \frac{\sqrt{MSE(\text{Modello})}}{\sqrt{MSE(\text{ModelloBenchmark})}} \\
 &= \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - y_{t-1})^2}}
 \end{aligned} \tag{2.39}$$

Come si può notare l'indice u di tail può prendere valori all'interno dell'intervallo $[0, \infty)$, inoltre se $U = 0$ significa che il modello è perfetto infatti $e_t = 0 \forall t$, il valore soglia che classifica il modello come peggiore del modello naive è $U = 1$ in questo caso il modello prevede male come il modello benchmark e nel caso in cui $U > 1$ il modello prevede peggio del modello benchmark.

L'ultimo test che verrà discusso in questa sezione prende il nome di **Test di Diebold - Mariano** che permette, attraverso un test statistico, di capire se le previsioni effettuate da due modelli sono significativamente differenti. Il test utilizza le 'funzioni di perdita' ovvero una funzione $L(e_t)$ applicata all'errore di previsione, le principali sono la funzione di perdita quadratica tale che $L(e_t) = e_t^2$ oppure la funzione di perdita assoluta $L(e_t) = |e_t|$, il test di Diebold - Mariano si basa sul calcolo della differenza tra le funzioni di perdita quindi è possibile considerare $d_t = L(e_t^{(1)}) - L(e_t^{(2)})$ la differenza delle due funzioni di perdita al tempo t degli errori calcolati nei modelli presi in considerazione, Formalmente il sistema di verifica d'ipotesi risulta essere:

$$\begin{cases} H0 : E(d_t) = 0 \\ H1 : E(d_t) \neq 0 \end{cases} \tag{2.40}$$

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

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{T} \sum_{k=-M}^M \hat{\gamma}_k}} \xrightarrow{d} N(0, 1) \tag{2.41}$$

in cui $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$, $M = T^{\frac{1}{3}}$ e $\hat{\gamma}_k$ corrisponde all'autocovarianza campionaria della serie d_t

Capitolo 3

Presentazione dei dati

La tesi in questione si occupa dell'applicazione di metodi statistici con lo scopo di prevedere i prezzi dell'energia elettrica nei principali mercati Europei, prima di proseguire con la modellazione dei dati è doveroso considerare il fatto che non è importante quanto il modello adattato sia buono se i dati che vengono utilizzati non sono di qualità, la presente sezione ha quindi lo scopo di presentare la fonte e i dati che verranno successivamente utilizzati.

Le sfide legate al possesso di dati di qualità sono relative all'individuazione di fonti appropriate, di combinazione di diverse fonti di dati (i dati sono forniti da molte istituzioni diverse e quindi spesso disperse in vari archivi e spesso formattati in modo diverso), alla classificazione e alla documentazione in modo standardizzato dei dati e alla risoluzione di problemi come incongruenze, errori evidenti, dati mancanti ecc.

La qualità dei dati si basa quindi sui criteri di Trasparenza, Comprensibilità e Tracciabilità che influiscono notevolmente sulla qualità scientifica degli studi basati su un approccio modellistico e previsionale, se quindi i dati utilizzati non soddisfano i criteri di cui sopra non sarà possibile estrarre alcuna informazione attendibile su di loro.

Alla luce di questi problemi il progetto di *Open Power System Data* (OPSD), che è la fonte utilizzata nella presente tesi, ha come obiettivo quello di creare una piattaforma aperta al pubblico in cui vengono raccolti, controllati, elaborati, aggregati, documentati e pubblicati i dati con lo scopo di soddisfare i criteri di Trasparenza, Comprensibilità e Tracciabilità.

Open Power System Data suddivide i dati del sistema elettrico Europeo in 5 pacchetti ma nel presente elaborato viene preso in considerazione solamente il pacchetto riguardante le serie temporali nel quale sono inclusi dati di carico (consumo di energia), dati relativi ai prezzi dell'energia elettrica formati nel

mercato del giorno prima e dati riguardanti la generazione di energia da fonti rinnovabili quali solare ed eolico.

A seconda dell'impostazione del mercato nazionale di riferimento le serie temporali hanno una frequenza di 15 minuti (Austria, Belgio, Germania, Ungheria, Lussemburgo, Olanda), 30 minuti (Cipro, Irlanda, Inghilterra) o 60 minuti (tutti gli altri paesi). Nel pacchetto considerato sono comunque presenti le serie storiche orarie ottenute attraverso l'aggregazione dei dati con frequenza di 15 e 30 minuti e saranno l'oggetto di analisi di questa tesi poiché consentono una copertura geografica massima. Nel complesso i dati coprono 32 paesi europei con un arco temporale che parte dal 2015 fino al settembre 2020, la fonte principale da cui *Open Power System Data* raccoglie i dati è ENTSO-E ovvero la rete europea dei gestori dei sistemi di trasmissione introdotta appunto nel 2015, altre fonti ufficiali sono relative direttamente ai gestori del sistema di trasmissione nazionale.

Ogni serie storica presente nel dataset viene specificata in base a 3 categorie ovvero regione, variabile e attributo. La regione indica l'ambito geografico secondo i codici ISO 3166, la variabile distingue la serie in base alla misurazione effettuata ovvero consumo, generazione e prezzi infine l'attributo dà ulteriori proprietà ai dati che sono specifici per la rispettiva variabile, questo è riassunto nella Figura 3.1. Per capire meglio la struttura con la quale vengono organizzate le variabili è possibile prendere in considerazione la colonna `IT_load_actua_entsoe_transparency`, com'è possibile notare IT viene utilizzato per indicare la regione Italia, secondo appunto i codici ISO 3166, successivamente, separato dal simbolo '_' c'è load il quale da come indicazione che la variabile misura il consumo di elettricità, infine, sempre separato dal simbolo '_' vi è l'ultima indicazione ovvero `actual_entsoe_transparency` che, come indicato da *attribute* indica il consumo totale di energia.

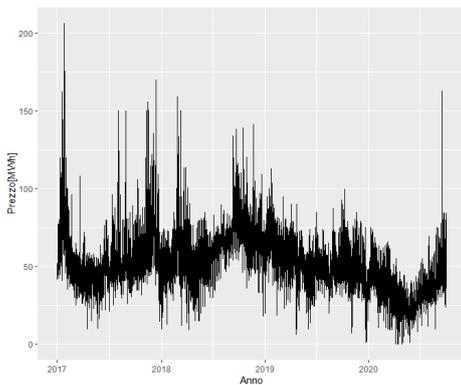
region	variable	attribute	Explanation
ISO 3166 area code and name or control area or bidding zone	load	actual_entsoe_power_statistics	Total load as published on ENTSO-E Data Portal
		actual_entsoe_transparency	Total load as published on ENTSO-E Data Portal/Power Statistics
		actual_tso	Total load as published by the TSO
		actual_net_consumption_tso	Total load excluding transmission losses as published only by the TSO
		actual_gross_generation_tso	Total power generation from national TSO
		day_ahead_forecast_entsoe_transparency	Day ahead total load forecast from ENTSO-E Transparency Platform
		solar / wind / wind_onshore / wind_offshore	generation
day_ahead_generation_forecast	Day ahead generation forecast		
generation_actual_dso	Electricity produced by power plants connected to distribution grid		
generation_actual_tso	Electricity produced by power plants connected to transmission grid		
capacity	Aggregated installed capacity of power plants (actual availability not accounted for)		
profile	Share of installed capacity producing		
price	day_ahead		

Figura 3.1: Classificazione delle variabili mediante una struttura a strati (Fonte: OPSD)

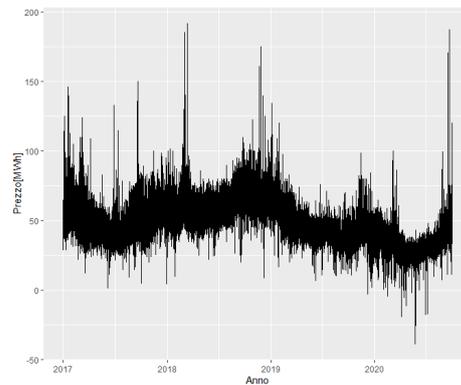
La presente tesi ha come obiettivo l'applicazione di modelli statistici per la previsione dei prezzi dell'energia elettrica nei principali paesi Europei, i principali paesi Europei che vengono considerati sono Gran Bretagna, Spagna, Francia, Italia (zona NORD), viene preso come periodo di riferimento l'intervallo 1/01/2017 fino a 29/09/2020. Viene poi dedicata una sezione alla Germania prendendo in considerazione i dati nell'intervallo di tempo 2021-01-01 00:00 fino al 2023-07-31 23:00.

Capitolo 4

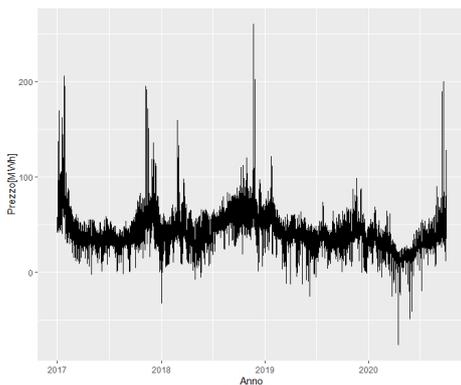
Analisi dati



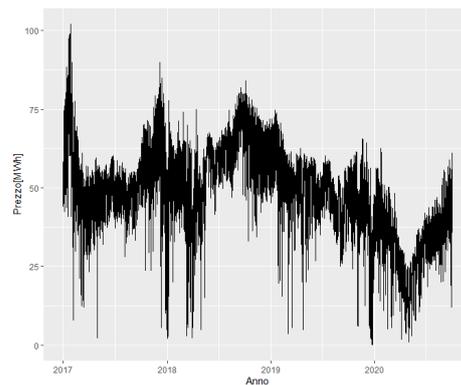
Serie storica oraria dei prezzi dell'energia nel mercato Italiano(Nord)



Serie storica oraria dei prezzi dell'energia nel mercato Inglese



Serie storica oraria dei prezzi dell'energia nel mercato Francese



Serie storica oraria dei prezzi dell'energia nel mercato Spagnolo

Figura 4.1: Rappresentazione grafica delle serie orarie del prezzo dell'energia a partire dal 2017-01-01 00:00 fino al 2020-09-29 23:00

Come spiegato nel capitolo 1 il prezzo di mercato dell'energia si genera dall'incontro tra la curva di domanda e la curva di offerta attraverso un meccanismo di asta doppia, per capire la struttura temporale del prezzo Figura 4.1 ci si può concentrare rispetto ai fattori che influenzano le due curve, dal lato della domanda vi sono i consumatori e quindi per analizzarne lo sviluppo temporale si può pensare a quelle che sono le abitudini delle famiglie, ad esempio quando la luce del giorno non c'è più si accende l'illuminazione, quando la temperatura è fredda o calda si accende il riscaldamento o il condizionatore, di notte le persone dormono e quindi spengono molti dispositivi e così via. Tra i consumatori di energia vi sono anche le industrie le quali riducono la loro produzione nei giorni festivi consumando meno energia mentre aumentano la loro produzione in situazioni economiche favorevoli consumando più energia, queste semplici considerazioni servono per introdurre alcune caratteristiche temporali della domanda. Alla luce di quanto detto è possibile notare che c'è un chiaro schema periodico della serie con diverse frequenze temporali (Cretì & Fontini, 2019)), in primo luogo si può notare un periodo settimanale infatti nel fine settimana la domanda di energia risulta inferiore rispetto ai giorni lavorativi, in secondo luogo è possibile notare una frequenza annuale in quanto la domanda di energia è più alta in inverno, diminuisce in primavera per poi risalire in estate e infine diminuire in autunno ma tendenzialmente ad un livello inferiore rispetto alla primavera. Uno schema periodico della domanda è anche quello giornaliero in quanto la domanda di energia elettrica tenderà a raggiungere un minimo di notte, aumentare durante il giorno avendo tipicamente 2 punti di massimo e infine tornare a scendere durante la notte, questo risulta coerente con un sistema paese caratterizzato da un carico industriale che segue un programma diurno. La curva di prezzo è anche influenzata dai fattori che a loro volta influenzano la curva di offerta come la presenza di fonti di energia rinnovabile, è stato detto nel capitolo 1 che in un'ottica di breve periodo la presenza di energia rinnovabile nella borsa elettrica sposta la curva di offerta verso destra facendo scendere il prezzo, in funzione di questo è possibile notare chiaramente come l'offerta risenta di una stagionalità di periodo giornaliero, se si pensa ad un impianto fotovoltaico tenderà a produrre maggiormente durante il giorno e meno durante la notte, nella curva di offerta è presente anche una stagionalità annuale in quanto in estate la produzione di rinnovabili aumenta per poi scendere nel periodo invernale. La struttura temporale del prezzo è dunque molto complicata e dipende da molteplici altri fattori come le condizioni economiche di un paese, condizioni meteorologiche, dalla tecnologia utilizzata dalle aziende produttrici, il costo dei combustibili e così via.

4.1 Relazioni con i predittori

Nella presente sezione verrà descritta la relazione esistente tra la serie storica dei prezzi e i predittori con riferimento all'Italia, non verranno trattati gli altri paesi in quanto le considerazioni descrittive che si possono fare risultano essere le stesse.

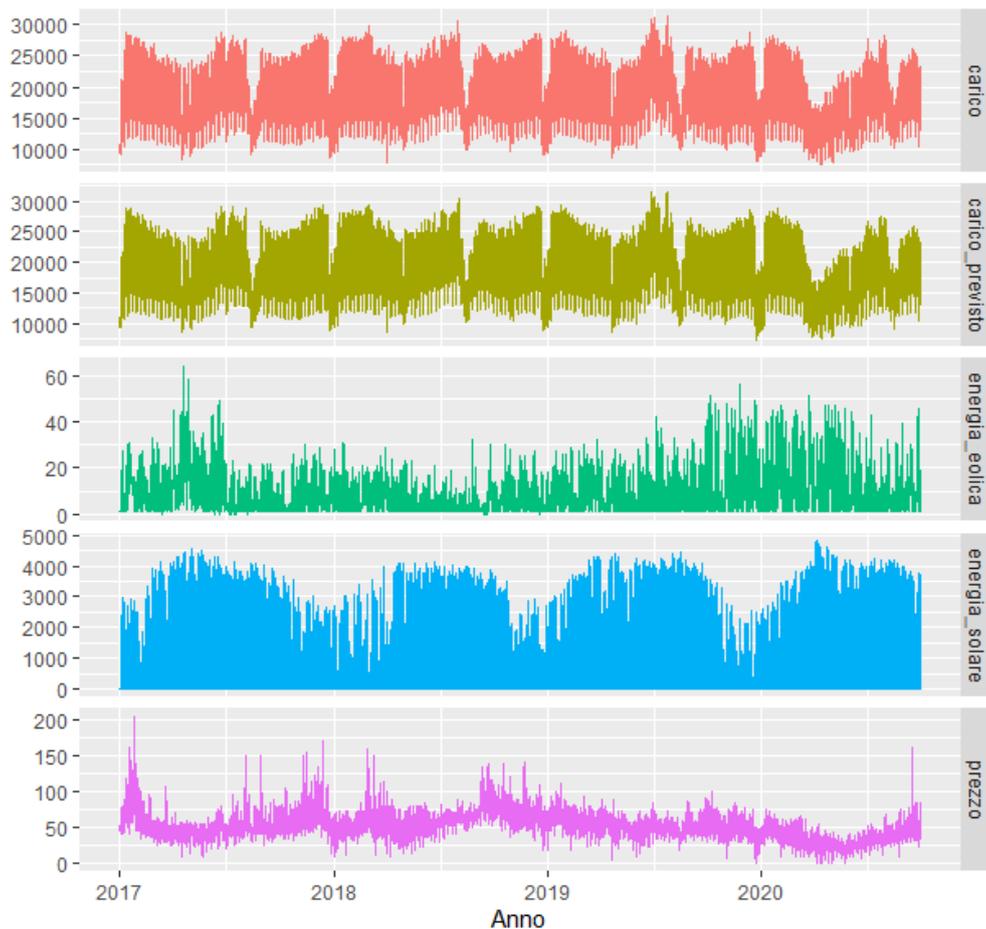


Figura 4.2: Rappresentazione grafica della serie storica dei prezzi dell'energia nel mercato Italiano zona Nord e dei predittori disposti in pannelli separati

Dalla Figura 4.2 è possibile fare delle considerazioni iniziali in merito al rapporto esistente tra i predittori e la variabile d'interesse ovvero il prezzo dell'energia, è possibile notare chiaramente tra l'anno 2017 – 2018 e nell'anno 2020 un calo iniziale della domanda e una lenta risalita della stessa è poi possibile notare che nello stesso periodo anche la curva di prezzo ha subito un andamento molto simile. Focalizzando l'attenzione nelle variabili indicanti

la produzione di energia rinnovabile oltre a notare la loro forte variabilità è possibile osservare come all'aumentare della generazione di energia rinnovabile il prezzo si abbassi infatti guardando la serie `energia_solare` risulta evidente l'incremento di energia elettrica proveniente da impianti fotovoltaici nei periodi centrali dei vari anni e in presenza di quei periodi si nota una diminuzione dei prezzi, per quanto riguarda la serie `energia_eolica` è difficile, per ora, notare un chiaro legame con i prezzi dell'energia nell'intero periodo osservato anche se all'inizio del 2017 in corrispondenza di una forte generazione di energia proveniente da impianti eolici vi è una chiara diminuzione del prezzo.

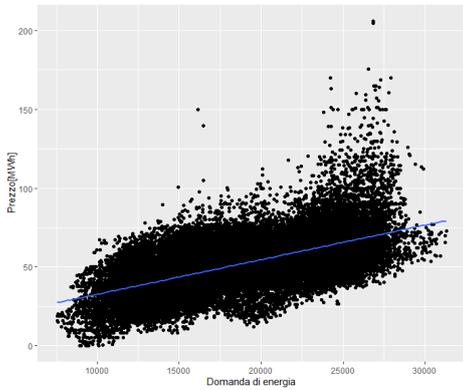


Diagramma di dispersione tra Prezzo e Domanda, correlazione pari a 0.60

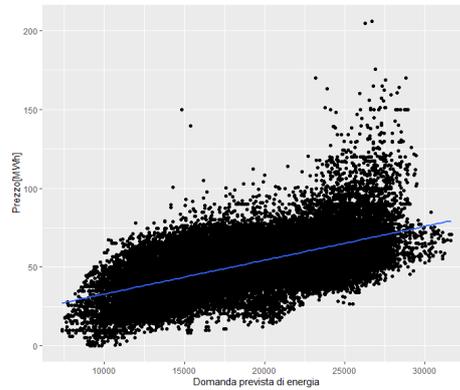


Diagramma di dispersione tra Prezzo e Domanda prevista, correlazione pari a 0.60

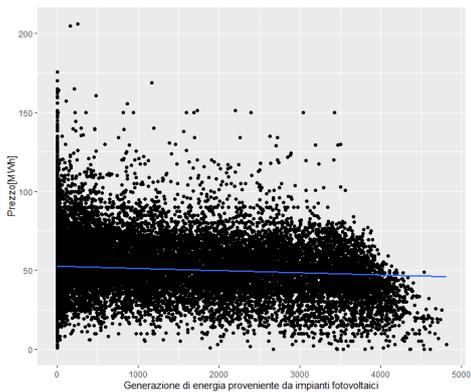


Diagramma di dispersione tra Prezzo e energia solare, correlazione pari a -0.09

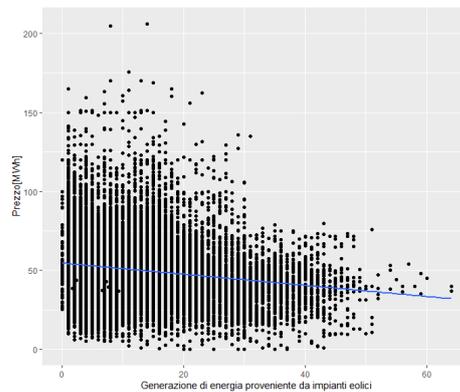


Diagramma di dispersione tra Prezzo e energia eolica, correlazione pari a -0.18

Figura 4.3: Rappresentazione grafica dei diagrammi di dispersione e della retta di regressione che indicano il legame lineare presente tra il prezzo dell'energia e i predittori

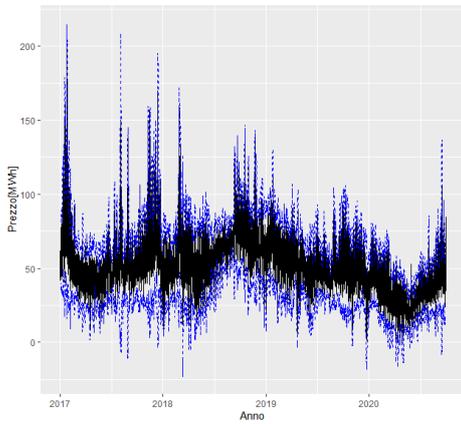
Dalla Figura 4.3 è possibile notare il legame lineare esistente tra il prezzo e le altre variabili presenti, si osserva una correlazione positiva sia tra il prezzo e la domanda sia tra il prezzo e la domanda prevista mentre si osserva una correlazione negativa e non particolarmente importante tra il prezzo e la generazione di energia solare infine si osserva una correlazione negativa tra il prezzo e la generazione di energia eolica.

4.2 Trattamento dei picchi

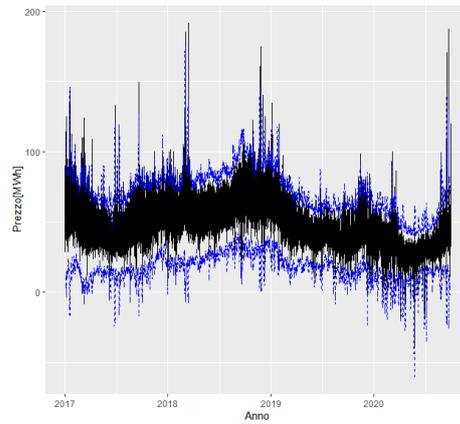
Come già ribadito in precedenza l'accuratezza delle previsioni dipende non solo dall'efficienza dei modelli impiegati ma anche dalla qualità dei dati analizzati, la serie del prezzo dell'energia possiede molte irregolarità e valori anomali (valori che si discostano dalla maggior parte dei dati), se il comportamento del prezzo risulta anomalo in una determinata ora questo potrebbe incidere sulla bontà delle previsioni future. L'applicazione dei modelli spiegati nel capitolo 2 avverrà solamente dopo una gestione dei valori anomali, un possibile metodo di trattamento di questi dati (Weron, 2006) consiste innanzitutto nell'individuare i valori anomali, il metodo utilizzato da Weron e riproposto in questo elaborato consiste in:

1. Calcolo della mediana y_t^{med} attraverso una finestra scorrevole di h ore;
2. Costruzione delle fasce che discriminano un valore anomalo da un valore non anomalo ovvero $B_t = y_t^{med} \pm 3 \cdot SD(y_t - y_t^{med})$ dove $SD(\cdot)$ corrisponde alla deviazione standard;
3. Identificare i valori anomali ovvero coloro che sono fuori dalle fasce B_t .

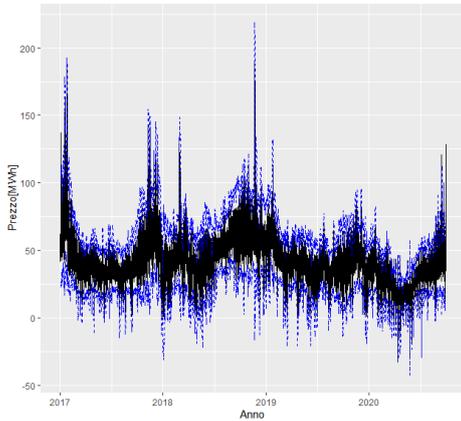
Nel metodo proposto viene impiegata la mediana in quanto più robusta della media, le bande B_t sono rappresentate nella Figura 4.4, il funzionamento di questa procedura si basa anche sul valore h che viene inserito cioè sulla larghezza della finestra temporale utilizzata per il calcolo della mediana e della deviazione standard infatti se i picchi presenti nella serie risultano essere isolati allora è possibile utilizzare anche $h = 5$ ma qualora ci fossero delle sequenze di picchi adiacenti tra loro h dev'essere aumentata così da permettere un'individuazione corretta di questi ultimi, è anche possibile utilizzare entrambe le finestre scorrevoli per catturare sia picchi isolati sia picchi adiacenti tra loro infatti nel caso dei prezzi dell'energia elettrica è stato utilizzato un valore $h = 49$ ed un valore $h = 5$. Successivamente all'individuazione dei dati anomali questi vengono classificati come dati mancanti e vengono sostituiti con la media delle osservazioni vicine.



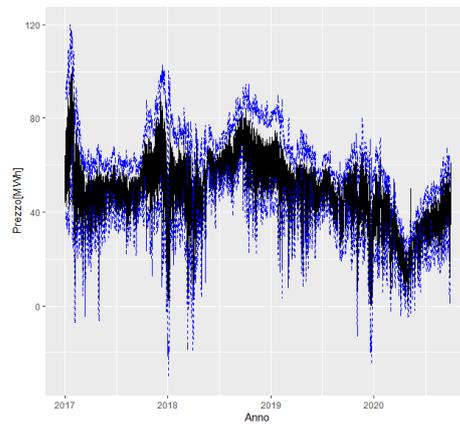
Individuazione valori anomali Italia zona Nord



Individuazione valori anomali Inghilterra



Individuazione valori anomali Francia



Individuazione valori anomali Spagna

Figura 4.4: Rappresentazione grafica del metodo utilizzato per l'individuazione dei valori anomali che consiste nel costruire delle fasce che discriminano i dati come anomali o non anomali

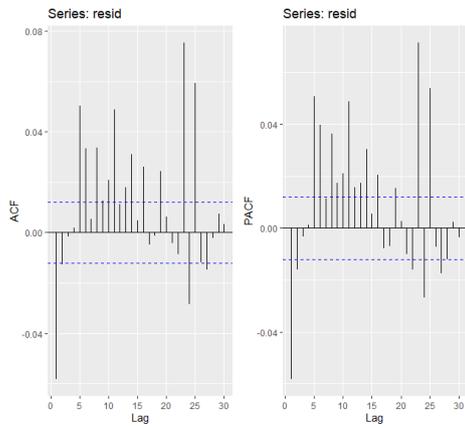
4.3 Adattamento dei modelli e previsioni Out of sample

Nella Tabella 4.1 vengono presentati i modelli che sono stati applicati alle serie storiche dei prezzi depurate dei valori anomali attraverso la procedura spiegata nella sezione 4.1, il metodo che è stato utilizzato per scegliere il modello è spiegato nella sezione 2.7 ovvero viene applicata una procedura iterativa automatizzata che consiste nell'adattare molte diverse strutture possibili del modello e nell'utilizzare un criterio statistico di bontà di adattamento, in questo caso il criterio di Akaike, per selezionare il modello migliore.

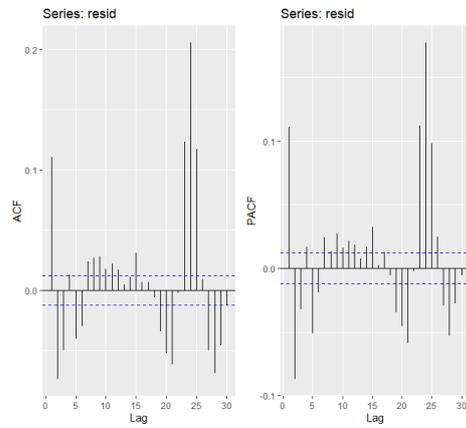
Tabella 4.1: Adattamento dei modelli

Nazione	Modello
Italia zona Nord	SARIMAX(1,1,3)(2,0,0)[24]
Inghilterra	SARIMAX(0,1,4)(0,0,2)[24]
Francia	SARIMAX(3,1,1)(0,0,2)[24]
Spagna	SARIMAX(4,1,1)(1,0,0)[24]

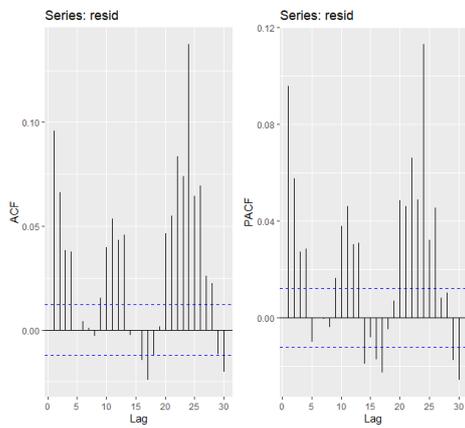
Nella Figura 4.5 è possibile osservare la funzione di autocorrelazione globale dei residui dei modelli, in particolare si può notare come i modelli precedentemente adattati non riescano a rappresentare il processo generatore dei dati, in quanto esistono chiaramente dei pattern non spiegati dal modello che poi portando inevitabilmente a delle previsioni inefficienti, come si può vedere in Figura 4.6 e attraverso i principali indicatori nella Tabella 4.3.



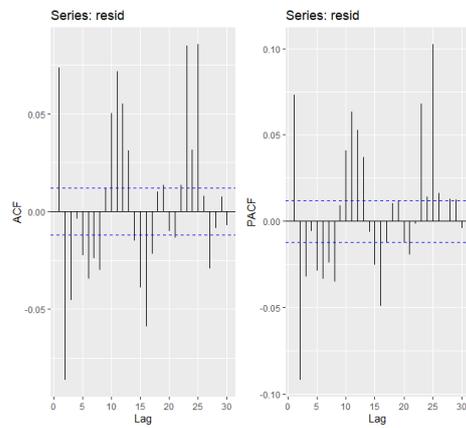
Italia zona Nord



Inghilterra

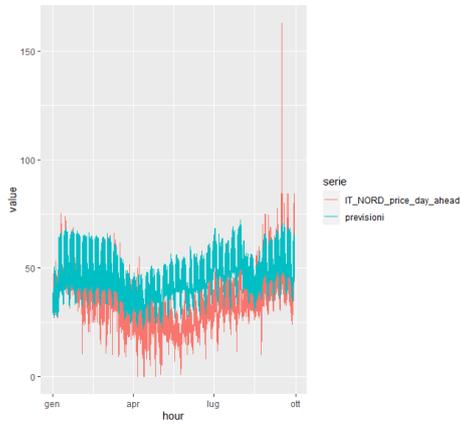


Francia

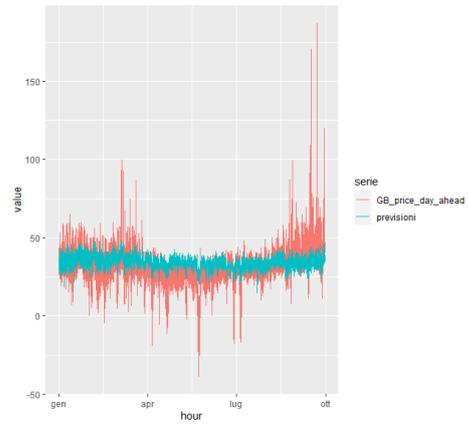


Spagna

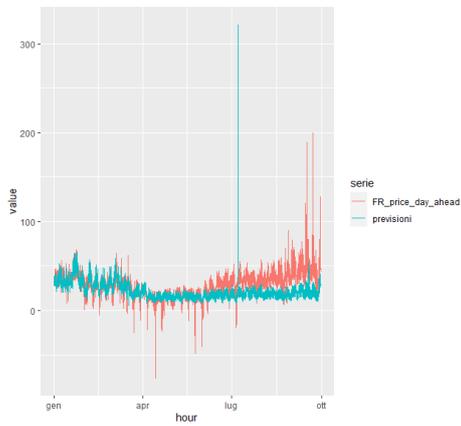
Figura 4.5: Rappresentazione grafica dei grafici della funzione di autocorrelazione globale (acf) e parziale (pacf)



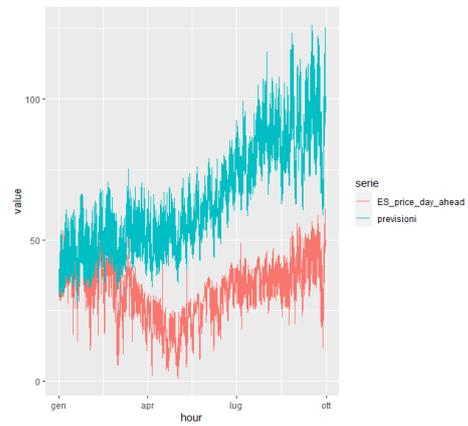
Confronto previsioni e valori osservati
Italia zona Nord



Confronto previsioni e valori osservati
Inghilterra



Confronto previsioni e valori osservati
Francia



Confronto previsioni e valori osservati
Spagna

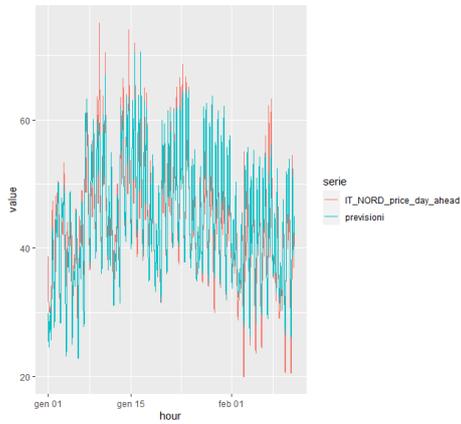
Figura 4.6: Rappresentazione grafica del confronto tra valori osservati e valori previsti dal modello nel periodo 2020-01-01 00:00:00 fino al 2020-09-29 23:00:00

Tabella 4.2: Valutazione dei modelli in termini di bontà delle previsioni

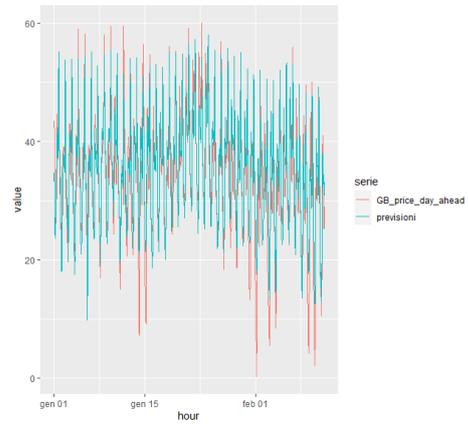
Indicatore	Italia zona Nord	Inghilterra	Francia	Spagna
MSE	126.8559	103.5137	183.3507	196.2250
% ES	0.3277	0.0640	0.2343	0.7292
% EV	0.0575	0.5194	0.1055	0.0188
% EC	0.6149	0.4166	0.6603	0.2520
RMSE	11.2630	10.1742	13.5407	14.0080
MAE	9.0845	7.1358	9.7534	12.4523
MSPE	19.6104	7824.1330	1847.1450	1.0819
MAPE	0.4821	5.1034	2.7826	0.5667
U di tail	2.8471	1.5082	2.6383	5.4879

4.4 Previsioni Rolling windows

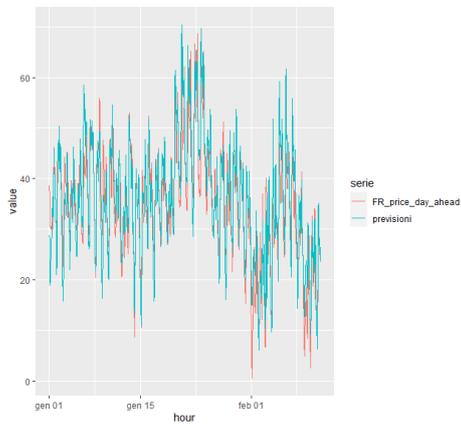
Alla luce di questo primo risultato si è provveduto a considerare un'altra tecnica di previsione utilizzando gli stessi modelli spiegati nel capitolo 2, in particolar modo ogni ora del giorno è stata divisa dalle altre formando quindi 24 dataset giornalieri, ad esempio un dataset può contenere tutte le ore 15:00 del dataset originario, a questo punto sono state fatte delle previsioni Rolling windows con tutti e 24 i dataset e successivamente le previsioni sono state raggruppate opportunamente in una serie oraria. Per questioni di difficoltà computazionale sono state fatte un totale di 1000 previsioni a partire sempre dal 2020-01-01 00:00. Questa modalità di previsione è sì migliore rispetto alla precedente, vedi Figura 4.7 e Tabella 4.4 inoltre è stato applicato il test di Diebol-Mariano confermando che le due metodologie forniscono previsioni significativamente diverse tra loro, ma ancora non basta per far sì che sia considerata accettabile, c'è però una precisazione da fare ovvero il periodo del 2020 è caratterizzato dalla diffusione della pandemia da COVID-19 che ha influito molto nei comportamenti dei singoli individui cambiando il processo generatore dei dati, è quindi difficile utilizzare un insieme di osservazioni con una propria struttura per prevedere bene dati che hanno modificato il loro comportamento.



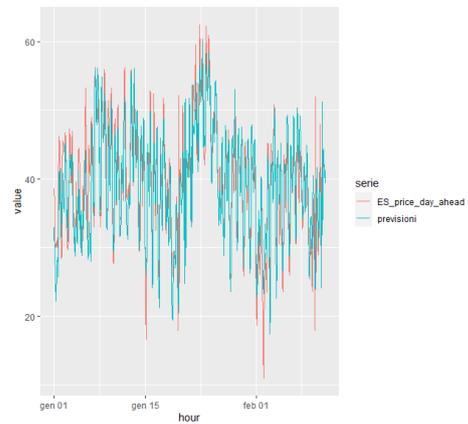
Confronto previsioni e valori osservati
Italia zona Nord



Confronto previsioni e valori osservati
Inghilterra



Confronto previsioni e valori osservati
Francia



Confronto previsioni e valori osservati
Spagna

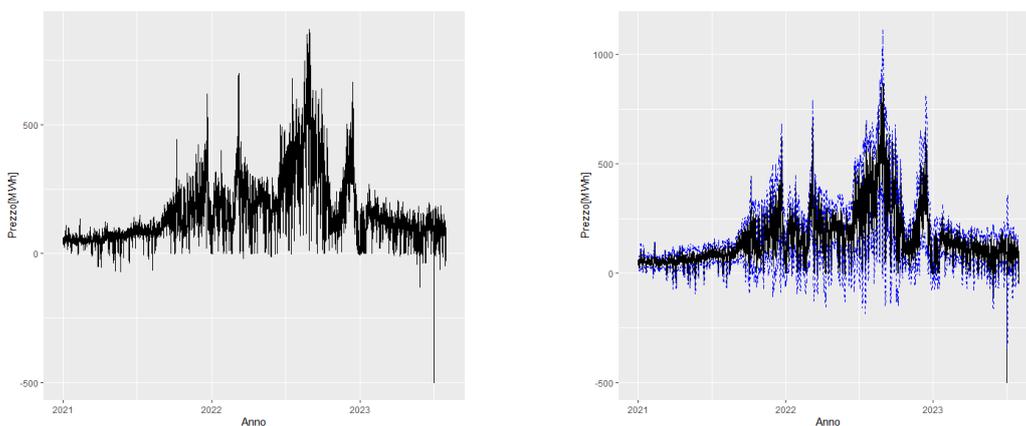
Figura 4.7: Rappresentazione grafica del confronto tra valori osservati e valori previsti attraverso la modalità Rolling windows dal modello nel periodo 2020-01-01 00:00:00 fino al 2020-09-29 23:00:00

Tabella 4.3: Valutazione dei modelli in termini di bontà delle previsioni

Indicatore	Italia zona Nord	Inghilterra	Francia	Spagna
MSE	15.2465	21.9507	25.3167	18.9120
% ES	0.0184	0.0282	0.0398	0.0244
% EV	0.0031	0.0670	0.0131	0.0127
% EC	0.9785	0.9048	0.9470	0.9629
RMSE	3.9047	4.6852	5.0316	4.3488
MAE	2.9215	3.5719	3.8508	3.0357
MSPE	0.0098	4.0228	0.8172	0.0156
MAPE	0.0699	0.2230	0.1702	0.0817
U di tail	0.9598	0.9595	1.2315	1.1742

4.5 Previsioni Rolling windows caso Germania

Tenendo in considerazione il fatto che dal 2020 il processo generatore dei dati è cambiato nella presente tesi è stata dedicata una sezione alla Germania, la quale possiede dati completi nel periodo che va dal 2021-01-01 00:00 fino al 2023-07-31 23:00, inoltre anche in questo caso, prima di procedere con l'adattamento del modello, si è provveduto alla sistemazione dei picchi sempre con la metodologia spiegata nella sezione 4.1, Figura 4.8



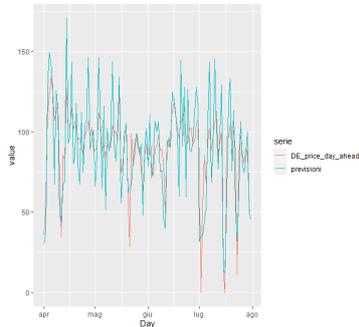
Serie storica oraria dei prezzi dell'energia nel mercato Tedesco

Individuazione dei valori anomali Germania

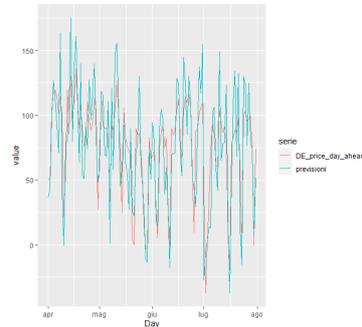
Figura 4.8: Rappresentazione grafica della serie storica dei prezzi della Germania nel periodo 2020-01-01 00:00 fino al 2022-07-31 23:00, in più viene rappresentato il grafico per l'individuazione dei valori anomali

Sono state calcolate le previsioni attraverso la metodologia Rolling windows avendo come primo periodo di stima del modello l'intervallo che va dal 2021-01-01 00:00 fino al 2023-03-31 23:00, in particolar modo sono state considerate le ore 3:00, 8:00, 12:00, 15:00 e 21:00 che, come nel caso precedente, sono state divise dalle altre formando 5 dataset indipendenti con frequenza giornaliera, successivamente, per ogni dataset, è stato applicato il metodo Rolling windows con il quale sono stati previsti i dati fino al 2023-07-31. Come si può vedere dalla Figura 4.9 e dalla Tabella 4.4 le previsioni sono considerate buone previsioni, questo conferma il fatto che non sono le metodologie utilizzate ad essere inefficienti per questo tipo di previsioni ma

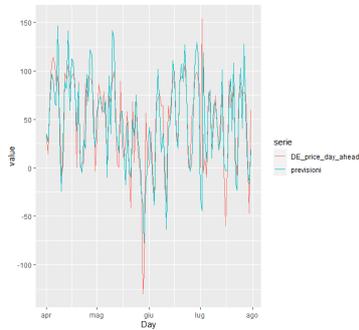
che con il cambiamento del processo generatore dei dati diventa difficile prevedere bene delle osservazioni basandosi su osservazioni con una struttura diversa.



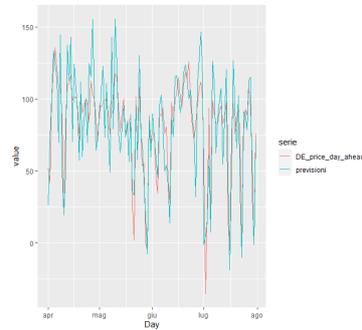
Confronto previsioni e valori osservati ore 3:00



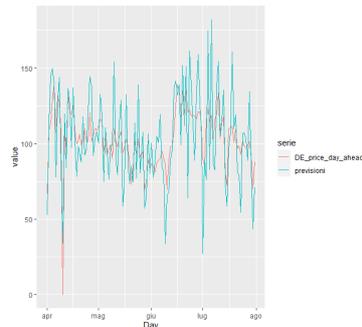
Confronto previsioni e valori osservati ore 8:00



Confronto previsioni e valori osservati ore 12:00



Confronto previsioni e valori osservati ore 15:00



Confronto previsioni e valori osservati ore 21:00

Figura 4.9: Rappresentazione grafica del confronto tra valori osservati e valori previsti attraverso la modalità nel periodo 2023-04-01 00:00:00 fino al 2023-07-31 23:00:00

Tabella 4.4: Valutazione dei modelli in termini di bontà delle previsioni

Indicatore	3:00	8:00	12:00	15:00	21:00
MSE	398.4805	685.3803	1132.2280	427.6158	466.9489
% ES	0.0001	0.0001	0.0005	0.0001	0.0001
% EV	0.1694	0.1613	0.0005	0.0945	0.3166
% EC	0.8305	0.8386	0.9990	0.9054	0.6832
RMSE	19.9620	26.1798	33.6486	20.6789	21.6090
MAE	15.9913	22.1618	23.4333	16.4668	16.9915
MSPE	3932.1730	5123.0940	4766.8880	758.8498	944.4685
MAPE	6.1016	7.0420	15.4891	2.8670	2.9445
U di tail	0.7766	0.6925	0.6274	0.6171	1.1254

Capitolo 5

Conclusione

Nella presente tesi è stato affrontato il complesso problema della previsione dei prezzi dell'energia elettrica, in particolar modo sono stati utilizzati dati storici dei prezzi, del consumo e della produzione di energia rinnovabile che, inseriti in modelli statistici in grado di catturare l'articolata struttura di correlazione dei dati, avevano lo scopo di prevedere efficacemente i prezzi futuri dell'energia. Nel presente elaborato è stato descritto il sistema elettrico concentrando l'attenzione nel processo di formazione del prezzo con lo scopo di capirne la struttura temporale e in che modo alcune variabili esogene possono influire sulla sua determinazione. Successivamente sono stati presentati i modelli statistici e i principali indicatori utili per valutarli in termini di bontà di adattamento e di previsione, questi modelli sono poi stati adattati e utilizzati per prevedere i prezzi di un periodo futuro attraverso un approccio Rolling windows e Out of sample. Inizialmente sono state fatte previsioni Out of sample e Rolling windows nell'intervallo 2020-01-01 00:00 fino al 2020-09-29 23:00 per i paesi Italia (zona nord), Inghilterra, Francia e Spagna rivelandosi inefficienti a causa del fatto che nel 2020 il processo generatore dei dati è cambiato vista la pandemia da COVID-19 anche se le previsioni basate su un approccio Rolling windows modellando le singole ore del giorno è risultato migliore delle previsioni Out of sample, ma comunque non idoneo. Quest'ultimo approccio è però stato utilizzato per prevedere il prezzo dell'energia elettrica del mercato Tedesco nel periodo 2023-04-01 00:00 fino al 2023-07-31 23:00 riscontrando essere una metodologia efficiente di previsione.

Bibliografia

Cretì, A., and Fontini, F. (2019), *ECONOMICS OF ELECTRICITY: Markets, Competition and Rules*, Cambridge university press.

Weron, R. (2006), *Modeling and forecasting electricity loads and prices: A statistical Approach*, John Wiley & Sons Ltd.

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

Beltrami, F., Fontini, F., Giulietti, M., and Grossi, L. (2021), "Environmental and Resource Economics", *The Zonal and Seasonal CO₂ Marginal Emissions Factors for the Italian Power Market*, 83, 381–411.

Fianu, S. E., Ahelegbey, F. D., and Grossi, L. (2022), "European Journal of Operational Research", *Modeling risk contagion in the Italian zonal electricity market*, 298, 656–679.

Ang, B. W., Tol, and R. S. J., Weyant, J. (2013), "Energy Economics", *Quantitative analysis of energy markets*, 35, 1-4.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., and Zareipour, H. (2020), "IEEE Open Access Journal of Power and Energy", *Energy Forecasting*, 7, 376 - 388.