



UNIVERSITA' DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

**"ANALISI DELLA SERIE STORICA DEL PREZZO E DEL RENDIMENTO
DEL TITOLO FCA"**

RELATORE:

CH.MO PROF. NUNZIO CAPPuccio

LAUREANDO/A: GIANMARCO LUZZI

MATRICOLA N. 1160772

ANNO ACCADEMICO 2019 – 2020

Indice

INTRODUZIONE.....	4
CAPITOLO 1: IL GRUPPO FCA.....	5
1.1 LA STORIA DEL GRUPPO	5
1.2 STRUTTURA E AZIONARIATO.....	6
1.3 LA FUSIONE FCA-PSA	7
CAPITOLO 2: CONCETTI TEORICI	9
2.1 INTRODUZIONE E BASI ALLE SERIE STORICHE.....	9
2.2 STAZIONARIETÀ E AUTOCORRELAZIONE.....	10
2.3 DATI E RAPPRESENTAZIONE GRAFICA	13
2.4 DIFFERENZIAZIONE	17
2.5 TEST DI RADICE UNITARIA.....	20
CAPITOLO 3: ANALISI E APPLICAZIONE DEI MODELLI	26
3.1 ANALISI DEI RENDIMENTI	26
3.2 IL MODELLO ARMA	30
3.2.1 $AR(p)$	30
3.2.2 $MA(q)$	31
3.2.3 $ARMA(p,q)$	32
3.3 IL MODELLO GARCH	34
3.4 DECOMPOSIZIONE E PREVISIONE	37
BIBLIOGRAFIA.....	42

Introduzione

Il presente elaborato ha come oggetto lo studio della serie storica del prezzo del titolo FCA, quotato alla borsa di New York. Il testo ha come obiettivo quello di generare un modello econometrico che riesca a spiegare l'andamento del prezzo giornaliero e mensile.

Nel primo capitolo viene presentato il gruppo FCA, i marchi che compongono il gruppo e la storia travagliata della fusione tra la FIAT e Chrysler Group LLC. Successivamente si presenta la struttura del gruppo, riportando i volumi di vendite per regione, e il suo azionariato. Infine, un piccolo riassunto della futura fusione del gruppo FCA con il gruppo francese PSA.

Nel secondo capitolo si introduce lo studio della serie storica spiegando il concetto stesso di serie storica, di come sia nato e si sia sviluppato nel corso del tempo soprattutto grazie all'apporto del modello Box-Jenkins. Il capitolo prosegue spiegando cosa è la stazionarietà e introducendo la prima rappresentazione grafica della serie storica. Infine, si spiega cosa è la differenziazione di una serie, concludendo con i test di radice unitaria, focalizzandosi sui test ADF.

Nel terzo capitolo si arriva all'analisi dei rendimenti, utilizzando i concetti riportati sopra. Si introducono poi i due modelli più importanti del testo, il modello ARMA e il modello GARCH. Come ultimo paragrafo si fa una piccola introduzione del concetto di decomposizione, utilizzato per generare un piccolo test di previsione.

Tutti i modelli e le rappresentazioni grafiche sono generati tramite il software R, il quale sarà coadiuvato da librerie di funzioni open source nate con lo scopo di arricchire l'utilizzo del software a fini statistici.

Capitolo 1: Il gruppo FCA

Fiat Chrysler Automobiles è una azienda automobilistica nata nel gennaio del 2014, che si occupa di progettare, produrre, distribuire e vendere veicoli per il mercato di massa. FCA è un gruppo che si espande in tutto il mondo ed è composto da molteplici marchi automobilistici come Abarth, Alfa Romeo, Chrysler, Dodge, Fiat, Fiat Professional, Jeep, Lancia, Ram, Srt e Maserati. Per completare il gruppo, FCA si avvale di imprese che si specializzano nel Servizio post-vendita come Mopar, oppure di imprese come Teksid e Comau (Consorzio Macchine Utensili) che rispettivamente si specializzano nella produzione di ghisa e alluminio, e nello sviluppo di processi di automazione e progettazione attraverso soluzioni di robotica.

Oltre a ciò, FCA offre soluzioni finanziarie attraverso accordi commerciali e joint venture come quella con Crédit Agricole, per creare FCA Bank la quale propone servizi finanziari, assicurativi, di prestito e noleggio di breve e lungo termine.

1.1 La storia del gruppo

La storia del gruppo inizia nel 2009 quando Fiat (Fabbrica italiana automobili Torino), data la situazione di forte crisi che colpì tutto il mercato automotive americano, annunciò l'acquisto del 20% della proprietà di Chrysler Group LLC comprandola dalla Cerberus Capital Management L.P, una tra le società di investimento più grande degli Stati Uniti, con l'obbiettivo di ristrutturare l'azienda americana e crearne un progetto innovativo.

Nel proseguire degli anni Fiat incrementò la propria quota azionaria in Chrysler Group fino a quando nel gennaio del 2014 decise di acquistare la quota rimanente dalla VEBA TRUST, un fondo per i pensionati della UAW (United Automobile Workers). Alla fine dello stesso anno i due gruppi si unirono sotto una stessa società, la Fiat Chrysler Automobiles.

FCA ha formato un'alleanza industriale, in base alla quale le parti collaborano su diversi fronti. Il loro punto di forza è la condivisione e lo sviluppo di prodotti e piattaforme che permette al gruppo di risparmiare nello sviluppo. Non escludendo poi la migliore distribuzione globale, l'approvvigionamento di materie prime, le infrastrutture informatiche condivise, i servizi di gestione e il miglioramento dei processi¹.

¹ Marketline, 19/9/2019. FCA US LLC SWOT Analysis, Marketline. p1-8.

1.2 Struttura e azionariato

La nuova società di diritto olandese ha sede a Londra da dove gestisce più di 192000 impiegati che lavorano in più di 130 mercati in tutto il mondo.

Nel 2019 il gruppo è riuscito a totalizzare una vendita di 4,6 milioni di veicoli, per un totale di 108 miliardi di euro in ricavi².

Come si può notare dalla tabella 1, il gruppo FCA divide il proprio operato in quattro macro mercati e uno che riguarda l'unico brand di auto di lusso rimasto, Maserati.

Tabella 1³ : Rapporto Vendite per regione

	2019	2018	2017
North America	2,5	2,5	2,4
LATAM	0,6	0,6	0,5
APAC	0,2	0,2	0,3
EMEA	1,3	1,4	1,5
Total mass market	4,6	4,7	4,7
Maserati	0,03	0,04	0,05
Total Worldwide	4,6	4,8	4,8

La maggiore vendita di veicoli del gruppo si registra nel mercato del nord America. Esso si suddivide rispettivamente in mercato statunitense e mercato canadese e si caratterizza per la vendita di veicoli di grossa taglia, come pick-up, minivan e macchine del segmento D ed E (automobili oltre i 4,5 metri).

Il mercato dell'America Latina (LATAM) è caratterizzato invece da una crescita costante del market share nel tempo con un primato in Brasile dove FCA si posiziona come primo gruppo per vendita di veicoli del segmento B, C e anche pick-up. Opposta è la situazione in Argentina, dove a causa dell'impatto della crisi economica del paese, il gruppo registra un declino del 43% delle vendite.

La regione asiatica (APAC) è invece quella che riporta la più bassa richiesta di automobili. Si tratta del mercato dove il gruppo è presente da minor tempo e dove vi è un continuo peggioramento della situazione, a causa della caduta della domanda di veicoli nei segmenti in cui FCA concorre (come i segmenti A, B, D ed E).

² Dati raccolti sul sito ufficiale <http://www.fcagroup.com/>

³ Dati disponibili nell'Annual Report del 2019 nel sito ufficiale <https://www.fcagroup.com>

Tutto ciò però offre una grande opportunità di crescita e sviluppo, soprattutto per il mercato dei veicoli elettrificati in Cina dove un rapido cambiamento sta mutando l'intero mercato dell'automotive.

In Europa la situazione è decisamente più matura e in peggioramento in seguito alla forte concorrenza dei più grandi gruppi automobilistici e dai regolamenti stringenti sulle emissioni, ma Fiat rimane leader per il segmento A e B (citycar).

Fiat Chrysler Automobiles successivamente alla fusione del 2014 decise di quotarsi in due borse differenti, quella di New York (New York Stock Exchange) sotto il simbolo FCAU, e nel Mercato Telematico Azionario sotto il simbolo FCA. Essa dispone di un capitale sociale di 20 milioni €, divisi in 1'573'949'512 azioni destinate al pubblico e 449'610'500 azioni con diritti di voto speciali, entrambe con valore nominale di 0,01€⁴.

Il principale azionista del gruppo è la Holding Exor N.v.: società di investimenti europea che si occupa di investimenti a medio e lungo termine in molteplici settori, ed è controllata a sua volta dalla famiglia Agnelli, tramite la società Giovanni Agnelli B.V.

Le restanti partecipazioni sono diffuse come mostrato in tabella 2 tra investitori istituzionali dell'area Euro ed Extra Euro, e in maggior parte diffuse sul mercato.

Tabella 2⁵: Principali azionisti gruppo FCA

Principali azionisti istituzionali	%
Exor N.V.	28,7%
Tiger Global Management LLC	4,79%
BlackRock , Inc.	4,07%
Ubs group AG	3,73%
Vanguard Groupe, Inc	2,31%
Baillie Gifford and Company	2,23%

1.3 La Fusione FCA-PSA

Nel dicembre del 2019, FCA e PSA⁶ (Peugeot S.A.), hanno deciso di fondere i due gruppi e creare una nuova Holding, chiamata Stellantis.

⁴ Dati disponibili nell'Annual Report del 2019 nel sito ufficiale <https://www.fcagroup.com>

⁵ Dati disponibili su Yahoo Finance nel sito <https://it.finance.yahoo.com/> [26/8/2020]

⁶ Nota: gruppo automobilistico francese formato dai marchi Peugeot, Citroën, DS Automobiles, Opel e Vauxhall Motors

La procedura di fusione durerà fino al primo trimestre del 2021 e porterà alla creazione del quarto gruppo per vendite più grande al mondo, dopo Volkswagen, Toyota e Renault-Nissan-Mitsubishi.

Questo atto porterà enormi vantaggi ad entrambi i gruppi, come per esempio il ritorno di Peugeot nel mercato americano dove non è mai riuscito a creare una rete di vendita locale, oppure la sfida per la conquista del mercato asiatico, grazie ai benefici portati da PSA nei veicoli elettrici. Ulteriore sfida sarà il riposizionamento dei diversi brand nel mercato a causa della concorrenza di alcune autovetture di entrambi i gruppi. Ma questo sarà superato grazie ai benefici della condivisione dei pianali delle autovetture che porterà enormi risparmi e avrà gli effetti maggiori sulle gamme dei diversi brand.

Capitolo 2: Concetti Teorici

2.1 Introduzione e basi alle serie storiche

Una raccolta di osservazioni presa a intervalli di tempo regolari in ordine cronologico per una certa variabile è chiamata serie temporale.

Gli intervalli di tempo delle serie storiche possono essere Giornalieri (prezzo giornaliero di titolo), Settimanali (prodotto fabbricato), Mensili (tasso di disoccupazione), Trimestrali (vendita di un certo prodotto), Annuali (prodotto interno lordo), Quinquennali (censimento dei prodotti), Decennali (censimento della popolazione). Le serie temporali possono essere applicate in vari campi, come le previsioni economiche, la previsione delle vendite, analisi del budget, analisi del mercato azionario, studi per investimenti, proiezioni di produzione, analisi di censimento e molti altro.

Il ruolo principale dello studio delle serie storiche è la previsione futura di un attributo, utilizzando modelli che guardano l'andamento storico dei dati, o più semplicemente per poter comprendere il futuro in funzione del passato.

Lo studio delle serie storiche nasce nei primi decenni del ventesimo secolo grazie all'interesse di studiosi che decisero di applicare modelli statistici a serie di dati riguardanti l'economia americana. I primi modelli prevedevano che il tempo venisse studiato come variabile discreta, e non si immaginava che i dati potessero dipendere l'uno dagli altri.

L'approccio classico per eccellenza, utilizzato e sviluppato all'inizio del ventesimo secolo da Warren M. Pearson (1919), vedeva distinguere e decomporre in componenti diversi la serie storica. Questi componenti non sono visibili a primo impatto, e venivano definiti indipendenti gli uni dagli altri e solamente quando questi venivano ricomposti generavano la serie come la si immagina. Si pensava poi che i componenti, in quanto indipendenti non potessero essere influenzati da un processo stocastico, ma che dovessero seguire una funzione del tempo.

Si distinguono così quattro diverse componenti:

- Trend: movimento o tendenza di lungo periodo della variabile, che si divide in tendenza al rialzo o positiva, o tendenza al ribasso o negativa.
- Business Cycle: sequenza di cambiamenti ricorrente ma non periodica che si sviluppa in un periodo lungo, solitamente più di anno.
- Seasonal Cycle: si riferisce ai cambiamenti prevedibili che si verificano nell'arco di un anno in un'economia.

- Residual: componente che racchiude i movimenti di natura stocastica e indipendente, sottratte le componenti sopra citate.

Tutto ciò non cambiò nei decenni successivi fino a che due statistici britannici, George E. P. Box e Gwilym M. Jenkins, scrissero il libro “Times Series Analysis: Forecasting and Control” (1970) nel quale introdussero un nuovo modo di guardare alle serie storiche, superando l’analisi deterministica a più componenti del metodo classico, e capendo la natura stocastica del processo. Non occorre più rifarsi a complessi modelli, composti da svariate equazioni, ma bastava utilizzare le informazioni contenute nei valori stessi per poter sviluppare un’analisi.

Nacque così l’analisi Box-Jenkins, un metodo che si fonda su una procedura a tre pilastri: Identificare il modello più adatto; Stimare i parametri del modello; Controllo o diagnosi attraverso test statistici per valutare la precisione delle stime.

È solo negli ultimi decenni che attraverso l’utilizzo di strumenti di calcolo e visualizzazioni più potenti è stato possibile riprendere in mano la teoria della divisione delle componenti di una serie storica e lo studio delle stesse attraverso modelli di visualizzazione per la previsione delle variabili.

2.2 Stazionarietà e Autocorrelazione

Una serie temporale stazionaria è una serie le cui proprietà non dipendono dal momento in cui essa viene osservata, e la sua distribuzione di probabilità non muta nel corso del tempo. Quindi si può facilmente evincere che le serie temporali che contengono componenti come Trend o Seasonal Cycle, sono non stazionarie, poiché l’andamento e la stagionalità influenzano il valore delle serie temporali in momenti diversi. Esistono due generi di stazionarietà, la prima detta *rigorosamente stazionaria* e la seconda *debolmente stazionaria*.

Per utilizzare le realizzazioni passate di una variabile, con il fine di prevederne i valori futuri, è necessario che il processo stocastico che ha originato le osservazioni sia stazionario.

In altri termini le serie temporali non stazionarie possono far sorgere previsioni distorte o poco efficienti.

Si definisca un processo stocastico $\{ \dots, X_{t-1}, X_t, X_{t+1}, \dots \}$ costituito da variabili casuali indicizzate al tempo t . Il comportamento di $\{X_t\}$ è determinato specificando le funzioni di densità $p(x_{t_1}, x_{t_2}, \dots, x_{t_m})$.

è *rigorosamente stazionaria* una serie $\{X_t\}$ se:

$$p(t_1 + \tau, t_2 + \tau, \dots, t_m + \tau) = p(t_1, t_2, \dots, t_m), \quad \forall \tau, \forall m, \forall (t_1, t_2, \dots, t_m).$$

Quindi una serie temporale è *rigorosamente stazionaria*, se la funzione di densità è uguale alla funzione di densità della stessa serie sotto una traslazione τ . Si dice così che la serie non varia sotto traslazioni temporali.

Si definisce una serie *debolmente stazionaria* invece, se rispetta le condizioni di stazionarietà in funzione della media e della covarianza.

- $E[x_t] = \mu_t = \mu$ costante per tutto t
- $Cov[x_t, x_s] = E[(x_t - \mu_t)(x_s - \mu_s)] = \gamma(|s - t|)$ un processo è detto a covarianza stazionaria quando la covarianza è solo funzione della distanza temporale tra le due serie.

La gran parte delle serie temporali però non presenta caratteristiche di stazionarietà, anzi è colpita da componenti che molte volte si spiegano solo con modelli di tipo stocastico.

Una serie temporale generata da un processo stazionario si può immaginare che possa fluttuare intorno alla media, la quale è costante nel tempo e dimostra che la serie si comporta in maniera molto particolare come se la sua memoria degli shock passati decadesse nel tempo.

Una classe fondamentale dei processi stazionari è l'elemento White Noise, un processo stocastico che è caratterizzato da proprietà simili ad una serie *debolmente stazionaria*.

Un processo per definirsi comunemente White Noise deve avere:

$WN(0, \sigma^2)$

- $E[x_t] = 0 \quad \forall t$
- $Var[x_t] = \sigma^2 \quad \forall t$
- $Cov[x_t, x_s] = E[x_t x_s] = 0 \quad \text{per tutti } t \neq s$

Per approfondire ulteriormente il principio di stazionarietà è utile introdurre due forme di autocorrelazione, la funzione di autocorrelazione (ACF) e la funzione parziale di autocorrelazione (PACF).

La funzione di autocorrelazione si riferisce al modo in cui le osservazioni in una serie temporale sono correlate tra loro, ed è misurata da una semplice correlazione tra l'osservazione corrente e l'osservazione τ periodi ritardati.

$$\rho_\tau = \frac{Cov(x_t, x_{t+\tau})}{\sqrt{Var(x_t) \cdot Var(x_{t+\tau})}}$$

In primo luogo, l'intervallo di confidenza cresce all'aumentare del ritardo perché il numero di punti dati n , utilizzati per stimare la correlazione, diminuisce di uno per ogni aumento di un ritardo.

In secondo luogo, bisogna fare attenzione quando si interpreta la "significatività" della correlazione a vari ritardi, perché ci si deve aspettare, a priori, che circa 1 correlazione su 20 sia significativa solo in base al caso.

Come per l'ACF, si vedrà più avanti come la PACF può essere utilizzata anche per aiutare a identificare l'ordine appropriato di p e q nel modello $ARMA(p, q)$.

Successivamente la funzione parziale di autocorrelazione misura la correlazione tra l'osservazione corrente e quella ritardata, quando tutti gli effetti della dipendenza lineare vengono rimosse. Si nota poi successivamente che anche il grafico PACF ha ancora una volta indici con valore reale per lo scarto temporale, ma non include alcun valore per il ritardo 0 perché è impossibile rimuovere qualsiasi autocorrelazione intermedia tra t e $t - k$ quando $k = 0$ e quindi la PACF non esiste a ritardo 0.

Nelle serie temporali di carattere economico molte volte si riscontra la presenza di autocorrelazione e ciò porta a confermare la natura non stazionaria delle serie, così da obbligare ad una trasformazione dei dati. Intervengono così Box e Jenkins che sostengono la rimozione del comportamento non stazionario di tendenza utilizzando il processo di differenziazione, al quale sottraggono i valori stimati dai valori originali. Per esempio, una differenziazione del primo ordine dovrebbe poter eliminare la componente trend se la serie presentasse un comportamento di trend lineare persistente nel tempo.

2.3 Dati e rappresentazione grafica

L'oggetto in esame di questo elaborato è il prezzo del titolo FCAU, quotato alla borsa di New York dal 2014 dopo il completamento della fusione con il gruppo Chrysler.

La serie si compone di 1429 osservazioni che iniziano dal 1° gennaio 2015 e finiscono al 30 settembre 2020. I prezzi sono stati resi disponibili dalla piattaforma *Yahoo finance* ed elaborati con il linguaggio informatico conosciuto come *R* e sviluppato all'interno del software *R studio*. Per rendere lo studio possibile sono state utilizzate librerie di carattere "open source" nate con lo scopo di analizzare le serie storiche e per fare previsione di dati.

Successivamente viene mostrato il grafico della serie storica del prezzo giornaliero di chiusura, aggiustato per dividendi e frazionamenti. Lo studio sarà diviso in concomitanza tra una serie a frequenza giornaliera e una serie a frequenza mensile. Quest'ultima sarà composta dalla media dei prezzi giornalieri, ed ha un totale di 69 osservazioni.

Figura 2.3.1: Grafico della serie storica giornaliera del prezzo del titolo FCAU

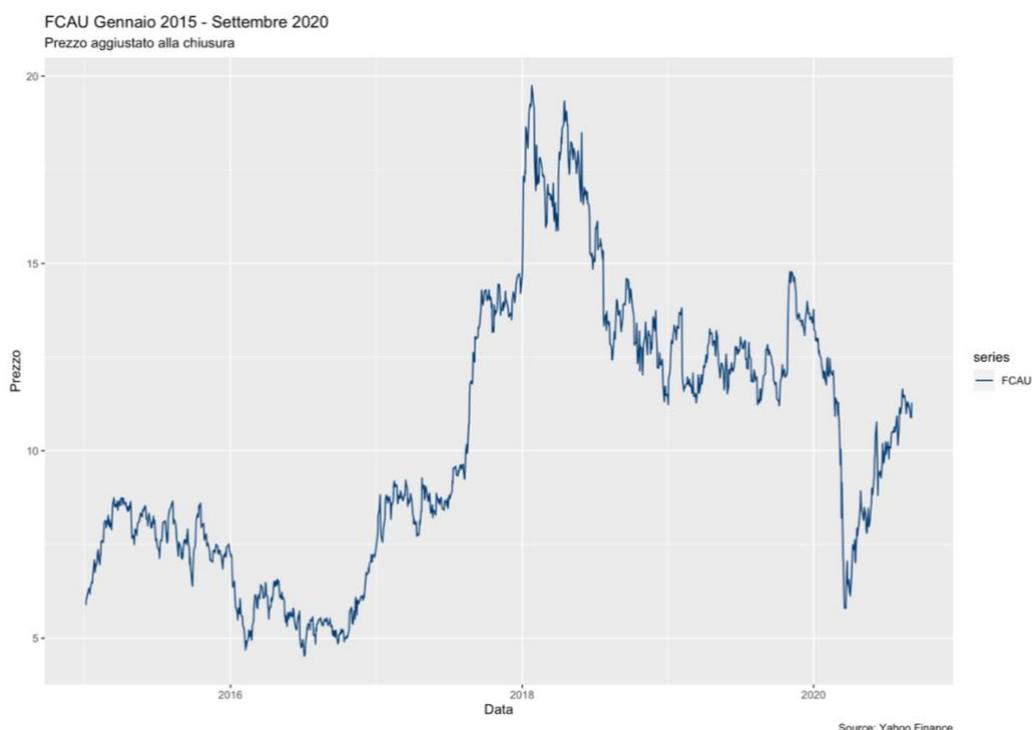
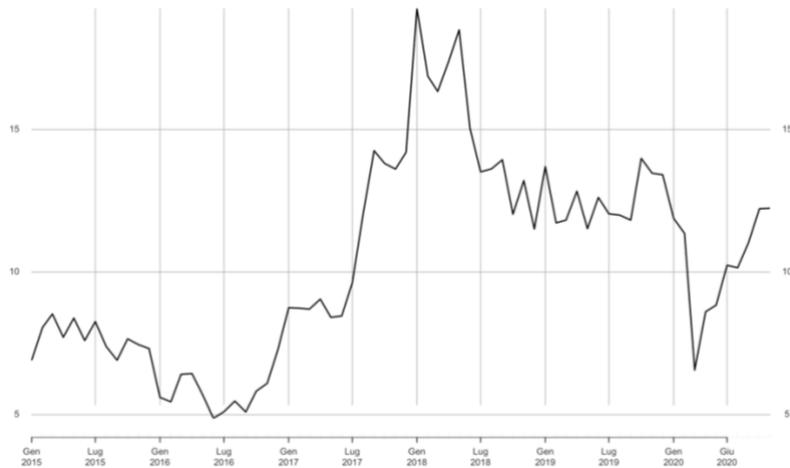


Figura 2: Grafico serie storica mensile del prezzo del titolo FCAU



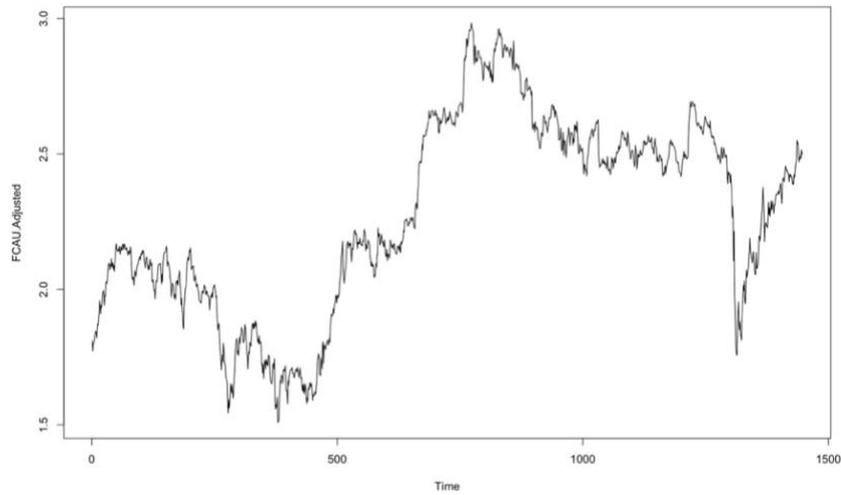
Per poter iniziare un'analisi di una serie storica è bene studiare il grafico di quest'ultima per riuscire a comprendere quelle proprietà che una serie mostra anche solo visivamente.

I dati mostrano una serie che si muove in maniera caotica, con i primi due anni (2015 e 2016) in cui il titolo mostra una trend discendente, per poi subire un grosso rimbalzo dal 2017 fino a metà del 2018 dimostrando di avere un trend crescente molto ripido in cui il titolo tocca il massimo storico di 19,75\$, a causa del raddoppio degli utili rispetto all'anno precedente. Nel 2019 il titolo soffre della caduta delle vendite delle auto, mantenendo però un trend costante. Successivamente la rapida caduta del 2020 mostra gli effetti della crisi del Covid-19 che ha colpito tutti i mercati, facendo così concludere la serie con un trend discendente.

D'altro canto, essa mostra un movimento di breve periodo molto pronunciato, nell'arco di 1 anno, mostrando la componente dei "Seasonal Cycle".

Come mostrato dalla figura 1, la volatilità del prezzo del titolo FCAU è pronunciata in tutti e cinque gli anni, perciò è bene fare una trasformazione logaritmica della serie come mostrato nella figura 3.

Figura 3: Grafico logaritmo naturale del prezzo



Di seguito si verifica la presenza di l'autocorrelazione grazie all'ausilio del correlogramma generato dalla funzione di autocorrelazione (ACF), e dalla funzione parziale di autocorrelazione (PACF), i quali verificheranno la stazionarietà o meno della serie. La funzione di autocorrelazione, come precedentemente citato, misura la relazione lineare tra i valori in ritardo di una serie temporale.

Figura 4: Correlogramma funzione di autocorrelazione

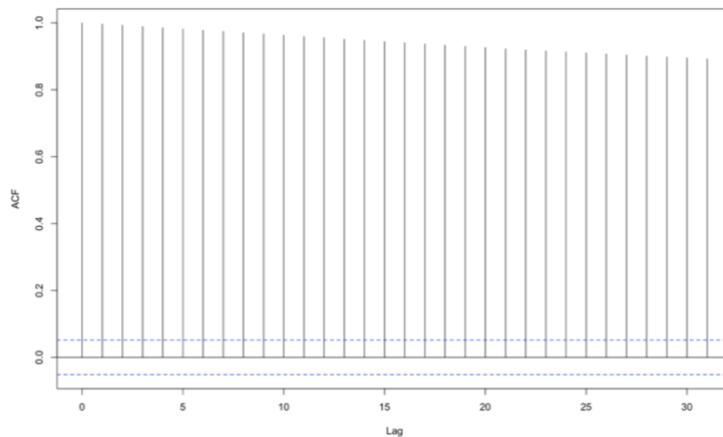


Figura 5: Correlogramma funzione parziale di autocorrelazione

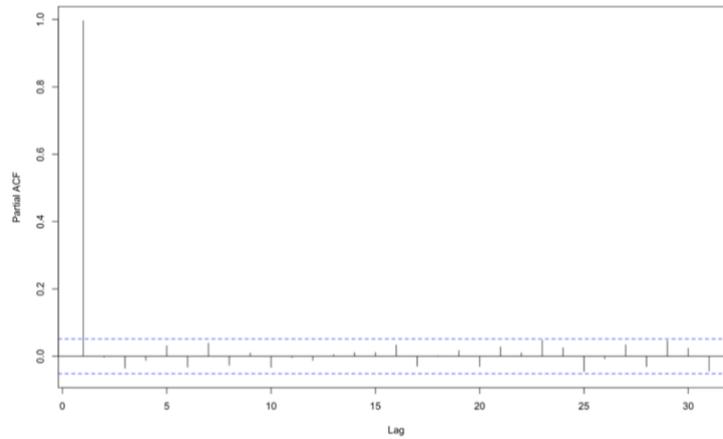


Figura 6: Correlogramma funzione di autocorrelazione serie mensile

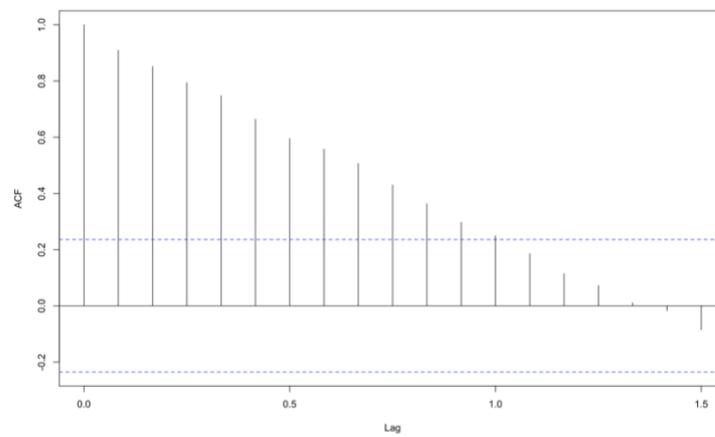
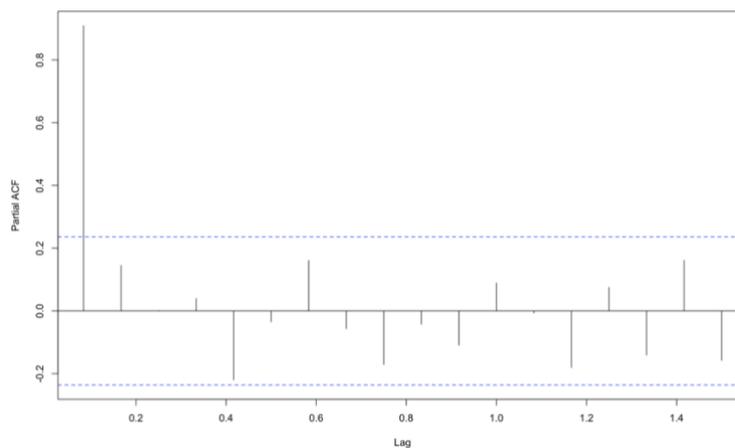


Figura 7: Correlogramma funzione parziale di autocorrelazione serie mensile



Il primo correlogramma in figura 4 mostra una forte correlazione positiva della serie che scende lentamente con l'aumentare dei ritardi. Tutto ciò mostra dati che presentano un trend, quindi le autocorrelazioni per i piccoli ritardi tendono ad essere grandi e positive, perciò le osservazioni

vicine nel tempo sono anche vicine per dimensioni. Quindi la ACF delle serie temporali con trend, tende ad avere valori positivi che diminuiscono lentamente con l'aumentare dei ritardi. Una serie che non mostra correlazione si può definire una serie *white noise*, di fatti ci aspettiamo che ogni autocorrelazione sia vicina allo zero. Naturalmente, non saranno esattamente uguali a zero in quanto c'è una qualche variazione casuale. Per una serie *white noise*, ci si aspetta che il 95% dei picchi della funzione di correlazione si trovi all'interno di una banda di confidenza $\pm 2/\sqrt{T}$ dove T è la lunghezza della serie temporale. È comune tracciare questi limiti su un grafico della funzione di correlazione, come quelli nelle figure, tramite una linea tratteggiata. Se uno o più dei grandi picchi sono al di fuori di questi limiti, o se sostanzialmente più del 5% dei picchi sono al di fuori di questi limiti, allora la serie probabilmente non è *white noise*.

Il correlogramma in figura 5 invece mostra la funzione parziale di autocorrelazione, la quale misura la correlazione quando la dipendenza lineare viene rimossa dalla serie. La figura non riporta alcuna correlazione in ogni ritardo, poiché in nessun ordine di ritardo la correlazione è significativamente diverso da 0.

Dunque, la serie non presenta alcuna caratteristica di stazionarietà, ma questo si può definitivamente dichiarare dopo che si passa da uno studio di tipo visivo a uno in cui la procedura è dettata da test statistici.

Tutt'altra cosa invece si può dire del correlogramma della serie mensile, poiché come si può notare dalla figura 6 la funzione di autocorrelazione scende con molta più rapidità dimostrando che la serie presenta caratteristiche di stazionarietà, anche se la poca numerosità delle osservazioni potrebbe portare a una efficienza ridotta del grafico.

2.4 Differenziazione

Il processo di differenziazione, come citato precedentemente, mostra un modo per rendere stazionaria una serie temporale non stazionaria, calcolando le differenze tra osservazioni consecutive.

La trasformazione logaritmica ha aiutato a diminuire l'eccessiva volatilità della serie temporale. La differenziazione invece, è utile per stabilizzare la media di una serie temporale cercando di rimuovere, o ridurre, i cambiamenti causati dalle componenti, come il trend o la stagionalità.

La differenziazione si divide in base al grado, ossia una differenziazione di primo grado equivale a sottrarre un'osservazione per la sua adiacente e può essere scritta come:

$$\nabla x_t = x_t - x_{t-1}.$$

La serie generata da una prima differenziazione è descrivibile come white noise, perché può essere scritta come di seguito,

$$x_t - x_{t-1} = \varepsilon_t,$$

dove ε_t è la componente white noise.

Riordinando l'ultima equazione in maniera tale da generare il modello “*random walk*”:

$$x_t = x_{t-1} + \varepsilon_t.$$

I modelli *random walk* sono ampiamente utilizzati per i dati non stazionari, grazie alla loro capacità di adattarsi ad un'ampia gamma di dati, in particolare per dati di carattere economico e finanziario. Come fa presagire il nome, il modello *random walk* presuppone che, in ogni punto del tempo, la serie si allontani in maniera aleatoria dalla sua ultima posizione registrata, con passi il cui valore medio è pari a zero. Un modello del genere solitamente presuppone che la serie abbia lunghi periodi di trend apparenti verso l'alto o verso il basso, ma soprattutto che presenti improvvisi e imprevedibili cambiamenti di direzione in tutta la serie temporale.

Figura 8: Grafico della serie dopo la differenziazione

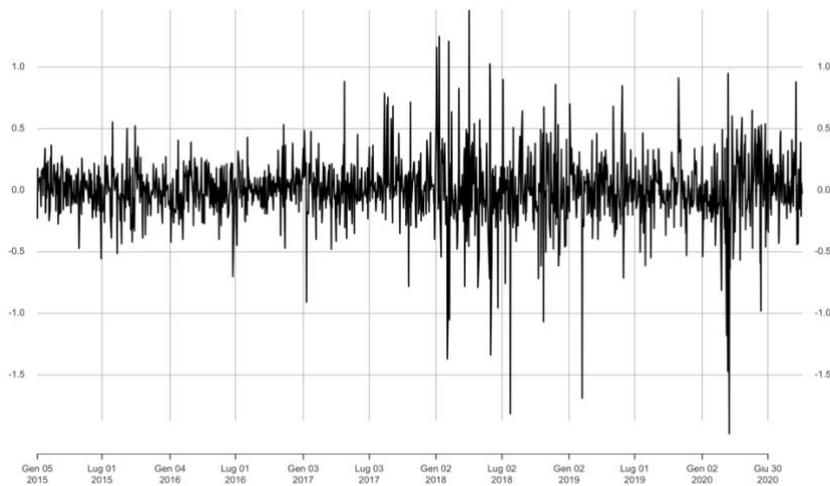


Figura 9: Funzione di autocorrelazione della differenziazione

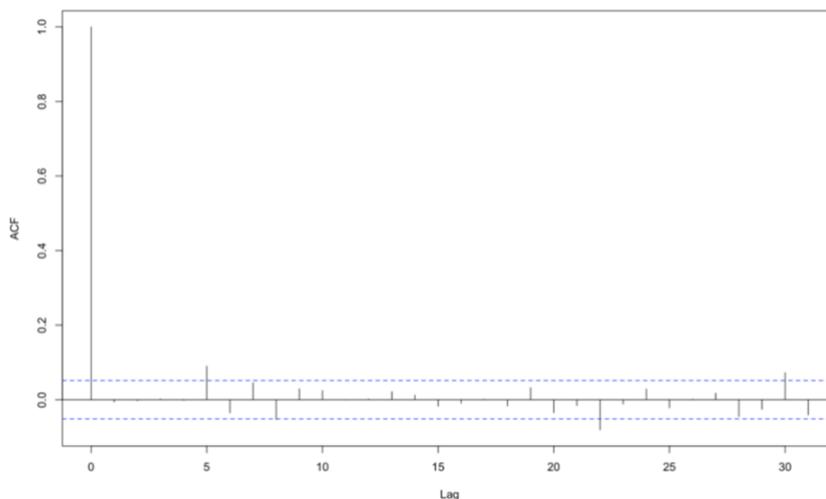
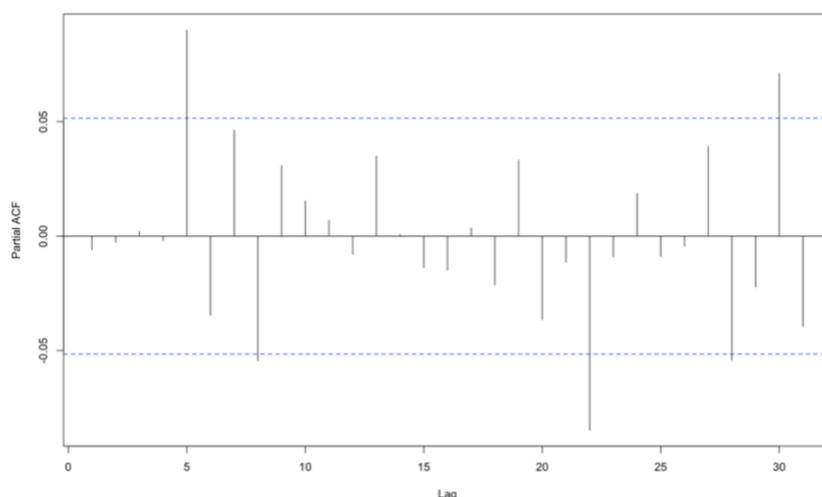


Figura 10: Funzione parziale di autocorrelazione della differenziazione



Guardando la figura 8, si nota che adesso la serie incominci a comportarsi approssimativamente come una serie stazionaria, avendo una media e una varianza più o meno costante. Calcolando poi la funzione di autocorrelazione, come mostrato in figura 9, si evince che siano presenti solo tre ordini di ritardo in cui la correlazione è leggermente positiva, nel ritardo 5 e 30, e leggermente negativa, nel ritardo 22. È importante sottolineare che il ritardo di ordine 0 nella funzione di autocorrelazione è sempre positivo per costruzione.

Per approfondire ulteriormente questo aspetto ci si può avvalere del test statistico Ljung-Box “Q” test, che serve a valutare le possibili autocorrelazioni all’interno dei dati, significativamente diverse da 0.

Il test si basa su due ipotesi ben distinte:

$$\begin{cases} H_0 : & \text{dati incorrelati} \\ H_1 : & \text{dati correlati} \end{cases}$$

Il test è

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

N è la dimensione dei dati, $\hat{\rho}_k^2$ è la funzione di autocorrelazione al ritardo k , e infine k e m sono il numero di autocorrelazioni ai vari ritardi. La statistica test si distribuisce asintoticamente come χ^2 con m gradi di libertà.

Di seguito il calcolo della statistica test tramite R.

Box-Ljung test

```
data: differenziazione
X-squared = 0.05131, df = 1, p-value = 0.8208
```

Dato il p-value della statistica test ben oltre la soglia limite, si dovrà accettare l'ipotesi nulla affermando così che i dati sono non autocorrelati tra di loro.

2.5 Test di radice unitaria

Nel campo economico come anticipato in precedenza le serie temporali sono sostanzialmente in gran parte non stazionarie. Questo è dovuto a componenti interne alla serie stessa che fanno sì che si muova seguendo un trend, un processo stagionale o un ciclo. Per poter adoperare il metodo Box-Jenkins si deve quindi rendere la serie stazionaria così da continuare lo studio. Dopo la differenziazione si passa così ai test di radice unitaria, i quali servono a capire se all'interno di una serie sia presente una radice unitaria, che può causare la sua non stazionarietà. Questo test è stato sviluppato nella metà degli anni Settanta da due professori di statistica americani, rispettivamente Wayne A. Fuller e David A. Deickey, i quali attraverso il test nominato DF, hanno studiato il modo per capire se all'interno della serie sia presente o meno un trend stocastico.

Il test suppone una serie X_t la quale segue un modello di genere AR (1):

$$X_t = \phi X_{t-1} + \varepsilon_t$$

Dove ε_t è un processo White Noise.

Il test proposto ha le seguenti ipotesi:

$$\begin{cases} H_0 : \phi = 1 \\ H_1 : |\phi| < 1 \end{cases}$$

Il test è di tipo unilaterale, e nell'ipotesi nulla H_0 prevede che la serie sia non stazionaria quindi è presente una radice unitaria, invece nell'ipotesi H_1 la serie è definita stazionaria.

Il test è

$$t = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

dove $\hat{\phi}$ è la stima dei minimi quadrati di ϕ , e $SE(\hat{\phi})$ è la stima dei minimi quadrati dell'errore standard di $\hat{\phi}$.

Questo test è corretto fino a che non si vogliono aggiungere dei ritardi AR(p) nello studio della serie, così da modificare il modello AR(1).

È così che Said e Dickey nel 1984, cambiano il test di base della radice unitaria per accogliere i modelli $ARMA(p, q)$ con ordini sconosciuti, il test così viene chiamato test Aumentato Dickey-Fuller (ADF).

La regressione da cui parte è

$$\Delta x_t = \alpha + \beta t + \phi x_{t-1} + \sum_{i=1}^k \theta_i \Delta x_{t-i} + \varepsilon_t$$

Il test ADF possiede le stesse ipotesi del test DF, e la statistica test è la medesima per entrambe. Fortunatamente la scelta del ritardo k da testare non sarà necessario calcolarlo poiché il software calcola automaticamente la statistica test nella maniera più efficiente.

A riguardo delle regressioni, si possono dividere in 3 tipologie di studio molto diverse.

La prima prevede che α e β siano uguali a 0, quindi un Dickey-Fuller test senza una costante e senza una variabile trend. La seconda opzione prevede che α sia diverso da 0 e β sia uguale a 0, quindi un Dickey-Fuller test con costante ma senza una variabile trend. L'ultima opzione invece ha entrambe le variabili differenti da 0 quindi un test aumentato Dickey-Fuller con costante e variabile trend.

Come si può notare nella tabella 3, la statistica test generata da questa libreria conferma l'ipotesi di presenza di radice unitaria, che si può notare dal valore della statistica test (0,4547), il quale non è un estremo dei valori critici, ma bensì un valore più grande. Il termine tau1 indica i valori critici della statistica test in cui si evince come ipotesi nulla che ϕ sia uguale a 0. Ulteriore conferma è la significatività del termine z.lag.1, il quale supporta l'assunzione di serie non stazionaria. Tutto ciò porta ad accettare l'ipotesi di presenza di radice unitaria e passare al test nella seconda opzione.

La stessa conclusione si può notare nella tabella 4, in cui il test genera due statistiche test (-1.6805 e 1.6753). Esse sono rispettivamente più grandi dei valori critici in tau2 e phi1, così da accettare l'ipotesi nulla e affermare la presenza di radice unitaria variabile trend. La conferma viene data anche dalla significatività dei termini z.lag.1 e dell'intercetta, il quale supporta l'assunzione di serie non stazionaria.

Come ultima opzione si ha il test generato in tabella 5, in cui si studia la presenza di radice unitaria, variabile trend e costante. Come i test precedenti si controlla la significatività dei coefficienti così da supportare la non stazionarietà della serie. Comparando i valori della statistica test (-1.7364, 1.2564, 1.6215) con i valori critici in tau3 phi3 e phi2 si può notare che nessun valore è un estremo minore rifiutandoli per ogni livello di significatività, quindi si accetta l'ipotesi nulla di presenza di radice unitaria.

Tabella 3: Test ADF tipo 1, serie giornaliera

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.242383 -0.012369 -0.000338  0.013588  0.106587

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1      0.0001393  0.0003063   0.455  0.649
z.diff.lag -0.0157059  0.0262898  -0.597  0.550

Residual standard error: 0.02683 on 1445 degrees of freedom
Multiple R-squared:  0.0003817, Adjusted R-squared:  -0.001002
F-statistic: 0.2759 on 2 and 1445 DF,  p-value: 0.759

Value of test-statistic is: 0.4547

Critical values for test statistics:
 1pct 5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Tabella 4: Test ADF tipo 2, serie giornaliera

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.242931 -0.012819  0.000063  0.013290  0.106025

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.008030   0.004529   1.773  0.0764 .
z.lag.1     -0.003305   0.001966  -1.681  0.0931 .
z.diff.lag  -0.014383   0.026281  -0.547  0.5843
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02681 on 1444 degrees of freedom
Multiple R-squared:  0.002201, Adjusted R-squared:  0.0008188
F-statistic: 1.592 on 2 and 1444 DF,  p-value: 0.2038

Value of test-statistic is: -1.6805 1.6753

Critical values for test statistics:
    1pct 5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

Tabella 5: Test ADF tipo 3, serie giornaliera

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.243848 -0.012628 -0.000039  0.013374  0.105176

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.187e-03  4.869e-03   1.887  0.0594 .
z.lag.1     -4.246e-03  2.445e-03  -1.736  0.0827 .
tt           1.360e-06  2.098e-06   0.648  0.5171
z.diff.lag  -1.376e-02  2.630e-02  -0.523  0.6009
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02681 on 1443 degrees of freedom
Multiple R-squared:  0.002491, Adjusted R-squared:  0.0004172
F-statistic: 1.201 on 3 and 1443 DF,  p-value: 0.308

Value of test-statistic is: -1.7364 1.2564 1.6215

Critical values for test statistics:
    1pct 5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Arrivati alla conclusione che la serie giornaliera del prezzo è non stazionaria dopo i test Aumentati Dickey-Fuller, ora si vede se dopo una prima differenziazione si arriverà alla stessa conclusione.

Tabella 6: Test ADF tipo 1, dopo la differenziazione

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-1.97996 -0.12012 -0.00004  0.13011  1.46357

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1      -1.007856   0.037312  -27.012 <2e-16 ***
z.diff.lag    0.002528   0.026305   0.096  0.923
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2715 on 1444 degrees of freedom
Multiple R-squared:  0.5027,    Adjusted R-squared:  0.502
F-statistic: 729.9 on 2 and 1444 DF,  p-value: < 2.2e-16

Value of test-statistic is: -27.0117

Critical values for test statistics:
    1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Tabella 7: Test ADF tipo 2, dopo la differenziazione

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-1.98436 -0.12445 -0.00431  0.12587  1.45929

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.004280   0.007143   0.599   0.549
z.lag.1     -1.008359   0.037330  -27.012 <2e-16 ***
z.diff.lag   0.002775   0.026314   0.105   0.916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2715 on 1443 degrees of freedom
Multiple R-squared:  0.5029,    Adjusted R-squared:  0.5022
F-statistic: 729.8 on 2 and 1443 DF,  p-value: < 2.2e-16

Value of test-statistic is: -27.0123 364.8326

Critical values for test statistics:
    1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

Tabella 8: test ADF tipo 3, dopo la differenziazione

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-1.98026 -0.12583 -0.00249  0.12647  1.45993

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.452e-03  1.431e-02   0.660   0.509
z.lag.1     -1.009e+00  3.734e-02 -27.008 <2e-16 ***
tt          -7.136e-06  1.711e-05  -0.417   0.677
z.diff.lag   2.891e-03  2.632e-02   0.110   0.913
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2716 on 1442 degrees of freedom
Multiple R-squared:  0.5029,    Adjusted R-squared:  0.5019
F-statistic: 486.3 on 3 and 1442 DF,  p-value: < 2.2e-16

Value of test-statistic is: -27.0078 243.1404 364.7106

Critical values for test statistics:
    1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Le tre tabelle 6, 7 e 8 compongono i test ADF dopo una prima differenziazione. Analizzando i risultati dei test in tutti i casi i valori sono estremi rispetto ai valori critici, e guardando la significatività del coefficiente z.lag.1 che in tutti i casi è significativamente non diverso da 0 per ogni livello di significatività, si può affermare che l'ipotesi nulla verrà rigettata concludendo così che la serie dopo una prima differenziazione presenta le caratteristiche di una serie stazionaria.

Capitolo 3: Analisi e applicazione dei modelli

3.1 Analisi dei Rendimenti

L'analisi dei rendimenti viene utilizzata per lo studio delle serie storiche perché a differenza della serie storica come si è vista precedentemente, nasce con le caratteristiche che si trovano in una serie stazionaria. L'analisi dei rendimenti è poi differente dalla differenziazione fatta precedentemente perché qua la differenza viene fatta prima dei logaritmi. I rendimenti in questo caso verranno calcolati come logaritmo della differenza dei prezzi.

$$R_t = \ln(p_t) - \ln(p_{t-1})$$

Lo studio della serie in questo caso inizierà nel medesimo modo della parte precedente, la conclusione come si potrà vedere sarà differente anche grazie ai modelli applicati in seguito.

Figura 11: Grafico rendimenti giornalieri

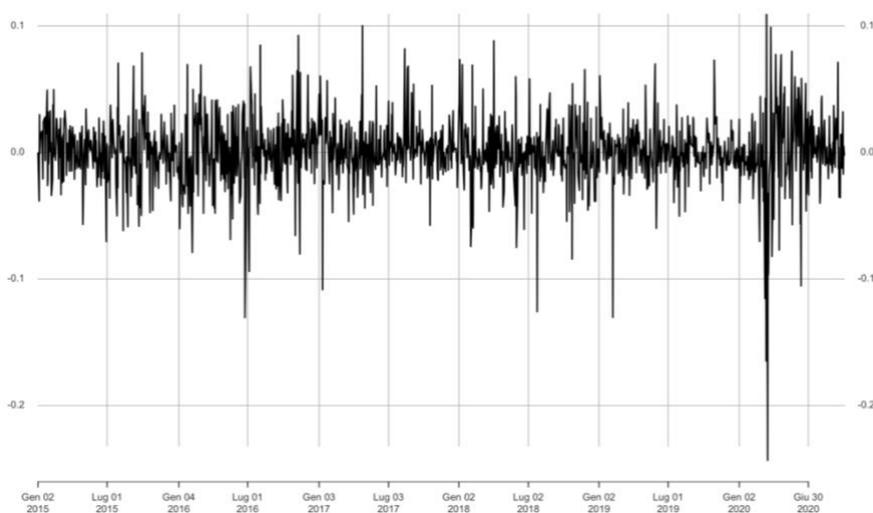
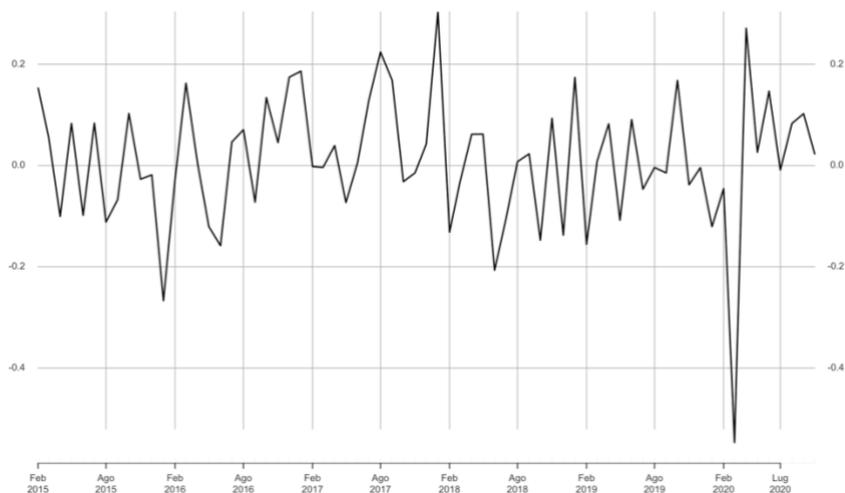


Figura 12: Grafico rendimenti mensili



Eseguendo una mera analisi visiva nelle figure 11 e 12 si può notare che la serie in entrambi in casi, sia quella giornaliera che quella mensile, non segue più un trend discendente o ascendente, ma rimane stazionaria nel lungo periodo. La varianza è però molto accentuata in tutta la sua lunghezza e necessiterà un migliore sviluppo nel seguito.

Figura 13: Funzione di autocorrelazione del rendimento giornaliero

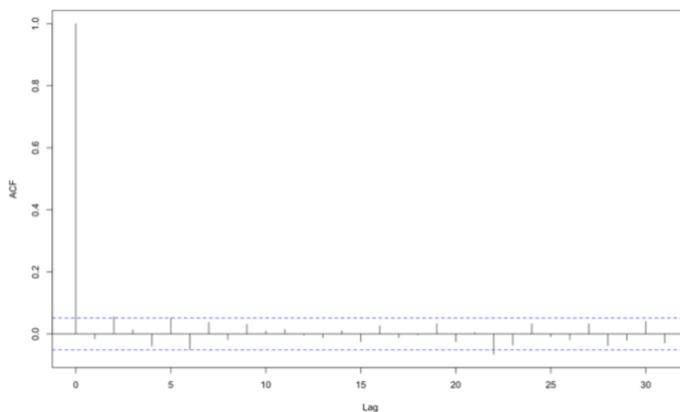


Figura 14: Funzione parziale di autocorrelazione del rendimento giornaliero

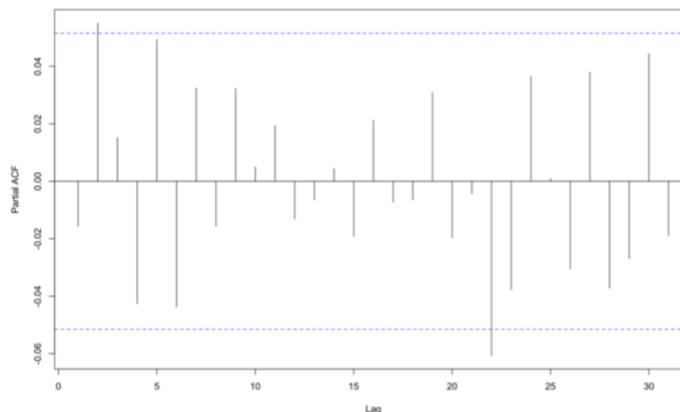


Figura 15: Funzione di autocorrelazione del rendimento mensile

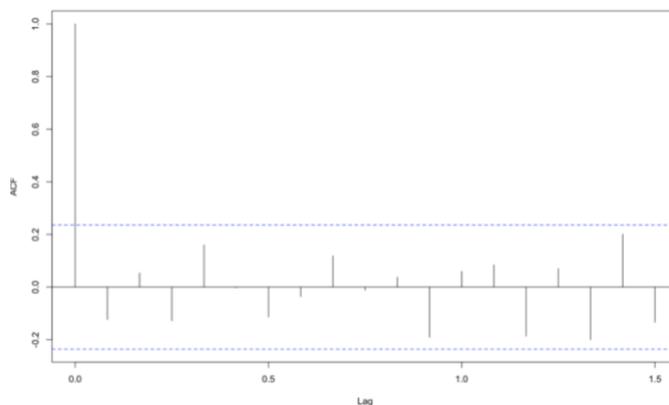
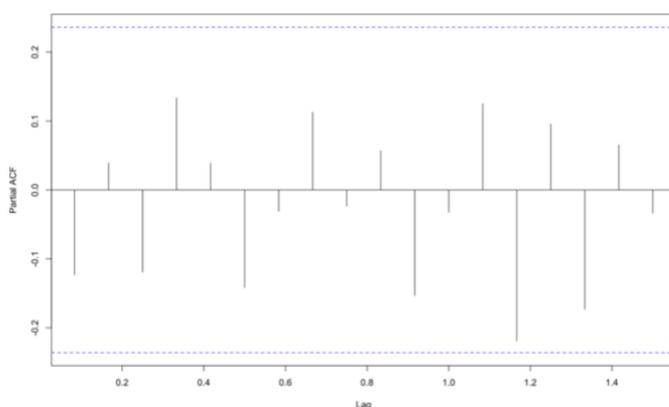


Figura 16: Funzione parziale di autocorrelazione del rendimento mensile



Si ricorda che la funzione di autocorrelazione si riferisce alle osservazioni che sono collegate tra di loro, e questo si misura tramite una semplice correlazione tra le osservazioni adiacenti.

Invece la funzione parziale di autocorrelazione misura il grado di associazione tra le osservazioni come la ACF, però togliendo la dipendenza lineare dalla serie.

Nella figura13 si può notare che la serie giornaliera mostra una leggera correlazione positiva nel ritardo 2 e una correlazione negativa al ritardo 22, invece la serie mensile nella figura 15 non mostra alcun grado di correlazione in nessun ritardo. Tutto ciò aiuta nel confermare il giudizio che sia la serie giornaliera che la serie mensile possano avere le caratteristiche di una serie stazionaria. La funzione parziale di autocorrelazione arriva alla stessa conclusione fatta nell'ACF, quindi da ora in poi si studia la stazionarietà delle due serie.

La stazionarietà si giudicherà attraverso i test di radice unitaria, più precisamente, si potrà misurare attraverso il test Aumentato Dickey Fuller.

Tabella 9: Test ADF rendimento giornaliero

```
#####  
# Augmented Dickey-Fuller Test Unit Root Test #  
#####  
  
Test regression none  
  
Call:  
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
  
Residuals:  
    Min       1Q   Median       3Q      Max  
-0.233077 -0.012099  0.000351  0.013786  0.109116  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
z.lag.1    -0.95922     0.03741  -25.644 <2e-16 ***  
z.diff.lag -0.05536     0.02625   -2.109  0.0351 *  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.02679 on 1445 degrees of freedom  
Multiple R-squared:  0.5096,    Adjusted R-squared:  0.5089  
F-statistic: 750.7 on 2 and 1445 DF,  p-value: < 2.2e-16  
  
Value of test-statistic is: -25.6439  
  
Critical values for test statistics:  
    1pct  5pct 10pct  
tau1 -2.58 -1.95 -1.62
```

Tabella 10: Test ADF rendimento mensile

```
#####  
# Augmented Dickey-Fuller Test Unit Root Test #  
#####  
  
Test regression none  
  
Call:  
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
  
Residuals:  
    Min       1Q   Median       3Q      Max  
-0.54902 -0.06663  0.00272  0.07923  0.30954  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
z.lag.1    -1.08412     0.18488  -5.864 1.66e-07 ***  
z.diff.lag -0.04051     0.12299   -0.329  0.743  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.1329 on 65 degrees of freedom  
Multiple R-squared:  0.5661,    Adjusted R-squared:  0.5528  
F-statistic: 42.41 on 2 and 65 DF,  p-value: 1.636e-12  
  
Value of test-statistic is: -5.864  
  
Critical values for test statistics:  
    1pct  5pct 10pct  
tau1 -2.6 -1.95 -1.61
```

Analizzando i valori in tabella 9 si può notare come il test abbia rigettato l'ipotesi nulla, poiché i coefficienti sono significativamente non differenti da 0, e la statistica test ha misurato un

valore (-25,6439) che è ben più estremo dei valori critici calcolati ad ogni livello di significatività.

Stessa valutazione può essere fatta alla tabella 10, la quale rigetta l'ipotesi nulla di non stazionarietà, grazie alla non significatività dei coefficienti diversi da 0 e la statistica test misura un valore (-5,864) che va oltre i valori critici per ogni livello di significatività. Rigettando l'ipotesi nulla in entrambi i test si arriva così a confermare che entrambe le serie sono stazionarie.

3.2 Il modello ARMA

Il modello ARMA (p,q), acronimo di *Auto-regressive moving average*, è un modello econometrico molto famoso nato con l'obbiettivo di riuscire a generare una funzione che possa identificare al meglio i processi stocastici. Da come fa intendere il nome, il modello nasce il modello nasce dalla combinazione di due sotto modelli: il modello autoregressivo e quello a media mobile.

3.2.1 AR(p)

In un modello a regressione multipla, si prevede la variabile di interesse utilizzando una combinazione lineare di predittori. In un modello di autoregressione, prevediamo la variabile di interesse utilizzando una combinazione lineare di valori passati della variabile.

Il termine autoregressione indica che si tratta di una regressione della variabile contro sé stessa.

Un modello autoregressivo di ordine p si può scrivere come:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \varepsilon_t$$

Oppure

$$x_t = \beta_0 + \sum_{j=1}^p \beta_j x_{t-j} + \varepsilon_t$$

dove ε_t è una sequenza detta *White noise* con media 0 e varianza σ^2 , e β_0 è una costante.

Il modello autoregressivo è come una regressione multipla, però composta da valori ritardati di x_t .

Esso è notevolmente flessibile nel gestire un'ampia gamma di modelli di serie temporali diversi, riuscendo ad adattarsi ai diversi modelli. Per il calcolo dei ritardi del modello autoregressivo ci si rifà all'utilizzo della funzione parziale di autocorrelazione, la quale calcola la diretta correlazione dei ritardi con l'osservazione corrente. L'ordine p quindi sarà il totale delle correlazioni significativamente diverse da 0.

Prendendo come esempio un modello $AR(1)$ si avrà:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t$$

Nel caso di:

- $\beta_1 = 0$ allora x_t è uguale a un processo *white noise*
- $\beta_1 = 1$ e $\beta_0 = 0$ allora x_t è uguale ad un processo *random walk*
- $\beta_1 = 1$ e $\beta_0 \neq 0$ allora x_t è uguale ad un processo *random walk* con tendenza
- $\beta_1 < 0$ allora x_t tende ad oscillare intorno alla media

3.2.2. MA(q)

Invece di utilizzare i valori passati della variabile di previsione in una regressione come nel modello autoregressivo, un modello a media mobile utilizza gli errori di previsione passati.

La media mobile $MA(q)$, può essere pensata come una media ponderata degli errori di previsione del passato u_t , ed essa si può scrivere come:

$$x_t = \mu + u_t + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_q u_{t-q}$$

Dove u_t è un processo *white noise* con media 0 e varianza σ^2 .

Come fatto per il modello $AR(p)$, il modello a media mobile si rifà alla funzione di autocorrelazione per il calcolo dei ritardi, ovviamente significativamente diversi da 0.

Un modello autoregressivo $AR(p)$ stazionario p può essere riscritto come un modello a media mobile di ordine infinito $MA(\infty)$, come dimostrato di seguito:

$$x_t = \phi_1 x_{t-1} + u_t$$

$$x_t = \phi_1(\phi_1 x_{t-2} + u_{t-1}) + u_t$$

$$x_t = \phi_1^2 x_{t-2} + \phi_1 u_{t-1} + u_t$$

$$x_t = \phi_1^3 x_{t-3} + \phi_1^2 u_{t-2} + \phi_1 u_{t-1} + u_t$$

E così a proseguire, a condizione che $-1 < \phi_1 < 1$. Il valore di ϕ_1^k sarà sempre più piccolo, al crescere di k . Quindi riscrivendo otterremo un modello $MA(\infty)$

$$x_t = u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \phi_1^3 u_{t-3} + \dots + \phi_1^k u_{t-k}$$

Il risultato inverso vale se imponiamo alcuni vincoli ai parametri del modello MA.

Allora il modello a media mobile è chiamato invertibile poiché possiamo scrivere qualsiasi processo $MA(q)$ come un processo $AR(\infty)$. I modelli non sono semplicemente introdotti per consentire la conversione da modelli a media mobile a modelli autoregressivi, ma essi hanno anche alcune proprietà matematiche desiderabili.

Come per esempio un processo $MA(1)$ $x_t = u_t + \phi_1 u_{t-1}$. Si può dimostrare che il suo processo $AR(\infty)$, che il suo errore più recente può essere scritto come funzione lineare delle osservazioni attuali e passate

$$u_t = \sum_{j=0}^{\infty} (-\phi)^j x_{t-j}$$

Quando $|\theta| > 1$, i pesi aumentano con l'aumentare dei ritardi, quindi più le osservazioni sono distanti, maggiore è la loro influenza sull'errore. Quando $|\theta| = 1$, i pesi sono di dimensioni costanti, e le osservazioni distanti hanno la stessa influenza delle osservazioni recenti.

Abbiamo quindi bisogno di $|\theta| < 1$ così che le due situazioni abbiano senso, quindi le osservazioni più recenti hanno un peso maggiore rispetto alle osservazioni del passato più lontano. Pertanto, il processo è invertibile quando $|\theta| < 1$.

3.2.3 ARMA(p,q)

Il modello $ARMA(p, q)$ non è altro che l'unione del modello autoregressivo e quello a media mobile, il quali combinandosi creano un processo più efficace e potente dei due presi singolarmente, così da trovare il miglior processo stocastico possibile.

Quindi il modello ARMA potrà essere scritto come somma dei due modelli sopra citati:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t + u_t + \phi_1 u_{t-1} + \dots + \phi_q u_{t-q}$$

Gli ordini dei ritardi p e q , sono scelti tramite i due processi sopra, e per prassi essi devono essere più piccoli possibili quando applicati alle serie temporali.

Altra caratteristica del modello è che il processo stocastico di partenza venga definito stazionario e che i residui siano distribuiti come un processo *white noise* $WN(0, \sigma^2)$.

Per capire quindi se i residui seguono un processo *white noise* si userà come fatto in precedenza il test Box-Ljung, il quale genererà una statistica test per controllare l'autocorrelazione tra i dati. Per il nostro obiettivo è bene che i residui accettino l'ipotesi nulla.

Una volta identificato l'ordine del modello (cioè i valori di p , d e q), si devono stimare i parametri. Questi saranno generati direttamente dalla piattaforma software R, il quale stimerà il modello ARIMA, utilizzando la stima della massima probabilità (MLE, *Maximum likelihood estimation*). Questa tecnica trova i valori dei parametri che massimizzano la probabilità di ottenere i dati che abbiamo osservato. In pratica, R riporterà il valore della probabilità dei logaritmi dei dati, cioè il logaritmo della probabilità che i dati osservati provengano dal modello stimato. Per determinati valori di p , d e q , R cercherà di massimizzare la probabilità del logaritmo nel momento di trovare le stime dei parametri.

È stato così inserito così il modello ARIMA, il quale non è altro che una trasformazione a sua volta del modello ARMA con l'aggiunta del termine I. Quest'ultimo sta per *Integrated*, ossia l'inverso di differenziato, poiché il modello $ARIMA(p, d, q)$, è utile per studiare le serie che non passano i test di stazionarietà. Nella serie in oggetto però non occorrerà utilizzare questo modello perché come provato in precedenza, la serie dei rendimenti giornalieri e mensili sono entrambe stazionarie, quindi rispettano la caratteristica del modello ARMA.

Partendo dai residui del modello dei rendimenti giornalieri si nota che

Box-Ljung test

```
data: residuals(arma(rendimento_giornaliero, order = c(2, 2)))  
X-squared = 0.83418, df = 1, p-value = 0.3611
```

il p-value della statistica test è ben oltre la soglia limite così da accettare l'ipotesi nulla e affermare che i residui sono non correlati tra di loro, che seguono un processo *white noise*.

Il modello ARMA ricercato è stato trovato tramite l'utilizzo del software R il quale attraverso il modello ARIMA, stimerà la migliore combinazione di ritardi. Il criterio per stimare la combinazione è detto criterio di informazione di Akaike (AIC), che è stato utile nella selezione dei predittori per la regressione, è utile anche per determinare l'ordine di un modello ARIMA. È importante notare che questi criteri informativi tendono a non essere buone guide per la selezione dell'ordine appropriato di differenziazione di un modello, ma solo per la selezione dei valori di p e q . Ciò è dovuto al fatto che la differenziazione modifica i dati sui quali viene

calcolata la probabilità, rendendo i valori AIC tra modelli con diversi ordini di differenziazione non comparabili.

Il miglior modello che stima il processo dei rendimenti giornalieri è il modello $ARMA(2,2)$, che ha come coefficienti:

```
arma(x = rendimento_giornaliero, order = c(2, 2), include.intercept = TRUE)
```

```
Coefficient(s):  
      ar1      ar2      ma1      ma2 intercept  
-1.075180 -0.656490  1.038805  0.655045  0.002064
```

Per quanto riguarda il modello dei rendimenti mensili, come si può notare dall'autocorrelazione e dalla funzione parziale di autocorrelazione, non ci sono ordini di ritardo significativamente diversi da 0.

Testando il modello i residui secondo il test sono incorrelati tra di loro, dato che il p-value della statistica test ricade ben oltre il limite, così da confermare che i residui seguono un processo di tipo *white noise*.

Box-Pierce test

```
data: residuals(arma(rendimento_mensile, order = c(1, 2)))  
X-squared = 0.00012849, df = 1, p-value = 0.991
```

Il modello ARMA qua trovato è diverso da quello precedente poiché l'ordine che meglio riproduce il processo stocastico, secondo il criterio sopra citato è il modello $ARMA(1,2)$.

Il quale ha come coefficienti

```
Coefficient(s):  
      ar1      ma1      ma2 intercept  
-0.811070  0.717463  0.013954  0.008565
```

3.3 Il modello GARCH

Le serie temporali finanziarie mostrano spesso un comportamento che è noto come clustering della volatilità: la volatilità cambia nel tempo e il suo grado mostra una tendenza a persistere, cioè ci sono periodi di bassa volatilità e periodi in cui la volatilità è elevata.

L'effetto chiamato clustering è ben visibile all'interno della figura 11, dove si notano zone di bassa volatilità alle quali susseguono zone di bassa volatilità, e altre zone in cui la volatilità è decisamente più persistente ed accentuata.

Il modello ARCH e GARCH è nato grazie agli studi del professor Robert F. Engle e del suo studente Tim P. Bollerslev. Entrambi sono famosi nella comunità internazionale proprio per le idee sulla misurazione e previsione della volatilità dei mercati finanziari. Il primo è famoso per lo studio del modello ARCH (*Autoregressive Conditional Eteroskedasticity*) nato nel 1982, e il secondo per aver implementato ed esteso il modello precedente sviluppando il processo GARCH (*Generalised Autoregressive Conditional Eteroskedasticity*) nato nel 1986.

Il processo in questione ha come scopo quello di generare il miglior modello possibile che riesca ad adattarsi allo studio della volatilità di una serie temporale.

Il contetto principale è che la volatilità registrata in un punto è affetta dai suoi residui al quadrato, i quali seguono un processo autoregressivo di ordine p .

Il modello ARCH

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2$$

dove ε_t^2 è il quadrato del residuo al tempo t , e il processo segue un modello di tipo $AR(p)$, e con $\alpha_j \geq 0, j = 0, 1, 2, \dots, p$, come vincolo a causa della non negatività della varianza.

Il motivo che muove il modello GARCH invece è che la volatilità registrata in un punto è affetta dalla volatilità registrata nel punto adiacente a quest'ultimo, così registrando una dipendenza tra le volatilità nel tempo.

Il modello GARCH

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

con $\alpha_i \geq 0$ e $\beta_j \geq 0, \forall j, \alpha + \beta < 1$.

Figura 17: Funzione di autocorrelazione dei rendimenti al quadrato

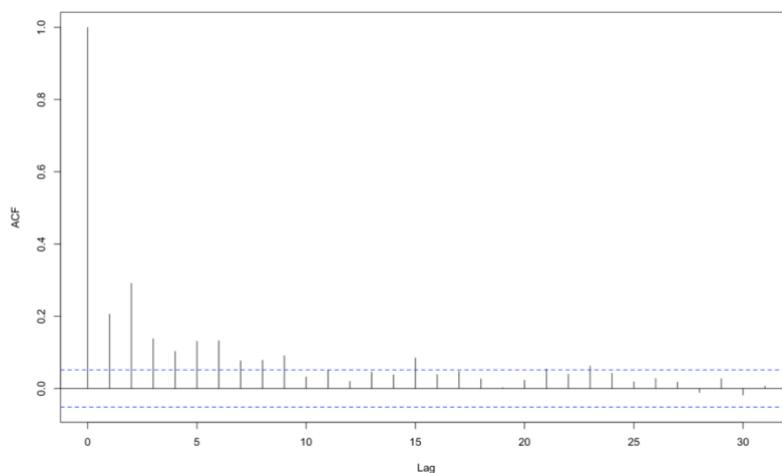
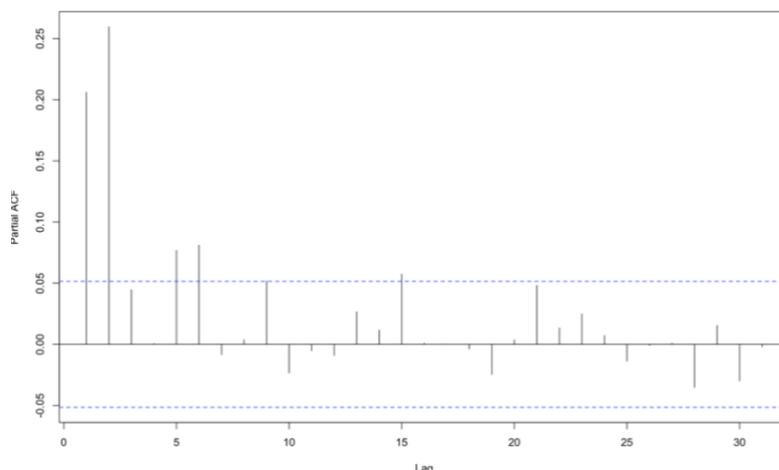


Figura 18: Funzione parziale di autocorrelazione dei rendimenti al quadrato



Nelle figure 16 e 17 è presente autocorrelazione positiva per i rendimenti al quadrato, portando così a investigare successivamente eseguendo il modello GARCH(1,1).

Questo modello, solitamente, è sufficientemente efficace per potersi applicare ad una svariata quantità di serie temporali differenti.

I coefficienti ottimali sono così generati automaticamente dal software.

```
*-----*
*          GARCH Model Fit          *
*-----*

Optimal Parameters
-----
      Estimate  Std. Error  t value  Pr(>|t|)
mu      0.000811   0.000616   1.3169  0.187859
omega   0.000093   0.000023   4.0778  0.000045
alpha1  0.131437   0.028152   4.6689  0.000003
beta1   0.733011   0.053332  13.7444  0.000000
```

Per approfondire ulteriormente se il modello è il più adatto possibile ai dati, si effettua così un Box-Ljung test sui residui per vedere se questi presentino caratteristiche di autocorrelazione.

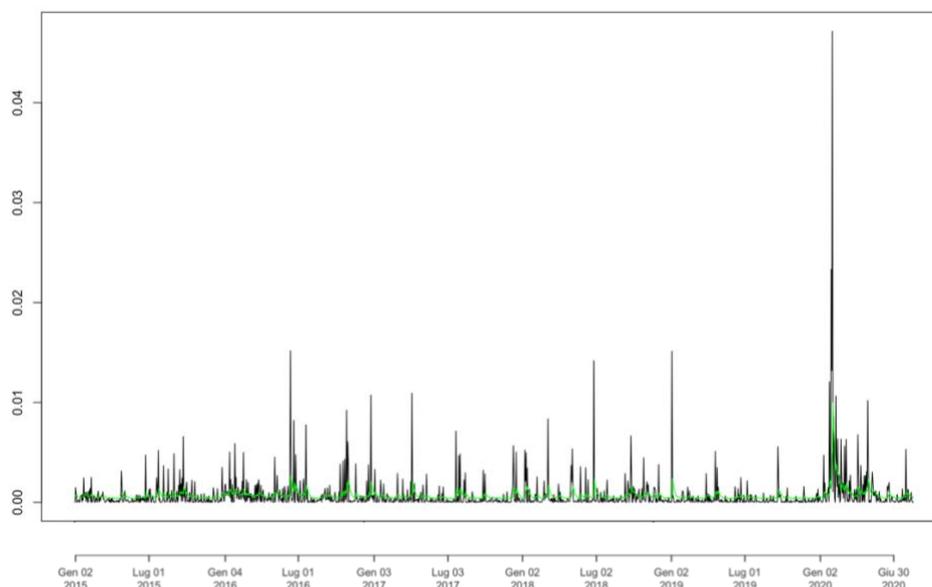
```
Weighted Ljung-Box Test on Standardized Residuals
-----
              statistic  p-value
Lag[1]                3.253e-05  0.9954
Lag[2*(p+q)+(p+q)-1][2] 1.225e-01  0.9042
Lag[4*(p+q)+(p+q)-1][5] 1.085e+00  0.8401
d.o.f=0
H0 : No serial correlation
```

Il test generato afferma che, dato il modello, non è presente autocorrelazione nei residui.

Ulteriore modo per capire se il modello generato riesca ad adattarsi al meglio alla serie, è analizzare visivamente se la varianza condizionata segua i picchi dei residui al quadrato.

La figura 18 mostra che nei momenti in cui la volatilità cresce, il modello riesce ad adattarsi di conseguenza, e nei momenti in cui i picchi sono meno accentuati il modello continua ad adattarsi.

Figura 19: Grafico residui al quadrato



3.4 Decomposizione e previsione

La previsione è un metodo statistico per prevedere un attributo utilizzando modelli storici nei dati. Tutte le aziende si affidano alle previsioni e traggono grandi benefici da un accurato valore di previsione. Ogni azienda o organizzazione applica diversi metodi di previsione in diverse situazioni, pertanto è imperativo identificare quale previsione e quale metodo di previsione dovrebbe essere utilizzato in diversi scenari in modo da ridurre al minimo il rischio di previsione.

I dati possono essere classificati in tre gruppi principali attraverso la natura temporale usata:

- **Dati trasversali:** I dati sono raccolti in un unico punto nel tempo o in una o più variabili, in questo caso i dati non sono sequenziali. Solitamente i punti dati sono indipendenti l'uno dall'altro;

- Dati di serie temporali: i dati univariati e multivariati sono osservati nel tempo in modo sequenziale a intervalli di tempo equidistanti a seconda del predatore;
- Combinazione di dati di sezioni trasversali e serie temporali: si tratta di un progetto di studio complesso in cui le informazioni sulle variabili di scena sono raccolte in vari punti di tempo. Molti campionamenti di indagini fanno uso dei dati dei pannelli.

Per lo studio fatto in questo elaborato lo studio ricade sui dati delle serie temporali.

I dati delle serie temporali hanno diverse caratteristiche che li rendono unici.

La prima riguarda la dipendenza delle osservazioni, di fatti nei dati delle serie temporali ogni osservazione dovrebbe dipendere dall'osservazione passata.

Altra caratteristica riguarda la mancanza di dati e la loro imputazione, poiché tutti i punti di dati sono sequenziali nelle serie temporali, se mancano dei punti di dati, devono essere imputati prima dell'inizio del processo di analisi vero e proprio, altrimenti non viene conservato l'ordine corretto.

Infine, non è possibile mescolare due diversi tipi di intervalli. I dati delle serie temporali vengono osservati sulla stessa variabile in un determinato periodo di tempo con intervalli di tempo fissi e regolari. Tuttavia, i dati possono essere raccolti a vari intervalli di tempo, ad esempio annuale, mensile, settimanale, giornaliero e orario, oppure a qualsiasi intervallo di tempo specifico. L'intervallo deve rimanere lo stesso per l'intero arco temporale stimato, per esempio le serie annuali non possono essere combinate con le serie trimestrali o mensili.

L'utilizzo principale dei dati a serie temporale ha come fine la previsione. Essa viene applicata per estrarre informazioni dalle serie storiche e viene utilizzata per prevedere il comportamento futuro delle stesse serie sulla base di modelli passati.

Ci sono due approcci principali per la previsione delle serie temporali, il primo è la Decomposizione che si basa sull'estrazione dei singoli componenti delle serie temporali. Il secondo è la regressione che si basa sulla regressione su osservazioni passate.

Il metodo della Decomposizione come visto in precedentemente prevede la divisione in componenti della serie che per semplicità si diminuiscono a tre: *Trend*, *Seasonal cycle* e *Irregular component*. Il primo prevede un incremento lineare o non lineare nel lungo periodo, il secondo invece, riguarda movimenti intra annuali simili a fluttuazioni che si ripetono in tutta la lunghezza della serie. L'ultimo raggruppa tutti i movimenti imprevedibili e imprevedibili nella serie. Esso non prende ruolo nell'analisi della previsione.

Date queste componenti, alcuni metodi comuni per estrarli da una serie temporale sono il metodo additivo e il metodo moltiplicativo.

La decomposizione additiva è la più appropriata se l'entità delle fluttuazioni stagionali, o la variazione intorno al trend, non varia con il proseguire della serie.

$$Y_t = T_t + S_t + I_t$$

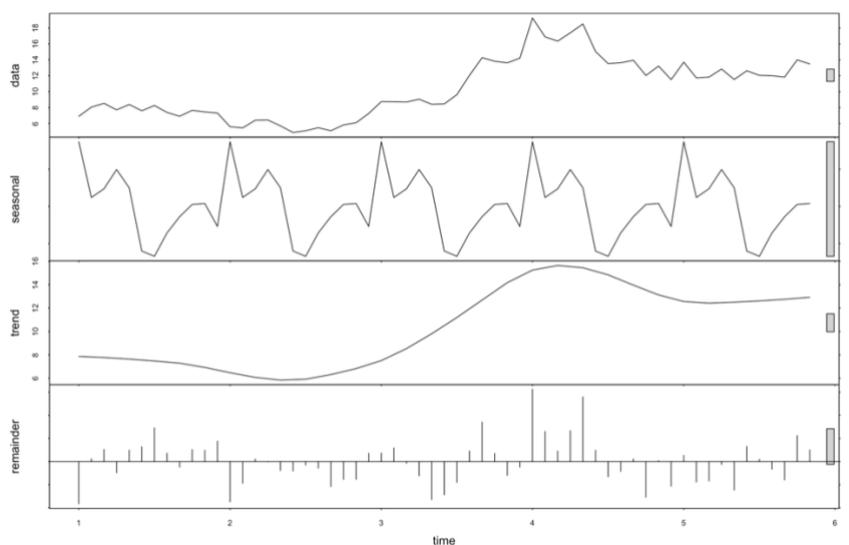
Quando la variazione del modello stagionale, o la variazione intorno al trend, sembra essere proporzionale al proseguire della serie, allora una decomposizione moltiplicativa è più appropriata.

$$Y_t = T_t \times S_t \times I_t$$

Le decomposizioni moltiplicative sono comuni con le serie temporali economiche. Un'ulteriore caratteristica è che quest'ultima può essere trasformata in una serie additiva prendendo il logaritmo delle componenti.

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(I_t)$$

Figura 20: Decomposizione serie temporale mensile con metodo moltiplicativo



In questo caso, si utilizzano dati mensili poiché così è più semplice rilevare l'effetto stagionale. La figura 20 mostra la scomposizione dei dati mensili secondo il metodo STL.

Questo è un metodo versatile e robusto per la decomposizione di serie temporali. STL è l'acronimo di "*Seasonal and Trend decomposition using Loess*", mentre *Loess* è un metodo per la stima delle relazioni non lineari. Il metodo STL è stato sviluppato da Cleveland, Cleveland, McRae, & Terpenning (1990).

Il modello è particolare perché consente alla componente stagionale di cambiare nel tempo.

Può essere robusto verso gli *outlier*, in modo che occasionali osservazioni insolite non influenzino le stime del ciclo di tendenza e le componenti stagionali

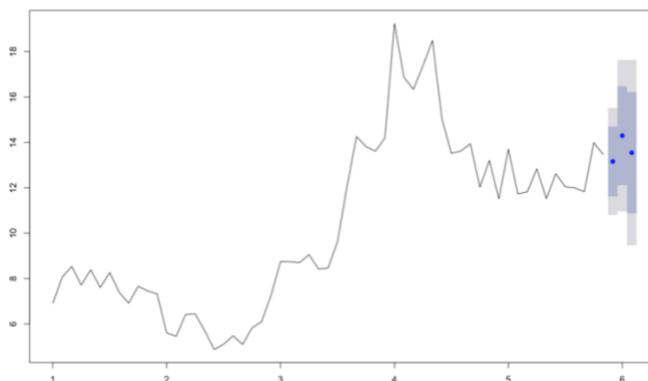
La serie mensile così rappresentata divide visivamente tutte le componenti l'una dalle altre mostrando un effetto stagionale molto pronunciato e un trend che accompagna l'intera serie.

Prima di proporre un metodo di previsione, il metodo deve essere valutato. A tal fine, i dati devono essere suddivisi in due gruppi: formazione e test. I dati di formazione aiutano ad identificare e ad adattare il modello giusto e i dati dei test vengono utilizzati per convalidare lo stesso. Il modello di previsione viene generato usando sia il modello STL sia il modello ARIMA.

Nel caso di dati di serie temporali i dati di prova sono la parte più recente della serie in modo da conservare l'ordine dei dati.

Quindi per testare l'accuratezza della previsione, si genera un grafico per mostrare la previsione degli ultimi tre mesi di dati, così da validare il modello.

Figura 21: Grafico previsione ultimi 3 mesi luglio 2020-settembre 2020

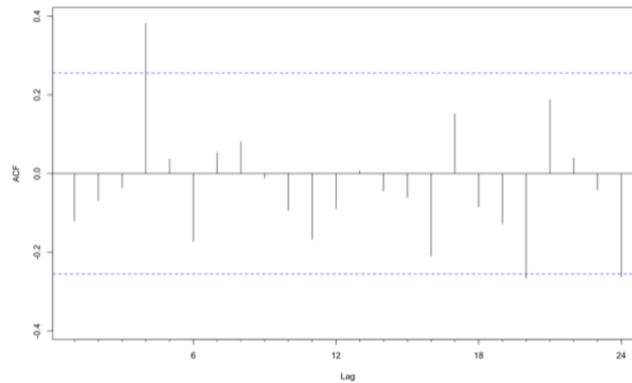


La figura 21 mostra un grafico mensile in cui si prevedono gli ultimi tre mesi di dati, con delle bande che indicano con quale confidenza i dati potranno posizionarsi in quei punti.

Per valutare la bontà del modello si valuta l'accuratezza della previsione, che misura e confronta i valori previsti con il valore osservato per quantificare il potere predittivo del modello proposto.

I residui devono essere non correlati, altrimenti nei dati rimangono delle informazioni che devono essere contabilizzate. Anche la media dei residui dovrebbe essere vicino allo zero, altrimenti la previsione è distorta.

Figura 22: Funzione di autocorrelazione dei residui



Le linee non si spezzano eccessivamente sopra o sotto le linee blu tratteggiate, ossia nel segno livello di confidenza, e quindi la mancanza di correlazione suggerisce che la previsione è buona. Per maggiore accuratezza si calcola media dei residui per capire quanto il valore generato sia vicino allo zero.

```
> mean(res)
[1] 0.1253019
```

I residui mostrano che il modello utilizzato può dichiararsi buono, data la loro media, anche se leggermente sovrastimano i valori predetti.

Bibliografia

Corrado Canali, 2020. Stellantis, tutte le sfide che attendono il nuovo gruppo nato dalla fusione fra FCA e PSA. Il sole 24 ore. Disponibile su <http://www.ilsole24ore.com/>

Stock J. e Watson M. M., 2016. *Introduzione all'econometria*. Milano: Pearsons

Said E. S. e Dickey D. A., 1984. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, vol. 71 (3), pp. 599-607

Kirchgässner G. and Wolters J., 2007. *Introduction to Modern Time Series Analysis*. : Springer

Fiat Chrysler Automobiles N.V SWOT Analysis. 6/18/2020, p1-7. 7p. Database: Business Source Complete. Disponibile su: <https://web.a.ebscohost.com>

Fox, Kevin J. White noise and other experiments on augmented Dickey-Fuller tests. *Applied Econometrics Letters*, Vol 4.

Robert B. Cleveland, William S. Cleveland, Jean E. McRae, Irma Terpenning, 1990. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, Vol.6 No.1, pp 3-73.

Zsuzsanna Horvath, Ryan Johnston. AR(1) Time series process Econometrics. Relazione finale, The university of Utah. Disponibile su: <https://www.math.utah.edu>

Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer, 2020. *Introduction to Econometrics with R*. disponibile su: <https://www.oek.wiwi.uni-due.de>.

Box G. E. P. e Pierce D. A., 1970. Distribution of Residual Autocorrelations in Autoregressive- Integrated Moving Average Time Series Models, *Journal of the American Statistical Association*, pp. 1509-1526.

Dati raccolti dal sito ufficiale di Fiat Chrysler Automobiles. Disponibile su: <http://www.fcagroup.com/>