

UNIVERSITÁ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Economia e Statistica

Corso di laurea in

SCIENZE STATISTICHE ED ECONOMICHE



ANALISI DELLA SERIE STORICA PER LA MODELLAZIONE  
E PREVISIONE DEL PREZZO SPOT DELL'ENERGIA  
ELETTRICA

Relatore: Prof. Andrea Marletta

Tesi di Laurea di:

Davide Fabio Loreti

Matr. N. 865309

Anno Accademico 2022/2023



# Indice

INTRODUZIONE .....	1
CAPITOLO 1 .....	2
1.1 STORIA DELL'ENERGIA ELETTRICA NEL MONDO E IN ITALIA .....	2
1.2 ZONE DI MERCATO .....	4
1.3 ARTICOLAZIONE DEL MERCATO ELETTRICO .....	5
1.4 PREZZO UNICO NAZIONALE .....	7
1.5 CARO ENERGIA: LE CAUSE E POSSIBILI SOLUZIONI .....	8
1.6 CARATTERISTICHE E PROPRIETA' DEI PREZZI DELL'ENERGIA ELETTRICA .....	9
CAPITOLO 2 .....	12
2.1 INTRODUZIONE ALLE SERIE STORICHE .....	12
2.2 PROCESSI STOCASTICI: DEFINIZIONE E PROPRIETA' .....	13
2.3 MODELLI DELL'APPROCCIO MODERNO: AR, MA E ARMA .....	14
2.4 LA NON-STAZIONARIETA' NELLE SERIE STORICHE .....	16
2.5 LA PROCEDURA DI BOX-JENKINS .....	17
2.5.1 ANALISI GRAFICA .....	17
2.5.2 VERIFICA DELLA STAZIONARIETA' DELLA SERIE .....	17
2.5.3 ANALISI DEI CORRELOGRAMMI .....	18
2.5.4 STIMA DEL MODELLO .....	19
2.5.5 VERIFICA DEL MODELLO .....	19
2.6 MODELLI CON COMPONENTE STAGIONALE .....	22
2.7 PREVISIONE NEI MODELLI ARIMA .....	23
CAPITOLO 3 .....	27
3.1 FOCUS SUI DATI E ANALISI GRAFICA DELLA SERIE STORICA .....	27
3.2 ANALISI DEI CORRELOGRAMMI E SCELTA DEL MODELLO .....	30
3.3 DIAGNOSTICA SUI REDISUI .....	32
3.4 PREVISIONE DEL P.U.N. ....	33
CONCLUSIONE .....	37
Bibliografia .....	38
Appendice .....	40

## INTRODUZIONE

La guerra in Ucraina e la ripresa dell'economia post COVID-19 hanno avuto un notevole impatto sul prezzo dell'energia e sui mercati dell'energia elettrica causando degli shock importanti.

In questo elaborato si andrà a modellare e prevedere il P.U.N. (prezzo unico nazionale) attraverso i modelli SARIMA(P,D,Q,p,d,q).

Nel capitolo 1 verrà descritta la storia del mercato elettrico in Italia, dal secolo scorso fino ad oggi.

Si entrerà nel dettaglio della sua articolazione, su come si divide in ambito nazionale e inoltre verrà fatto un focus sul prezzo dell'energia e delle sue principali caratteristiche e proprietà.

Nel capitolo 2 verrà descritta dettagliatamente la metodologia statistica usata per la modellazione e previsione facendo un focus sull'approccio classico per le serie storiche.

Il capitolo 3 è dedicato alla modellazione del PUN: verrà illustrata l'analisi della serie con la scrittura di un algoritmo implementato in ambiente R e verranno analizzati i risultati di stima e previsione, verrà, inoltre usato il software Gretl per un'analisi grafica della serie storica.

# **CAPITOLO 1**

## **PANORAMICA DEL MERCATO ELETTRICO E DEL PREZZO SPOT DELL'ENERGIA ELETTRICA**

### **1.1 STORIA DELL'ENERGIA ELETTRICA NEL MONDO E IN ITALIA**

La diffusione dell'energia elettrica ha rivoluzionato le abitudini dell'uomo portando un miglioramento sostanziale dello stile di vita.

Negli anni '30 del XIX secolo il fisico inglese Michael Faraday creò un prototipo per trasformare la corrente elettrica continua in corrente alternata. In quegli anni il problema principale fu che l'erogazione dell'energia elettrica in modalità continua favoriva il fenomeno della dissipazione di energia producendo calore nel cavo conduttore. Questo problema fece in modo che fosse necessario aumentare l'intensità della corrente causando un aumento degli sprechi e contemporaneamente un aumento dei costi.

L'italiano Antonio Pacinotti diede un contributo fondamentale alla risoluzione del problema, infatti nel 1860 ottimizzò l'invenzione di Faraday realizzando un commutatore ad anello, chiamato l'anello di Pacinotti, che permise di trasformare la corrente alternata. Le prime centrali elettriche del mondo sorsero nel 1882 a Londra e a New York, dopo che Thomas Edison ebbe perfezionato la dinamo (1879) (Treccani.it).

In Italia nel 1883 a Milano, in via Santa Radegonda, nacque la prima centrale elettrica italiana (e la seconda in Europa dopo la centrale di Londra), la con una potenza di 350 kW, alimentava i locali più in vista nel centro della città meneghina e soprattutto il teatro della Scala (9colonne.it).

Mentre il primo impianto idroelettrico d'Italia fu quello di Isoverde a Genova, seguito negli anni successivi da quello sull'Adda (1898) e sul Ticino (1901).

I primi grandi stabilimenti elettrici in Italia furono le centrali termoelettriche a carbone, solo in seguito fu possibile lo sfruttamento del bacino idroelettrico delle Alpi arrivando al 1914, anno in cui il 74% della potenza elettrica fu di origine idrica (Energit.it, 2020). A partire degli anni '50 del XX secolo in Italia, non avendo risorse energetiche sufficienti, si inizia a puntare sull'energia dell'atomo. La prima centrale nucleare viene costruita a Latina (Borgo Sabotino) nel 1963, otto mesi più tardi la seconda a Sessa Aurunca, presso Caserta e meno di un anno dopo

l'impianto di Trino, in provincia di Vercelli. Una quarta centrale venne inaugurata solo nel 1978 a Caorso in provincia di Piacenza e nel 1982 infine viene messo in cantiere l'impianto di Montalto di Castro in provincia di Viterbo, mai entrato in funzione. I rischi connessi all'installazione del nucleare e la psicosi causata dalla più grande catastrofe nella storia del nucleare il 26 aprile 1986, ovvero l'esplosione del reattore 4 di Chernobyl, fecero sì che l'8 Novembre 1987 con un Referendum abrogativo l'80% dei votanti si esprime a favore abrogando una serie di norme in ambito dell'energia nucleare e orientando di fatto le scelte future del paese in ambito energetico (La Repubblica, 2022.).

A partire dal 1962 fino alla fine degli anni '90 il mercato elettrico era caratterizzato da uno stato di monopolio, con l'ENEL come unica azienda statale incaricata di produrre e fornire energia elettrica al paese (La Repubblica, 1999).

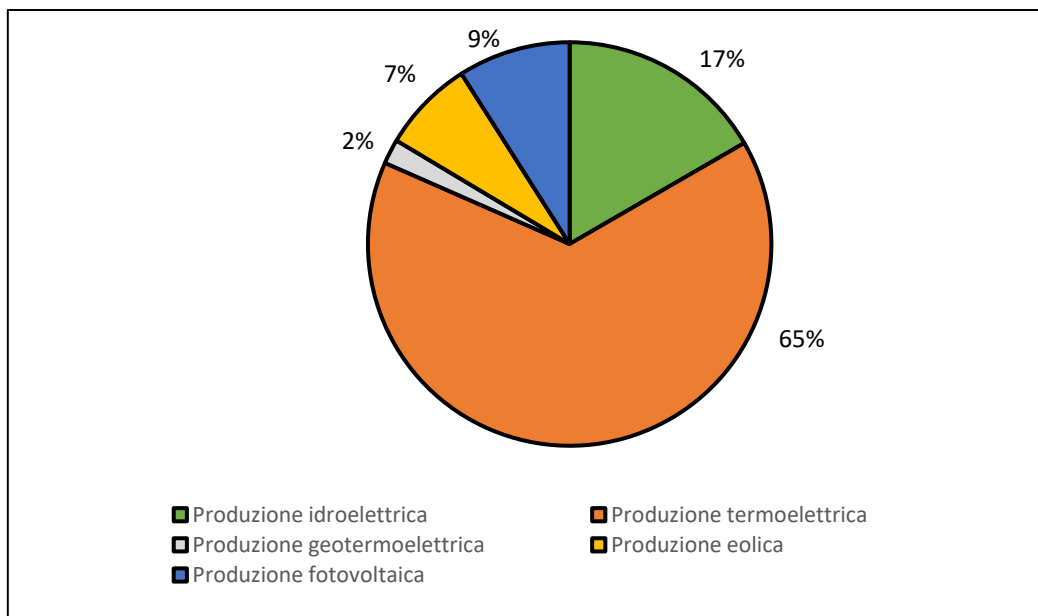
La liberalizzazione del settore elettrico in Italia si ebbe con il Decreto Legislativo 16 marzo 1999 n.79, comunemente chiamato Decreto Bersani, poiché fu un provvedimento dell'allora Ministro dell'Industria, del Commercio e dell'Artigianato Pier Luigi Bersani.

Tale norma fu l'attuazione della riforma del settore elettrico voluta dal Parlamento e Consiglio europeo e ufficializzata attraverso la DIRETTIVA 96/92/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 19 dicembre 1996 (Gazzetta Ufficiale).

Il mercato elettrico, chiamato anche "Borsa elettrica italiana", regolato dal Gestore dei Mercati Energetici S.p.A., consente a produttori, consumatori e grossisti di stipulare contratti orari di acquisto e vendita di energia elettrica.

Al Gestore dei Mercati Energetici S.p.A. è affidata la gestione e organizzazione economica del mercato elettrico.

Negli ultimi anni la produzione di energia elettrica in Italia è diventata sempre più green, grazie al costante aumento delle rinnovabili; tuttavia Terna S.p.A. ha stimato nel 2020 che in Italia il 57,6% dell'energia elettrica veniva ancora generata attraverso fonti non rinnovabili tramite le centrali termoelettriche. Secondo i dati di Terna S.p.A., nel 2022 il fabbisogno di energia elettrica in Italia è stato pari a 317 GWh di cui 276,4 GWh di produzione nazionale e 47,4 GWh importati a cui però vanno sottratti 44,4 GWh esportati e 2,5 GWh di pompaggio (Dati Provvisori di Esercizio del sistema elettrico nazionale 2021, Terna.it).



**Figura 1** Produzione nazionale di energia elettrica per fonte nel 2021

## 1.2 ZONE DI MERCATO

Il sistema elettrico viene suddiviso in porzioni di territorio, chiamate zone di mercato.

Le zone di mercato sono caratterizzate ciascuna da un prezzo zonale dell'energia.

La rete di trasmissione è interconnessa con l'estero attraverso 22 linee così distribuite: 4 con la Francia, 12 con la Svizzera, 1 con l'Austria e 2 con la Slovenia ed 1 cavo in corrente continua con la Grecia (Allegato a.24 al codice di rete individuazione zone della rete rilevante, Terna.it).

Le zone di mercato sono suddivise nel seguente modo (Figura 2):

- Zone geografiche: Centro – Nord; Nord; Centro – Sud; Sud; Sicilia; Sardegna; Calabria. (Prima del 2021 la Calabria faceva parte della zona di mercato Sud, successivamente è diventata una Zona di mercato a tutti gli effetti).
- Zone virtuali estere: Francia; Svizzera; Austria; Slovenia; Corsica; Corsica AC; Grecia; Malta e Montenegro.



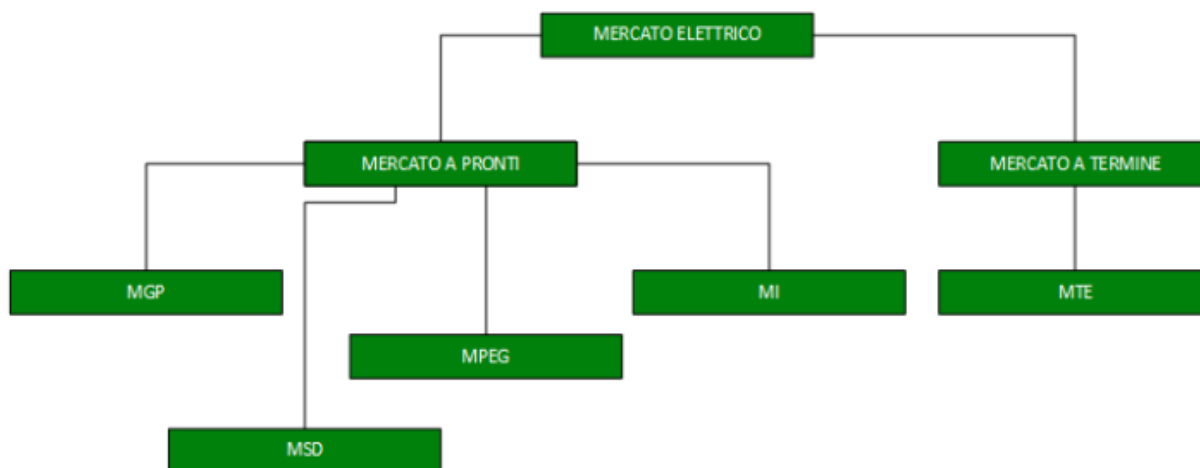
**Figura 2:** Zone di mercato dal 2021 (fonte Terna S.p.A.)

### **1.3 ARTICOLAZIONE DEL MERCATO ELETTRICO**

Il mercato elettrico, gestito e organizzato dal Gestore dei Mercati Energetici S.p.A., si articola nel seguente modo:

- Mercato Elettrico a Pronti (MPE)
- Mercato Elettrico a Termine dell'energia elettrica con obbligo di consegna e ritiro (MTE)
- Piattaforma per la consegna fisica dei contratti finanziari conclusi sull'IDEX (CDE)





**Figura 3** Articolazione del mercato elettrico in Italia nel 2023 (fonte GME)

Il Mercato Elettrico a Pronti (MPE) si articola a sua volta in:

- Mercato del Giorno Prima (MGP), ospita la maggior parte delle transazioni di compravendita di energia elettrica.  
Gli operatori partecipano presentando offerte nelle quali si indicano la quantità ed il prezzo massimo/minimo al quale si è disposti comprare o vendere. Le offerte sono accettate dopo la chiusura della seduta di mercato, sulla base del merito economico e nel rispetto dei limiti di transito tra le zone. Il MGP è quindi un mercato d'asta e non un mercato a contrattazione continua (Gestore dei Mercati Energetici.org).
- Mercato dei Prodotti Giornalieri (MPEG), è la sede per la negoziazione dei prodotti giornalieri con obbligo di consegna dell'energia. Sul MPEG sono negoziabili prodotti giornalieri con "differenziale unitario di prezzo", ovvero che il prezzo che si determina in fase di negoziazione è il differenziale, rispetto al P.U.N. al quale gli operatori sono disposti a negoziare tali prodotti. Inoltre sul medesimo mercato sono negoziabili anche prodotti giornalieri definiti a "prezzo unitario pieno" il prezzo dei quali è l'espressione del valore unitario di scambio dell'energia elettrica oggetto dei contratti negoziati.
- Il Mercato Infragiornaliero (MI), consente agli operatori di apportare modifiche ai programmi definiti nel MGP attraverso ulteriori offerte di acquisto o vendita. Le negoziazioni sul MI avvengono attraverso lo svolgimento di tre sessioni d'asta MI-A e una sessione di negoziazione continua MI-XBID (Gestore dei Mercati Energetici.org).
- Il Mercato del Servizio di Dispacciamento (MSD), attraverso il quale il Gestore della Rete di Trasmissione Nazionale si approvvigiona delle risorse per il servizio di dispacciamento.

E' la sede delle contrattazioni di energia tramite la selezione, sulla base di un ordine di merito economico, di offerte di aumento o riduzione dell'energia da immettere o prelevare in ogni ora (Mercatoelettrico.org).

Il Mercato Elettrico a Termine è la sede per la negoziazione di contratti a termine dell'energia elettrica con obbligo di consegna e ritiro.

Nel Mercato a Termine gli operatori partecipano presentando proposte nelle quali indicano tipologia e periodo di consegna dei contratti, numero dei contratti e prezzo al quale sono disposti ad acquistare/vendere (Gestore dei Mercati Energetici.org).

Il CDE, ovvero la Piattaforma per la consegna fisica dei contratti finanziari conclusi sull'IDEX, è dove vengono eseguiti i contratti finanziari derivati sull'energia elettrica sull'IDEX relativamente ai quali l'operatore abbia richiesto di esercitare l'opzione di consegna fisica sul mercato elettrico, dell'energia sottostante il contratto stesso (Gestore dei Mercati Energetici.org).

## 1.4 PREZZO UNICO NAZIONALE

Il Prezzo Unico Nazionale è l'indicatore del prezzo dell'energia che viene scambiata sul mercato nazionale della Borsa Elettrica Italiana, il quale viene calcolato facendo una media dei prezzi zionali del mercato del giorno prima (MGP) ponderata con gli acquisti totali, al netto di quelli provenienti dalle zone estere (Sorgenia.it, 2023).

$$P.U.N. = \frac{\sum_z P_z \cdot Q_z}{\sum_z Q_z}$$

dove  $P_z$  è il prezzo zonale e  $Q_z$  è la quantità acquistata, entrambi riferiti alla zona  $z$ .

Ci sono diversi fattori che possono influenzare il prezzo all'ingrosso dell'energia elettrica come il costo del petrolio, il costo del gas e le diverse performance economiche poiché un'economia in espansione è causa di un aumento dei consumi energetici i quali a loro volta fanno aumentare il valore P.U.N. (Sorgenia.it, 2023).

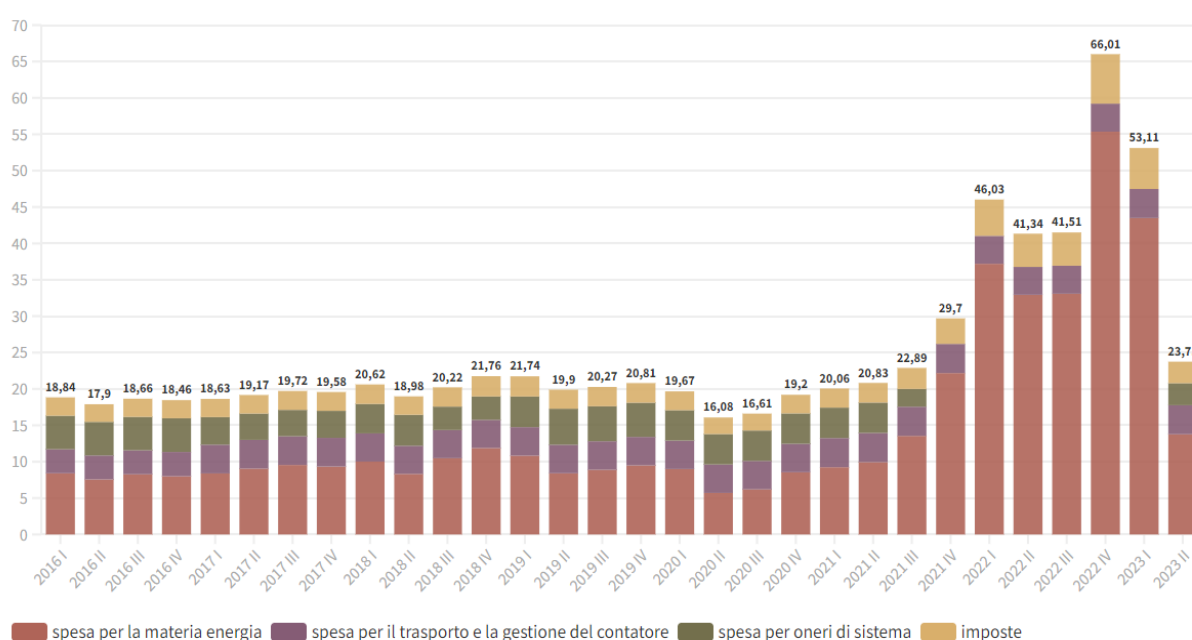
Il P.U.N. può essere scomposto in più voci che ne determinano il valore:

- Spesa per la materia energia
- Spesa per il trasporto e la gestione del contatore
- Spesa per oneri di sistema

## - Imposte

Per avere un'idea di quanto queste voci incidano sul P.U.N. prendiamo in considerazione il P.U.N. del primo trimestre del 2023 pari a 23,75 c€/kWh di cui:

- il 58% (13,77 c€/kWh) di spesa per la materia energia.
- il 17% (3,99 c€/kWh) di spesa per il trasporto e la gestione del contatore.
- il 13% (3,03 c€/kWh) di spesa per oneri di sistema.
- il 12% (2,96 c€/kWh) di imposte.



**Figura 4** Andamento del prezzo dell'energia elettrica per il consumo domestico in Italia dal 2016 al 2023 (Fonte Arera.it)

## 1.5 CARO ENERGIA: LE CAUSE E POSSIBILI SOLUZIONI

La crisi energetica che ha caratterizzato i mesi invernali del 2022 e i primi mesi del 2023 trova origine in diversi fattori, soprattutto geopolitici e di mercato.

La rapida riapertura delle attività economiche dopo il lockdown infatti ha generato un'inflazione che ha caratterizzato gli ultimi due anni.

La guerra in Ucraina è stato un ulteriore fattore scatenante della crisi energetica, poiché ha influito sulla reperibilità del gas.

La Russia infatti, il maggior esportatore di gas naturale in Europa, in risposta alle sanzioni ha diminuito il flusso di gas verso i paesi europei (LuceGas.it, “Caro Bollette Energia: come frenare il raddoppio dei prezzi?”).

L’andamento del Prezzo Unico Nazionale dell’energia elettrica è legato a quello del gas, specialmente in Italia, poiché buona parte consistente dell’energia elettrica prodotta nel nostro paese proviene dalla combustione del gas metano.

Il prezzo del gas quindi, influenza molto quello della luce nelle dinamiche che avvengono nella borsa elettrica (LuceGas.it, “TTF Gas (Title Transfer Facility): cos’è, a cosa serve e qual è il prezzo?”).

Per far fronte a questi aumenti il governo ha stanziato fondi e previsto misure per contenere i rincari nel breve periodo, tuttavia questi interventi hanno lo scopo di tamponare gli aumenti e non sono risolutivi né sostenibili a lungo termine. Oltre agli interventi citati pocanzi va affrontata la causa principale dei rincari, ovvero l’aumento della materia prima come si poteva notare nella figura 4.

Per contrastare il caro energetico le soluzioni migliori sono quelle di sbloccare lo sviluppo delle rinnovabili, favorire la creazione di comunità energetiche e migliorare l’efficienza energetica nazionale (legambiente.it, 2022).

La lentezza della burocrazia in Italia ha influito anche sullo sviluppo delle rinnovabili, le quali darebbero un grosso contributo al paese per essere meno dipendente dalle fonti fossili e conseguire gli obiettivi climatici europei del 2030.

Le comunità energetiche prevedono che l’energia prodotta da fonti rinnovabili possa essere condivisa da consumatori come privati, piccole e medie imprese e amministrazione pubblica, i quali ne trarrebbero benefici importanti in termini di risparmio sui costi di approvvigionamento dell’energia. Le politiche di efficienza apportate sugli edifici, residenziali e non, potrebbero far sì che i consumi si riducano fino al 50%, parametri in linea con le proposte europee (legambiente.it, 2022).

## **1.6 CARATTERISTICHE E PROPRIETA’ DEI PREZZI DELL’ENERGIA ELETTRICA**

Il prezzo dell'energia elettrica gode di alcune proprietà dovute sia al mercato che alla natura del prodotto.

Considerato che l'energia elettrica non può essere immagazzinata in grandi quantità la domanda e l'offerta devono essere in una situazione di equilibrio, la produzione di energia, quindi, deve corrispondere alla domanda di energia.

La particolare attenzione posta al cambiamento climatico degli ultimi anni ha incrementato la produzione di energia tramite le fonti rinnovabili, specialmente eolico e solare, rendendo i prezzi dell'energia più volatili, dal momento che la produzione di energia rinnovabile è influenzata delle condizioni meteorologiche (Graßl, Markus, and Gunther Reinhart, 2014).

Un'altra caratteristica dei prezzi dell'energia sono i Jumps, ovvero balzi improvvisi verso l'alto (Graßl, Markus, and Gunther Reinhart, 2014).

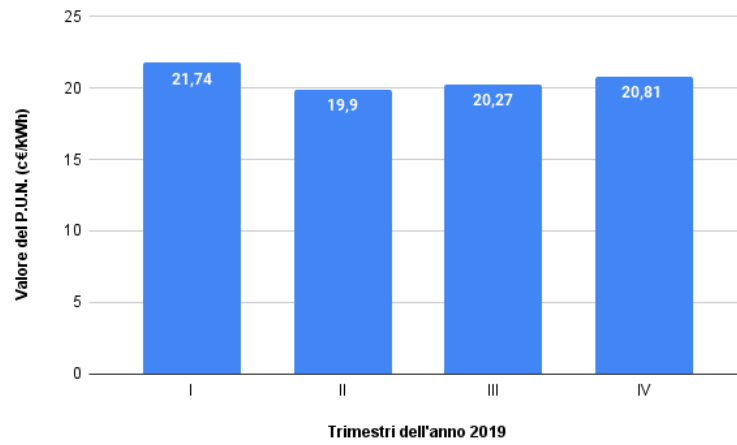
I Jumps vengono tuttavia compensati dal fenomeno del mean-reversion, letteralmente “ritorno alla media”; infatti si osserva che nei periodi successivi ai picchi il prezzo dell'energia torna a muoversi in un intorno del prezzo medio (Graßl, Markus, and Gunther Reinhart, 2014).

Il prezzo spot dell'energia elettrica ha una forte componente stagionale. Nel primo trimestre dell'anno, il più freddo, a causa del prezzo del gas ai massimi annuali, il prezzo dell'energia elettrica è tra i più alti dell'anno.

Nel secondo trimestre, al minor costo del gas si aggiunge anche la produzione idroelettrica dovuta allo scioglimento delle nevi invernali.

Nel terzo trimestre, caratterizzato dalle alte temperature estive, il prezzo dell'energia torna a salire rispetto al secondo trimestre in risposta all'aumento di domanda dovuto all'uso di condizionatori e a dispositivi di raffreddamento.

Nel quarto trimestre il calo delle temperature fa aumentare nuovamente il prezzo dell'energia (yem-energy.com, 2021).



**Figura 5** L'andamento del P.U.N. per trimestre nel 2019 (Fonte ARERA.it)

## CAPITOLO 2

### MODELLI PER LE SERIE STORICHE

#### 2.1 INTRODUZIONE ALLE SERIE STORICHE

Per serie si intende una collezione di osservazioni di un fenomeno. Una serie storica, o serie temporale, esprime la dinamica di un certo fenomeno  $X_t$ , con  $t \in \mathbb{Z}$ , registrato istantaneamente (variabile di stato) o conteggiato nel periodo definito (variabile di flusso). Si parla di serie storiche univariate, bivariate o multivariate a seconda che i fenomeni oggetto di indagine siano 1,2 o molteplici (Domenico Piccolo, 2010).

Una serie storica, quindi è una successione di dati riferiti ad un fenomeno, nella quale ogni dato è associato un istante temporale.

L'analisi statistica delle serie storiche può essere svolta con due metodologie diverse, note come "Approccio Moderno" e "Approccio Classico".

Per comprendere le differenze tra i due approcci formalizziamo il processo generatore dei dati con la seguente funzione:

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

L'approccio moderno non pone l'attenzione sulla parte sistematica del modello, ovvero  $f(t)$  la quale deve essere opportunamente stimata, quanto alla componente accidentale del modello  $u_t$  con  $u_t \sim \text{WN}(0, \sigma^2)$ .

L'approccio classico assume che il processo rappresentato dalla serie, comprenda una parte deterministica che consente di stimare le componenti virtuali della serie (trend/ciclo/stagionalità) e una componente di disturbo casuale.

## 2.2 PROCESSI STOCASTICI: DEFINIZIONE E PROPRIETA'

Un processo stocastico può essere definito come una sequenza infinita di variabili casuali. L'analisi di una serie storica è condotta attraverso l'individuazione del processo stocastico generatore della serie, e la ricerca del processo avviene sulla base di un suo insieme finito di realizzazioni, la serie storica, dove la t-esima componente  $x_t$  è l'unica realizzazione della variabile casuale  $X_t$ .

I momenti più utilizzati sono:

- Media:  $\mu_1(t) = E[X_t]$

- Varianza:  $\mu_2 = E[X_t - \mu_1(t)]^2 = \gamma(0)$

- Autocovarianza (momento misto centrale del secondo ordine):

$$\mu_{11} = E[X_{t_j} - \mu_1(t)][X_{t_i} - \mu_1(t)] = \gamma(t_i; t_j)$$

- Autocorrelazione globale:

$$\rho(k) = \frac{\text{Cov}(X_t; X_{t-k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)}$$

- Autocorrelazione parziale:

$$\text{Corr}(X_t; X_{t+k} | X_{t+1}, \dots, X_{t+k-1})$$

Le proprietà dei processi stocastici sono la stazionarietà, l'invertibilità e l'ergodicità. Il processo stocastico  $X_t$  è detto stazionario in senso forte se le distribuzioni congiunte  $(F_{t1}, \dots, F_{tn})$  e  $(F_{t1+k}, \dots, F_{tn+k})$  sono uguali per ogni  $t_n$  e per ogni  $k$ . Poiché questa ipotesi è molto limitativa e difficilmente dimostrabile, per l'analisi delle serie storiche si preferisce far ricorso alla stazionarietà in senso debole.

Un processo stocastico è detto stazionario in senso debole se sono stazionarie la media e la varianza, ovvero che non dipendono dal tempo e la covarianza dipende solo dallo sfasamento dei due tempi considerati.

Ovvero:

- $E[X_t] = \mu$  per  $\forall t$ .
- $\text{Var}(X_t) = \gamma(0)$  per  $\forall t$ .



- $\text{Cov}(X_t, X_{t+k}) = \gamma(k)$  per  $\forall t, \forall k$ .

Un processo stocastico è invertibile se si può esprimere in funzione delle variabili casuali del passato con l'aggiunta di una variabile casuale, ovvero:

$$X(t)=[X_{t-1}, X_{t-2}, \dots, a_t]$$

Data la definizione dei momenti di un processo stocastico, sembra naturale calcolarne una stima sulla base di un numero finito N di risultati, tuttavia nelle serie storiche si conosce solo una delle possibili realizzazioni.

Il teorema ergodico permette, quando ci si trova in queste particolari condizioni, di sostituire al calcolo degli stimatori usuali, l'applicazione di stimatori temporali.

## 2.3 MODELLI DELL'APPROCCIO MODERNO: AR, MA E ARMA

Il processo AR(p) consiste nell'esprimere  $Z_t$  come funzione delle sue realizzazioni passate:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t$$

Introducendo l'operatore ritardo B è possibile riscrivere il processo AR(p) nel seguente modo:

$$Z_t = \phi(B)Z_t + a_t$$

Dove  $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ .

Il processo AR(p) è sempre invertibile poiché  $Z_t$  è espressa in funzione dei suoi valori passati ed è anche stazionario se tutte le radici  $B_i$  dell'equazione caratteristica  $\phi(B)=0$  in valore assoluto sono maggiori di uno, ovvero  $|B_i| > 1$  per  $i=1, \dots, p$ .

I momenti di un processo AR(p) sono i seguenti:

- Varianza:  $\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p) + \sigma_a^2$
- Autocovarianza:  $\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2) + \dots + \phi_p\gamma(k-p)$  per  $k \neq 0$ .
- Autocorrelazione globale:  $\rho(k) = \frac{\phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p) + \sigma_a^2}{\phi_1\gamma(k-1) + \phi_2\gamma(k-2) + \dots + \phi_p\gamma(k-p)}$  per ogni  $k \leq p$ ,  $\rho(k)=0$  altrimenti.

Il modello a media mobile MA(q) può essere espresso come:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Con l'operatore ritardo B l'espressione del modello a media mobile diventa:

$$Z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t$$

Il processo MA(q) è sempre stazionario poiché media, varianza e covarianza di  $Z_t$  sono costanti e non dipendono da t e risulta stazionario se tutte le radici  $B_i$  dell'equazione caratteristica  $\phi(B)=0$  in modulo sono maggiori di uno, ovvero  $|B_i| > 1$  per  $i=1, \dots, q$ .

I momenti di un processo MA(q) sono:

- Varianza:  $\gamma(0) = \sigma_a^2 + \theta_1^2 \sigma_a^2 + \dots + \theta_q^2 \sigma_a^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2$
- Autocovarianza:  $\gamma(k) = (-\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_q \theta_{k-q}) \sigma_a^2$
- Autocorrelazione globale:  $\rho(k) = \frac{-\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_q \theta_{k-q}}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)}$  per ogni  $k \leq q$ ,  $\rho(k)=0$  altrimenti.

Tra i modelli AR(p) e MA(q) esiste una relazione, infatti ogni processo stazionario AR(p) di ordine finito si può scrivere come un modello MA di ordine infinito e ogni processo MA(q) di ordine infinito si può scrivere come un modello AR di ordine infinito.

Questa relazione duale tra i modelli MA e AR si ritrova anche nelle rispettive funzioni di autocorrelazione e di autocorrelazione parziale. Infatti un processo MA presenta una funzione di autocorrelazione che si annulla dopo il ritardo corrispondente all'ordine del processo, mentre la sua funzione di autocorrelazione parziale decresce verso lo zero con andamenti diversi che dipendono ancora all'ordine del processo e dal valore i suoi parametri.

Un modello ARMA(p,q) si presenta con la seguente forma:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Usando gli operatori polinomiali  $\phi(B)$  e  $\theta(B)$ , la precedente formulazione può essere riscritta in forma compatta come:

$$\phi_p(B)Z_t = \theta_q(B)a_t$$

Il processo ARMA(p,q) risulta stazionario se tutte le radici  $B_i$  dell'equazione caratteristica  $\phi(B)=0$  hanno modulo maggiore di uno, ovvero  $|B_i|>1$  per  $i=1,\dots,p$ .

Il processo risulta invece invertibile se tutte le radici  $B_j$  dell'equazione caratteristica  $\theta(B)=0$  hanno modulo maggiore di uno, ovvero  $|B_j|>1$  per  $j=1,\dots,q$ .

## 2.4 LA NON-STAZIONARIETA' NELLE SERIE STORICHE

Gran parte delle serie storiche presentano un andamento non stazionario. Nei modelli ARMA la non stazionarietà di un modello si manifesta quando una o più radici del polinomio che rappresenta la parte AR sono inferiori a uno in valore assoluto.

La non stazionarietà può essere di due tipi:

- Non stazionarietà in media, che si verifica quando il livello della serie non rimane costante.
- Non stazionarietà in varianza si ha quando la varianza della serie non è costante nel tempo.

Per trattare serie storiche con non-stazionarietà in media vengono usati i modelli ARIMA (AutoRegressive, Integrated, Moving Average), l'operatore differenza del primo ordine viene usato per trattare la non stazionarietà in media ma con un trend lineare. Il metodo che viene più usato per trattare la non stazionarietà in varianza è quello introdotto da Box e Cox nel 1964 ed è conosciuto come la "trasformazione Box-Cox".

Per un processo  $Z_t$  questa trasformazione è fornita dal processo  $Z_t(\lambda)$  definito da:

$$Z_t(\lambda) = \begin{cases} \frac{Z_t^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \ln(Z_t), & \text{se } \lambda = 0 \end{cases}$$

## 2.5 LA PROCEDURA DI BOX-JENKINS

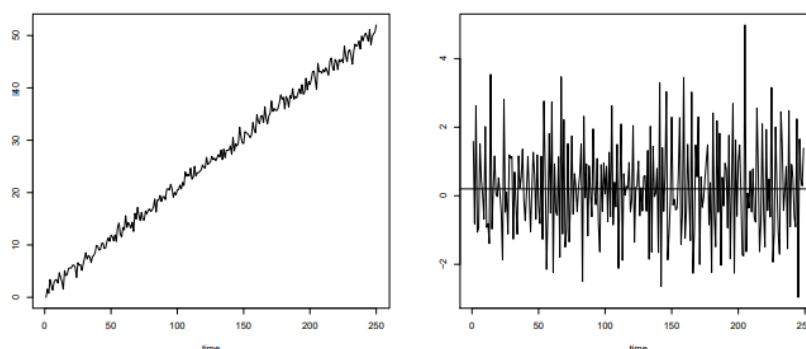
L'identificazione del modello viene effettuata tramite degli step elaborati da Box e Jenkins nel 1979, chiamata per l'appunto Procedura di Box-Jenkins (Box e Jenkins, 1979) .

La procedura di Box-Jenkins può essere riassunta nelle seguenti 6 fasi:

- 1) Analisi grafica della serie storica
- 2) Verifica della stazionarietà
- 3) Analisi dei correlogrammi
- 4) Stima del modello
- 5) Scelta e verifica del modello
- 6) Previsione

### 2.5.1 ANALISI GRAFICA

Rappresentare graficamente la serie è utile per identificare l'eventuale presenza di trend, stagionalità e non stazionarietà in media ed in varianza.



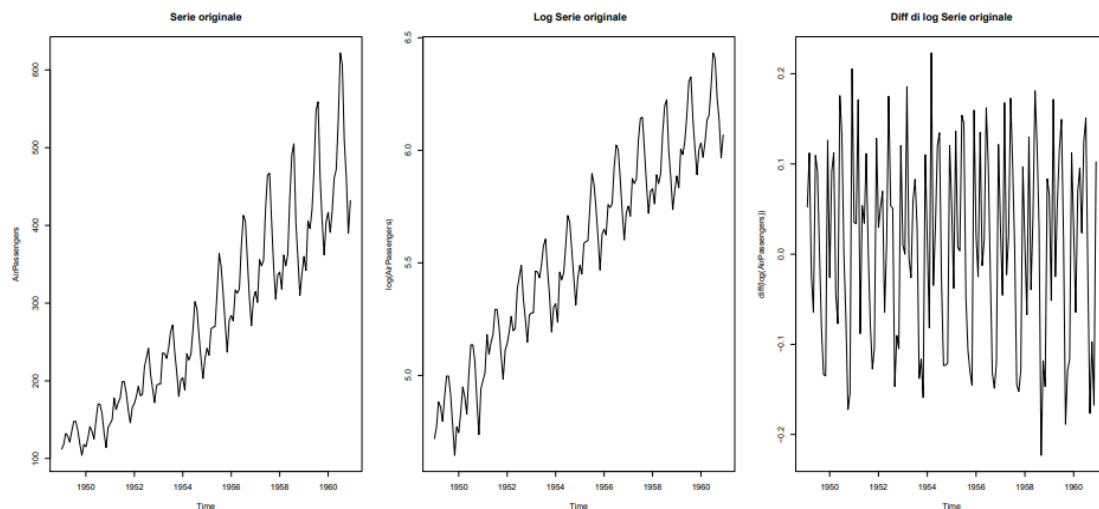
**Figura 6** A sinistra una serie non-stazionaria, a destra la stessa serie resa stazionaria.

### 2.5.2 VERIFICA DELLA STAZIONARIETÀ DELLA SERIE

Nel caso in cui dall'analisi grafica emergesse un andamento non stazionario la prima cosa da fare è capire che tipo di non stazionarietà è presente nella serie.

Nel caso in cui la serie fosse non stazionaria in media bisogna applicare la differenziazione, invece se la serie fosse non stazionaria in varianza si fa ricorso alle trasformazioni di Box-Cox.

Esiste anche il caso in cui una serie presenti la non stazionarietà sia in media che in varianza si applica prima la trasformazione di Box-Cox, solitamente la trasformazione logaritmica, e solo successivamente si differenzia, poiché una differenziazione può dare valori negativi i quali darebbero problemi con i logaritmi.



**Figura 7** Passaggi per trattare una serie non stazionaria sia in media che in varianza.

### 2.5.3 ANALISI DEI CORRELOGRAMMI

Si calcola ed esamina l'ACF e la PACF della serie differenziata per identificare l'ordine  $p$  e  $q$  dei polinomi AR e MA.

L'ordine dei polinomi si decide in base a questa regola generale:

PROCESSI	ACF	PACF
AR( $p$ )	Decresce esponenzialmente	Nulla dopo il lag $p$
MA( $q$ )	Nulla dopo il lag $q$	Decresce esponenzialmente
ARMA( $p, q$ )	Decresce esponenzialmente	Decresce esponenzialmente

**Tabella 1** Regola generale per stabilire l'ordine del Modello

### 2.5.4 STIMA DEL MODELLO

Si calcolano e si esaminano l'ACF e la PACF al fine di identificare l'ordine degli operatori AR e MA per prendere in considerazione qualche modello, generalmente 3 o 4, che soddisfino l'analisi fatta al punto precedente.

### 2.5.5 VERIFICA DEL MODELLO

La verifica del modello si svolge nei seguenti passaggi:

- Verifica della significatività dei parametri stimati
- Test sui residui, per verificarne la natura di White Noise

I parametri vengono stimati massimizzando la funzione di verosimiglianza dei dati.

Ipotizzando di aver osservato una realizzazione finita del processo data da  $z_1, z_{t-1}, \dots, z_1$ , dobbiamo calcolare la funzione di probabilità:  $f_{z_t, z_{t-1}, \dots, z_1}(z_t, z_{t-1}, \dots, z_1; \beta)$  che può essere vista come la probabilità di aver osservato questo particolare campione. Una volta stimati i parametri è necessario accertarsi che siano significativi attraverso un test statistico. Sia  $\beta = (\phi, \beta)$  la stima di  $\hat{\beta}$  si distribuisce come una normale multivariata con vettore delle medie  $\beta$  e con matrice di varianze e covarianze  $\Sigma$ .

L'ipotesi da verificare è quindi:

$$H_0: \beta_j = 0$$

contro l'ipotesi alternativa:

$$H_1: \beta_j \neq 0$$

Dall'ipotesi di normalità la statistica test da utilizzare è una T di Student con gradi di libertà pari a  $n-(p+q+1)$  costruita come segue:

$$t = \frac{\hat{\beta}_{jj} - 0}{\sqrt{\hat{\Sigma}_{jj}}} \sim T_{n-(p+q+1)}$$

Ponendo  $\alpha=0.05$  i valori soglia risultano essere quindi  $t_{\alpha/2}= 2$  e  $-t_{\alpha/2}= -2$ , quindi si accetta l'ipotesi che il coefficiente  $\beta$  non sia significativamente diverso da 0 se il valore della statistica test, in valore assoluto è minore di 2 oppure il coefficiente  $\beta$  in valore assoluto è minore di due volte lo standard error della stima:

$$|\hat{\beta}_j| \leq 2\sqrt{\hat{\Sigma}_{jj}}$$

Quando si cerca di individuare il processo generatore di una serie cui si può trovare di fronte ad una moltitudine di modelli che descrivono bene l'andamento dei dati.

La scelta del modello non può essere condotta tramite il criterio dell'indice  $R^2$ . L'indice  $R^2$  infatti può essere riscritto nel modo seguente:

$$R^2 = 1 - \frac{Var(a_t)}{Var(Z_t)} = \frac{\sum_{j=1}^{\infty} \psi_j}{1 + \sum_{j=1}^{\infty} \psi_j^2}$$

Da questa scomposizione è facile notare come l'indice  $R^2$  dipenda unicamente dai pesi del White Noise e non dai dati. I criteri maggiormente usati sono il criterio di Akaike (AIC) e il criterio di Schwartz (SBC). Entrambi i criteri penalizzano la presenza di molti parametri, i quali rendono il modello complesso, quindi a parità di adeguatezza si sceglie il modello con AIC o SBC minori.

Ponendo  $M$  il numero di parametri stimati e  $t$  la numerosità della serie il criterio di Akaike e il criterio di Schwartz sono così definiti:

$$AIC(M) = 2M - 2\log(L(\beta)) \approx t \log(\sigma_a^2) + 2M$$

$$SBC(M) = t \log(\sigma_a^2) + M \log(t)$$

Una volta scelto il modello, attraverso delle verifiche diagnostiche, bisogna accertarsi che il modello sia adeguato. Per verificare l'adeguatezza del modello si attua un'analisi sui residui  $\hat{a}_t$ , ovvero le stime dei White Noise non osservati  $a_t$ , i quali devono soddisfare due ipotesi:

- Normalità
- Incorrelazione

Per verificare la normalità dei residui si fa ricorso al test di Jarque-Bera (1979) che si basa sull'asimmetria e sulla curtosi della distribuzione oggetto di verifica.

Una distribuzione normale è caratterizzata da una asimmetria nulla e ad un indice di curtosi pari a 3.

La statistica test proposta da Jarque e Bera è la seguente:

$$JB = \frac{n}{6} \left( S^2 + \frac{1}{4}(k - 3)^3 \right)$$

dove S è l'indice di asimmetria, k l'indice di curtosi e n il numero delle osservazioni.

Sotto l'ipotesi nulla  $H_0$  il test si distribuisce asintoticamente come un chi quadro con 2 gradi di libertà. Per valori troppo grandi della statistica test si rifiuta l'ipotesi di normalità (Jarque, Bera, 1987).

Per verificare l'incorrelazione dei residui si utilizza il test di Portmanteau che utilizza tutte le autocorrelazioni campionarie dei residui fino al lag k per verificare l'ipotesi nulla:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

con la statistica test:

$$Q = t(t + 2 \sum_{k=1}^K (t - k)^{-1} \rho_k^2)$$

$$Q \sim \chi_{M-K}^2$$

Sotto l'ipotesi nulla la statistica Q si distribuisce come una chi-quadrato con M-K gradi di libertà, dove M è il numero di parametri stimati dal modello (Box e Pierce; 1970).

L'eventuale violazione dell'ipotesi di normalità potrebbe essere data dalla presenza di outliers, i quali solitamente sono individuabili da subito nella serie originale.

La presenza di valori anomali può essere trattata sostituendoli con dei valori interpolati oppure se gli effetti si limitano alla non normalità dei residui, si può considerare comunque il modello buono a meno degli outliers.



## 2.6 MODELLI CON COMPONENTE STAGIONALE

La maggior parte delle serie storiche economiche è affetta da stagionalità. La stagionalità è una componente di breve periodo costituita da variazioni che si riscontrano con intensità simile negli stessi periodi di anno in anno ma con intensità diversa nel corso di uno stesso anno. Un processo ARMA stagionale si indica con la sigla  $ARMA(P,Q)^s$ , dove  $P$  è l'ordine della parte autoregressiva della serie,  $Q$  è l'ordine della parte a media mobile della serie e  $s$  indica la periodicità stagionale.

Un processo autoregressivo stagionale di ordine  $P$ ,  $AR(P)^s$ , risulta avere la seguente formula generale:

$$Z_t = \Phi_1 Z_{t-s} + \Phi_2 Z_{t-2s} + \dots + \Phi_P Z_{t-Ps} + a_t$$

Un processo costituito da un  $AR(p)$  e un  $AR(P)^s$  è possibile rappresentarlo nel seguente modo con l'operatore ritardo:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})Z_t = a_t$$

Un processo a media mobile stagionale di ordine  $Q$ ,  $MA(Q)^s$ , ha invece la seguente notazione:

$$Z_t = a_t - \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - \dots - \Theta_Q a_{t-Qs}$$

Usando l'operatore ritardo è possibile rappresentare un processo costituito da una componente  $MA(q)$  e da una componente  $MA(Q)^s$  nel modo che segue:

$$Z_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})a_t$$

Una serie storica può essere generata da componenti autoregressive e a media mobili stagionali e non stagionali, formalmente siamo in presenza di un processo  $SARIMA(p,d,q)(P,D,Q)^s$  formalizzabile con l'uso dell'operatore ritardo nel modo seguente:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})Z_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})a_t$$

Dove:

- $(1 - \phi_1 B - \dots - \phi_p B^p)$  rappresenta il processo AR(p) non stagionale
- $(1 - \Phi_1 B^s - \dots - \Phi_p B^{ps})$  rappresenta il processo AR(P)<sup>s</sup> stagionale
- $(1 - \theta_1 B - \dots - \theta_q B^q)$  rappresenta il processo MA(q) non stagionale
- $(1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})$  rappresenta il processo MA(Q)<sup>s</sup> stagionale

Le funzioni di correlazione di serie stagionali seguono le stesse regole dei processi non stagionali, tuttavia bisogna porre particolare attenzione ai picchi stagionali, in corrispondenza di  $s$  e ai multipli di  $s$ .

PROCESSI	ACF	PACF
AR(p)	Decresce ai lag $k*s$	Nulla dopo il lag $p*s$
MA(p)	Nulla dopo il lag $q*s$	Decresce ai lag $k*s$
ARMA(p,q)	Decresce ai lag $k*s$	Decresce ai lag $k*s$

**Tabella 2** Regola generale per stabilire l'ordine del modello con stagionalità

Identificare il modello quando si è in presenza di stagionalità richiede degli accorgimenti in più rispetto alle serie senza componente stagionale.

Per trattare la non stazionarietà in caso di serie stagionale bisogna attuare una differenziazione di ordine  $s$ ; durante l'analisi dei residui il test di incorrelazione va valutato nei lag multipli di  $s$ .

## 2.7 PREVISIONE NEI MODELLI ARIMA

Ipotizzando che il modello ARIMA sia stato correttamente identificato e che la diagnostica sui residui non abbia evidenziato problematiche si procede con l'ultima fase della procedura, ovvero la previsione, che rappresenta l'obiettivo finale dell'analisi della serie storica.

Il problema della previsione consiste nel trovare un'approssimazione ottimale  $\hat{Z}_{t+1}$  di una variabile casuale  $Z_{t+1}$  tramite una funzione  $g(Z_t)$  dove  $Z_t = (z_1, \dots, z_t)$  è l'insieme delle variabili casuali disponibili fino al momento  $t$ .

La scelta della funzione  $g(x)$  assume un ruolo cruciale per l'attendibilità della previsione, si sceglie infatti la funzione che minimizza l'errore quadratico medio che si commette sostituendo la variabile da prevedere  $Z_{t+1}$  con la funzione scelta

$\hat{Z}_{t+1} = g(Z_t)$ . Formalmente:

$$MSE = E\{[Z_{t+1} - g(Z_t)]^2 | Z_t\} = \min$$

L'errore di previsione è dato dalla differenza della variabile da prevedere  $Z_{t+h}$  e la sua approssimazione  $\hat{Z}_{t+h}$ , ovvero:

$$e_t = Z_{t+h} - \hat{Z}_{t+h}$$

I momenti dell'errore di previsione sono:

$$- E(e_t) = E[Z_{t+h} - \hat{Z}_{t+h}] = 0$$

$$- \text{Var}(e_t) = \text{Var}(Z_{t+h} - \hat{Z}_{t+h}) = E[Z_{t+h} - \hat{Z}_{t+h}]^2 = \sigma_a^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2)$$

La varianza dell'errore di previsione, per  $h$  che tende ad infinito, converge alla varianza del processo  $Z_t$ , formalmente:

$$\lim_{h \rightarrow \infty} [\sigma^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2)] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

Per processi normali, dalla varianza dell'errore di previsione è possibile trovare i valori soglia dell'intervallo di confidenza per le previsioni:

$$\hat{Z}_{t+h} - \frac{z_{\alpha/2}}{2} \sqrt{\hat{v}_k} \leq Z_{t+h} \leq \hat{Z}_{t+h} + \frac{z_{\alpha/2}}{2} \sqrt{\hat{v}_k}$$

dove  $z_{\alpha/2}$  è il quantile della normale standard al livello  $\alpha$  e  $\hat{v}_k$  è la varianza dell'errore di previsione.

Un processo AR(1) non stagionale può essere scritto come segue:

$$Z_t = \phi_1 Z_{t-1} + a_t$$

Ipotizzando di voler prevedere il valore  $Z_{t+1}$  si definisce il previsore ottimo:

$$\hat{Z}_{t+1} = E(Z_{t+1} | Z_t, Z_{t-1}, \dots) = E(\phi_1 Z_t + a_{t+1})$$

poiché  $E(a_{t+1}) = 0$  posso riscrivere la formula nel modo seguente:

$$\hat{Z}_{t+1|1} = \hat{\phi}_1 Z_t$$

analogamente è possibile calcolare le previsioni per  $t+k$  periodi avanti:

$$\hat{Z}_{t+k|1} = \hat{\phi}_1^k Z_t$$

La stima della varianza dell'errore di previsione, per il modello AR(1), necessaria per la costruzione dell'intervallo di confidenza risulta essere:

$$\hat{v}_k = \hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2 = \hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\phi}^{2j} = \hat{\sigma}_a^2 \frac{1 - \hat{\phi}^{2k}}{1 - \hat{\phi}^2}$$

Generalizzando ai processi AR(p), per ogni periodo avanti k, la previsione è funzione esclusivamente dei parametri del processo e delle ultime p osservazioni.

Ricordando che un modello MA(1) si formalizza nel seguente modo:

$$Z_t = a_t - \theta_1 a_{t-1}$$

Nell'ipotesi di voler prevedere il valore al periodo t+1, il previsore ottimo in questo caso è:

$$\hat{Z}_{t+1} = E(Z_{t+1} | Z_t, Z_{t-1}, \dots) = E(a_{t+1} - \theta_1 a_t)$$

poiché  $E(a_{t+1}) = 0$  si ha quindi:

$$\hat{z}_{t+1|t} = -\hat{\theta}_1 a_t$$

In un modello MA(q) è possibile fare previsioni non nulli per i periodi q.

La stima della varianza dell'errore di previsione nel caso di un MA(1) è:

$$\hat{v}_k = \hat{\sigma}_a^2 (1 + \hat{\theta}_1^2)$$

Ipotizzando di aver adattato un modello ARMA(1,1), formalizzabile come segue:

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1}$$

Le previsioni ottimali sono quindi:

$$\hat{z}_{t+1|t} = \hat{\phi}_1 z_t - \hat{\theta}_1 a_t$$

$$\hat{z}_{t+k|t} = \hat{\phi}_1^k z_t - \hat{\theta}_1 \hat{\phi}_1^{k-1} a_t$$

La varianza dell'errore di previsione per il modello in esame risulta:

$$\hat{v}_k = \hat{\sigma}_a^2 \left\{ 1 + \sum_{j=1}^{k-1} [\hat{\phi}_1^{j-1} (\hat{\phi}_1 - \hat{\theta}_1)]^2 \right\}$$

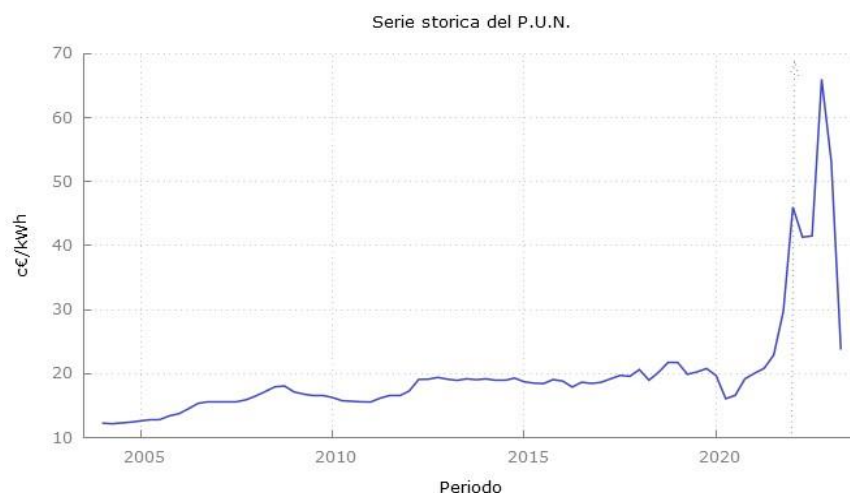
## CAPITOLO 3

### ANALISI DELLA SERIE STORICA DEL PREZZO UNICO NAZIONALE

In questo capitolo si descriverà la procedura della scelta e verifica del modello SARIMA che meglio spiega l'andamento del prezzo unico nazionale al fine di poter fare previsione. La formalizzazione del modello verrà eseguita con delle funzioni viste nel corso di Serie Storiche ed economiche, funzioni che essendo state implementate in ambiente R non fanno parte di nessun pacchetto.

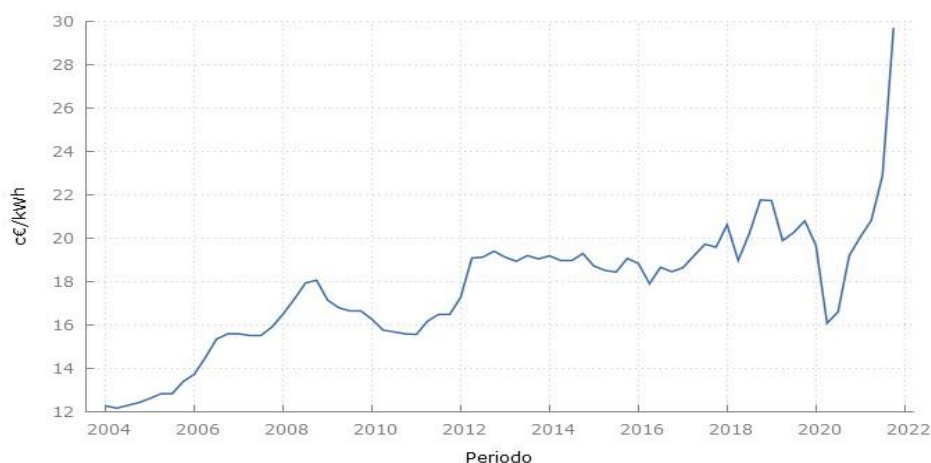
#### 3.1 FOCUS SUI DATI E ANALISI GRAFICA DELLA SERIE STORICA

I dati oggetto di analisi sono stati reperiti dal sito Arera.it e vengono raccolti con cadenza giornaliera, poiché come spiegato nel capitolo 1 il P.U.N. è una grandezza che non solo varia di giorno in giorno ma varia anche all'interno della giornata in base alle fasce orarie. L'Autorità di regolazione per energia reti e ambiente (Arera) ha reso disponibili i valori medi dei trimestri del P.U.N. dal primo trimestre del 2004 fino all'attuale trimestre, ovvero il II trimestre 2023.



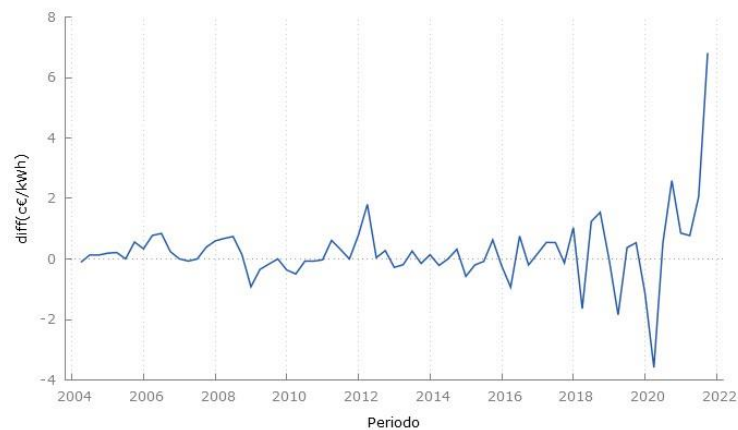
**Figura 8** Andamento del P.U.N. dal I trimestre 2004 al II trimestre 2023 (grafico costruito con il software Gretl)

La presenza di valori anomali causati da fattori esogeni, come la guerra in Ucraina citata nel capitolo precedente, rende sensata la rimozione dei valori dalla serie relativi ai periodi dal I trimestre 2022 al II trimestre 2023. Consideriamo quindi i valori del Prezzo Unico Nazionale dal I trimestre del 2004 al IV trimestre del 2021, ottenendo quindi la seguente serie storica:



**Figura 9** Andamento del P.U.N. depurato dai valori che causano lo shock

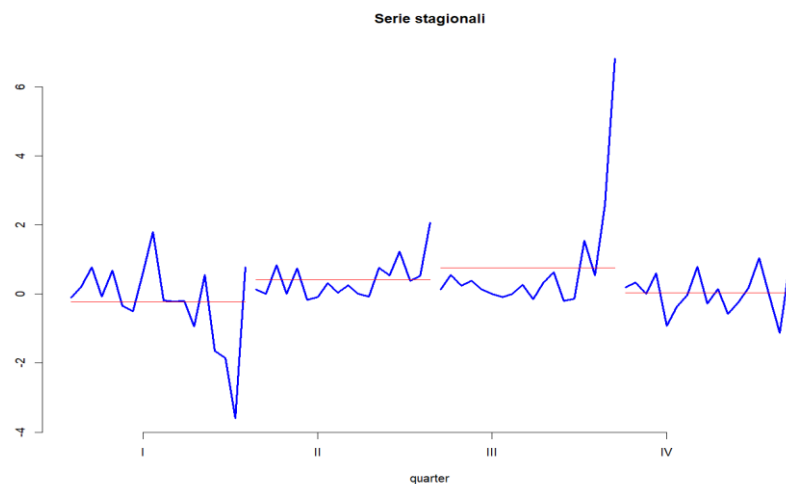
Facendo un'analisi grafica della serie depurata dai valori che danno luogo allo shock, si può notare come il cambiamento di scala metta in evidenza come la serie sia non stazionaria in media, avendo un trend crescente, ma risulta tutto sommato stazionaria in varianza. La non stazionarietà in media, come detto nel capitolo 2, si tratta attuando una differenziazione della serie, tuttavia la differenziazione di ordine 1 può non essere sufficiente se la serie differenziata risulta avere ancora un trend. La serie storica del P.U.N. differenziata risulta avere quindi questo andamento:



**Figura 10** Serie del P.U.N. differenziata

La serie differenziata non sembra avere un trend, la differenza prima quindi risulta essere sufficiente per trattare la non stazionarietà in media.

Nel capitolo 1 è emerso come il Prezzo Unico Nazionale fosse influenzato da una componente stagionale, in questo caso se con la funzione *seasplot*, che come output ha un grafico che fa emergere la stagionalità, emergono dei salti rilevanti, la serie andrebbe differenziata ulteriormente 4 volte poiché sono dati trimestrali.



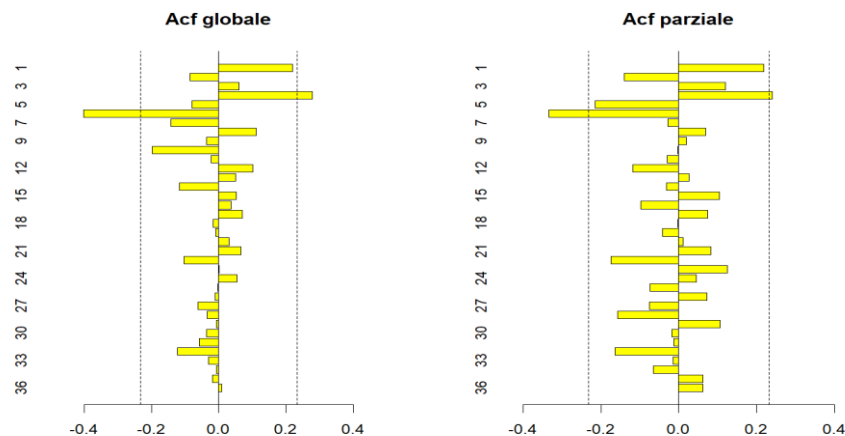
**Figura 11** Output della funzione *seasplot* applicata ai dati differenziati

L'output della funzione *seasplot* evidenzia come la differenziazione di ordine 1 abbia trattato anche la componente stagionale, infatti non emergono sostanziali differenze tra i trimestri. La differenziazione di ordine 4 sui dati differenziati per trattare la stagionalità, quindi, risulta essere inutile.



### 3.2 ANALISI DEI CORRELOGRAMMI E SCELTA DEL MODELLO

Il passo successivo per l'analisi della serie storica è l'analisi dei correlogrammi, seguendo le regole generali della Tabella 1 e della Tabella 2, presenti nel secondo capitolo, utili per stimare i parametri della componente stagionale e della componente non stagionale. Per visualizzare i correlogrammi si è fatto uso della funzione *correlogrammi* la quale, applicata alla serie differenziata oggetto di analisi, ha l'output che segue:



**Figura 12** Correlogramma totale e parziale della serie differenziata del Prezzo Unico Nazionale

Focalizzandosi sui multipli di 4 per la componente stagionale i correlogrammi evidenziano una ACF con valori che escono dalle bande ai lag 1 e 5, facendo propendere per una componente  $AR(1)^s$  o  $AR(2)^s$ . Guardando invece i lag non stagionali i notiamo che sia nella ACF che nella PACF l'autocorrelazione si annulla definitivamente dopo il lag 6.

La funzione *arimaest*, la quale prende in input gli ordini dei modelli ipotizzati, restituisce in output le stime dei parametri e AIC, è utile per valutare i modelli. I modelli che hanno i parametri significativi, con i relativi output, risultano essere i seguenti:

```
- AR(1)xAR(1)s
Coefficients:
      ar1      sar1
    0.5011  0.5769
s.e.  0.1295  0.1236

sigma^2 estimated as 0.9351:  log likelihood = -99.35,  aic = 204.71

$statistics
      ar1      sar1
3.870469 4.668234
```

- MA(1)xAR(1)<sup>s</sup>

```

Coefficients:
      ma1      sar1
      0.4131  0.5135
s.e.    0.1344  0.1391

sigma^2 estimated as 0.9914:  log likelihood = -101.13,  aic = 208.26

$tstatistics
      ma1      sar1
3.074000 3.690094

```

- ARMA(2,2)xAR(1)<sup>s</sup>

```

Coefficients:
      ar1      ar2      ma1      ma2      sar1
      1.7891 -0.9300 -1.4582  0.5488  0.6815
s.e.    0.0793  0.0722  0.1485  0.1375  0.1229

sigma^2 estimated as 0.834:  log likelihood = -96,  aic = 204.01

$tstatistics
      ar1      ar2      ma1      ma2      sar1
22.558642 -12.884916 -9.820751  3.992707  5.543211

```

- MA(1)xMA(1)<sup>s</sup>

```

Coefficients:
      ma1      sma1
      0.4249  0.3700
s.e.    0.1326  0.1158

sigma^2 estimated as 1.051:  log likelihood = -102.9,  aic = 211.79

$tstatistics
      ma1      sma1
3.204345 3.196078

```

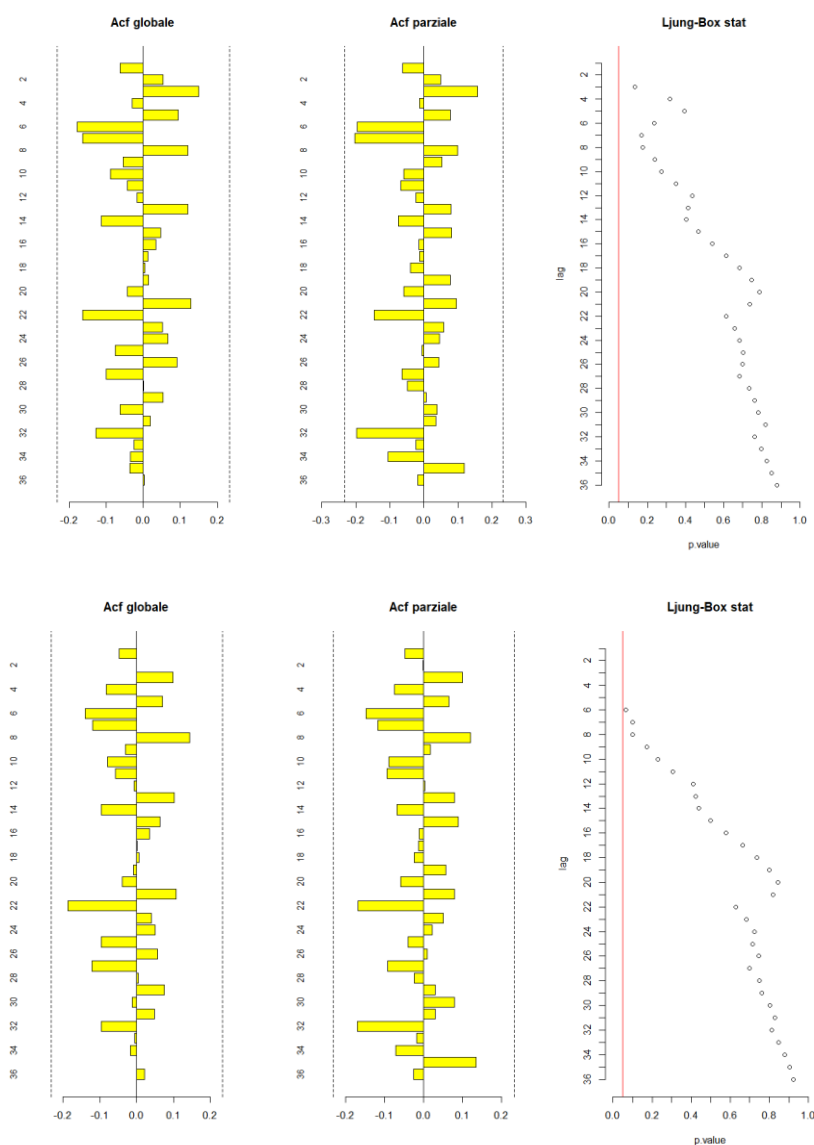
Per scegliere il modello tra i 4 candidati si guarda il valore dell'AIC, i modelli scelti risultano essere quindi:

- Modello 1: AR(1)xAR(1)<sup>s</sup>
- Modello 2: ARMA(2,2)xAR(1)<sup>s</sup>

Da qui in avanti si procederà con l'analisi di entrambi i modelli poiché le differenze in termini di AIC sono minime, inoltre un solo particolare modello potrebbe avere delle previsioni più in linea con l'andamento reale del P.U.N. .

### 3.3 DIAGNOSTICA SUI REDISUI

Il passo successivo per l'analisi della serie storica, seguendo la procedura di Box-Jenkins, è la diagnostica sui residui. I residui del Modello 1 e del Modello 2, come spiegato nel capitolo 2, devono essere incorrelati tra di loro e avere un andamento normale. Per fare queste verifiche si farà utilizzo delle funzioni *correlresidui*, per verificare l'incorrelazione, e *testnorm*, per verificare la normalità. Partendo con la verifica dell'incorrelazione dei residui, gli output della funzione *correlresidui*, che in input prevede i residui dei modelli candidati, sono rispettivamente:

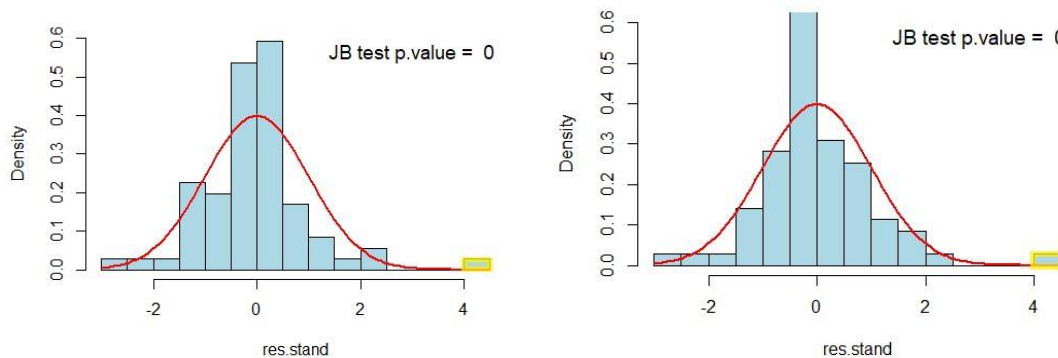


**Figura 13** Output della funzione *correlresidui* per il Modello 1 (in alto) e per il Modello 2 (in basso) che evidenzia l'assenza di autocorrelazione dei residui standardizzati

La funzione *correlresidui* mette in evidenza come i residui di entrambi i modelli siano incorrelati, infatti è possibile notare come le bande di autocorrelazione siano tutte dentro le bande di

confidenza e il test di Ljung-Box (terzo grafico) per ogni lag dia un p-value superiore a 0.05, ovvero si accetta l'ipotesi per cui i residui siano tra di loro incorrelati.

Una volta verificata l'incorrelazione dei residui non resta che verificarne la normalità, che si andrà a constatare con la funzione *testnorm*, la quale permette di fare sia tramite un'analisi grafica, con il grafico di una funzione di densità, sia un'analisi con un test statistico, ovvero il test di Jarque-Bera. Gli output della funzione *testnorm* per i modelli considerati sono i seguenti:



**Figura 14** Istogramma dei residui standardizzati della serie del P.U.N. per il Modello 1(a sinistra) e per il Modello 2 (a destra)

Dagli output della funzione *testnorm* emerge un p-valore pari a 0, quindi andrebbe rifiutata l'ipotesi per cui i residui siano normalmente distribuiti. Tuttavia il test di Jarque-Bera è un test molto sensibile ai valori anomali e, ad esclusione dell'ultima classe evidenziata in giallo nei grafici, possiamo notare come tuttavia i residui siano in linea con l'andamento della funzione di densità di una normale standard. Una volta appurata l'incorrelazione dei residui, sebbene con qualche dubbio sul loro andamento normale, non resta che fare la previsione.

### 3.4 PREVISIONE DEL P.U.N.

L'ultimo step previsto dalla procedura di Box-Jenkins consiste nell'usare i modelli candidati per fare una previsione del prezzo unico nazionale. Il vantaggio di aver tenuto conto di 2 modelli è che in base alle previsioni è possibile scegliere il modello con i valori previsti del P.U.N. che effettivamente ne rispecchiano l'andamento reale.

Per calcolare i valori previsti è necessario usare ancora una volta la funzione *arimaest*.

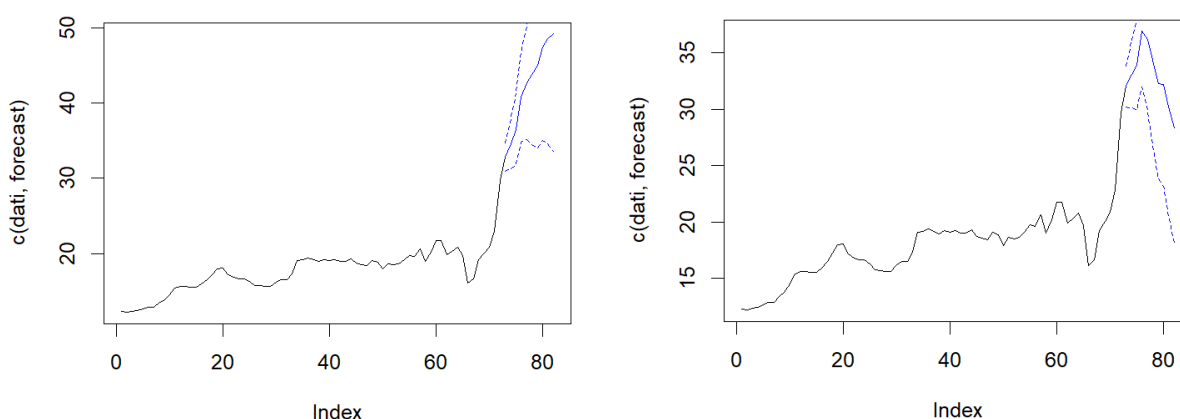
```
fit1<- arimaest(dati,nsorder=c(1,1,0),sorder=c(1,0,0),periodo=4)$stime
fit2<- arimaest(dati,nsorder=c(2,1,2),sorder=c(1,0,0),periodo=4)$stime
```

La funzione *forecastplot*, con *fit* come input, restituisce in output i valori previsti, i relativi standard error e un grafico che aggiunge alla serie storica i valori previsti con le linee di confidenza a livello  $\alpha = 0,05$ .

I valori previsti sono: 32.85741 34.63800 36.49520 40.75910 42.74857 43.85995 44.97355 47.45460 48.61294 49.25942 per il Modello 1.

32.03263 33.13300 33.93620 36.96897 36.24039 34.33828 32.29700 32.19880 30.23654 28.33136 per il modello 2.

I grafici con le previsioni stimate dai due modelli sono i seguenti:



**Figura 15** A sinistra la serie con la previsione del Modello 1, a destra la serie con la previsione del Modello 2

Il Modello 1, essendo autoregressivo, prevede molto bene lo shock grazie alle ultime osservazioni tuttavia non prevede una fine allo shock del P.U.N. , che, essendo un prezzo spot, ragionevolmente non cresce all'infinito ma per previsioni per un futuro prossimo il modello funziona bene. Guardando invece il Modello 2, salta subito all'occhio come il modello riesca a prevedere lo shock e la sua fine, tuttavia il massimo dello shock previsto dal modello, ovvero 36.96897 c€/kWh, risulta essere decisamente più basso del massimo reale ovvero 66,01 c€/kWh.

MODELLO PERIODO	VALORE REALE	MODELLO 1	MODELLO 2
<b>I trimestre 2022</b>	46,03	32,85	32,03
<b>II trimestre 2022</b>	41,34	34,63	33,13

<b>III trimestre 2022</b>	41,51	36,49	33,93
<b>IV trimestre 2022</b>	66,01	40,75	36,96
<b>I trimestre 2023</b>	53,11	42,74	36,24
<b>II trimestre 2023</b>	23,75	43,85	34,33
<b>III trimestre 2023</b>	23,85	44,97	32,29

**Tabella 3** Confronto tra i valori reali e i valori previsti dai modelli arrotondati al secondo valore decimale

Dalla Tabella 3 è possibile vedere come il modello 1 preveda meglio il Prezzo Unico Nazionale nonostante cresca molto più lentamente del P.U.N. reale. Tuttavia dal I trimestre del 2023 il Modello 2 sembra quello più adatto a descrivere il Prezzo Unico Nazionale.

In definitiva quindi il Modello 1 è preferibile se consideriamo le previsioni di breve periodo e, invece, optiamo sul Modello 2 quando siamo interessati a fare previsioni sul lungo periodo.



## CONCLUSIONE

Questo elaborato aveva come obbiettivo di trovare un modello SARIMA(P,D,Q,p,d,q) che meglio spieghi, e quindi preveda, l'andamento del Prezzo Unico Nazionale. Tale previsione risulta particolarmente importante sia per gli operatori economici che per i privati cittadini poiché il Prezzo Nazionale Unico rappresenta il prezzo di riferimento dell'energia elettrica nel nostro Paese. Il periodo temporale considerato per l'analisi parte dal 2004 arrivando fino al 2023 con osservazioni trimestrali.

Per la costruzione del modello sono stati presi in considerazione 2 modelli che, in termini di AIC e significatività dei parametri, risultavano i migliori. I due modelli tuttavia, sebbene abbiano caratteristiche differenti, spiegano il P.U.N. abbastanza bene ma bisogna contestualizzare in che modo.

Il Modello 1 infatti risulta essere un buon modello ma solo per previsioni del P.U.N. in un futuro prossimo poiché cresce esponenzialmente senza fermarsi mai, il che va in contrasto con il concetto di prezzo. Il Modello 2 invece cattura meglio l'andamento sul lungo periodo.

Le problematiche di questi modelli è che non possono prevedere gli shock o eventi geopolitici, infatti bisogna notare come il modello 2 non preveda le politiche messe in atto da governo per calmierare il prezzo dell'energia elettrica. Il modello 2 prevede una diminuzione del P.U.N. che, però, decresce meno velocemente di quanto non sia successo nella realtà.

La costruzione di tali modelli predittivi risulta essere impegnativa poiché non esiste un solo modello che, una volta superati tutti i passi della procedura di Box-Jenkins, risulta essere un buon modello. Il compito dell'analista nel valutare quale sia meglio tra i modelli candidati risulta quindi fondamentale per avere previsioni che siano il più precise possibili.



## Bibliografia

9colonne.it, “La prima centrale termoelettrica europea”.

Borriello E. ,1999 ,“Addio al monopolio dell’Enel”, La Repubblica.

Box G. E. P. e Jenkins G. M. , 1979, “Time Series Analysis: Forecasting and Control”, Holden-Day

Box, G. E. P.; Pierce, D. A. ,1970 , “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models”. (332): 1509–1526.

Coen E. , 2022 “L’Italia e il nucleare, dal primo impianto ai referendum: storia di un rapporto complicato”, La Repubblica.

Energit.it, “Quando è arrivata l’energia elettrica in Italia?”, 16 ottobre 2020.

Gazzetta Ufficiale, Decreto Legislativo 16 marzo 1999 n.79.

Gestore dei Mercati Energetici.org

Graßl M. and Reinhart G. , 2014, “Evaluating measures for adapting the energy demand of a production system to volatile energy prices.” Procedia CIRP 15 (2014): 129-134.

Jarque, C. M. & Bera, A. K.; 1987, “A test for normality of observations and regression residuals”, International Statistical Review 55, 163–172

legambiente.it, 2022, “Cosa Causa Il Caro-Bollette?”.

Masci F. , 2023, LuceGas.it, “TTF Gas (Title Transfer Facility): cos'è, a cosa serve e qual è il prezzo?”.

Mercatoelettrico.org, Vademecum Della Borsa Elettrica Italiana.

Pasculli F. , LuceGas.it, “Caro Bollette Energia: come frenare il raddoppio dei prezzi?”.

Piccolo D. , 2010, Statistica, Il Mulino, Terza Edizione, 53.

Sorgenia.it, 2023, “Pun: Cos'è e come determina il costo dell'elettricità”.

Terna.it, Allegato a.24 al codice di rete individuazione zone della rete rilevante.

Terna.it, Dati Provvisori di Esercizio del sistema elettrico nazionale 2021.

Treccani.it, “Centrale elettrica”.

Yem-energy.com, 2021, “Impatto sui prezzi della stagionalità di gas ed energia elettrica”.

## Appendice

### Script in ambiente R

```
#pulizia del Global Enviroment
rm(list = ls())
#caricamento dei dati
library(readxl)
serie_energia <- read_excel("C:/Users/HP/Desktop/serie energia.xlsx")
View(serie_energia)
#preparazione dei dati
dati<-ts(serie_energia$`COSTO c€/kwh`,start=c(2004,1),frequency = 4)
dati<-dati[1:72] #levo i valori che causano lo shock
ts.plot(dati,gpars=list(xlab="ANNO", ylab="c€/kwh"))
abline(h=mean(dati),lty=3)
seasplot(dati,periodo = 4) #verifico la stagionalità
ddati<-diff(dati,1)
ts.plot(ddati)
seasplot(ddati,4) #non sembra necessario fare una differenziazione
#di ordine 4 per trattare la stagionalità
correlogrammi(ddati)
arimaest(ddati,nsorder = c(1,0,0), sorder = c(1,0,0), periodo= 4) #AIC:204.71
arimaest(ddati,nsorder = c(0,0,1), sorder = c(1,0,0), periodo= 4) #AIC:208.26
arimaest(ddati,nsorder = c(2,0,2), sorder = c(1,0,0), periodo= 4) #AIC:204.01
arimaest(ddati,nsorder = c(0,0,1), sorder = c(0,0,1), periodo= 4) #AIC:211.79
res1 <- arimaest(ddati,nsorder = c(1,0,0), sorder = c(1,0,0), periodo= 4,res=T)
correlresidui(res1, npar=2)
testnorm(res1 , nclassi=20) # dal grafico ha andamento normale
#ma la presenza di outlier fa sì che il p-value venga 0 (molto sensibile alla
#presenza di outlier)
res2<-arimaest(ddati,nsorder = c(2,0,2),sorder = c(1,0,0),periodo = 4, res =T)
correlresidui(res2,npar = 5)
testnorm(res2,nclassi=20)
fit1 <- arimaest(dati,nsorder=c(1,1,0),sorder=c(1,0,0),periodo=4)$stime
forecastplot(dati,fit1,passi=10) #previsioni di 10 trimestri avanti, dal IV
2021
fit2 <- arimaest(dati,nsorder=c(2,1,2),sorder=c(1,0,0),periodo=4)$stime
forecastplot(dati,fit2,passi=10)

#####
#
#funzioni usate nello script
#
#####
#####funzione
correlogrammi#####
correlogrammi<-function(dati, lm=36){
  n.obs <- length(dati)
```

```

par(mfrow=c(1,2))
acfs<-acf(dati,lag.max = 1m, plot=F)
m.acfs<-max(abs(acfs$acf[-1])) + 0.1
pacfs<-pacf(dati,lag.max = 1m, plot=F)
m.pacfs<-max(abs(pacfs$acf)) + 0.1

#plot acf
barplot(rev(as.matrix(acfs$acf[-1])), beside = T, col= "yellow", xlim = c(-
m.acfs, m.acfs), horiz = T, main = "Acf globale", ylab = "", cex.names = 0.9,
names.arg = rev(seq(1:1m)))
abline(v=0)
abline(v=c(-1.96/n.obs^(1/2),1.96/n.obs^(1/2)), lty = 2)
#plot pacf
barplot(rev(as.matrix(pacfs$acf)), beside = T, col="yellow", xlim=c(-
m.pacfs,m.pacfs), horiz = T,
main= "Acf parziale", ylab = "", cex.names = 0.9, names.arg =
rev(seq(1:1m)))

abline(v=0)
abline(v=c(-1.96/n.obs^(1/2),1.96/n.obs^(1/2)), lty = 2)

par(mfrow=c(1,1))
}

```

```

##### funzione arimaest
#####
arimaest <- function(dati, nsorder= c(0,0,0), sorder= c(0,0,0), periodo=NA,
vf=F, res=F){
  ris<-arima(dati, order=nsorder, seasonal= list(order=sorder, period=periodo),
include.mean = vf)
  coef.t<-ris$coef/diag(ris$var.coef)^(1/2)
  out<- if (res==F) list (ris, coef.t) else ris$residuals
  names(out)<-list("stime", "tstatistics")
  out
}

```

```

##### funzione testnorm
#####
testnorm<-function(residui, nclassi){
  res.stand<- (residui - mean(residui))/sd(residui) #residui standardizzati
  hist(res.stand, freq = F, main="", col="lightblue", ylim = c(0,0.6),
nclass=nclassi) #istogramma
  curve(dnorm, add=T, col="red", lwd=2) # adattiamo curva normale per confronto
library(tseries)
  jb<-jarque.bera.test(res.stand) #test di Jarque-Bera
  legend("topright", legend = paste("JB test p.value = ", round(jb$p.value,3)),
bty = "n", cex=1.2)

```

```
jb
}
```

```
##### funzione correlresidui
#####
correlresidui<-function(residui,lm=36, npar=0){
  par(mfrow=c(1,3))
  n.obs<-length(residui)
  acfs<-acf(residui, lag.max = lm, plot = F)
  m.acfs <- max(abs(acfs$acf[-1])) + 0.1
  pacfs<-pacf(residui, lag.max = lm, plot = F)
  m.pacfs<-max(abs(pacfs$acf)) + 0.1

  #plot acf
  barplot(rev(as.matrix(acfs$acf[-1])), beside = T, col= "yellow", xlim = c(-
m.acfs, m.acfs), horiz = T,
          main = "Acf globale", ylab = "", cex.names = 0.9, names.arg =
rev(seq(1:lm)))

  abline(v=0)
  abline(v=c(-1.96/n.obs^(1/2),1.96/n.obs^(1/2)), lty = 2)

  #plot pacf
  barplot(rev(as.matrix(pacfs$acf)), beside = T, col="yellow", xlim=c(-
m.pacfs,m.pacfs), horiz = T,
          main= "Acf parziale", ylab = "", cex.names = 0.9, names.arg =
rev(seq(1:lm)))

  abline(v=0)
  abline(v=c(-1.96/n.obs^(1/2),1.96/n.obs^(1/2)), lty = 2)

  #grafico statistica Ljung-Box
  LBv<-rep(0,lm)
  for (i in ((npar+1):lm)){
    LBv[i]<-Box.test(residui, lag=i, type = "Ljung-Box",fitdf = npar)$p.value
  }
  plot(LBv, rev(seq(1:lm)),main = "Ljung-Box stat", axes = F, ylab = "lag",
xlab="p.value",
       xlim = c(0,1), type = "n")
  points(LBv[(npar+1):lm], rev(seq(1:(lm-npar))))
  abline(v=0.05, col="red")
  axis(1,seq(0,1,0.1))
  axis(2,at=seq(1:lm),labels = rev(seq(1:lm)))

  par(mfrow=c(1,1))
}
```

```
##### funzione seasplot
#####
seasplot<-function(dati, periodo){
  #dati: dati da plottare
  #periodo: 12 se serie mensile, 4 se serie trimestrale

  nomi<-if (periodo==12) list ("G", "F", "M", "A", "M", "G", "L", "A", "S",
"O", "N", "D")
  else list ("I", "II", "III", "IV")

  n<-length(dati)
  li<-rep(0, periodo)
  mi<-rep(0, periodo)

  li[1]<-length(dati[seq(1, length(dati),by=periodo)]) #serie gennaio
  labx<-if (periodo==12) "month" else "quarter" #asse delle ascisse

  par(mfrow=c(1,1), cex=0.8)
  plot(seq(1:li[1]),dati[seq(1, length(dati), by=periodo)], ylab="", ylim =
c(min(dati),max(dati)), xlim=c(1,n),
      xlab=labx, main="Serie stagionali",type = "l",col="blue", lwd=2, axes=F)
#grafico spezzata gennaio
  mi[1]<-mean(dati[seq(1, length(dati),by=periodo)]) #media gennaio
  segments(1, mi[1], li[1], mi[1], col="red")#aggiungo media gennaio sul
grafico

  for (i in 2:periodo){
    li[i]<-length(dati[seq(i, length(dati),by=periodo)])
    lines(cumsum(li)[i-1] + seq(1:li[i]), dati[seq(i,
length(dati),by=periodo)], col="blue", lwd=2) #altri grafici
    mi[i]<-mean(dati[seq(i, length(dati),by=periodo)]) #medie altri mesi
    segments(1+cumsum(li)[i-1], mi[i],li[i] + cumsum(li)[i-1], mi[i],
col="red") #aggiungo medie successive grafico
  }
  axis(1, at = (li[i])*seq(1:periodo) - li[1]/2, labels= nomi)
  axis(2)
}
```

```
##### funzione forecastplot
#####
forecastplot<-function(dati, stime, passi){
  #dati = dati originali
```

```

#stime = stime arimaest a partire da serie originale con d, D diversi da 0
#passi = numero di istanti di tempo da prevedere

forecast <- predict(stime, passi)$pred
forecast.se <- predict(stime, passi)$se

plot(c(dati,forecast), type = "l")
lines(forecast, col = "blue")
lines(forecast + 1.96*forecast.se, col="blue", lty=2)
lines(forecast - 1.96*forecast.se, col="blue", lty=2)

list(forecast, forecast.se)
}

```