



Università  
Ca' Foscari  
Venezia

Corso di Laurea  
magistrale

in Economia e Finanza

Tesi di Laurea

# Analisi del mercato e modelli predittivi per serie storiche finanziarie

**Laureando**

Ludovico Breda

Matricola 879761

**Relatore**

Ch.mo Prof. Claudio Pizzi

**Anno Accademico**

2023 / 2024

# Sommario

|                                                                                            |           |
|--------------------------------------------------------------------------------------------|-----------|
| <b>Introduzione .....</b>                                                                  | <b>3</b>  |
| 1.1 Efficienza del mercato: realtà o utopia? .....                                         | 4         |
| 1.2 Mercato inefficiente: l'interpretazione "behavioural prone" .....                      | 8         |
| 1.2.1 La teoria dei prospetti .....                                                        | 9         |
| 1.2.2 Le anomalie di mercato .....                                                         | 11        |
| <b>Capitolo I – Concetti di statistica e storia delle previsioni .....</b>                 | <b>14</b> |
| 2.1 Cos'è un processo stocastico .....                                                     | 14        |
| 2.2 Proprietà di un processo stocastico .....                                              | 15        |
| 2.2.1 La stazionarietà .....                                                               | 15        |
| 2.2.2 L'invertibilità e l'ergodicità .....                                                 | 16        |
| 2.3 L'exponential smoothing .....                                                          | 17        |
| 2.3.1 Il simple exponential smoothing (SES) .....                                          | 17        |
| 2.3.2 Il metodo esponenziale con trend lineare .....                                       | 18        |
| 2.3.3 Il metodo esponenziale con stagionalità .....                                        | 19        |
| 2.3.4 Generalizzazione dell'exponential smoothing: fattore additivo e moltiplicativo ..... | 21        |
| 2.4 I modelli ARIMA .....                                                                  | 24        |
| 2.5 I modelli ARCH .....                                                                   | 29        |
| 2.6 I modelli di machine learning basati su reti neurali .....                             | 32        |
| <b>Capitolo II – Proprietà delle reti neurali artificiali .....</b>                        | <b>34</b> |
| 3.1 Le reti feed forward .....                                                             | 36        |
| 3.1.1 Il Convolutional Neural Network .....                                                | 39        |
| 3.2 Le reti neurali ricorrenti .....                                                       | 42        |
| 3.2.1 Le Long Short Term Memory .....                                                      | 43        |
| 3.2.2 Le Gated Recurrent Unit .....                                                        | 45        |
| 3.3 Gli aspetti negativi delle reti neurali .....                                          | 47        |
| <b>Capitolo III – Confronto operativo tra modelli neurali .....</b>                        | <b>49</b> |
| 3.1 Previsioni utilizzando dati giornalieri .....                                          | 50        |
| 3.1.1 Previsione a n passi .....                                                           | 59        |
| 3.2 Previsioni utilizzando dati infragiornalieri .....                                     | 61        |
| 3.2.1. Previsione a n passi .....                                                          | 71        |
| <b>Capitolo IV – Strategia di trading .....</b>                                            | <b>75</b> |
| <b>Conclusioni .....</b>                                                                   | <b>80</b> |

Bibliografia e Sitografia ..... 82

## Introduzione

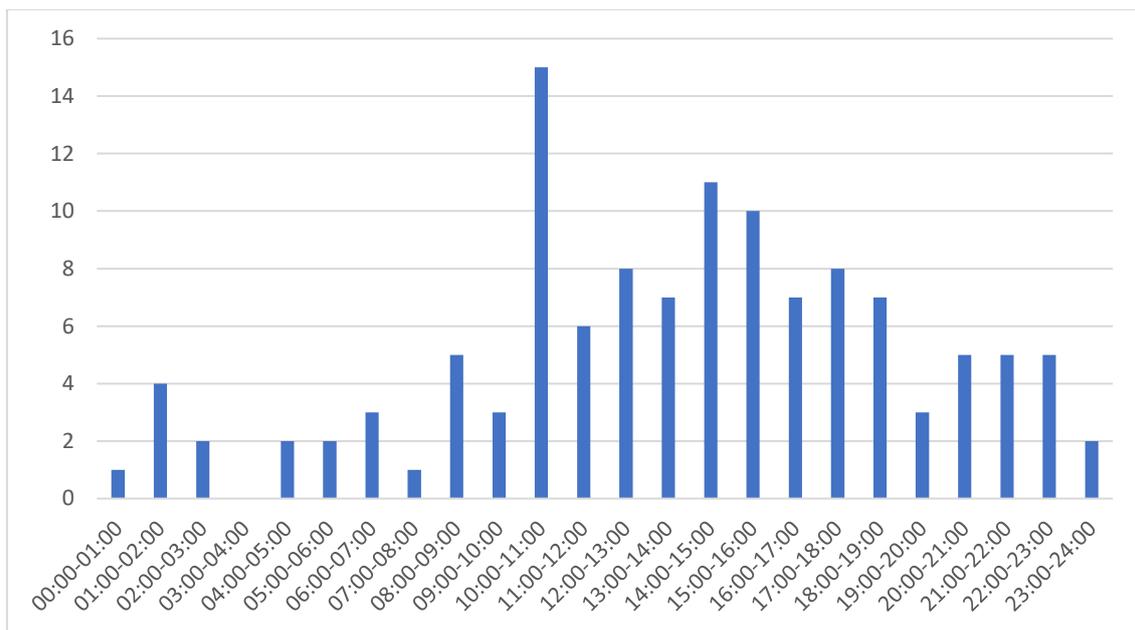
L'interesse per ciò che accadrà in un futuro sempre più incerto, abbinato alla curiosità nell'ambito del machine learning ma più in generale dell'innovazione tecnologica, mi ha avvicinato a questo tema che ben si presta ad un'interpretazione in ambito statistico-finanziario.

L'obiettivo di questa tesi è quello di analizzare e ad approfondire tutti gli aspetti che ruotano attorno al mondo delle previsioni in ambito finanziario, andando ad analizzare e definire, oltre al comportamento e alla reazione degli investitori nei confronti delle nuove notizie a disposizione, il metodo previsivo che maggiormente si adatta ad un contesto finanziario in costante evoluzione.

L'elaborato, nella prima parte, andrà a soddisfare la necessità di un inquadramento dal punto di vista storico, come richiede un argomento di questo tipo, per poi vedere analizzate le diverse metodologie di previsione delle serie storiche finanziarie in contesto di low frequency data e di high frequency data. Nell'ultima parte della tesi verrà presentato un sistema di trading automatico basato sul modello previsionale ritenuto più adeguato a trattare e a manipolare dati ad alta frequenza, per verificarne l'adeguatezza e l'adattabilità nel contesto di più asset finanziari, i quali costringono gli investitori a monitorare maggiormente le loro posizioni sul mercato di riferimento per modificare o confermare le scelte d'investimento. Sarà inoltre interessante comprendere e studiare la struttura su cui si basano gli algoritmi di machine learning atti alla gestione dei dati finanziari e all'approssimazione del processo generatore dei dati delle serie storiche finanziarie e in particolare il meccanismo che caratterizza le reti neurali.

L'evoluzione degli strumenti informatici e la possibilità di accedere a sempre più dati e informazioni hanno permesso, anche nel mondo finanziario, di individuare un'applicazione in termini di monitoraggio dell'andamento dei prezzi dei titoli; terminali come Bloomberg, Profit o TradingView hanno iniziato ad accompagnare gli investitori nel loro processo di investment decision e ad accorpate le notizie di mercato di interesse, rendendo democratica la fruizione di contenuti storici o in tempo reale.

Grafico 1. Distribuzione oraria delle informazioni finanziarie pubbliche (28 marzo 2024)



Fonte: Reuters, Borsa&Finanza

Anche se non indica una situazione rappresentativa, il Grafico 1 permette di comprendere approssimativamente il numero (seppur soggetto ad una oscillazione che dipende da fattori macroeconomici e di contesto) di informazioni che un investitore deve gestire dai principali siti di aggiornamento finanziario. Un contesto di frenetico aggiornamento delle informazioni, sulla scia di una sempre maggiore potenza computazionale, impone l'utilizzo di nuove tipologie di framework (o architettura logica di un modello) che si possano adattare per riuscire ad ottenere delle stime sufficientemente affidabili su cui si possano basare delle strategie di investimento. Per poter intercettare i flussi informativi e gli shock contenuti nelle migliaia delle transazioni nel mercato finanziario che avvengono ogni giorno non sono più sufficienti dei modelli parametrici statici, ma come vedremo risultano più efficaci dei sistemi di apprendimento autonomi che possano ricevere in input dati strutturati e non, per poi poterne estrarre le informazioni più significative.

### 1.1 Efficienza del mercato: realtà o utopia?

Prima di definire i modelli di previsione è necessario specificare in che modo il mercato, nella sua componente fondamentale ovvero il prezzo di un titolo, possa reagire alle notizie che lo coinvolgono (dato il comportamento degli investitori); il concetto di efficienza dei

mercati fonda le proprie radici su uno studio dell'economista Eugene Fama<sup>1</sup> (1970), che definì l'efficienza come la capacità dei prezzi azionari di assorbire le notizie di mercato; questo assunto implica che sia impossibile approfittare di temporanei disallineamenti (o comunque molto difficile in quanto le inefficienze vengono immediatamente assorbite sfruttando l'opportunità di arbitraggio) nel prezzo rispetto al fair value, o valore fondamentale, degli asset finanziari. Questo tipo di teoria implica che il prezzo rifletta il valore fondamentale del titolo di riferimento in base alle informazioni a disposizione degli investitori; in virtù di quest'ultime si possono distinguere, secondo Fama, tre forme di efficienza: la forma debole, secondo cui i prezzi riflettono tutte le informazioni contenute nei prezzi passati, la forma semi-forte, che si differenzia dalla precedente in base alla capacità dei prezzi di riflettere anche le notizie pubbliche (come bilanci, annunci su dividendi, aspettative di crescita/decrecita di un settore) e la forma forte, che rappresenta il grado di efficienza ultimo, dove oltre alle precedenti nozioni il prezzo riflette anche le informazioni non accessibili pubblicamente (quelle conosciute dagli insider di mercato).

È inoltre fondamentale riportare un'altra teoria del mercato efficiente, presentata da James Tobin (1984)<sup>2</sup>. Tobin propone più interpretazioni del significato di efficienza a differenza della teoria proposta da Fama, che secondo il quale ha una sola forma. L'efficienza si veste dell'aspetto informativo poiché i prezzi riflettono in ogni momento tutte le informazioni disponibili, sia pubbliche che private; è di tipo valutativo in quanto le informazioni concorrono alla determinazione corretta del valore dell'impresa, in base al suo valore fondamentale (che viene riflesso nei prezzi). L'efficienza del mercato inoltre ricopre la caratteristica della completezza, in quanto c'è la possibilità di realizzare scambi relativi a tutte le scadenze che coprono tutte le possibilità future, è funzionale poiché non è fine a sé stessa ma contribuisce all'allocazione del risparmio e alla domanda di capitale.

---

<sup>1</sup> Si fa riferimento al primo paper intitolato *"Efficient capital markets: a review of theory and empirical work (1970)"*, mentre nell'articolo *"Market efficiency, long-term returns, and behavioral finance (1998)"* E. Fama risponde alle critiche sulla sua teoria dimostrando, nell'intento di smentire la teoria di inefficienza dei mercati proposta da alcuni economisti comportamentali, che le over/under reaction sulle notizie del mercato dei prezzi dei titoli si neutralizzano nel lungo periodo, andando via via a "sgonfiarsi" e a riportare i prezzi al loro valore fondamentale.

<sup>2</sup> Teoria presentata nel paper *"On the Efficiency of the Financial System (1984)"*.

Infine, l'efficienza è operativa giacché permette di utilizzare un insieme di procedure grazie alle quali il mercato svolge le proprie operazioni a costi minimi.

A supporto di questo ragionamento è stato dimostrato da Blake, Elton e Gruber (1993) che le performance di fondi comuni di investimento, ovvero fondi di gestione attiva che hanno l'obiettivo di "battere il mercato" in termini di rendimenti intercettati rispetto alla media del benchmark, sono inferiori rispetto al benchmark di mercato di riferimento di un ammontare percentuale pari alle commissioni caricate nel fondo (perciò mediamente riescono al massimo a riprodurre l'andamento dell'indice e non a batterlo).

La casualità nel movimento dei prezzi nei mercati finanziari portò gli economisti, nella seconda metà del '900, alla conclusione che il modello statistico che più rispecchia questo comportamento si sarebbe dovuto basare sulla teoria random walk.

Questo modello implica che i valori di scambio futuri di un titolo non possano essere predetti conoscendo le informazioni sui prezzi passati (Porter, 1905); la previsione del prezzo di un titolo domani (ipotizzando come riferimento  $t$  il tempo oggi), infatti, sarà pari a:

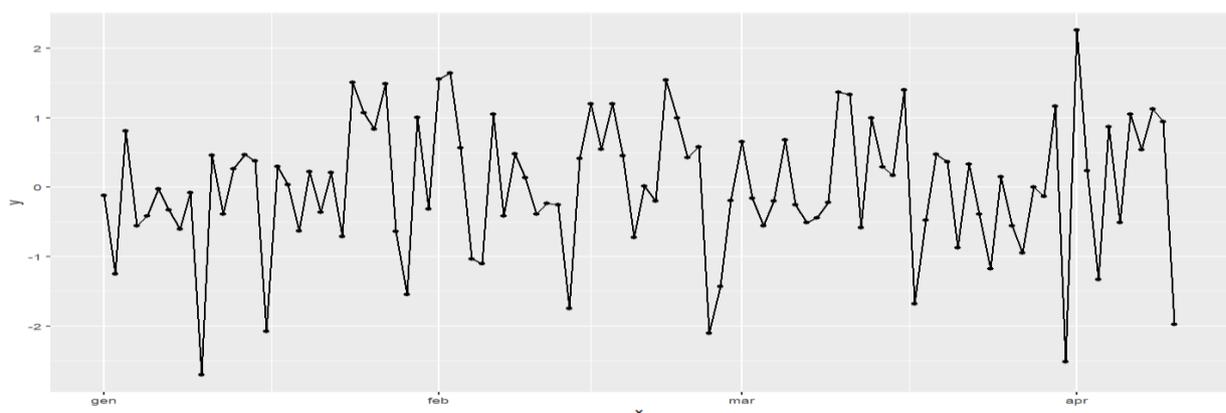
$$p_{t+1} = p_t + \varepsilon_{t+1}$$

(1.0)

dove:

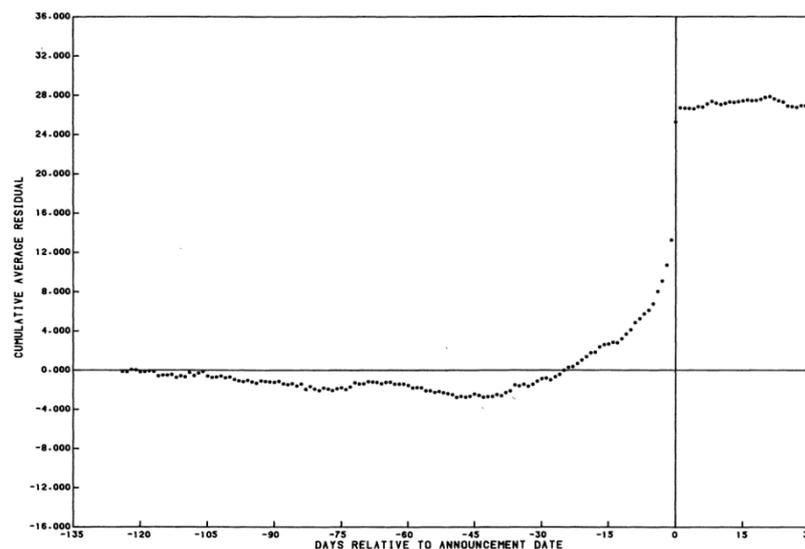
- $\varepsilon_{t+1}$  è una variabile causale di tipo white noise

Grafico 2. Esempio dell'andamento di una variabile white noise con media 0 e varianza 1



Il termine rappresenta la realizzazione al tempo  $t+1$  di un processo white noise, cioè un processo a componenti indipendenti ed identicamente distribuite con media 0 e varianza 1. Questa componente stocastica,  $\varepsilon_{t+1}$ , che al tempo  $t$  non è nota poiché è una variabile casuale (v.c.) il cui valore sarà osservabile al tempo  $t+1$  (domani), rappresenta ciò che di incognito e imprevedibile esprime la teoria del mercato efficiente, e si può scomporre da  $p_{t+1}$  solo nel momento in cui viene osservato quest'ultimo. Le variabili di un processo stocastico white noise si distribuiscono con media 0 e varianza  $\sigma^2$  (tipicamente 1) e descrivono gli shock informativi che nel processo random walk differenziano i prezzi passati da quelli futuri.

Grafico 3. Reazione degli investitori a notizie di acquisizione della società target.



Fonte: Merger announcements and insider trading Activity: An Empirical investigation, Keown, A. J. & Pinkerton, J. M. (1981)

Il concetto di efficienza dei mercati finanziari può essere compreso osservando il grafico 3, proposto nell'ambito di uno studio di Keown e Pinkerton (1981). E' possibile, infatti, notare l'andamento crescente dei rendimenti anomali cumulati, da cui emerge l'assorbimento graduale delle notizie man mano che ci si avvicina l'ufficialità dell'acquisizione; dopo quest'ultima però, sembra persistere una nuova stabilizzazione dei rendimenti anomali ad un livello superiore dato il processo di accumulazione precedente, questo fenomeno può essere visto nell'ottica di un adattamento e progressivo esaurimento delle nuove informazioni, relativamente a questo case study. La prospettiva

di un graduale aumento dei rendimenti anomali avvenuti prima dell'annuncio di acquisizione (informazione pubblica) mostra come gli investitori non si fossero allineati immediatamente in quanto il grado di conoscibilità della futura notizia di acquisizione (giorno 0 nel grafico) rimaneva in capo agli insider di mercato, che hanno iniziato ad investire in tale titolo aumentandone l'appetibilità e il prezzo secondo la legge domanda-offerta. L'andamento dei residui del modello di mercato quindi suggerisce la presenza di un mercato efficiente in forma semi-forte, dove i prezzi incorporano le informazioni pubbliche e sui prezzi passati.

## **1.2 Mercato inefficiente: l'interpretazione "behavioural prone"**

Alla teoria del mercato efficiente si contrappone una visione per alcuni aspetti più pragmatica, proposta da alcuni esponenti della finanza comportamentale; secondo questa teoria le notizie di mercato inaspettate vengono assorbite in modo non coerente dagli investitori a causa di loro over-reaction che generano anomalie ricorrenti, e perciò prevedibili, nel mercato.

La base su cui si fonda questo tipo di dottrina, che unisce la psicologia di scelta all'ambito economico-finanziario, presuppone che gli individui assumano decisioni d'investimento in situazioni in cui gli stessi sono influenzati da diversi fattori, non come viene invece proposto da modelli normativi come quello basato sulla massimizzazione dell'utilità attesa, dove le scelte economiche vengono fatte in base all'avversione al rischio dell'individuo sulla base di pura razionalità.

"Se i prezzi delle azioni superano sistematicamente i limiti, la loro inversione dovrebbe essere prevedibile solo sulla base dei dati sui rendimenti passati, senza l'utilizzo di dati contabili come gli utili. Nello specifico, vengono suggerite due ipotesi: i movimenti estremi dei prezzi delle azioni saranno seguiti da successivi movimenti dei prezzi nella direzione opposta. Quanto più estremo è il movimento iniziale del prezzo, tanto maggiore sarà il successivo aggiustamento. Entrambe le ipotesi implicano una violazione dell'efficienza del mercato in forma debole" (tradotto da De Bondt e Taler, 1985, p.795).

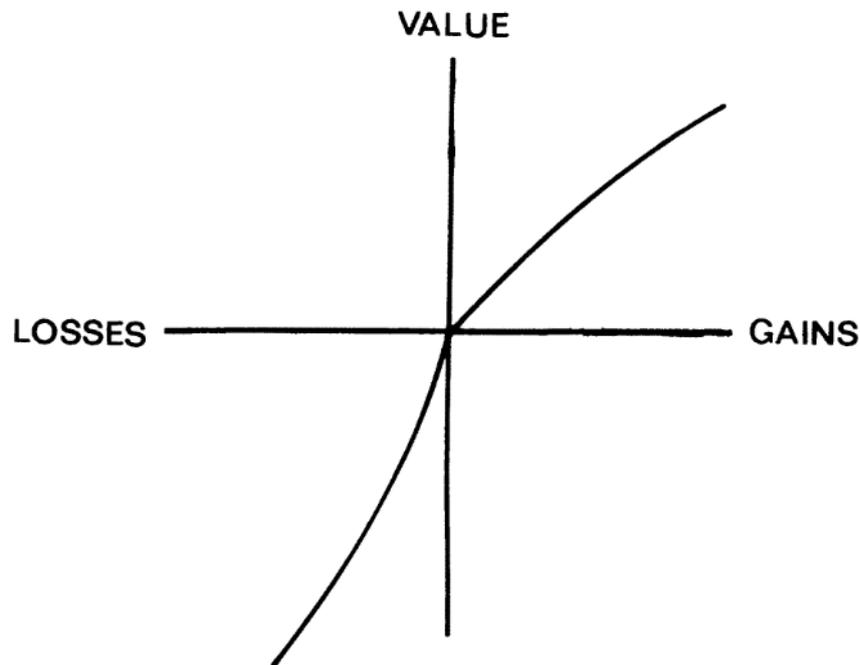
Il pensiero di De Bondt e Taler mostra come, secondo la teoria comportamentista, i mercati non raggiungono una forma di efficienza, perciò è possibile realizzare extra-

rendimenti sfruttando delle anomalie in quanto gli investitori non accedono tutti nello stesso modo alle notizie, oltre che ad essere influenzati da bias cognitivi che limitano un'interpretazione prettamente razionale delle scelte economiche, come invece suggerito dalla teoria dell'utilità attesa. Secondo questa linea di pensiero è quindi sensato elaborare tecniche basate su modelli predittivi che possano anticipare i movimenti del mercato in base ai comportamenti che, seppur irrazionali, seguono uno specifico pattern come vedremo. La discriminazione informativa permette a un determinato gruppo di investitori di accedere alle informazioni e utilizzarle per battere il mercato; in questo caso quindi, a differenza della teoria del mercato efficiente, i prezzi non scontano tutte le informazioni e per questo non rappresentano il valore fondamentale del titolo. È possibile acquistare uno stock sottovalutato/sopravvalutato e generare un profitto dal suo futuro aumento/diminuzione di prezzo, ipotizzando che il prezzo tenda al fair value nel futuro.

### **1.2.1 La teoria dei prospetti**

La teoria dei prospetti (Kahneman e Tversky, 1979) rappresenta l'evoluzione nel campo della finanza comportamentale della cosiddetta teoria normativa dell'utilità attesa; questa teoria è di tipo descrittivo, perché va a spiegare i comportamenti non razionali che l'individuo mette in atto quando è di fronte a delle scelte economiche. Al contrario delle teorie normative e dei modelli costruiti sull'ipotesi di mercati efficienti questa scuola di pensiero non prescinde da bias cognitivi che influenzano i processi decisionali degli individui, che interpretano le scelte economiche in base ad un reference point di partenza: "il reference point corrisponde solitamente alla situazione patrimoniale corrente, in cui gli utili e le perdite coincidono con gli importi effettivi ricevuti o pagati. Tuttavia, la posizione del reference point, e la conseguente codifica dei risultati come guadagni o perdite, possono essere influenzati dalla formulazione delle prospettive offerte e dalle aspettative del decisore" (Kahneman e Tversky, 1979, p.12). Il piano cartesiano, secondo questa visione, viene diviso in due parti in base al "punto di partenza" che è soggettivo per ogni individuo e dipende altresì dalla ricchezza di partenza e di quanto l'eventuale guadagno conseguito vada ad incrementare il patrimonio.

Grafico 4. La prima funzione del valore individuata da Kahneman e Tversky



Fonte: The prospect theory, Kahneman e Tversky (1979)

Come è evidente dal grafico 4, il comportamento assunto dagli individui differisce a seconda che ci si trovi di fronte ad un guadagno o ad una perdita. La funzione del valore assume una forma concava nel dominio dei guadagni, il che esprime che gli individui, quando si trovano in una situazione di guadagno tendono a diventare maggiormente avversi al rischio poiché all'aumentare del guadagno il valore attribuito all'esito aumenta meno che proporzionalmente. Nel dominio delle perdite gli individui tendono ad attribuire un peso maggiore all'evento negativo (lo allontana dal reference point) e infatti la funzione diventa più ripida e di forma convessa, che rappresenta una maggiore propensione al rischio per cercare di tornare al punto iniziale.

$$V(x) = \begin{cases} x^{0,88} & \text{se } x \geq 0 \\ -2,25(-x)^{0,88} & \text{se } x < 0 \end{cases}$$

(1.1)

L'espressione 1.1 conferma quanto detto in precedenza: viene attribuito un valore maggiore alle situazioni in perdita. Come emerge dall'equazione, se  $x$  è minore di zero il

valore del risultato negativo viene ulteriormente "aggravato" dall'individuo e questo viene espresso dal fattore moltiplicativo di -2,25. L'esponente minore di uno invece esprime la non proporzionalità dell'aumento del valore attribuito dal singolo rispetto a guadagni/perdite ricorrenti.

### 1.2.2 Le anomalie di mercato

Le cosiddette anomalie di mercato sono pattern di rendimento ripetuti in cui si può trovare una spiegazione comportamentale e su cui si basa la stessa branca dell'economia per dimostrare le inefficienze del mercato; questi fenomeni rendono il mercato meno imprevedibile di quanto presentato dalla teoria del mercato efficiente di Fama.

Di seguito verranno presentate le anomalie di mercato più rilevanti.

- *Il January Effect*

"Il January effect è una delle anomalie che sembrano essere incoerenti con l'ipotesi del mercato efficiente. Se il rendimento delle azioni a gennaio è effettivamente superiore a quello del resto dell'anno, secondo l'EMH gli investitori dovrebbero iniziare a comprare azioni a dicembre e rivenderle alla fine di gennaio, annullando efficacemente il January effect" (Kampman, 2012).

Nonostante la presenza di questo tipo di inefficienza e la possibilità da parte degli investitori di assorbirla, il fenomeno sembra persistere. Come riportato da Rozeff e Kinney (1976) il mese di gennaio rappresenta l'inizio e la fine di importanti eventi finanziari e informativi, come l'inizio dell'anno fiscale per gli investitori e l'inizio dell'esercizio contabile per molte imprese, nonché il periodo durante il quale vengono annunciati gli utili conseguiti dalle società dell'esercizio precedente. Alcune delle spiegazioni di questo fenomeno sono state fornite da Hiller e Marshall (2002): gli scambi a gennaio sono maggiori per via del ribilanciamento dei portafogli da parte dei gestori; gli investitori invece tendono a riacquistare i titoli dopo averli venduti a dicembre per recuperare l'eventuale credito fiscale dovuto a minusvalenze (tax-lossing effect e turn of the year effect).

- *Il rapporto price/earnings*

Per monitorare questa anomalia viene utilizzato il rapporto P/E, che è uno degli indicatori più utilizzati in dottrina per valutare la redditività di un'impresa nel titolo che posiziona nel mercato per finanziarsi. Questo rapporto indica la capacità dell'impresa, espressa dai dividendi distribuiti per unità di azione, di rimborsare l'investimento in termini di costo attuale dell'azione. Basu (1977), utilizzando questo indicatore, individua performance migliori per imprese con basso P/E ipotizzando quindi che il mercato prezzi erroneamente queste realtà. I titoli con un P/E basso vengono chiamati "value" e sono riconosciuti dagli analisti di mercato come titoli poco performanti in termini di profittabilità futura e di crescita; le imprese con questo tipo di aspettative future operano solitamente in ambienti di mercato maturi. Vista la possibile sottovalutazione di questa tipologia azionaria, la durata senza precedenti della fase di overperformance del fattore "growth", (che invece rappresenta valutazioni elevate e forti aspettative di crescita) dal 2007 al 2020 e la reale possibilità che i tassi d'inflazione nei prossimi anni saranno più alti che nel passato (anche se ad oggi le prospettive della BCE sono quelle di un abbassamento graduale dei tassi d'interesse dato che l'inflazione si sta sgonfiando), potrebbe essere possibile sfruttare questa inefficienza del mercato andando ad acquistare i titoli con un P/E basso che in questo momento hanno una valutazione underperforming e per questo sottovalutati.

- *Small firm effect*

In ultima istanza verrà preso in considerazione il cosiddetto "small firm effect", che presuppone che l'ammontare di informazioni dipenda dalla dimensione dell'impresa in oggetto (Banz, 1981) e che, secondo il CAPM (Capital Asset Pricing Model), le "small firms" abbiano rendimenti aggiustati per il rischio più alti in media rispetto alle "large firms". Nello studio inoltre, le considerazioni fatte da Basu relativamente al P/E ratio vengono legate allo small firm effect, in quanto secondo Banz il P/E ratio è una proxy della dimensione di un'impresa e non viceversa, e perciò esprimerebbe in altri termini questo effetto.

"L'evidenza indica che quasi il cinquanta per cento della media del premio corretto per il rischio delle piccole imprese rispetto alle grandi imprese in questo periodo è dovuto ai rendimenti anomali di gennaio. Inoltre, oltre il ventisei per cento del premio dimensionale

è attribuibile a rendimenti anomali durante la prima settimana di negoziazione dell'anno e quasi l'undici per cento è attribuibile al primo giorno di negoziazione. I dati non rivelano un comportamento stagionale significativo in nessun altro mese.”

Keim (1983) in questo inciso attribuisce l'eccesso di rendimento delle small firms ai rendimenti anomali di gennaio (January effect), con la maggiore influenza dei rendimenti nella prima settimana del mese.

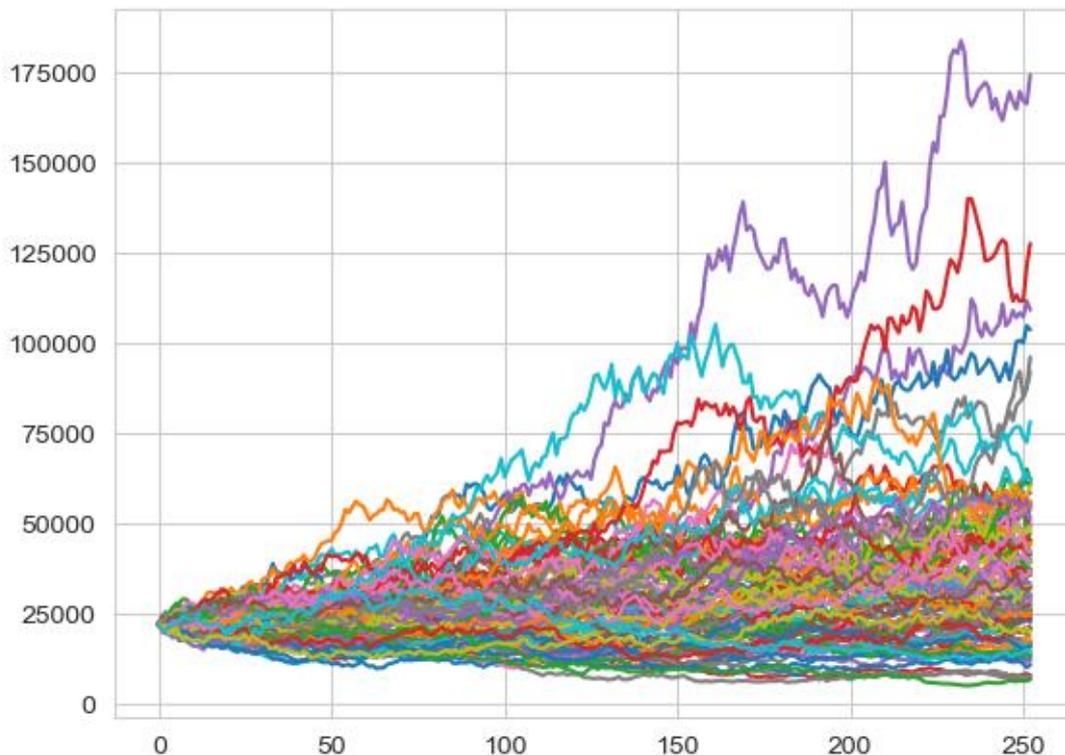
## Capitolo I – Concetti di statistica e storia delle previsioni

Conclusa l'introduzione sul mercato finanziario e le sue diverse interpretazioni, assume ora importanza andare a definire quali siano stati, nel corso della storia, i modelli predittivi funzionali per poter ragionare sul possibile andamento futuro del mercato finanziario in generale e in particolare su un determinato valore di stock, ma prima di fare ciò è importante definire alcuni concetti fondamentale nell'ambito della statistica finanziaria.

### 2.1 Cos'è un processo stocastico

“Un processo stocastico è la successione di variabili aleatorie con la quale si rappresenta un sistema che si sviluppa nel tempo e nello spazio secondo leggi probabilistiche. Un processo stocastico è solitamente indicato con il simbolo  $\{X_t; t \in T\}$  dove  $X_t$  è la generica variabile aleatoria della successione, cioè la variabile che descrive i possibili valori al tempo  $t$ , e  $T$  è un insieme discreto o continuo” (Treccani, 2013).

Grafico 5. Simulazione Montecarlo del prezzo di Bitcoin ipotizzando che sia generato da un processo random walk



Questa simulazione dà l'idea di cosa possa essere un processo stocastico finanziario nel concreto. In questo caso è stato selezionato uno stock di Bitcoin e, dopo aver importato i dati degli ultimi tre anni di serie storica di questa criptovaluta è stato calcolato il valore atteso della variabile casuale rendimento e la sua varianza, dimodoché i primi due momenti della distribuzione possano "guidare" i sentieri possibili dei prezzi in un intervallo successivo di 252 giorni di trading (un anno solare); il numero di simulazioni per motivazioni prettamente computazionali sono state limitate a mille, il che porta questa rappresentazione ad essere un sotto-insieme del processo stocastico vero e proprio che rimane un concetto di riferimento teorico.

## 2.2 Proprietà di un processo stocastico

### 2.2.1 La stazionarietà

La stazionarietà è un elemento fondamentale per poter effettuare analisi predittive in modo efficace, poiché permette di assumere che la variabile casuale oggetto della previsione o dello studio sul processo generatore delle sue realizzazioni (GDP) possa avere le stesse caratteristiche nel tempo e questo è l'unico modo per adattarne i modelli statistici e fare inferenza sui dati a disposizione.

- Un processo stocastico  $y(t)$  si dice stazionario in senso forte se:

$$F(y_{t_1} < y_1, \dots, y_{t_k} < y_k) = F(y_{t_1+\tau} < y_1, \dots, y_{t_k+\tau} < y_k) \quad \forall (t+1, t+2, \dots, t+k) \in T \text{ e per } \forall k, \tau \in Z$$

La condizione sopracitata presuppone che tutti i momenti della distribuzione della variabile casuale non dipendano dal tempo e siano perciò costanti.

Solitamente gli studiosi svolgono le analisi statistiche sui rendimenti piuttosto che sui prezzi; questo a causa della suddetta stazionarietà dei rendimenti rispetto ai prezzi (che hanno un andamento tipicamente non stazionario), il cui GDP come visto in precedenza viene rappresentato adeguatamente da un processo random walk. La caratteristica di randomness dei prezzi non ne permette un corretto studio in quanto il prezzo di uno stock è spiegato da una componente legata al prezzo del giorno precedente e dalle notizie di

mercato che difficilmente sono pronosticabili, mentre i residui del modello rappresentano ciò che quest'ultimo non è riuscito a prevedere; questo tipo di processo è un caso specifico del modello auto-regressivo di ordine 1, che verrà presentato in seguito.

Dato che è difficile verificare la non stazionarietà forte è stato introdotto il concetto di stazionarietà in senso debole.

La stazionarietà in senso debole (o in senso lato) comprende vincoli sui primi due momenti della distribuzione delle variabili casuali del processo stocastico, il valore atteso ( $\mu_t$ ) e la varianza ( $\sigma^2_t$ ) che sono costanti e non dipendono dal tempo, mentre l'auto covarianza, che è data dalla covarianza tra due variabili aleatorie appartenenti allo stesso processo stocastico distanziate temporalmente da uno sfasamento pari a  $k = t_2 - t_1$ , dipende solo da  $k$  e quindi dalla distanza temporale delle due variabili casuali.

I test d'ipotesi più utilizzati per verificare la presenza o l'assenza di tale caratteristica sono i test di Dickey-Fuller e di Phillip-Perron.

### 2.2.2 L'invertibilità e l'ergodicità

Oltre alla stazionarietà ci sono altre proprietà che, se verificate, permettono di fare inferenza sul GDP a partire dalla serie storica osservata: l'invertibilità e l'ergodicità.

- Un processo stocastico si può definire invertibile se può essere espresso in funzione dei suoi valori passati e di un processo white noise:

$$Y_{(t)} = f(Y_{(t-1)}, Y_{(t-2)}, \dots) + \alpha_{(t)}$$

dove  $\alpha_{(t)}$  è il processo white noise che può essere stimato come differenza tra  $Y_{(t)}$  e i suoi valori passati

- Un processo stocastico viene definito ergodico rispetto ad una funzione momento  $E[v(Y_t)]$  se la media temporale di tale funzione  $v(Y_t)$  converge in probabilità alla funzione momento  $E[v(Y_t)]$ :

$p \lim_{n \rightarrow +\infty} n^{-1} \sum v(Y_t) = E(v(Y_t))$  dove al crescere della dimensione campionaria il momento calcolato sulla serie storica converge in probabilità allo stesso valore a cui tende il valore atteso delle singole variabili casuali, perciò la media delle

realizzazioni di una variabile casuale, se il processo è ergodico, tende al valore atteso della distribuzione da cui provengono le variabili stesse.

### 2.3 L'exponential smoothing

La prima vera definizione di metodo predittivo a filtraggio ricorsivo fu fornita da Robert Brown nel 1956 e da Charles Holt nel 1957, a dimostrazione che il mondo delle previsioni delle serie storiche economico-finanziarie ha un'origine che si spinge alla metà del '900 ma che ancora oggi è soggetta a revisioni e adattamenti in base alle nuove tecnologie, che vengono utilizzate per migliorare e rendere attuali i modelli pensati nel passato e per elaborarne di più efficaci.

Come definito da Holt (1957), il cosiddetto "exponential smoothing" è un metodo che presuppone, con lo scopo di prevedere il valore futuro di una variabile (si cercò di prevedere il valore futuro delle vendite di un determinato bene non specificato), di parametrare il peso di un'osservazione passata in base all'influenza che può avere nel futuro. Questo metodo non è altro che una media mobile ponderata esponenzialmente, dove il valore della media al tempo  $t+1$  si ottiene andando a calcolare la media ponderata di due componenti, la media al tempo  $t$  e il valore della serie storica al tempo  $t$ . Questo metodo permette di effettuare una previsione a breve termine che si può facilmente dimostrare essere uguale al valore dell'ultima osservazione a cui si aggiunge l'ultimo errore di previsione moltiplicato per il parametro di lisciamiento.

#### 2.3.1 Il simple exponential smoothing (SES)

Il SES è ottimale nel caso in cui si voglia effettuare una previsione ad un passo in assenza di trend e di stagionalità.

Prendiamo ad esempio il semplice caso in cui la media mobile delle vendite di uno stock dipenda dalla previsione calcolata al periodo precedente e dall'ultimo valore osservato

$$\bar{S}_t = B * S_{t-1} + (1 - B) * \bar{S}_{t-1}$$

(1. 2)

dove:

- $\bar{S}_t$  è la previsione del prossimo valore della serie storica secondo il modello exponential smoothing
- $S_{t-1}$  sono le vendite osservate al tempo t-1
- $\bar{S}_{t-1}$  sono le vendite stimate al tempo t-1
- $B$  è una costante tra 0 e 1 e rappresenta il fattore di proporzionalità nel prezzo oggi che non dipende dalle informazioni passate ed è pari a:

$$B = \frac{\bar{S}_t - \bar{S}_{t-1}}{S_{t-1} - \bar{S}_{t-1}}$$

(1.3)

Nel caso in cui la media della distribuzione subisse forti shock, la componente passata dovrebbe conseguentemente pesare di meno, poiché non è stata capace di prevedere lo shock stesso (e l'errore di previsione sarebbe alto), e  $1-B$  dovrebbe restituire un valore basso per ridurre l'impatto della media precedente nel calcolo della nuova media della distribuzione.

L'intuizione che sta alla dietro alla formula si basa sul mantenere una componente legata all'andamento della serie storica, espressa dalle previsioni nei periodi precedenti (ponderata esponenzialmente per un fattore  $1-B$ ), e una legata alle nuove informazioni fornite dall'osservazione corrente.

### 2.3.2 Il metodo esponenziale con trend lineare

Per prevedere il valore futuro di una serie storica in presenza di un trend lineare, e ipotizzando ancora l'assenza di fluttuazioni legate alla stagionalità, è stato proposto da Holt un metodo SES che potesse includere anche la fattispecie del trend.

$$\bar{S}_t = B * S_t + (1 - B) * R_t * \bar{S}_{t-1}$$

(1.4)

dove:

- $R_t$  è la componente aggiustata per il trend e si scompone ulteriormente in:

$$C * \frac{\bar{S}_t}{\bar{S}_{t-1}} + (1 - C) * R_{t-1}$$

(1.5)

dove:

- $C$  è un parametro di liscio esponenziale e determina quanto velocemente i pesi esponenziali applicati all'indice del trend decrescono col passare dei periodi:

$$C = \frac{R_t * \bar{S}_{t-1} - R_{t-1}}{\bar{S}_t - R_{t-1}}$$

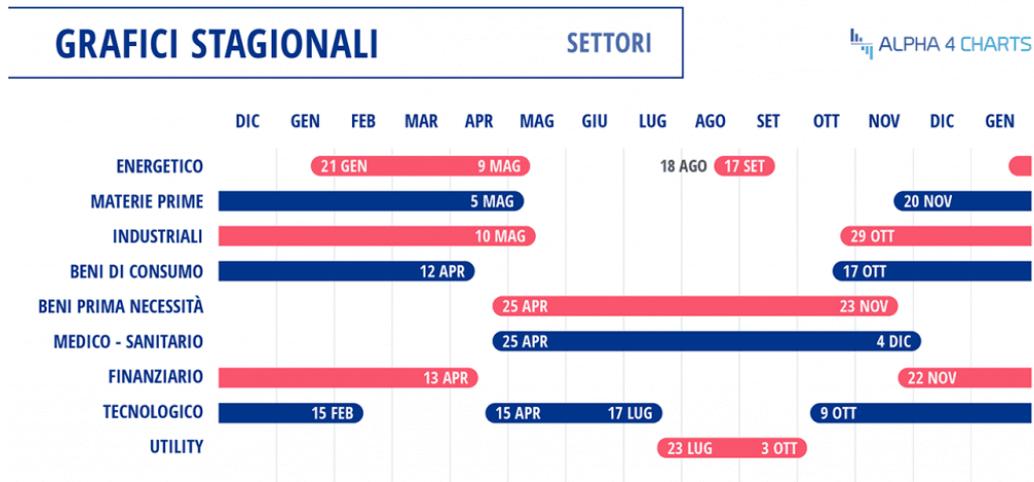
(1.6)

A differenza del simple exponential smoothing, il metodo di Holt con trend lineare include la variabile  $R_t$ , che rappresenta la componente del trend. In questo caso, la previsione sulla media  $\bar{S}_t$  dipende sia dall'influenza dalle previsioni passate  $\bar{S}_{t-1}$  sia dalla presenza di una tendenza nella serie (che possa essere crescente, con massimi e minimi relativi crescenti o viceversa). Tale componente viene espressa combinando l'indice trend corrente con l'indice del precedente.

### 2.3.3 Il metodo esponenziale con stagionalità

La stagionalità viene riconosciuta come la caratteristica di alcuni fenomeni osservati nel tempo, che si evidenzia sotto forma di fluttuazioni a cadenza periodica più o meno regolare; queste possono essere legate a cambiamenti meteorologici, al calendario, o a particolari tempi decisionali (Treccani). In virtù di questa caratteristica delle serie storiche finanziarie (e non solo) è possibile formulare previsioni attraverso i modelli predittivi in grado di catturare tale variabilità.

Grafico 6. Stagionalità settoriale



Fonte: <https://charts.alpha4all.it>

Come emerge dal grafico 6, per il settore finanziario è presente una stagionalità cross-the-year che va da novembre ad aprile e, senza un adattamento, i modelli presentati precedentemente non permettono di intercettarla.

Se ipotizziamo l'assenza di un trend e un solo lag, la logica di costruzione del modello è la medesima del caso semplice, con l'aggiunta di una componente che ha l'obiettivo di intercettare questo fattore di variabilità e di renderlo neutrale (o destagionalizzato).

$$\bar{S}_t = A * P_t * S_{t-1} + (1 - A) * \bar{S}_{t-1}$$

(1.7)

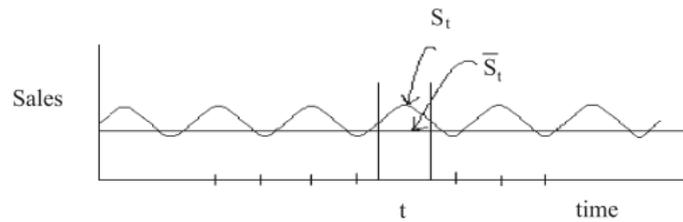
dove (vedi modelli precedenti):

- $A$  rappresenta il parametro di "smoothing" esponenziale che esprime come i pesi decrescono dopo ogni periodo, ed è compreso tra 0 e 1
- $P_t$  rappresenta la componente di aggiustamento stagionale e si esprime combinando il livello corrente di vendite con quello aggiustato per la stagionalità di un anno prima:

$$B * \frac{\bar{P}_t}{P_t} + (1 - B) * P_{t-N}$$

(1.8)

Grafico 7. Andamento di una variabile casuale soggetta a stagionalità



### 2.3.4 Generalizzazione dell'exponential smoothing: fattore additivo e moltiplicativo

I modelli con presenza di diversi tipi di trend e stagionalità rappresentano la base per la previsione di serie storiche più complesse che sono state riprese dettagliatamente da Pegels, che integra nove modelli considerando come possono interagire le componenti di trend e stagionalità nell'exponential smoothing.

Pegels (1969) imposta il suo intervento presentando una formula standardizzata da cui poi si possono derivare modelli in base alla fattispecie in oggetto:

$$\bar{S}_t = w_0 * d_{S,t} + (1 - w_0) * d_{T,t} \tag{1.9}$$

dove:

- $\bar{S}_t$  è la componente legata alla previsione sulle vendite nel periodo t
- $w_0$  rappresenta i pesi esponenziali attribuiti in base alla lontananza delle osservazioni precedenti
- $d_{S,t}$  è la componente influenzata dal tipo di stagionalità mentre  $d_{T,t}$  è la componente che varia in base al tipo di trend presente. Questi parametri rappresentano fattori variabili in base alla presenza e ai tipi di tendenza e stagionalità spiegati da sotto-funzioni come  $A_t, B_t, G_t, Z_t$ .

$$A_t = w_A * (\bar{S}_t - \bar{S}_{t-1}) + (1 - w_A) * A_{t-1} \tag{2.0}$$

dove:

- $A_t$  è la sotto-funzione che esprime la componente trend additiva (quindi che interviene andando a sommare alla serie storica originale)
- $w_A$  è la ponderazione del trend attribuita alla differenza tra la previsione precedente e quella attuale (come si è mossa la serie storica nell'ultimo periodo), mentre la ponderazione complementare  $(1 - w_A)$  esprime quanto pesa il trend individuato nel periodo precedente.

$$B_t = w_B * \frac{\bar{S}_t}{S_{t-1}} + (1 - w_B) * B_{t-1}$$

(2.1)

dove:

- $B_t$  è la sotto-funzione che esprime la componente trend moltiplicativa (quindi che interviene andando a moltiplicare l'effetto del trend)
- $w_B$  è la ponderazione del trend attuale (come si è mossa la serie storica nell'ultimo periodo), mentre la ponderazione complementare  $(1 - w_B)$  esprime quanto pesa il trend individuato nel periodo precedente.

$$G_t = w_G * (\bar{S}_t - S_t) + (1 - w_G) * G_{t-L}$$

(2.2)

dove:

- $G_t$  è la sotto-funzione che rappresenta la componente stagionale additiva
- $w_G$  è la ponderazione del fattore stagionale corrente mentre  $1 - w_G$  esprime la ponderazione dell'indice stagionale della stessa stagione ma dell'anno precedente ( $G_{t-L}$ ).

$$F_t = w_F * \frac{S_t}{\bar{S}_t} + (1 - w_F) * F_{t-L}$$

(2.3)

dove:

- $F_t$  è la sotto-funzione che rappresenta la componente stagionale moltiplicativa

- $w_F$  è la ponderazione del fattore stagionale corrente mentre  $1 - w_F$  esprime la ponderazione dell'indice stagionale della stessa stagione ma dell'anno precedente ( $F_{t-L}$ ).

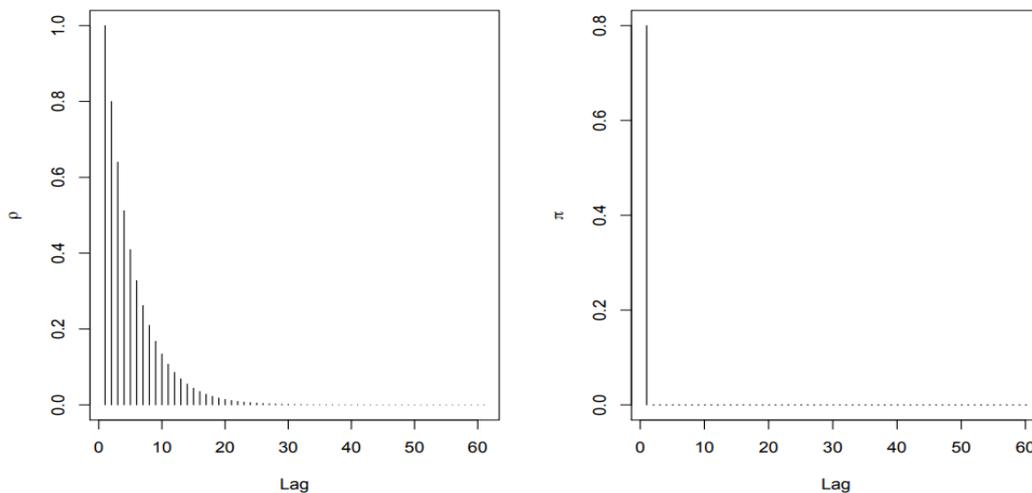
I parametri proposti da Holt e Pegels vengono utilizzati per espandere il concetto di exponential smoothing in base al tipo di particolarità che assume una serie nelle sue realizzazioni storiche. La presenza di trend e stagionalità implica una intrinseca non stazionarietà nelle serie storiche, in quanto le caratteristiche osservabili della variabile casuale cambiano sistematicamente nel tempo; il problema che sorge è legato al fattore di ponderazione (nei modelli proposti da Holt assume la nomenclatura "A", "B", "C" mentre nei modelli di Pegels è "w") dei valori passati di stagionalità, trend e osservazioni storiche. Estraniandosi dalla variabile casuale utilizzata per proporre questi modelli e trattando le serie storiche dei prezzi dei titoli finanziari, si pone il problema di come questi parametri di "smoothing", che hanno l'obiettivo di differenziare il peso attribuito a determinate caratteristiche della serie storica in base alla loro distanza nel tempo, possano rendere il modello predittivo sostenibile e credibile a priori. Nel concetto di media mobile esponenziale, infatti, questi pesi differiscono in base alla capacità delle informazioni passate di prevedere quelle future. Se ipotizziamo un mercato efficiente dove i prezzi scontano tutte le informazioni e non dipendono dalle informazioni passate, questo tipo di modello predittivo difficilmente può risultare efficace poiché non si è in grado di estrapolare un grado di previsione in base ad una struttura di dati storico-finanziari ad alta frequenza. Nel caso in cui si dovessero scegliere dei valori bassi del parametro di "lisciamiento", questo impedirebbe al modello predittivo di adattarsi rapidamente agli shock della serie, rendendo poco efficace le stime; il focus di questa tesi in ambito finanziario impedisce di considerare affidabile tale metodo predittivo poiché le serie dei prezzi variano molto più rapidamente di come i parametri dei modelli esponenziali possano fare (e che dovrebbero essere aggiustati manualmente), oltre che a tipi di stagionalità più complessi e difficilmente intercettabili utilizzando modelli parametrici di questo tipo.

## 2.4 I modelli ARIMA

I modelli auto-regressive integrated moving average, la cui diffusione è dovuta al successo del metodo fornito da Box e Jenkins (1970) per l'analisi statistica, sono metodi di approssimazione del processo generatore di dati i cui parametri, in base al tipo di caratteristiche della serie storica, vanno ad assumere valori differenti; i parametri del modello sono scelti in base alla struttura di autocorrelazione tra le osservazioni, la quale esprime il grado di dipendenza tra i valori assunti dal processo stocastico nel corso del tempo. Poniamo il caso in cui la struttura di autocorrelazione mostri un risultato significativo nei primi due ritardi della variabile casuale; ciò significa che quest'ultima ne è influenzata, perciò verrà costruito un modello statistico di approssimazione del GDP considerando solo i primi due parametri di ritardo.

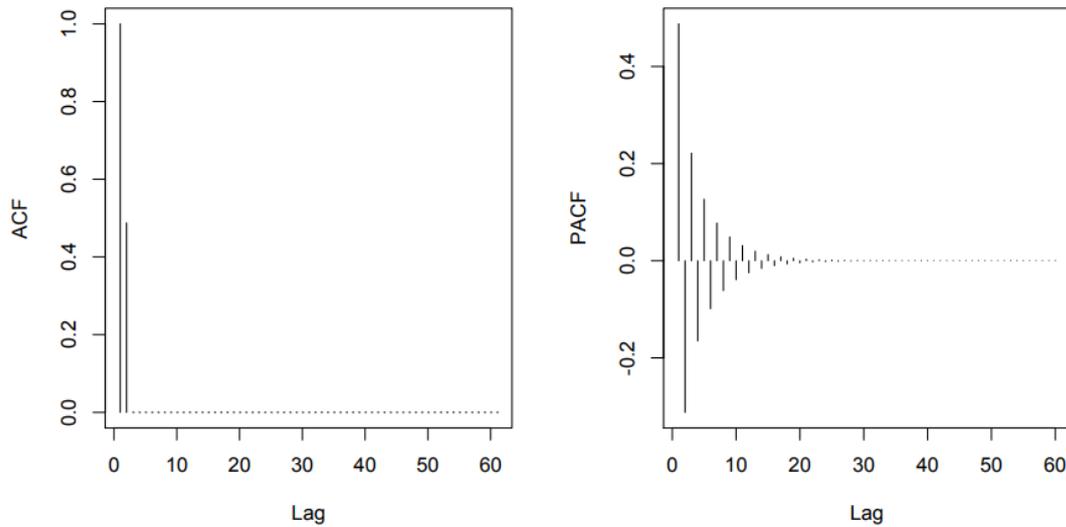
Questo tipo di approccio si adatta maggiormente a pattern di dati che assumono comportamenti non stazionari, come possono essere le serie finanziarie dei prezzi.

Grafico 8. Funzioni di ACF e PACF di un processo auto-regressivo di ordine uno AR (1)



La componente auto-regressiva, qualora dovesse essere verificata, esprime la capacità delle osservazioni di essere spiegate da osservazioni precedenti, in base a dei parametri di ponderazione; questa fattispecie può essere identificata dalla significatività dei ritardi nella funzione di autocorrelazione parziale, che verifica la dipendenza tra le diverse variabili andando ad eliminare le componenti di disturbo delle variabili intermedie.

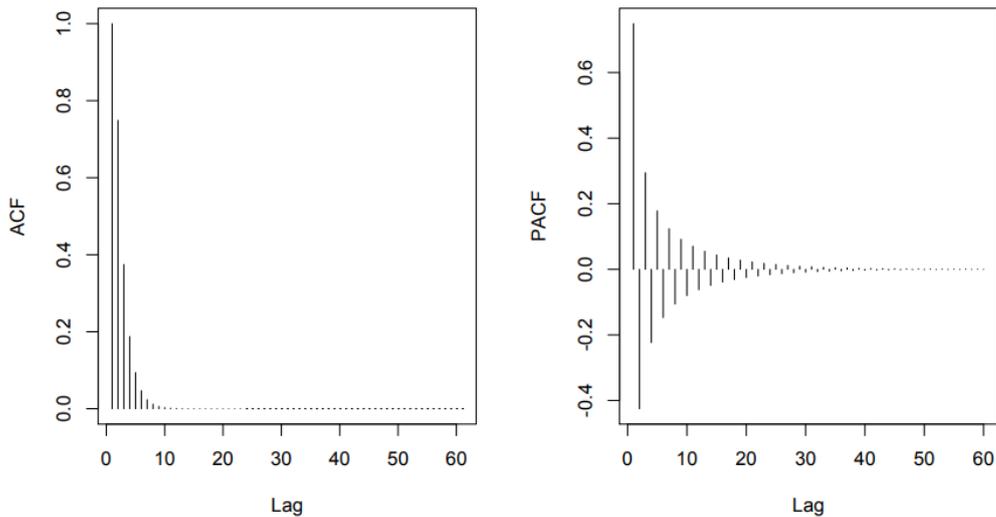
Grafico 9. Funzioni di ACF e PACF di un processo a media mobile di ordine uno MA(1)



La componente a media mobile invece, spiega il comportamento dell'osservazione corrente in base alle componenti casuali (white noise) passate; si ipotizza infatti che il GDP possa essere approssimato utilizzando un modello statistico dove il comportamento dei rendimenti dipende dalle innovazioni passate  $\varepsilon_{t-n}$ . I ritardi significativi del modello MA possono essere individuati attraverso la funzione di autocorrelazione.

Il limite dei modelli a media mobile è di non avere la capacità di prevedere oltre all'ordine del modello stesso specificato, poiché man mano che il numero di passi in avanti della previsione cresce c'è bisogno delle innovazioni future, che non disponibili al tempo  $t$ . Inoltre, è possibile che la stima dei parametri del modello, considerando le osservazioni della finestra temporale precedente, possano non risultare validi per i nuovi dati, risultando quindi inadeguati per fare previsioni.

Grafico 10. Funzioni di ACF e PACF di un processo misto ARMA (1,1)



È possibile inoltre, dopo un'attenta analisi delle autocorrelazioni e un adeguato studio di significatività dei parametri stimati, identificare un processo misto che possa incorporare entrambe le caratteristiche sopracitate.

I modelli ARIMA derivano dai modelli ARMA, e la dicitura "integrated" indica la natura non stazionaria dei processi, ed esprime la possibilità di descrivere la serie originale differenziandola di uno o più ordini per poterla rendere stazionaria e manipolabile (tipicamente in finanza viene applicata una differenziazione in scala logaritmica per utilizzare, a partire dai prezzi di un titolo, i suoi rendimenti).

Questa metodologia viene utilizzata principalmente per identificare modelli per la media dei rendimenti, il cui scopo è quello di parametrizzare la ponderazione dei coefficienti delle componenti auto-regressive e a media mobile del processo generatore dei dati al fine di poterli utilizzare per scopi predittivi. In base al modello individuato, infatti, si possono effettuare previsioni a breve termine:

$$\widehat{r_{t+1}} = E[r_{t+1}|I_t] \tag{2.4}$$

L'equazione esprime la previsione ad un passo come il valore atteso della realizzazione condizionata alle informazioni disponibili al tempo t.

Il secondo membro dell'equazione poi può essere riscritto sostituendo  $r_{t+1}$  con il modello identificato basato sulla procedura Box-Jenkins: ad esempio si ipotizzi di utilizzare un

modello ARMA (2,1) per rappresentare il GDP della media dei rendimenti. Il modello comprende due parametri auto-regressivi e uno legato alla componente di media mobile. Il previsore può essere scritto come segue:

$$\widehat{r_{t+1}} = E[\theta_1 * r_t + \theta_2 * r_{t-1} + \theta_3 * \varepsilon_{t-1} + \varepsilon_t | I_t] \quad (2.5)$$

per le proprietà dell'operatore valore atteso:

$$\widehat{r_{t+1}} = E[\theta_1 * r_t + \theta_2 * r_{t-1} | I_t] + E[\theta_3 * \varepsilon_{t-1} + \varepsilon_t | I_t] \quad (2.6)$$

scomponendo ulteriormente si ottiene:

$$\widehat{r_{t+1}} = E[\theta_1 * r_t | I_t] + E[\theta_2 * r_{t-1} | I_t] + E[\theta_3 * \varepsilon_{t-1} + \varepsilon_t | I_t] + E[\varepsilon_t | I_t] \quad (2.7)$$

possono essere estratti dall'operatore valore atteso i parametri  $\theta_1, \theta_2, \theta_3$  in quanto costanti:

$$\widehat{r_{t+1}} = \theta_1 * E[r_t | I_t] + \theta_2 * E[r_{t-1} | I_t] + \theta_3 * E[\varepsilon_{t-1} + \varepsilon_t | I_t] + E[\varepsilon_t | I_t] \quad (2.8)$$

ipotizzando che i residui siano descritti da un processo white noise con media 0 e varianza  $\sigma^2$  e considerando che  $r_t$  e  $r_{t-1}$  all'istante di tempo t siano noti, la previsione può essere riscritta come:

$$\widehat{r_{t+1}} = \theta_1 * r_t + \theta_2 * r_{t-1} \quad (2.9)$$

Ora che tutti i valori presenti nell'espressione sono noti, possono essere sostituiti e calcolato il valore della previsione al passo n=1.

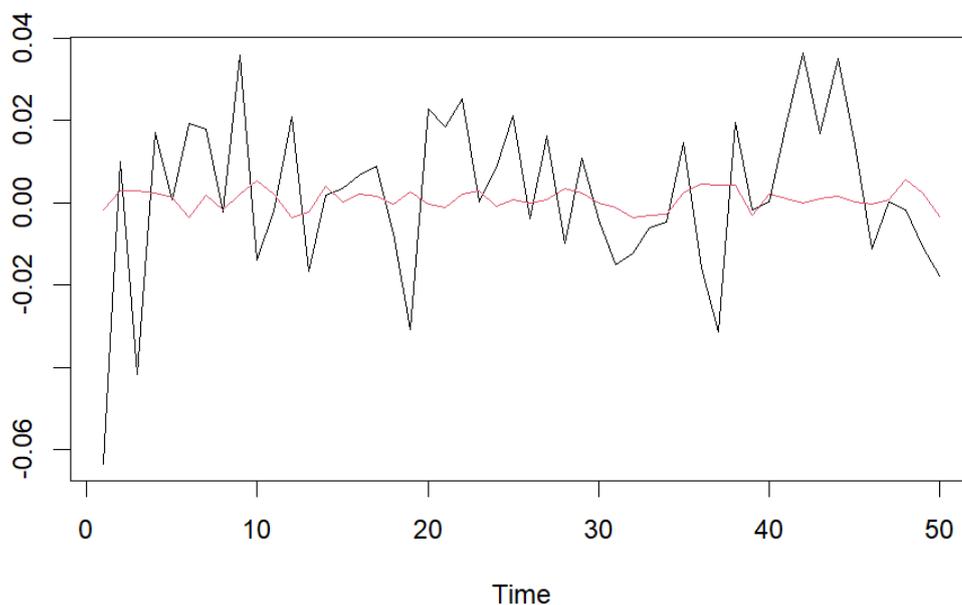
Dall'espressione precedente è evidente che il valore di previsione dipenda da alcuni elementi come la stima dei parametri  $(\theta_1, \theta_2)$ , dalle ipotesi fatte sui residui  $\varepsilon_t$  e dalla specificazione del modello.

La presenza di forte rumore all'interno delle serie storiche, come nel caso di quelle finanziarie, rende questo tipo di modelli difficili da interpretare ed implementare in questo ambito; molto spesso l'analisi rappresenta solamente la prima fase di uno studio

più approfondito, dove vengono utilizzati modelli per la varianza dei rendimenti, la quale risulta analizzabile per via delle sue caratteristiche di persistenza.

A titolo esemplificativo è stato svolto lo studio sui rendimenti del titolo Netflix e, dopo aver validato i dati attraverso opportuni test, aver identificato il modello, i parametri dello stesso e averlo addestrato e validato, per dare una idea della capacità predittiva è stato elaborato un grafico in cui sono stati inseriti i dati osservati e i dati previsti.

Grafico 11. Rendimenti osservati del titolo Netflix (nero) e previsti (rosso)



Per il titolo analizzato è stato identificato un processo AR (5) con i primi quattro ritardi non significativi; in rosso il rumore intercettato dal modello identificato è relativamente basso e spiega poco dell'andamento dei rendimenti, questo dimostra come ricercando pattern di rendimento nel passato questi non possano garantire una spiegazione esaustiva dei pattern futuri. La variabilità dei dati finanziari impone l'utilizzo di un modello che possa spiegare tale fenomeno e senza questa integrazione pensare di stimare questo tipo di dati appare quantomeno irrealistico.

## 2.5 I modelli ARCH

Lo studio relativo ai residui del modello per la media dei rendimenti, seguendo la procedura Box-Jenkins, rappresenta il primo passo per l'identificazione di un modello per la varianza dei rendimenti, infatti è possibile calcolare:

$$\hat{\varepsilon}_t = y_t - \hat{y}_t$$

(3.0)

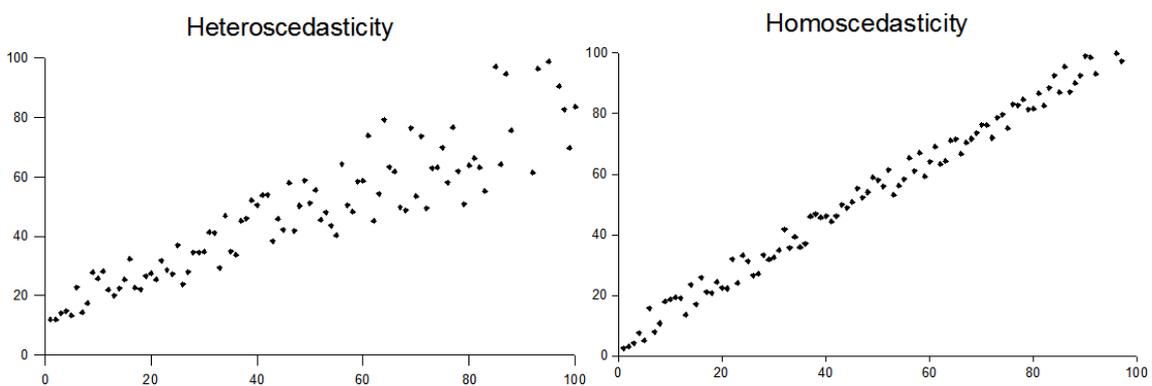
dove:

- $\hat{\varepsilon}_t$  rappresenta i residui del modello sulla media dei rendimenti ( $\hat{y}_t$ ) rispetto ai rendimenti effettivamente osservati ( $y_t$ ).

I residui  $\hat{\varepsilon}_t$  rappresentano la distanza tra il dato osservato e la sua approssimazione, e sono oggetto di uno studio che vada a completare quello fatto sui rendimenti. I modelli ARIMA, che per le serie dei rendimenti solitamente permettono di ottenere un modello lineare per la media del processo, sono insufficienti.

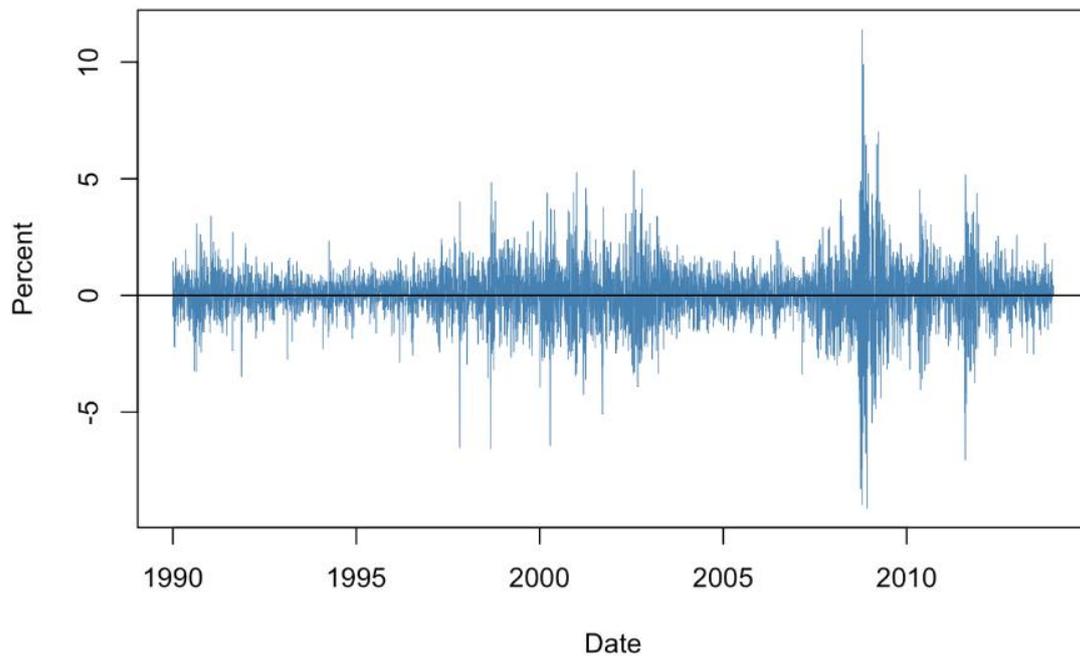
I modelli ARCH (Auto-regressive Conditional Heteroskedasticity) permettono di considerare l'eteroschedasticità condizionata dei residui del modello, in altri termini permettono di modellare il legame tra la variabilità di ogni osservazione con quelle precedenti.

Grafico 12. Andamento di dati soggetti ad omoschedasticità ed eteroschedasticità



Molto frequentemente, nelle serie finanziarie, si verifica il fenomeno del cosiddetto “volatility clustering”, ovvero la presenza di pattern di volatilità che differenziano i periodi storici, i quali si susseguono con caratteristiche di variabilità che sembrano essere spiegabili (Mandelbrot, 1963). Il fenomeno del volatility clustering prevede che si susseguano periodi di volatilità crescente e decrescente, dimostrando delle caratteristiche di persistenza.

Grafico 13. Volatility clustering dell'indice Wilshire 5000



Fonte: <https://www.econometrics-with-r.org/>

Com'è possibile notare dal grafico sono presenti 4 diversi periodi oscillatori (due a variabilità medio-bassa fino alla metà degli anni '90 e dal 2004 al 2008, uno a variabilità media all'inizio degli anni 2000 e uno ad alta variabilità in corrispondenza della crisi finanziaria dei subprime) dell'indice in analisi, facendo intuire un margine di predicibilità di questo fenomeno che presenta caratteristiche di stazionarietà, se il processo viene scomposto in dei sottoperiodi.

Il modello a varianza condizionata, introdotto da Engle nel 1982, presuppone che la riproduzione di un pattern di rendimenti sia rappresentato dalla seguente espressione:

$$r_t = \mu_t + \varepsilon_t$$

(3.1)

dove:

- $\mu_t$  è la componente legata alla media e può essere intercettato e spiegata attraverso una corretta specificazione di un modello ARIMA (ma non fornisce una spiegazione esaustiva del comportamento dei rendimenti).  $\varepsilon_t$  (nella sua formulazione al quadrato) rappresenta una proxy della varianza del processo, che è oggetto di analisi nella fattispecie ARCH. La componente  $\varepsilon_t$  viene scomposta ulteriormente in:

$$\varepsilon_t = h_t * \eta_t$$

e (3.2)

$$h^2_t = \text{var}(r_t|I_{t-1}) = \text{var}(\varepsilon_t|I_{t-1}) = \omega + \alpha * \varepsilon^2_{t-n}$$

(3.2.1)

dove:

- $h_t$  rappresenta la varianza condizionata ai lag precedenti, il fattore che i modelli ARCH hanno l'obiettivo di intercettare
- $\eta_t$  evidenzia le innovazioni dei residui, indipendenti e identicamente distribuiti a media 0 e varianza 1.
- $\omega, \alpha$  e sono i coefficienti del modello ARCH

Successivamente è stato introdotto un altro modello per lo studio della varianza dei rendimenti (Bollerslev, 1986), il modello GARCH, (Generalized Auto-Regressive Conditional Eteroschedasticity) che rappresenta una versione evoluta del precedente ARCH, poiché contiene una componente legata ai valori ritardati della varianza condizionata:

$$h^2_t = \omega + \alpha * \varepsilon^2_{t-n} + \beta * h^2_{t-n}$$

(3.2.2)

Output Rstudio 1. Stima dei parametri per un modello GARCH (1,1) per la volatilità dei rendimenti del titolo Netflix

| Error Analysis: |           |            |         |          |
|-----------------|-----------|------------|---------|----------|
|                 | Estimate  | Std. Error | t value | Pr(> t ) |
| omega           | 3.807e-05 | 2.310e-05  | 1.648   | 0.0993   |
| alpha1          | 4.696e-02 | 2.229e-02  | 2.107   | 0.0351   |
| beta1           | 8.937e-01 | 5.211e-02  | 17.149  | <2e-16   |

Dall'output del procedimento di stima dei parametri GARCH si può notare come i valori dei coefficienti dei ritardi siano trascurabili, sintomo di una difficoltà del modello nel dare una spiegazione della variabilità dei rendimenti in base a valori passati.

Il concetto di rischio-rendimento è alla base di ogni tipo di ragionamento finanziario, e rappresentare un modello che possa prevedere i livelli futuri di oscillazione dei rendimenti e di conseguenza dei prezzi di un titolo, rappresenta un valore aggiunto nell'analisi. Tuttavia, le ipotesi fatte sulle funzioni di autocorrelazione e autocorrelazione parziale per la media e la varianza dei rendimenti come visto in precedenza restituiscono dei risultati che sembrano confermare l'ipotesi di efficienza dei mercati, almeno nella forma debole; il segnale recuperato dai modelli ARIMA e GARCH spesso risulta insufficiente a sviluppare delle previsioni attendibili.

## 2.6 I modelli di machine learning basati su reti neurali

Mentre i modelli visti in precedenza prevedono input sul tipo e sul numero di parametri, le reti neurali sono uno dei possibili metodi non parametrici di apprendimento al servizio del machine learning, utilizzati anche a scopi predittivi.

Il funzionamento dei modelli di reti neurali ha l'obiettivo di replicare la struttura fisiologica del cervello umano, sfruttando poi la capacità computazionale per ampliare le interconnessioni attraverso le elaborazioni automatiche di un sistema informatico. Il focus su questa metodologia di previsione, data la maggiore capacità di adattamento ai costanti cambiamenti nel mondo della finanza, implica la rappresentazione e lo studio approfondito dei modelli che vengono utilizzati in questo ambito.

Il precursore storico di questa tipologia di elaborazione dei dati fu Rosenblatt con il suo lavoro sul Perceptron nel 1958, riconosciuto come il primo modello di rete neurale a classificazione binaria; nell'idea di Rosenblatt le sinapsi biologiche venivano sostituite da dei parametri (o pesi): in base alla funzione di retro-propagazione dell'errore, pensata per minimizzare lo stesso andando a modificare i pesi delle connessioni del modello, questo avrebbe restituito un valore, 0 o 1 in base ad una funzione di conversione.

Una delle prime applicazioni di reti ad apprendimento automatico in ambito economico fu proposta da Swanson e White nel 1997, mostrando in questo campo che “quando utilizziamo i nostri modelli econometrici per includere informazioni disponibili solo in tempo reale, le nostre previsioni superano le previsioni SPF<sup>3</sup> per molte delle variabili, basate sull'errore di previsione quadratico medio, sull'errore di previsione assoluto medio e sulle misure di deviazione assoluta media”.

Dallo studio emerge che il modello di rete neurale artificiale utilizzato si comporta ottimamente con i dati a disposizione; una delle caratteristiche che ha avvantaggiato questa tipologia di sistema predittivo rispetto a quelli basati sui modelli parametrici è la capacità di generalizzare meglio i comportamenti non lineari delle serie storiche finanziarie.

Tra i modelli di reti neurali che hanno ottenuto un discreto successo in ambito finanziario e che verranno approfonditi nel corso del prossimo capitolo si possono segnalare i Long Short Term Memory (LSTM), le Reti Neurali Ricorrenti (RNN) e i Convolutional Neural Network (CNN).

---

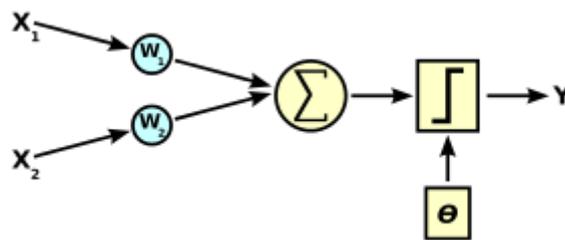
<sup>3</sup> Survey of Professional Forecasters

## Capitolo II – Proprietà delle reti neurali artificiali

Prima di poter utilizzare questo tipo di modelli in ambito previsionale è necessario specificarne struttura, caratteristiche e proprietà.

In precedenza, è stata definita l'origine delle reti neurali in termini storici, il perceptrone: questa è la più semplice modellazione di un network a cui si possa pensare, in cui uno o più input (in cui vengono introiettati i dati) sono collegati ad una funzione di conversione.

Grafico 14. Struttura del perceptrone



Com'è possibile notare dal grafico 14, il modello più semplice di una rete neurale prevede, oltre che agli input e agli output, una funzione di attivazione (in questo caso una somma pesata) in base alle ponderazioni attribuite ai vari dati in ingresso; purtroppo il primo prototipo aveva dei grossi problemi di classificazione, dati i pochi nodi presenti.

Definiamo  $e$  come la funzione di attivazione (step function) del perceptrone e ipotizziamo che si possano verificare due condizioni in base al risultato della somma pesata di due nodi e di un valore bias ( $z$ ).

$$e = w_1 * x_1 + w_2 * x_2 - z$$

(3.3)

$$output(y) = \begin{cases} 1 & \text{se } e > 0 \\ 0 & \text{altrimenti} \end{cases}$$

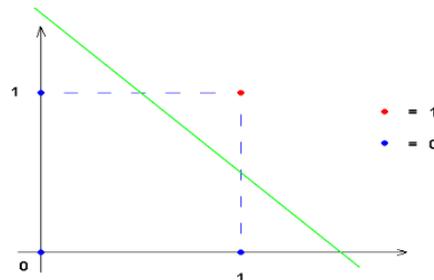
(3.4)

In base alla funzione 3.4 l'algoritmo sarà capace di classificare un pattern di dati esclusivamente nel caso in cui questo sia separabile linearmente (è possibile tracciare una retta per distinguere l'appartenenza ad una delle due categorie).

In particolare, se ipotizziamo  $z=1$  (come nel grafico 14) e i pesi anch'essi = 1 possiamo rappresentare le possibili soluzioni (in base ai valori di  $x_1$  e  $x_2$ ) in un piano cartesiano,

dove l'unico caso che prevede l'accensione del perceptrone è il caso in cui  $e > 0$ , che si verifica nel momento in cui entrambi i nodi sono "accesi".

Grafico 15. Soluzioni del modello basato sul perceptrone



La retta verde rappresenta la soglia di attivazione, che è il valore al di sotto del quale il perceptrone non si "accende". Questo aspetto viene regolato dal parametro indicato da  $z$ , dove se  $z > w_1 * x_1 + w_2 * x_2$  l'output sarà appunto 0.

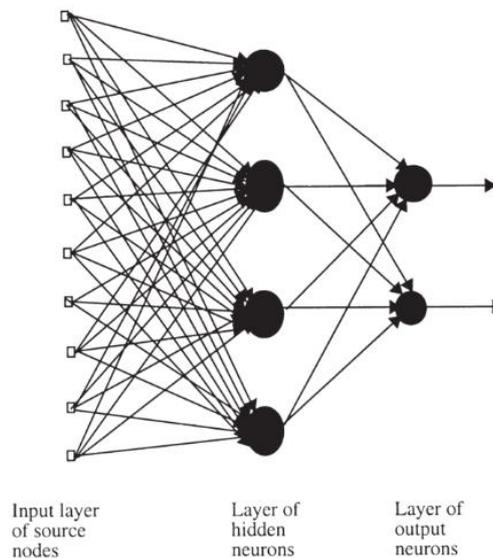
L'apprendimento della microstruttura del perceptrone, ma di qualsiasi altro tipo di rete neurale, può avvenire sostanzialmente in tre modalità differenti:

- Supervisionato: prevede la presenza di un soggetto che mostri alla rete la soluzione che vuole ottimizzare e fornisce i mezzi per poterla raggiungere; vengono quindi fornite le coppie già classificate di input-output al sistema informatico e successivamente lo stesso andrà ad aggiustare i pesi degli input con un allenamento in modo da raggiungere ottimamente gli output nel testing set.
- Non supervisionato: prevede che all'algoritmo venga fornito esclusivamente il vettore di dati in input non strutturati, senza una esplicita spiegazione sulle relazioni che deve avere con l'output; si richiede al sistema una classificazione non conosciuta a priori.
- Semi-supervisionato: vengono forniti all'algoritmo dati classificati e non.

In ambito previsionale viene utilizzato il metodo supervisionato, poiché al sistema informatico devono essere forniti gli input e i corrispondenti output su cui effettuare le previsioni; l'algoritmo inizierà a determinare i pesi ottimali, minimizzando la funzione d'errore; la grande mole di notizie finanziarie permette di soddisfare le caratteristiche che più si conformano al tipo di apprendimento: i dati sono strutturati e abbondanti.

### 3.1 Le reti feed forward

Schema 1. Rappresentazione di un Feed Forward Network



La necessità di sviluppare un archetipo più articolato di rete neurale, per poter raggiungere dei modelli di elaborazione dei dati più complessi del perceptrone, ha portato alla proposta dei Feedforward Neural Network. Un FFN è generato dalla sovrapposizione di più perceptron network, con l'introduzione di un terzo layer di nodi, l'hidden layer, che a differenza dei layer di input e output non è interpretabile in termini matematici in quanto non utilizza algoritmi definiti; questo tuttavia permette al sistema di gestire una maggiore quantità di input per poter effettuare calcoli più complessi.

Supponiamo di avere una serie finanziaria e utilizzare una rete neurale come modello per ottenere un fitting adeguato dei dati a disposizione. L'allenamento supervisionato del modello si suddivide in due fasi: la forward fase che prevede, dopo che l'istruttore ha fornito alla rete un vettore di input (come i prezzi passati di un determinato titolo) e uno di output, che vengano fissati i pesi del modello. Tale scelta poi si propagherà in un'unica direzione in avanti nella struttura neurale producendo il primo vettore output, dal quale viene calcolato una funzione di perdita, corrispondente alla differenza tra il valore passato alla rete come output (il dato empirico già manifestato su cui si vuole calcolare la precisione dell'algoritmo) e il valore calcolato dalla rete. Il passo successivo dell'apprendimento segue la direzione opposta e viene chiamato backward fase: l'algoritmo permette di propagare l'errore di previsione a ritroso nella rete, in questa fase

vengono applicati degli aggiustamenti ai pesi dei nodi con l'obiettivo di minimizzare la funzione di perdita, che si può ad esempio basare sul Mean Squared Error (MSE).

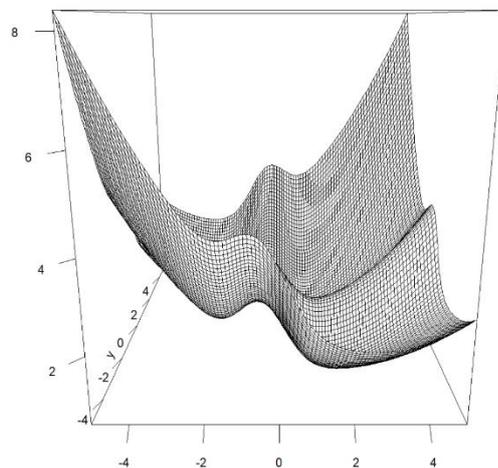
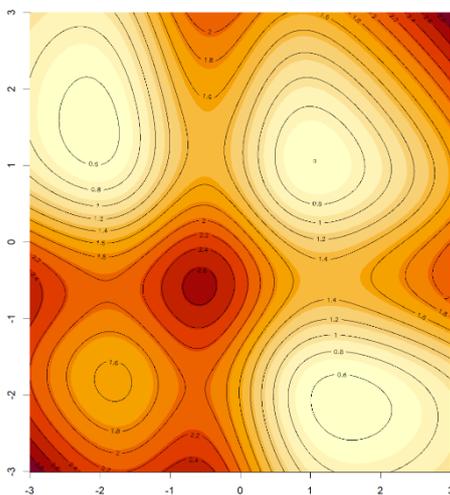
$$MSE = \frac{1}{N} * \sum_{i=1}^N (d_i - y_i)^2$$

(3.5)

Questa misurazione sintetica di adattamento considera la media degli errori al quadrato per ottenere una grandezza confrontabile.

La fase di apprendimento "a ritroso", che ha l'obiettivo di minimizzare la funzione di perdita, si basa sul metodo del gradiente; quest'ultimo non è altro che il campo vettoriale attraverso il quale la funzione cresce più rapidamente. Seguendo iterativamente la direzione di tale vettore, ma con verso opposto, si possono raggiungere i punti di minimo della funzione, giungendo così all'obiettivo dell'apprendimento supervisionato nel modello Feed Forward Network.

Grafici 16 e 17. Heat-map e visione 3D per un problema di ottimizzazione funzionale con due variabili



Come si può notare dai grafici 16 e 17, in questo semplice problema di ottimizzazione a due variabili, sono presenti più minimi relativi, attraverso la discesa del gradiente si percorre lo spazio vettoriale all'indietro quando si giunge ad un punto vicino al minimo (zone più chiare).

L'attivazione di uno strato neurale artificiale può essere differente in base all'architettura scelta per la rete, e alcune tra le più utilizzate sono la funzione sigmoide, la tangente iperbolica e la funzione threshold (vista nel caso del perceptrone).

Grafico 18. Andamento funzioni di attivazione sigmoide (verde), tangente (viola) e threshold (rosso)



Com'è possibile notare dal grafico 18, la funzione threshold restituisce valori pari a uno nel caso la funzione in input sia maggiore di zero mentre pari a zero se è minore o uguale a zero. La funzione sigmoide assume invece valori compresi tra zero e uno e la funzione tangente iperbolica tra meno uno e uno.

Merita una menzione anche la funzione softmax, che deriva dalla funzione sigmoidea e viene utilizzata per attribuire all'output calcolato dalla rete, la probabilità che questo appartenga ad una delle classi definite dall'utente in fase di istruzione del modello.

Durante la fase di apprendimento della rete e dell'algoritmo di back-propagation però, a causa dell'utilizzo di queste funzioni può sorgere un problema di ottimizzazione, ovvero la scomparsa del gradiente.

Al fine di ridurre la distanza tra dato osservato e predetto, la funzione di perdita viene derivata in base ai pesi del modello, in questo modo viene calcolata la variazione della funzione di perdita in virtù di una variazione dei pesi del modello. Il problema che sorge è legato al gradiente delle funzioni di attivazione, che assume valori bassi sia nel caso della funzione sigmoide (tra 0 e 1) che nel caso della funzione tangente iperbolica (tra -1 e 1), poiché i gradienti di ogni livello della rete vengono moltiplicati tra loro secondo la regola

della catena<sup>4</sup> e qualora i valori fossero così compressi per via della tipologia di funzione di attivazione il gradiente diminuirebbe fino a svanire, rendendo la rete non più in grado di apprendere.

Per superare questo problema è stata introdotta una funzione di attivazione alternativa chiamata ReLu, in cui prevede che se la somma pesata dei neuroni ha un valore positivo allora verrebbe fatto procedere nella rete un valore pari a:

$$f(x) = \begin{cases} x & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

(3.6)

In questo modo viene velocizzata la convergenza della discesa del gradiente verso il minimo globale della funzione di perdita, come descritto da Hahnloser et al. (2000).

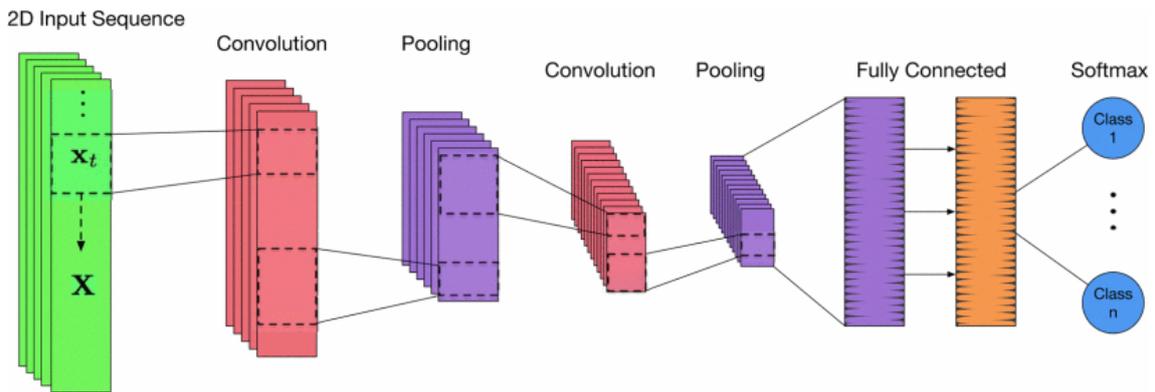
### 3.1.1 Il Convolutional Neural Network

Le cosiddette reti convoluzionali rappresentano una tipologia di feed forward network, e sono caratterizzate dalla presenza di alcuni hidden layer, chiamate feature maps, che vengono utilizzati affinché la mole di dati in input possa essere ridotta attraverso operazioni di convoluzione e di pooling, che coinvolgono gli input del modello e dei filtri chiamati kernel. Il principale ambito di applicazione di questa rete è il riconoscimento e la classificazione automatica di immagini ad alta risoluzione, ma la sua architettura si può prestare anche all'ambito delle previsioni.

---

<sup>4</sup> La regola della catena è una regola di derivazione che permette di calcolare la derivata della funzione composta di due funzioni derivabili. Nel problema di minimizzazione della funzione di perdita, la sua derivata in funzione dei pesi del modello può essere scomposta andando a considerare le derivate di ogni nodo del modello; in questo modo viene calcolata la sensibilità degli hidden layer su variazioni dei pesi, che concorrono alla generazione di un output per ogni istante temporale

Figura 1. Funzionamento di una rete convoluzionale



Il funzionamento di questa tipologia di rete neurale si basa sulle due operazioni descritte poc'anzi: la convoluzione e il pooling. L'algoritmo di convoluzione prevede che la matrice di input, i cui singoli elementi con i rispettivi pesi risultano essere poco gestibili all'interno del network, venga scomposta in delle matrici più piccole e per ognuna venga applicato un filtro kernel che infine crea una sotto-matrice meno pesante chiamata feature map; la scelta del kernel è fondamentale in quanto la dimensione della nuova matrice di dati dipenderà dalla dimensione della matrice kernel, maggiore è quest'ultima più piccola sarà la matrice "convoluzionata".

L'espressione di una tipica convoluzione di una matrice 5 x 5 applicata ad un kernel 3 x 3 è:

$$y = \sum_{i=1}^3 \sum_{j=1}^3 w_{i,j} * x_{i,j}$$

(3.7)

dove

- $w_{i,j}$  rappresentano i singoli elementi di riga (i) e colonna (j) della matrice input
- $x_{i,j}$  rappresentano i singoli elementi di riga (i) e colonna (j) della matrice kernel

Ogni volta che viene applicato il filtro kernel viene effettuato uno slide laterale della matrice di input 5 x 5 per selezionare la successiva sottomatrice 3 x 3.

Successivamente viene utilizzato l'algoritmo di pooling, che può essere di due tipi, in base al tipo di estrazione che viene effettuata: il max pooling sceglie il valore reale più alto all'interno della matrice mentre l'average pooling individua il valore medio della stessa; infine si ottiene il flattening layer, che è direttamente connesso ad un classificatore di output che restituisce la classe più probabile a cui appartiene la matrice di input, dopo un'adeguata fase di training del modello.

L'adattabilità alle serie storiche finanziarie delle reti convoluzionali, pensate inizialmente per la classificazione delle immagini, necessita una precisazione; questo tipo di rete lavora solitamente con matrici a due dimensioni, mentre le serie storiche sono composte da un vettore unidimensionale. Sarà quindi utile andare ad incorporare nell'analisi anche altre informazioni utili all'apprendimento del modello oltre al prezzo di chiusura, come il prezzo giornaliero più alto, quello più basso oppure ancora i volumi di vendita e la volatilità implicita, rendendo la serie multivariata.

Rimanendo nel dominio finanziario, è esplicativo dell'efficacia di tale framework lo studio di Tsantekidis et al. (2017), in cui viene utilizzato un algoritmo di Convolutional Network per prevedere il movimento al rialzo o al ribasso dei prezzi. La variabile in oggetto sono i prezzi bid-ask di un libro ordini, forniti dalla compagnia Nasdaq Nordic. Il libro ordini (book) è un prospetto che raccoglie le proposte di acquisto e di vendita relative ad un determinato strumento. Il book è fondamentale per gli operatori perché consente di capire se e a quale livello di prezzo e quantità un determinato ordine può essere eseguito<sup>5</sup>.

Dallo studio emerge come questa tecnica risulti più precisa di modelli come il multi-layer perceptron o il Support Vector Machine, portando alla conclusione che può essere un modello valido per le previsioni di serie storiche finanziarie; questa rete neurale però in letteratura viene utilizzata molto spesso per classificare i movimenti dei prezzi e prevedere trend a più passi, perciò può essere affidabile non tanto se si vuole una previsione puntuale e quantitativa sui prossimi valori della serie storica ma piuttosto per avere un riferimento qualitativo sulle sue oscillazioni.

---

<sup>5</sup> Definizione di Treccani



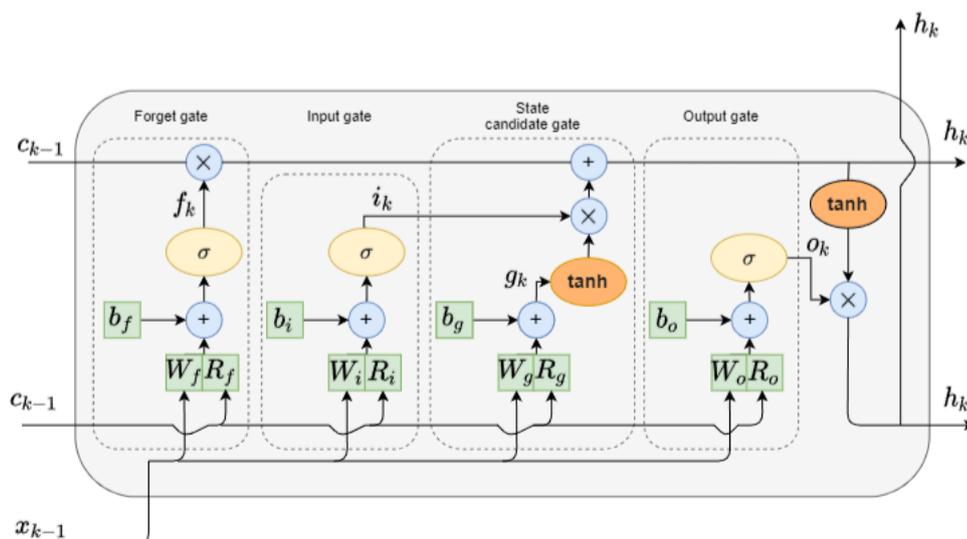
Un concetto che differenzia questa architettura dalle precedenti descritte è quello di condivisione dei pesi: dato che l'algoritmo viene propagato nel corso dei periodi è come se l'architettura fosse ripetuta di volta in volta esattamente allo stesso modo, reiterando così anche i pesi all'interno del modello.

A differenza delle reti feed forward, non è presente un'unica direzione di avanzamento della rete; sono infatti previsti dei "loop di apprendimento" dove l'output pesato al tempo  $t-1$  (che al contempo contiene tutti gli altri output processati precedentemente) venga utilizzato per istruire il modello a far emergere il miglior output di previsione. È perciò necessario fornire le esperienze passate della rete stessa ai nodi successivi che processano i dati, affinché possa attribuire un significato alla successione di output.

### 3.2.1 Le Long Short Term Memory

Le Long Short Term Memory sono una tipologia di reti ricorrenti caratterizzate dalla capacità di gestire informazioni anche molto lontane nel tempo; questo avviene grazie all'architettura deputata a gestire i flussi in ingresso della rete e ad individuarne l'influenza nella memoria del sistema.

Figura 3. Struttura di una rete long-short term



Come rappresentato in figura, la struttura della rete include dei nodi non trattati nelle generiche reti ricorrenti.

La cella si suddivide in forget gate, input gate, output gate e state candidate gate (cella di memoria a lungo termine), dove:

- $f_k$  rappresenta il valore del forget gate al tempo  $k$  e viene calcolato attraverso la seguente espressione  $f_k = \sigma(W_f * [h_{k-1}, x_k] + b_f)$ .  $\sigma$  è la funzione di attivazione sigmoideale mentre  $W_f * [h_{k-1}, x_k] = W_f R_f$  è la somma pesata dell'input  $x_k$  e dell'hidden layer precedente della rete  $h_{k-1}$ .  $b_f$  è il termine di bias del forget gate.
- $i_k$  rappresenta il valore dell'input gate al tempo  $k$  e viene calcolato attraverso la seguente espressione  $i_k = \sigma(W_i * [h_{k-1}, x_k] + b_i)$ .  $b_i$  è il termine di bias dell'input gate.
- $o_k$  rappresenta il valore dell'output gate al tempo  $k$  e viene calcolato attraverso la seguente espressione  $o_k = \sigma(W_o * [h_{k-1}, x_k] + b_o)$ .  $b_o$  è il termine di bias dell'output gate.
- La cella di memoria state candidate gate invece viene calcolata come  $C_k = f_k \otimes C_{k-1} + i_k \otimes C_t$  dove  $\otimes$  rappresenta il prodotto elemento per elemento fra le matrici.

Nella figura 3,  $x_k$  rappresenta il dato che viene utilizzato dalla rete come variabile di input; questo parametro, assieme allo stato precedente della cella  $c_{k-1}$ , verrà utilizzato per aggiornare la memoria della rete.

Nello specifico il forget gate elabora, in base al nuovo input  $x_k$  e allo stato precedente della cella  $c_{k-1}$ , quanto della memoria a lungo termine non è più necessario conservare per gli step successivi della rete o in altri termini la quantità di memoria a lungo termine che recorderà la rete.

L'input gate invece modula il livello di ingresso dei nuovi input nella rete, e in base all'utilità di tali informazioni il gate aggiorna lo stato della cella a lungo termine "state candidate gate" con i nuovi input a breve termine. Questo tipo di rete è ottimale per gestire serie storiche dove si verificano fenomeni di trend e stagionalità, poiché la capacità della rete di ricordare a lungo i movimenti precedenti della serie storica permette una comprensione maggiore dei fenomeni periodici in termini di previsioni future.

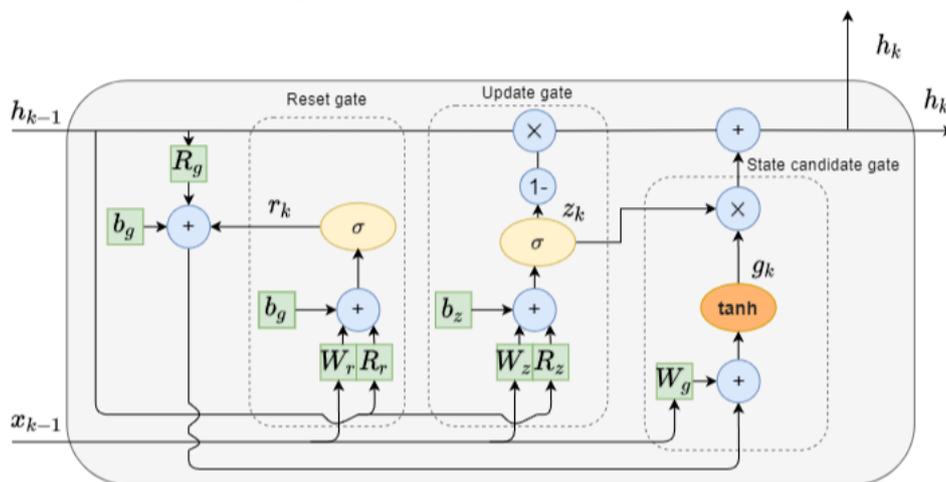
L'output gate infine è necessario per poter filtrare il livello di informazioni dello state candidate gate in base a ciò che potrebbe servire alla rete nell'input successivo. Infine, verrà calcolato il vero e proprio output della rete all'istante di tempo t per poi procedere ricorsivamente con i successivi step temporali per l'allenamento della rete.

Il recente studio di Yuping et al. (2024) dimostra come si possano raggiungere livelli soddisfacenti di previsione di una serie storica finanziaria utilizzando assieme ad un modello di non-linear regression (utilizzato per prevedere il livello medio dei prezzi) un modello LSTM (utilizzato per prevedere i residui del modello basato su una regressione non lineare).

### 3.2.2 Le Gated Recurrent Unit

La struttura delle unità gated recurrent differisce dalle long short term memory poiché l'algoritmo che modella la memoria a lungo termine della rete (che non viene conservata in una cella specifica) viene sviluppato direttamente dal nodo di output, attraverso tre soli gate: il reset gate, l'update gate e lo state candidate gate.

Figura 4. Struttura di una rete gated recurrent



La cella della figura 4 si suddivide in reset gate, update gate e state candidate gate, dove:

- $r_k$  indica il valore del reset gate al tempo k e viene calcolato attraverso la seguente espressione  $r_k = \sigma(W_r * [h_{k-1}, x_k] + b_g)$  dove  $b_g$  rappresenta il bias del reset gate.

- $z_k$  indica il valore dell'update gate al tempo k e viene calcolato attraverso la seguente espressione  $z_k = \sigma(W_z * [h_{k-1}, x_k] + b_z)$  dove  $b_z$  rappresenta il bias dell'update gate.
- $g_k$  indica il valore dello state candidate gate al tempo k e viene calcolato attraverso la seguente espressione  $g_k = \tanh(W_g * [x_k] + b_g)$  dove  $b_g$  rappresenta il bias dello state candidate gate e  $\tanh$  la funzione di attivazione dell'espressione tra parentesi.

Come mostrato nella figura 4 e anticipato in precedenza, sarà l'output di ogni fase del modello (state candidate gate) che raccoglie le informazioni rilevanti per ogni passo in base al reset gate e all'update gate; in questo modo, successivamente, lo stesso output verrà riutilizzato e riaggiornato con le nuove informazioni disponibili nell'istante di tempo seguente.

Per entrare maggiormente nello specifico è necessario analizzare le funzioni del reset gate e dell'update gate. Il primo passo, all'interno della rete ricorrente, è quello di utilizzare l'input  $x_k$  e lo stato precedente della rete  $h_{k-1}$  (state candidate gate); a queste due informazioni, che vengono moltiplicate tramite prodotto di Hadamard<sup>6</sup>, viene associata una ponderazione a cui viene sommato il valore in input della rete. Questo flusso infine passa attraverso una funzione di attivazione a tangente iperbolica che determinerà il valore del reset gate. A questo hidden layer va sommato, con opportuni pesi, il layer deputato alla componente della memoria da aggiornare con le nuove informazioni, conosciuto come update gate: questo strato dipende allo stesso modo da  $x_k$  e  $h_{k-1}$  e viene sommato al reset gate, arrivando così ad ottenere l'output della rete al tempo t.

Nell'ambito di un confronto condotto da Ghadimpour (2022) fra LSTM e GRU, viene dimostrata l'efficacia delle previsioni finanziarie di quest'ultime, portandomi a includerle nel confronto operativo.

---

<sup>6</sup> Prodotto che avviene elemento per elemento tra le matrici corrispondenti

### 3.3 Gli aspetti negativi delle reti neurali

I vantaggi delle reti neurali sono quelli citati precedentemente, solitamente questo tipo di machine/deep learning permette di cogliere alcuni pattern non lineari nascosti tra i dati, fattore che i metodi tradizionali non riescono a cogliere perché si basano su modelli con parametri statici che non fittano i dati delle serie storiche finanziarie; dopo aver suddiviso il campione di dati osservati in training set, validation set e testing set, attraverso i metodi iterativi di ottimizzazione della funzione di perdita l'algoritmo di rete neurale individua un modello non parametrico sulla base dell'addestramento compiuto, successivamente poi il modello viene validato e testato su dati non ancora utilizzati (ma osservati perché si sono già verificati), e infine confrontati i risultati.

Nel metodo di apprendimento delle reti neurali esiste però un importante aspetto che probabilmente è il più grande problema di questi algoritmi di machine learning, ovvero l'overfitting. Questo fenomeno potrebbe minare l'affidabilità dei risultati del modello, e si verifica quando la rete impara perfettamente i pattern del sotto-campione di training e riesce a riprodurli fedelmente, ma ciò impedisce di generalizzare i comportamenti dei dati al di fuori del training set, che è la fase che più interessa per verificare se la rete è utile a prevedere i movimenti del mercato.

Un metodo utile a ridurre al minimo questa problematica è verificare e confrontare i tassi di errore del training e del testing set; qualora il modello si adattasse bene al campione di training ma male al campione di test, questo è un sintomo di potenziale overfitting e perciò si dovrebbe intervenire in fase di ideazione e addestramento del modello riducendo il set di dati al servizio dell'apprendimento o ridurre le stesse epoche di addestramento. È necessario trovare il giusto equilibrio tra sub-set e scegliere la combinazione giusta di training, validation e test, aiutandosi con grafici di perdita sulla base di funzioni come il mean squared error (nel caso delle reti neurali) che mostrino quando è utile stoppare l'addestramento.

Concetti fondamentali che permettono la corretta adattabilità e addestramento del modello reti neurali ai dati osservabili sono le epoche e il batch size. In un contesto di costruzione della rete, questi sono iper-parametri fondamentali, ed è anche grazie a questi che la funzione di perdita viene minimizzata. Ogni epoca rappresenta il percorso,

all'interno della serie storica, che l'algoritmo ripete nella fase di addestramento mentre il batch size è il sottogruppo di dati che viene analizzato volta per volta all'interno di ogni epoca; banalmente se dovessimo ipotizzare di costruire una rete neurale utilizzando i parametri "epochs" e "batch size" uguali a 1 significherebbe che la rete dovrebbe analizzare tutti i dati a disposizione una volta sola, rendendo così il processo di apprendimento sostanzialmente inutile, poiché è grazie all'iterazione di questo processo che i pesi dei neuroni vengono aggiustati al fine di minimizzare la funzione di perdita. Per la specificazione di un buon modello è necessario, inoltre, che la funzione di perdita sia simile tra il campione di addestramento e di test: attraverso la corretta individuazione del parametro "epochs" è possibile aiutare il modello ad evitare l'overfitting che si verifica quando la funzione di perdita è molto più alta per il test set che per il training set.

### Capitolo III – Confronto operativo tra modelli neurali

Archiviato l'aspetto teorico e strutturale delle reti neurali, è ora utile verificare se e in che modo i modelli non parametrici definiti in precedenza possano intercettare segnali predittivi dalle serie storiche finanziarie. Per fare questa verifica è necessario implementare un algoritmo di previsione per ogni rete neurale specificata; i network che verranno utilizzati per il confronto sono le reti Long Short Term Memory, Convolutional Neural Network e Gated Recurrent Unit. In questo modo sarà possibile testare sia un tipo di rete neurale feed forward come le CNN sia una tipologia a nodi ricorrenti come le LSTM e le GRU.

La prima fase che occorre affrontare nell'addestramento di una rete neurale è la corretta specificazione dei dati che verranno utilizzati nell'analisi; la serie storica di riferimento sarà infatti quella dei rendimenti, in quanto più semplice da studiare essendo una serie tipicamente stazionaria.

La seconda fase consiste nel preparare i dataset pronti per essere completati con i risultati del modello, utilizzando la variabile indipendente  $x$  che va ad indicare i parametri che il modello utilizza per andare a minimizzare la funzione di perdita e la variabile dipendente  $y$  che andrà a contenere i valori finali di previsione sul campione di test.

La terza fase è legata alla vera e propria specificazione e addestramento del modello di rete neurale, andando ad analizzare poi la differenza tra le funzioni di perdita dei due campioni (training e test), verificando il livello ottimale di periodi attraverso cui svolgere l'addestramento, per evitare il problema dell'overfitting.

La quarta fase dell'analisi di un modello predittivo è quella di verificare, dopo la fase di addestramento svolta sul training set, se e come i risultati dell'algoritmo di rete neurale sul campione di test si adattano rispetto ai dati empirici.

La quinta e ultima fase del modello è effettuare una previsione a  $n$  passi di un campione out-of-sample che si deve ancora verificare empiricamente, andando ad utilizzare gli output ad ogni passo precedente come input del modello.

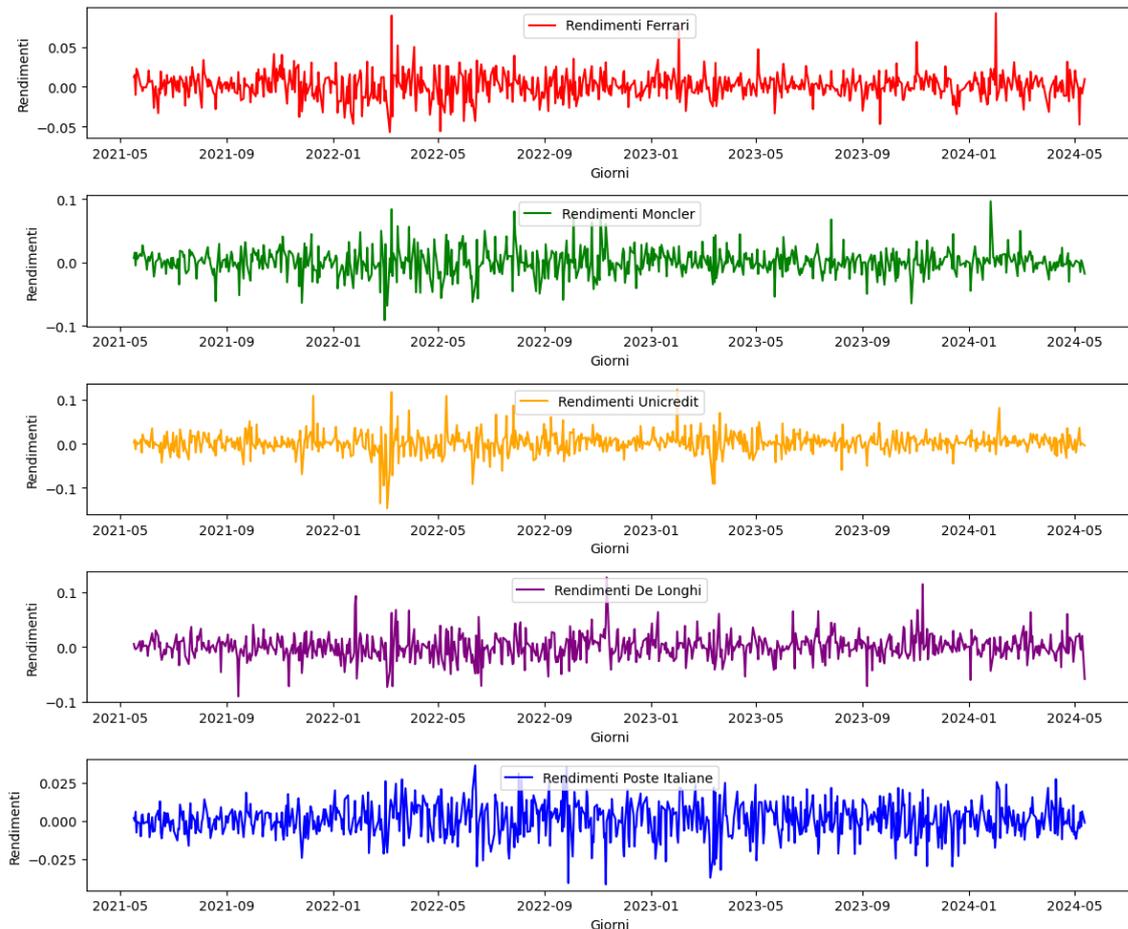
### 3.1 Previsioni utilizzando dati giornalieri

L'analisi è stata condotta utilizzando il software Python e sono stati utilizzati i dati giornalieri di cinque titoli esemplificativi per verificare la bontà delle previsioni: Ferrari, Moncler, Unicredit, De Longhi e Poste Italiane.

La prima rete analizzata è la rete Long Short Term Memory e di seguito verranno rappresentati i passi necessari al corretto sviluppo del programma.

Innanzitutto, sarà necessario (questo per tutti i modelli di reti neurali) importare le librerie utili a svolgere tutti i passaggi che sono richiesti, oltre alle librerie che contengono i metodi di costruzione delle singole reti.

Grafico 19. Rendimenti dei cinque titoli



Fonte: [www.alphavantage.co](http://www.alphavantage.co)

Per svolgere un'analisi predittiva consistente, è necessario prelevare un campione di dati su cui l'algoritmo possa esercitarsi a sufficienza ma non in eccesso, ecco perché è stata scelta una finestra temporale di tre anni.

Com'è possibile vedere dai vari grafici, è presente una differente variabilità dei rendimenti dei diversi titoli, con il rendimento del titolo Poste Italiane che è compreso in una fascia più ristretta.

Il passo successivo consiste nell'individuare la dimensione dei sotto-campioni e degli iperparametri che intervengono nell'analisi.

Figura 5. Impostazione di parametri

```
test_size = 0.15  
validation_size = 0.15  
epochs = 50  
batch_size = 32
```

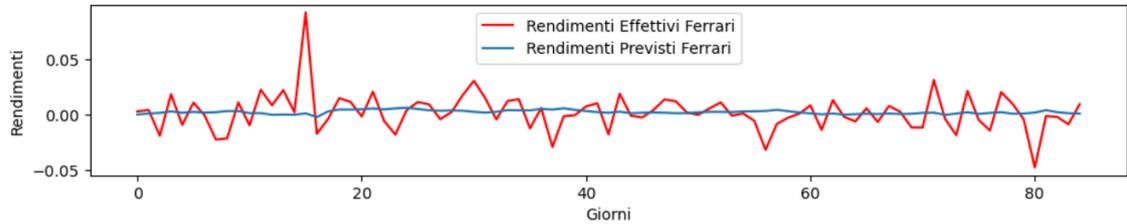
Nella figura 5 sono riportati i valori assegnati ai parametri al fine di raggiungere dei risultati soddisfacenti in fase di addestramento. Il campione è stato suddiviso in tre porzioni di dati: il settanta per cento è stato destinato all'apprendimento e addestramento della rete neurale (training set), mentre per ognuna delle fasi di validazione e di test del modello è stata utilizzata una porzione del quindici per cento dei dati a disposizione. È stato impostato un valore di epoche di addestramento pari a cinquanta, e il sotto-campione di dati che verrà analizzato consecutivamente è di trentadue. Un altro parametro utilizzato in questo programma è il numero di time steps utilizzati come input dal modello in fase di addestramento, impostati a trenta: per prevedere ogni rendimento nel testing set, perciò, la rete utilizzerà le trenta osservazioni precedenti.

I principi di modellazione della rete neurale utilizzati sono stati l'utilizzo del mean squared error come funzione di perdita del modello, il settaggio di un numero di hidden layer pari a cinquanta per permettere alla rete neurale artificiale di immagazzinare una memoria a lungo termine sufficiente allo sforzo computazionale e la restituzione di un output di rendimento predetto per ogni step temporale del campione di test.

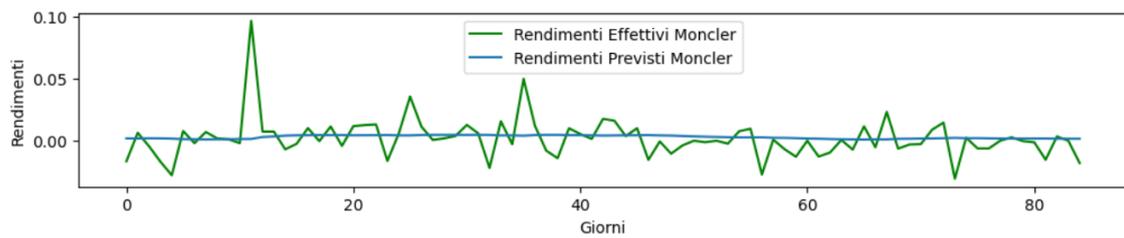
Per controllare ulteriormente la validità dei modelli è stato creato un algoritmo di accuratezza dei modelli che confrontasse per ogni rendimento osservato il rendimento

predetto nel testing set per i diversi modelli, non tanto in termini di valore assoluto ma di segno del rendimento, che esprime l'aumento o la diminuzione del prezzo.

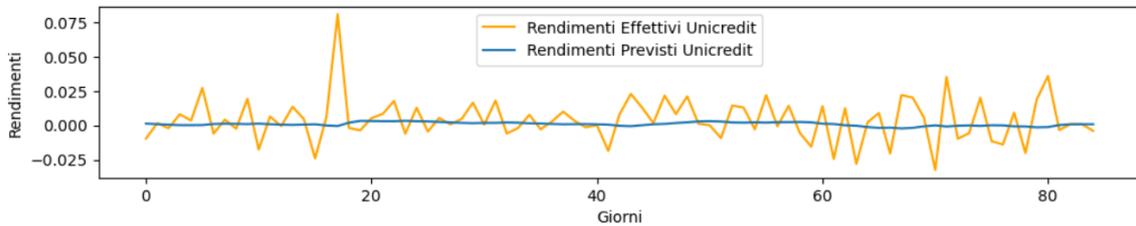
Grafico 20. Confronto tra rendimenti effettivi e predetti dalla rete neurale con LSTM



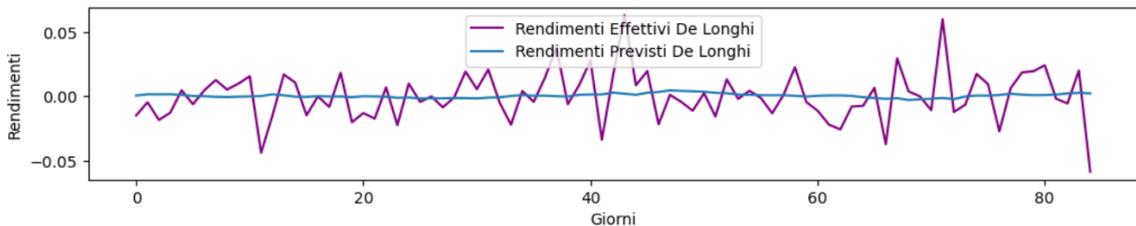
Numero di previsioni corrette: 51  
Numero totale di previsioni: 85  
Accuratezza modello LSTM: 60.00%



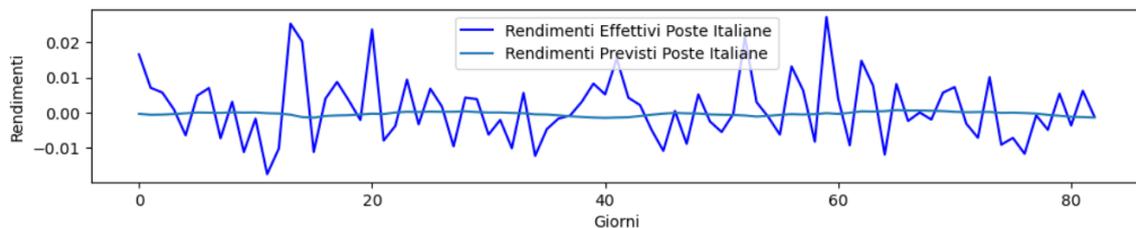
Numero di previsioni corrette: 42  
Numero totale di previsioni: 85  
Accuratezza modello LSTM: 49.41%



Numero di previsioni corrette: 52  
Numero totale di previsioni: 85  
Accuratezza modello LSTM: 61.18%



Numero di previsioni corrette: 43  
Numero totale di previsioni: 85  
Accuratezza modello LSTM: 50.59%

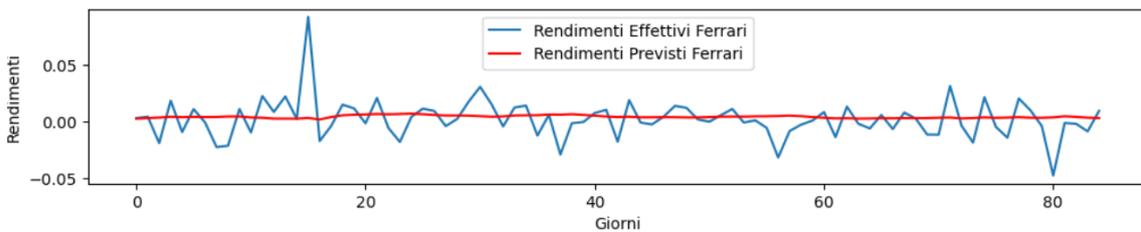


Numero di previsioni corrette: 38  
Numero totale di previsioni: 83  
Accuratezza modello LSTM: 45.78%

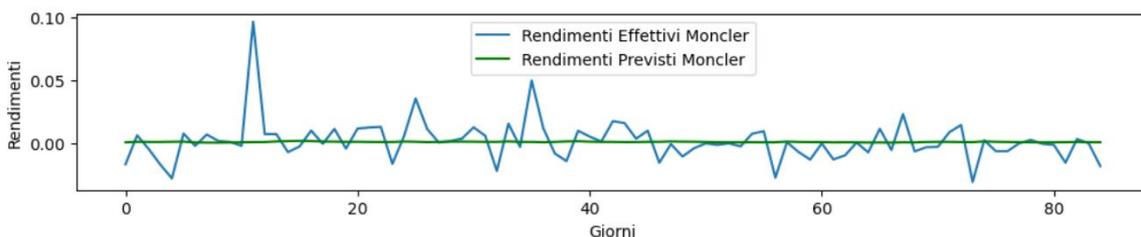
Dai grafici emerge come il segnale intercettato attraverso il sistema di rete neurale LSTM sia non particolarmente rilevante, questo conferma come i rendimenti in un contesto di dati giornalieri siano difficilmente prevedibili e il mercato raggiunga una forma di efficienza tale da non permettere grossi spazi di manovra. Il risultato non sembra migliorare andando a modificare gli iper-parametri e il numero di hidden layers della rete.

Il secondo modello utilizzato è il Gated Recurrent Unit e i fattori di costruzione del modello sono i medesimi utilizzati per la rete Long Short Term, tranne il numero di hidden layer e il numero di strati neurali. È stato scelto di istruire un modello che avesse due strati di neuroni artificiali ognuno composto da dieci strati nascosti e di seguito verranno riportati i risultati ottenuti dall'analisi tramite l'utilizzo della seconda rete neurale con dati giornalieri.

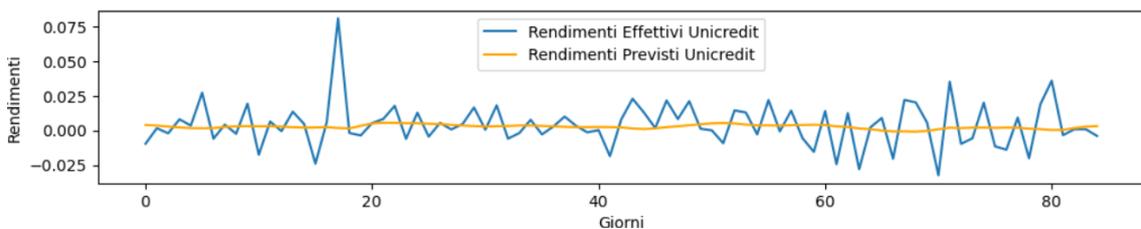
Grafico 21. Confronto tra rendimenti effettivi e predetti dalla rete neurale con GRU



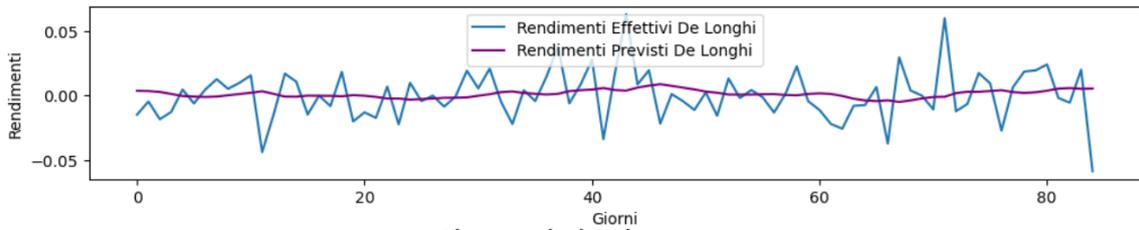
Numero di previsioni corrette: 44  
 Numero totale di previsioni: 85  
 Accuratezza modello GRU: 51.76%



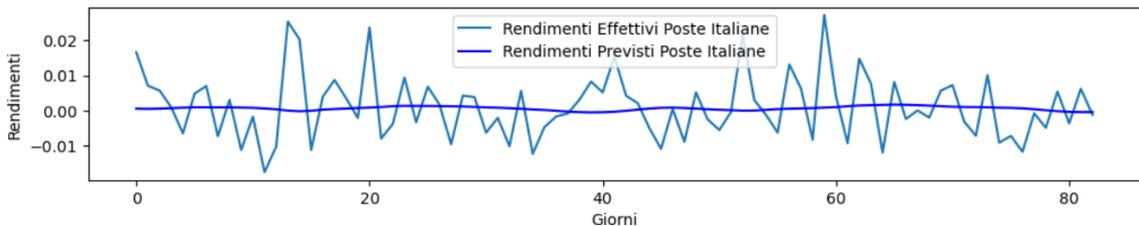
Numero di previsioni corrette: 48  
 Numero totale di previsioni: 85  
 Accuratezza modello GRU: 56.47%



Numero di previsioni corrette: 38  
 Numero totale di previsioni: 85  
 Accuratezza modello GRU: 44.71%



Numero di previsioni corrette: 40  
 Numero totale di previsioni: 85  
 Accuratezza modello GRU: 47.06%



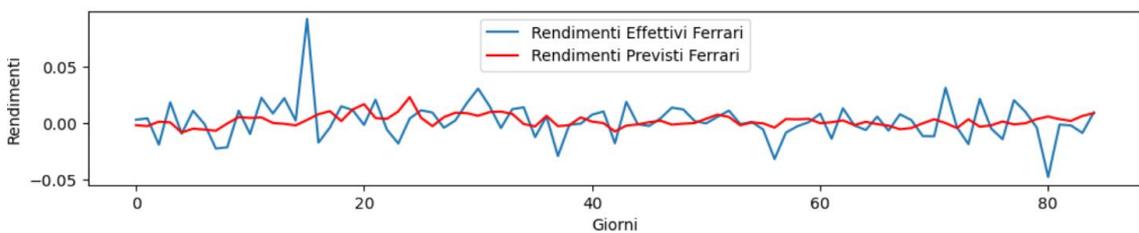
Numero di previsioni corrette: 42  
 Numero totale di previsioni: 83  
 Accuratezza modello GRU: 50.60%

I risultati sembrano essere in linea con quanto detto in precedenza nell'ambito di una rete neurale LSTM, con poco segnale intercettato nonostante questi modelli, a differenza di quelli parametrici, riescono a cogliere legami non lineari nei dati.

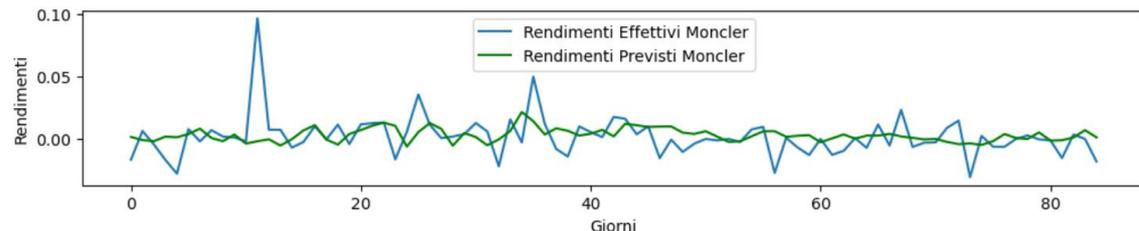
L'ultimo caso preso in esame è la fattispecie delle reti CNN. In questo caso è stata costruito un network costituito da una tecnica convoluzionale a una dimensione con dei filtri kernel.

Di seguito i risultati sullo studio dei rendimenti:

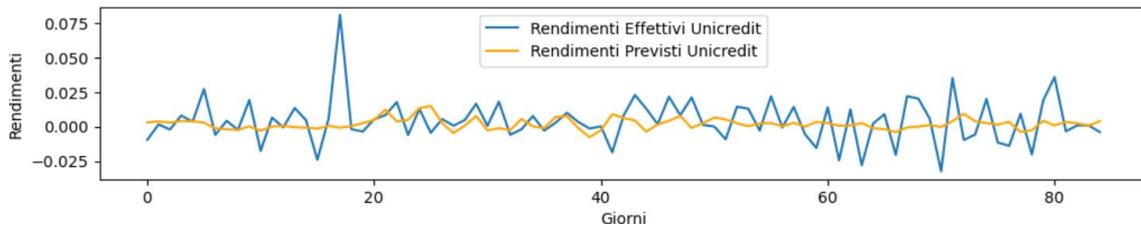
Grafico 22. Confronto tra rendimenti effettivi e predetti dalla rete neurale con CNN



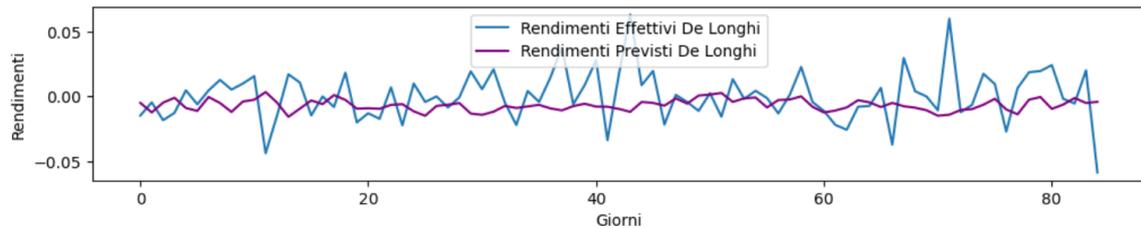
Numero di previsioni corrette: 44  
 Numero totale di previsioni: 85  
 Accuratezza modello CNN: 51.76%



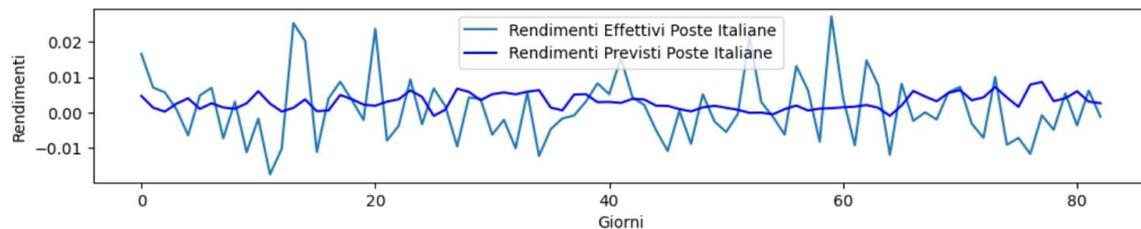
Numero di previsioni corrette: 46  
 Numero totale di previsioni: 85  
 Accuratezza modello CNN: 54.12%



Numero di previsioni corrette: 44  
 Numero totale di previsioni: 85  
 Accuratezza modello CNN: 51.76%



Numero di previsioni corrette: 36  
 Numero totale di previsioni: 85  
 Accuratezza modello CNN: 42.35%

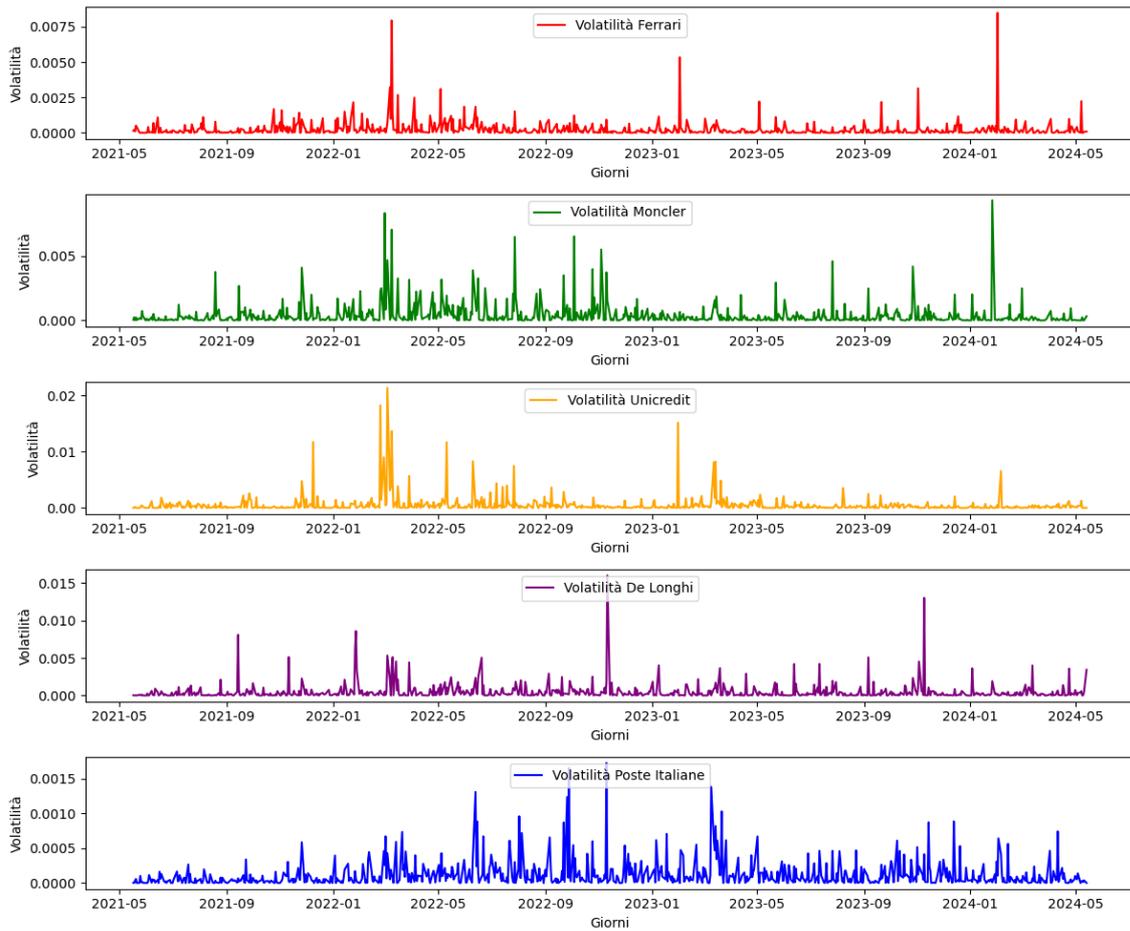


Numero di previsioni corrette: 35  
 Numero totale di previsioni: 83  
 Accuratezza modello CNN: 42.17%

Dai risultati dei grafici sembra che, a parità di parametri impostati (con la sola eccezione del titolo poste, in cui è stato utilizzato un numero di epoche minore per evitare overfitting), questo modello sia in grado di seguire maggiormente l'andamento dei rendimenti effettivi, con una maggiore capacità (seppur ancora insufficiente) di captare i segnali degli investitori.

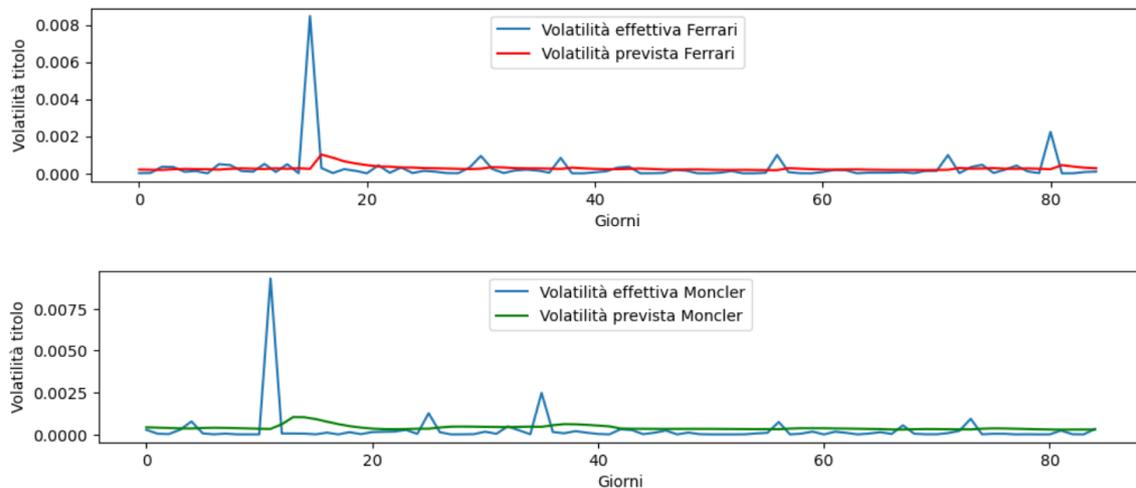
Un altro elemento su cui provare ad effettuare una stima è la volatilità, o meglio una proxy della volatilità, i rendimenti al quadrato. Precedentemente è stata menzionata una proprietà della volatilità per serie storiche finanziarie, la persistenza; questa caratteristica implica la presenza di pattern di volatilità con caratteristiche simili in diversi periodi storici ed è un elemento maggiormente prevedibile rispetto ai rendimenti, e per questo è possibile effettuare un'analisi attraverso le reti neurali.

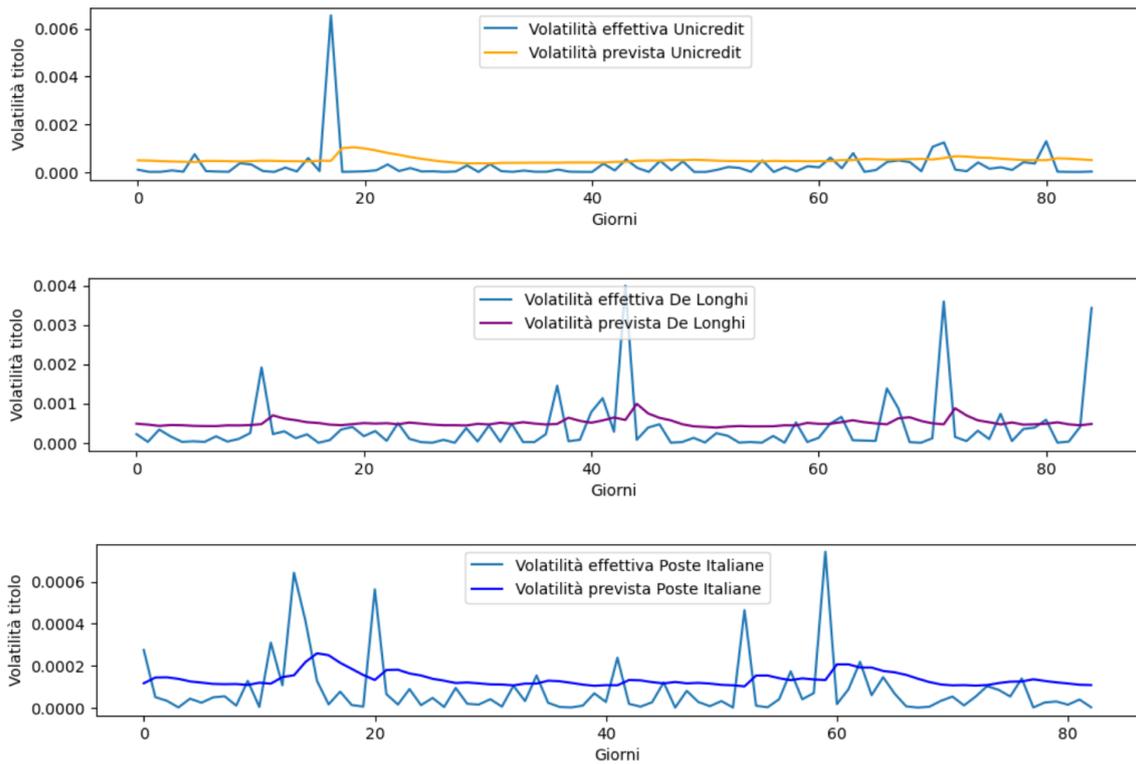
Grafico 23. Volatilità dei cinque titoli



Gli step necessari all'addestramento nel campione di training e alla previsione della rete nel campione di testing sono i medesimi rispetto alla procedura sui rendimenti, utilizzando come input al posto di quest'ultimi la proxy della volatilità.

Grafico 24. Confronto tra volatilità effettive e predette dalla rete neurale con LSTM

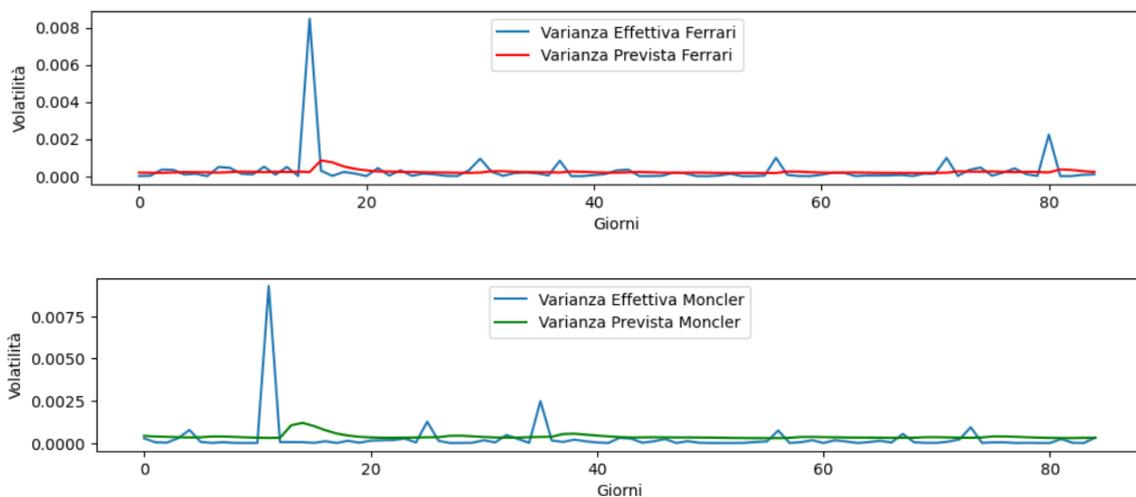


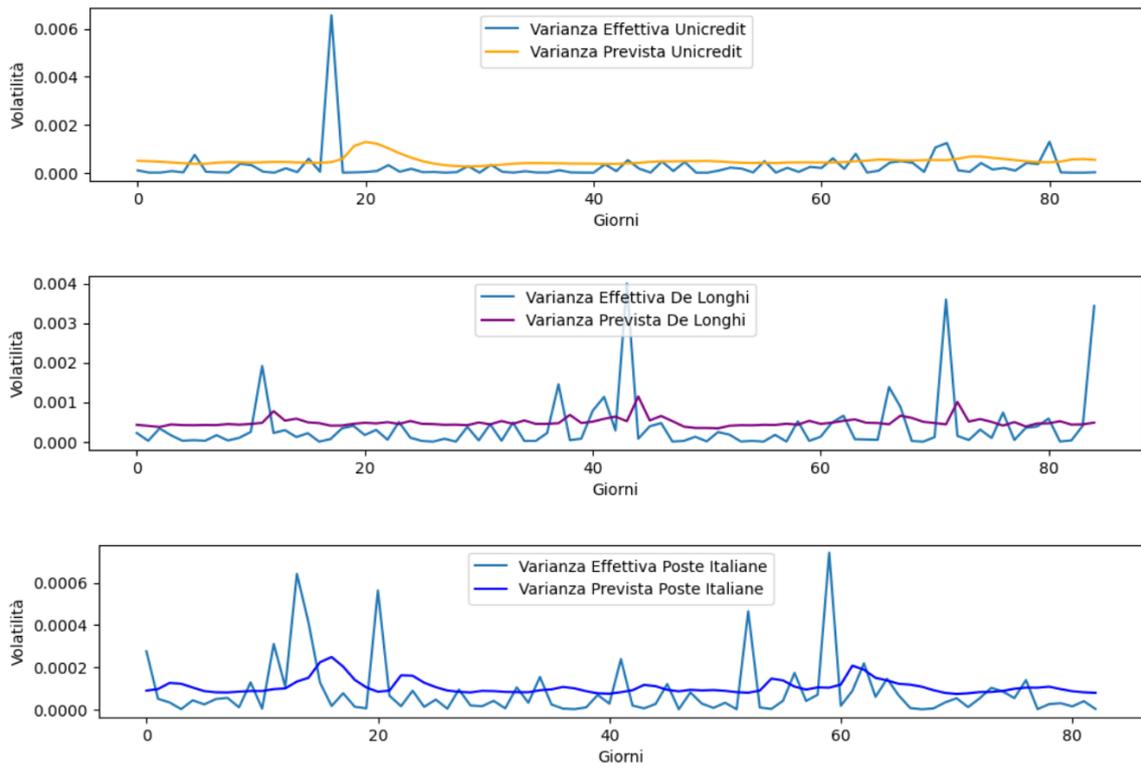


Da quanto emerge dal grafico 24, i picchi di volatilità vengono percepiti dalla rete neurale come persistenti ma difficilmente è presente una reazione immediata a questi picchi nella volatilità prevista, rimane comunque presente una capacità di intercettare i segnali maggiore rispetto alle serie dei rendimenti degli stessi titoli sottoposti ad analisi in precedenza.

Di seguito i risultati relativi all'analisi predittiva sulla volatilità con il modello GRU.

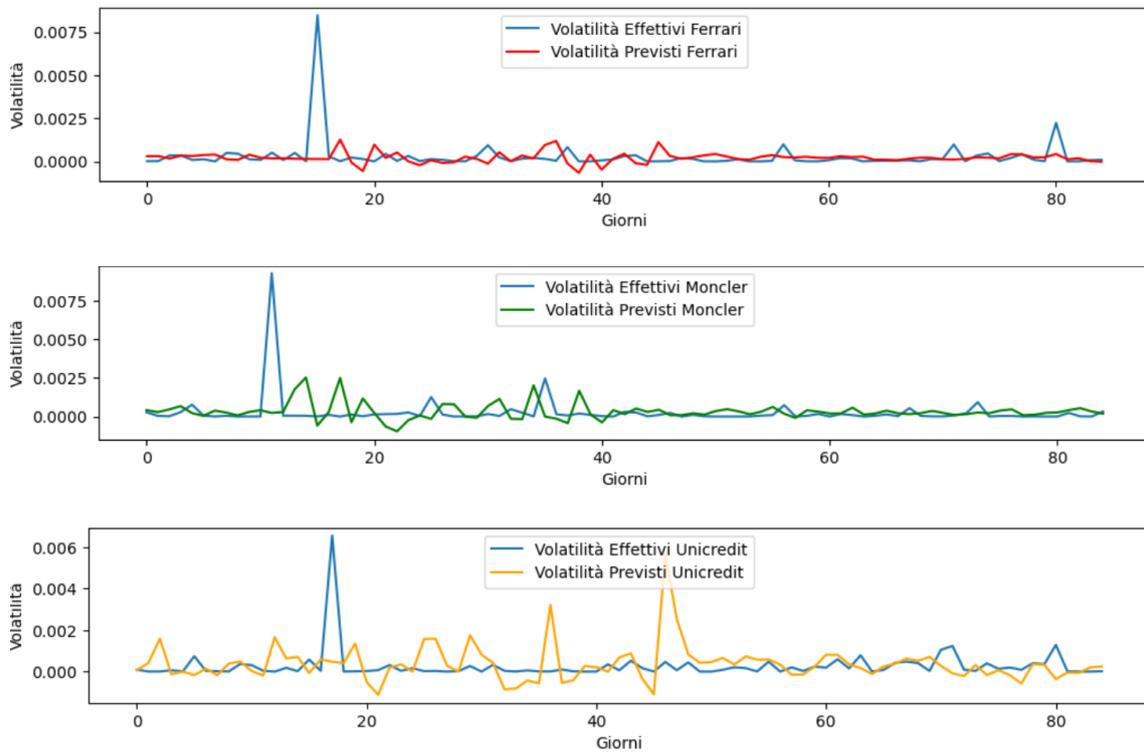
Grafico 25. Confronto tra volatilità effettive e predette dalla rete neurale con GRU

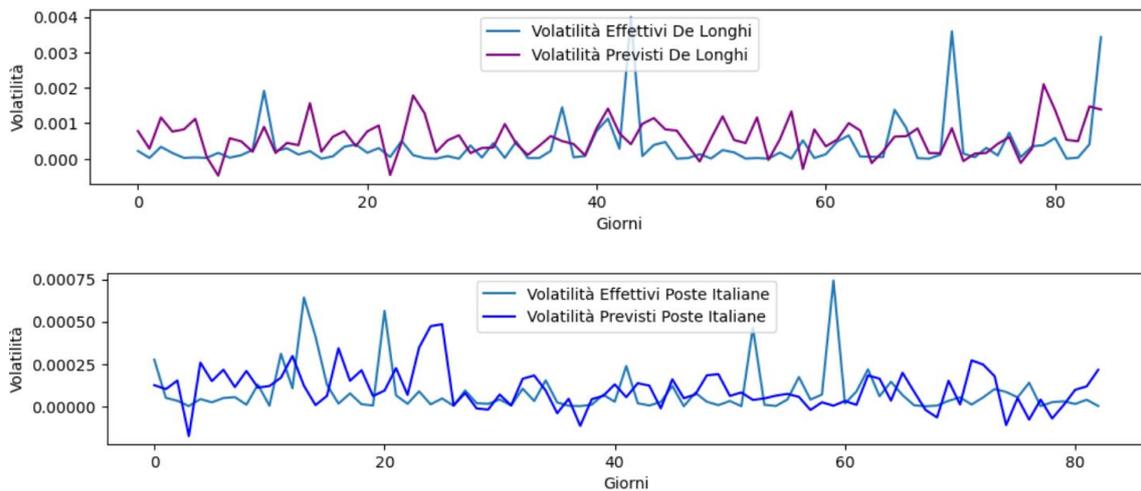




I risultati non si discostano molto da ciò che è emerso nel caso di utilizzo di rete LSTM.

Grafico 26. Confronto tra volatilità effettive e predette dalla rete neurale con CNN





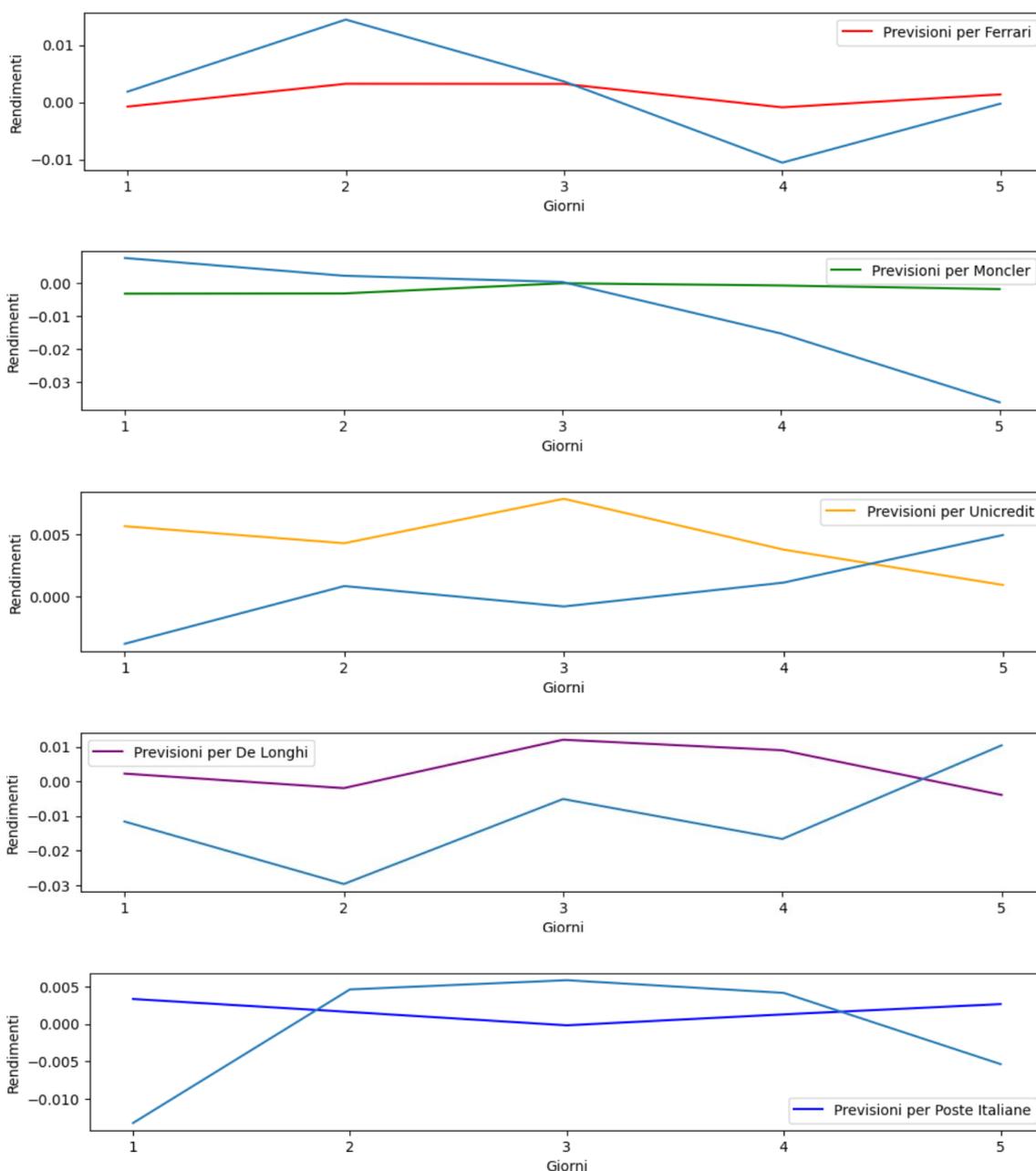
È interessante notare come, a differenza dei primi tre titoli, di cui le informazioni sulla volatilità non sono rilevanti, per De Longhi e Poste Italiane la rete CNN effettui delle previsioni che oscillano molto; questo significa che l'algoritmo di rete neurale comprende che queste due serie storiche sono diverse dalle prime tre e si adatta formulando delle previsioni più variabili.

### 3.1.1 Previsione a n passi

Il prossimo step dell'analisi sui rendimenti è la previsione a n passi, costruita utilizzando la previsione step by step sulle prossime potenziali realizzazioni della serie storica dei rendimenti. Il mio obiettivo, per ciò che riguarda i dati giornalieri, è quello di fare delle previsioni a breve termine, data la difficoltà di partenza dei modelli a captare informazioni sui rendimenti con un campione già in possesso. Il modello utilizzerà perciò le istruzioni di partenza aggiungendo degli strati in input, che corrispondono agli stessi rendimenti che il modello passo dopo passo prevede che si manifestino

I risultati della rete Long Short Term Memory e Gated Recurrent Unit non verranno mostrati poiché non risultano soddisfacenti; le previsioni a n passi fornite da questi due modelli non seguono l'andamento dei rendimenti out-of-sample di nessuno dei cinque titoli selezionati, ma sembrano fornire delle linee rette con variazioni minime dei rendimenti.

Grafico 27. Previsione a cinque passi utilizzando il modello CNN



Nel caso delle previsioni offerte dalla rete CNN è stato riscontrato un buon livello nella capacità di prevedere non tanto i rendimenti, quanto i rialzi e ribassi del titolo nel caso della società De Longhi e Ferrari; risulta infatti sufficiente la capacità di prevedere questo aspetto più che il valore estemporaneo del rendimento. Rimane ad ogni modo il problema che questo modello non si adatta ugualmente bene agli altri titoli, non confermando quanto detto per il titolo De Longhi e Ferrari.

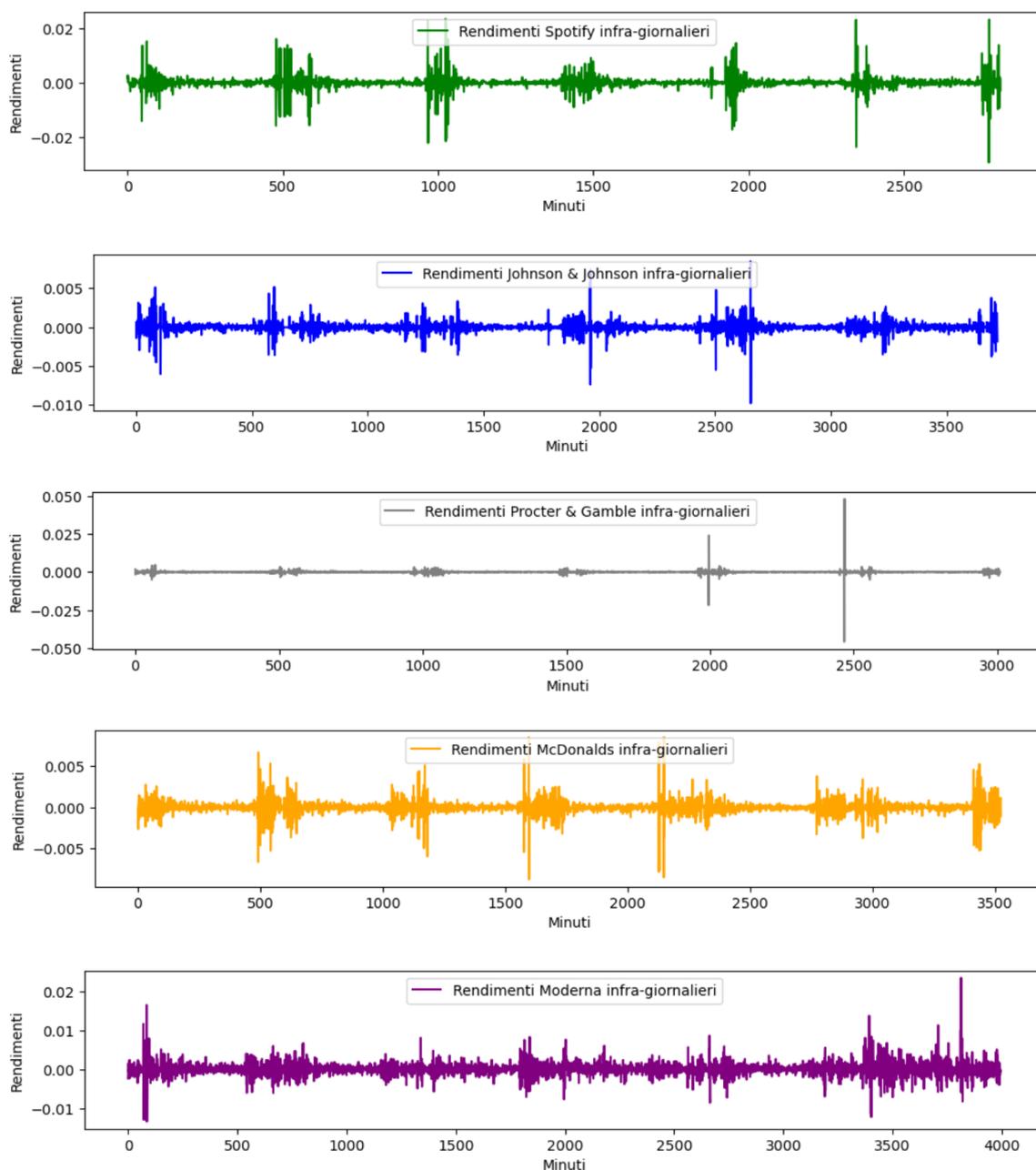
In considerazione dell'analisi fatta fino a questo punto, ritengo di poter confermare la difficoltà anche di un modello non parametrico come le reti neurali artificiali di determinare il rendimento o movimento futuro di uno specifico stock di mercato in un contesto di dati giornalieri, su cui vi è libero accesso in siti come "Yahoo Finance"; sembra sussistere inoltre un'ottima capacità di assorbimento delle informazioni finanziarie e conseguente aggiustamento dei prezzi in un arco temporale sicuramente inferiore al giorno lavorativo, rendendo più difficile intercettare extra-profitti che i modelli potrebbero prevedere. La rapidità e l'imprevedibilità del mercato finanziario, nel cui contesto anche modelli che superano i limiti di algoritmi parametrici per l'identificazione del GDP di serie storiche finanziarie sembrano faticare nel riconoscere dei pattern prevedibili, rendono questo tipo di analisi statistica di difficile utilità pratica.

### **3.2 Previsioni utilizzando dati infragiornalieri**

Nell'ambito di un panorama finanziario sempre più aggiornato e rapido nei cambiamenti (anche a causa di un'incertezza di prospettive per il futuro), i dati ad alta frequenza hanno iniziato a prendere sempre più piede nella considerazione dell'evoluzione delle serie storiche finanziarie; è da verificare se sia possibile individuare qualche tipo di inefficienza di mercato in questa categoria di transazioni finanziarie, ipotizzando che sia più probabile riconoscere pattern di rendimenti in questo tipo di frequenze rispetto ai dati giornalieri. Il tracciamento in tempo reale e l'immediatezza di informazioni sui prezzi di migliaia di titoli finanziari quotati è permesso dalla sempre maggiore capacità di calcolo garantita da software e terminali sempre più potenti.

Per motivi di liquidità dei titoli e di accessibilità di quest'ultimi sono stati selezionati cinque titoli su cui fare inferenza diversi dai precedenti ovvero Spotify, Johnson & Johnson, Procter & Gamble, McDonalds e Moderna.

Grafico 28. Rendimenti dei cinque titoli



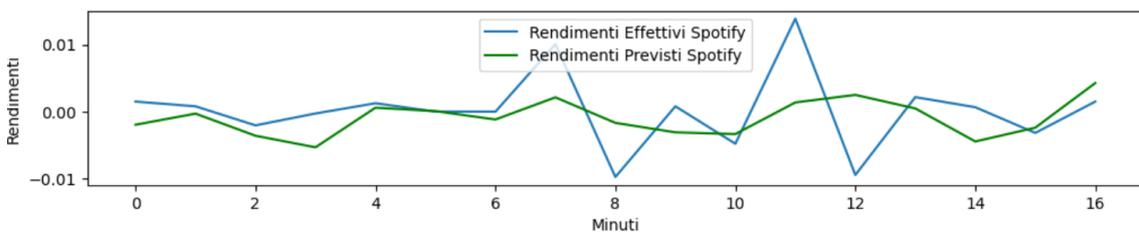
Per reperire le informazioni sui titoli finanziari è stato utilizzato il sito “alphavantage”, che permette una comoda fruizione di contenuti di questo tipo, nello specifico di questo elaborato sono stati selezionati di prezzi dei titoli intra-giornalieri sopracitati con intervallo scelto di un minuto.

Com'è possibile vedere dal grafico 28, sembra essere presente una ripetitività nei picchi dei rendimenti scandita circa ogni cinquecento osservazioni (8 ore), il che coincide con la fase di apertura e chiusura dei mercati finanziari; sono state collezionate le informazioni

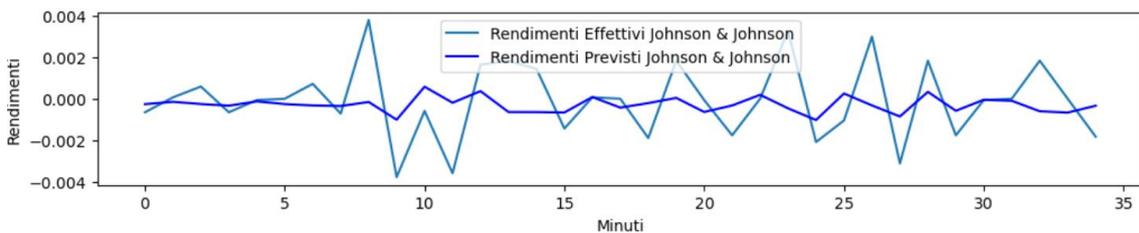
dei titoli che vanno dalle ore 04:00 alle 20:00. Questi orari coincidono con la fase di pre e post asta (04:00-09:00 e 16:00-20:00), ovvero la fase che precede l'asta di apertura e posticipa la chiusura sui mercati gestiti nella borsa statunitense, e la fase di negoziazione a mercato aperto, dalle 09:00 alle 16:00.

A questo riguardo, di seguito vengono mostrate le previsioni sul testing set sfruttando il modello convoluzionale, per ogni titolo selezionato; le timeline differiscono in quanto i diversi titoli possono essere stati scambiati ad orari e frequenze diverse e talvolta è necessario aspettare qualche minuto perché il prezzo venga aggiornato. Il titolo più liquido è Microsoft, dove per ogni minuto viene registrato almeno uno scambio di titoli.

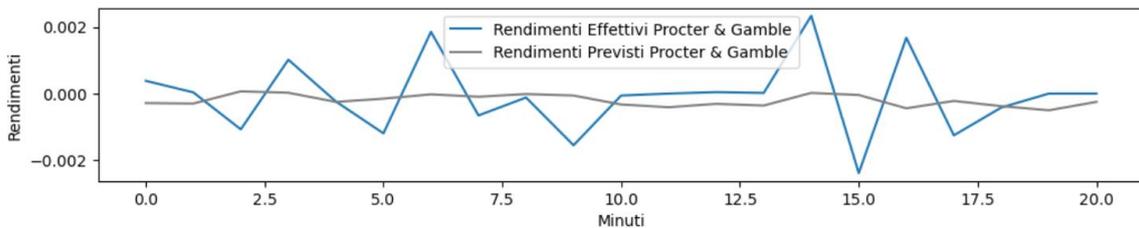
Grafico 29. Confronto tra rendimenti effettivi e predetti dalla rete neurale con CNN e accuratezza dei modelli



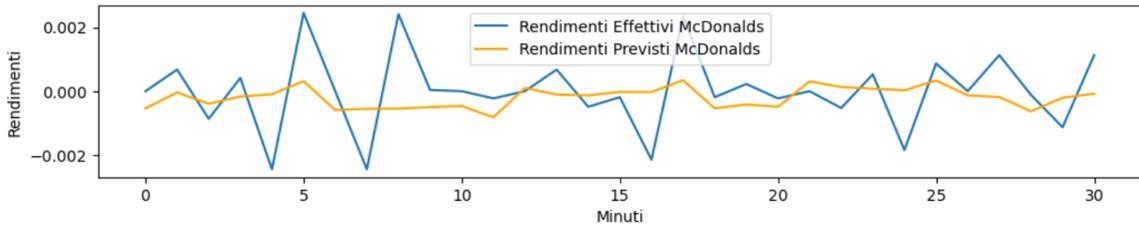
Numero di previsioni corrette: 12  
 Numero totale di previsioni: 17  
 Accuratezza modello CNN : 70.59%



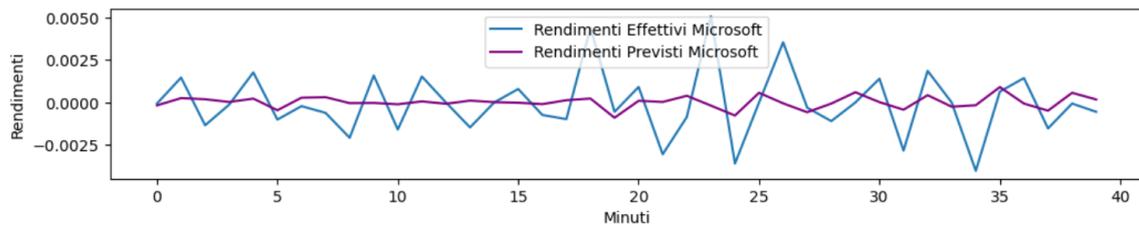
Numero di previsioni corrette: 23  
 Numero totale di previsioni: 35  
 Accuratezza modello CNN: 65.71%



Numero di previsioni corrette: 14  
 Numero totale di previsioni: 21  
 Accuratezza modello CNN: 66.67%



Numero di previsioni corrette: 19  
 Numero totale di previsioni: 31  
 Accuratezza modello CNN: 61.29%

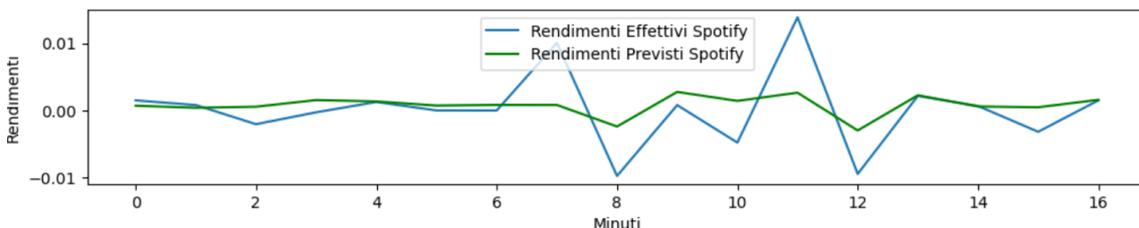


Numero di previsioni corrette: 22  
 Numero totale di previsioni: 40  
 Accuratezza modello CNN: 55.00%

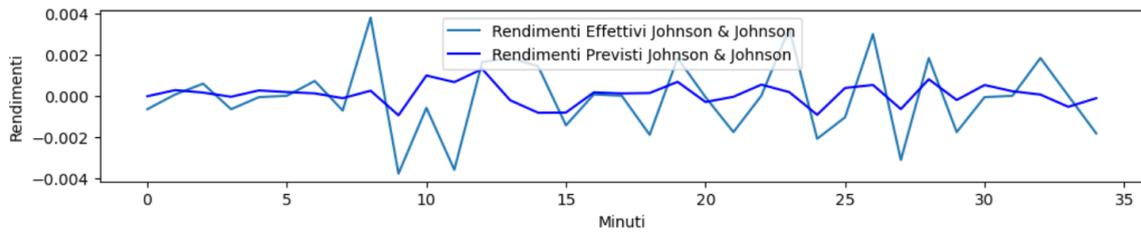
Il modello mostra una buona capacità di adattamento ai dati empirici restituendo statistiche di accuratezza altrettanto buone, raggiungendo nel caso del titolo Spotify un'accuratezza dei movimenti dei rendimenti del settanta per cento.

È da evidenziare il fatto che, stando ai risultati in figura, sembra esserci una relazione tra accuratezza e dimensione effettiva del testing set (basandosi su prelevamenti di campioni di diversa dimensione); minore è il numero dei rendimenti da prevedere maggiore è l'accuratezza. Questo potrebbe accadere perché il modello può perdere efficacia in un contesto di previsione a maggiore distanza temporale dalla fine del training set e perciò dell'addestramento.

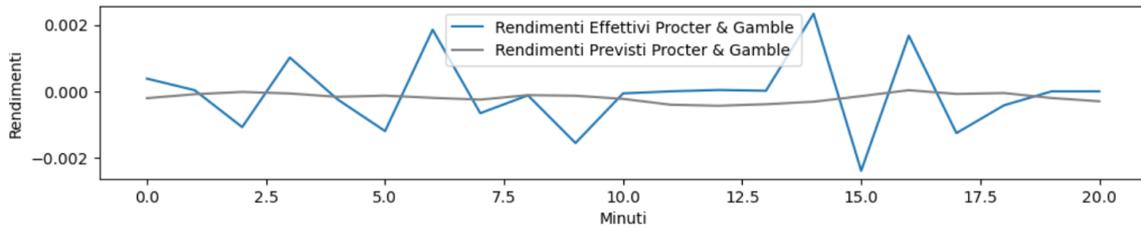
Grafico 30. Confronto tra rendimenti effettivi e predetti dalla rete neurale con LSTM e accuratezza dei modelli



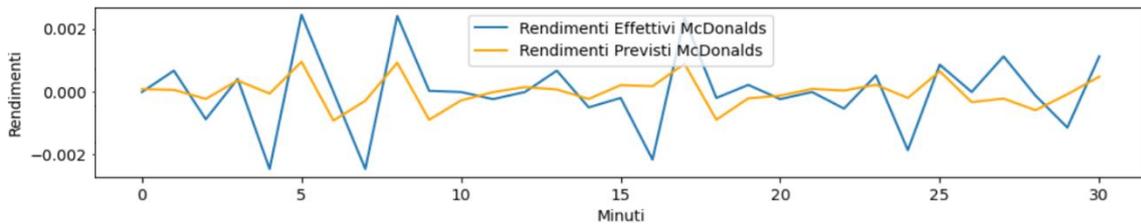
Numero di previsioni corrette: 12  
 Numero totale di previsioni: 17  
 Accuratezza modello LSTM: 70.59%



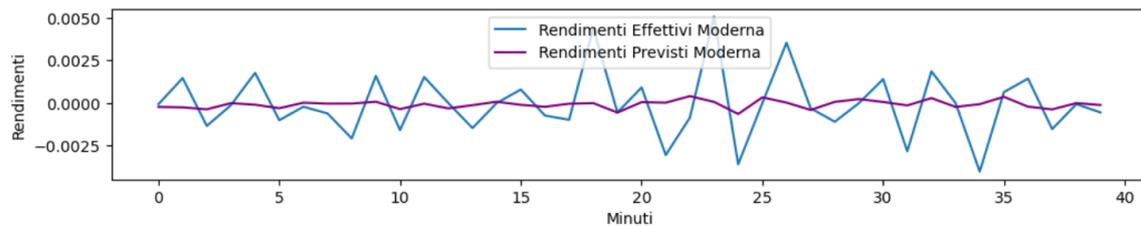
Numero di previsioni corrette: 23  
 Numero totale di previsioni: 35  
 Accuratezza modello LSTM: 65.71%



Numero di previsioni corrette: 14  
 Numero totale di previsioni: 21  
 Accuratezza modello LSTM: 66.67%



Numero di previsioni corrette: 22  
 Numero totale di previsioni: 31  
 Accuratezza modello LSTM: 70.97%

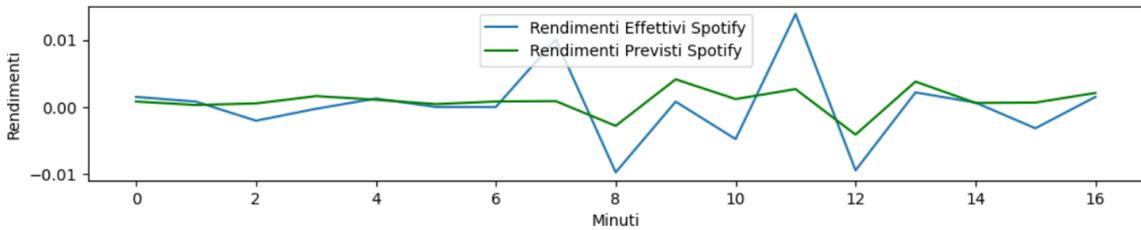


Numero di previsioni corrette: 27  
 Numero totale di previsioni: 40  
 Accuratezza modello LSTM: 67.50%

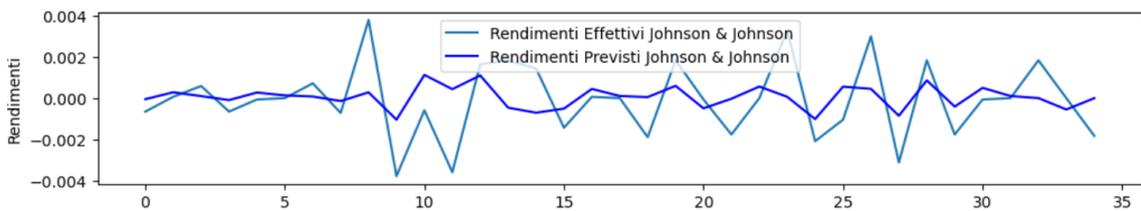
Com'è ben visibile dai grafici e dai risultati di accuratezza, questo modello ha performance leggermente migliori del precedente modello CNN, attestandosi mediamente sopra un sessantacinque per cento di media di previsioni corrette considerando tutti i titoli. Ad eccezione del titolo Procter & Gamble e Moderna, la rete neurale segue abbastanza bene anche i picchi (positivi e negativi) di rendimento, mentre per i due titoli citati precedentemente la forma è maggiormente smussata attorno al valore 0.

Concludendo l'analisi sulle previsioni sui rendimenti verranno presentati i risultati raggiunti utilizzando il modello Gated Recurrent.

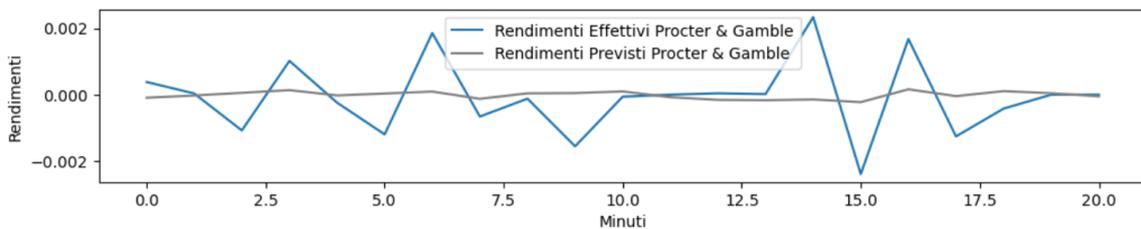
Grafico 31. Confronto tra rendimenti effettivi e predetti dalla rete neurale con GRU e accuratezza dei modelli



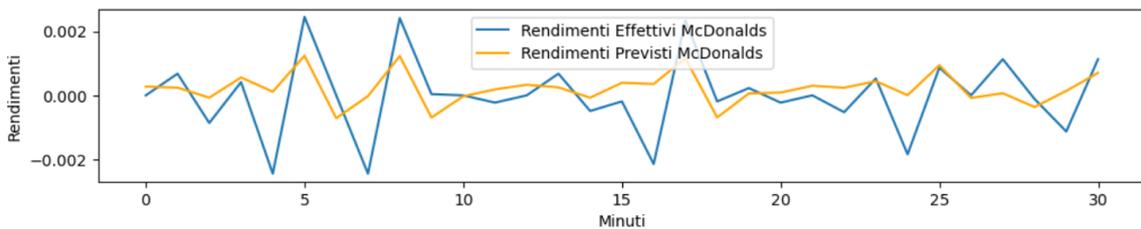
Numero di previsioni corrette: 12  
 Numero totale di previsioni: 17  
 Accuratezza modello GRU : 70.59%



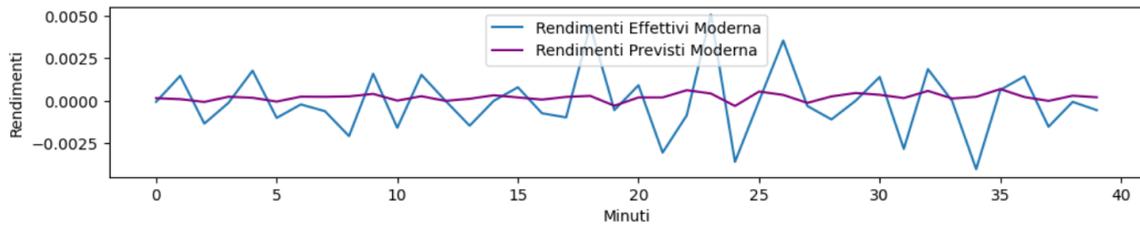
Numero di previsioni corrette: 23  
 Numero totale di previsioni: 35  
 Accuratezza modello GRU : 65.71%



Numero di previsioni corrette: 9  
 Numero totale di previsioni: 21  
 Accuratezza modello GRU : 42.86%



Numero di previsioni corrette: 19  
 Numero totale di previsioni: 31  
 Accuratezza modello GRU : 61.29%

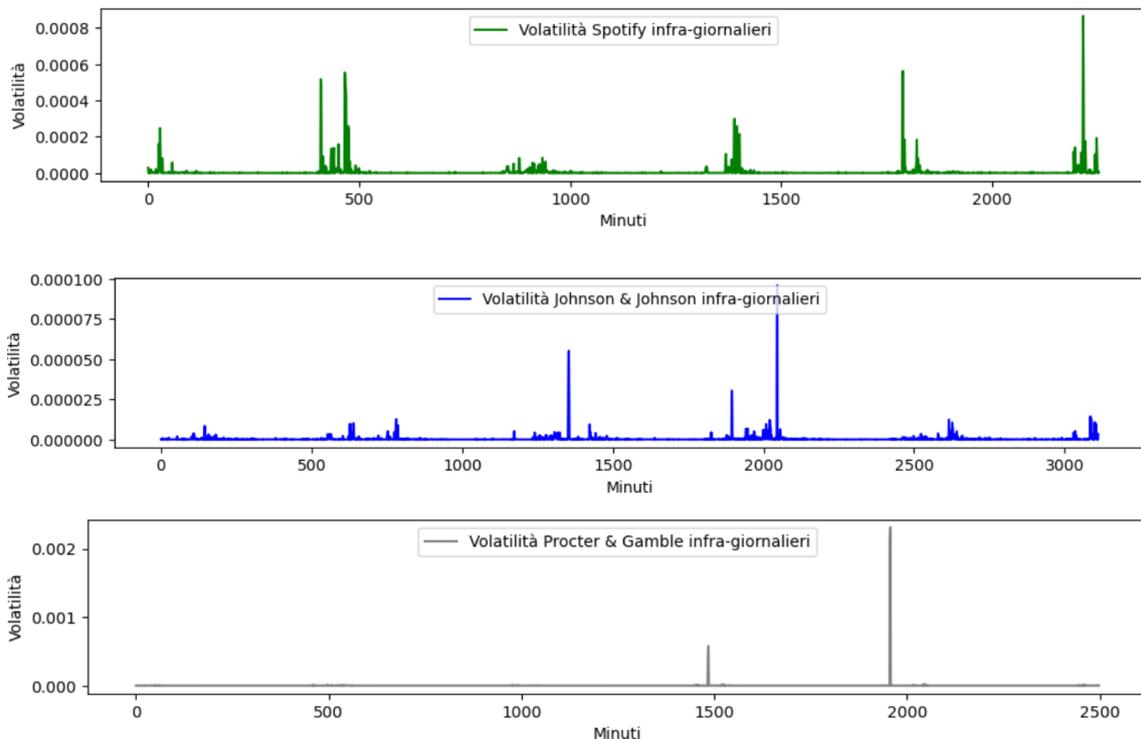


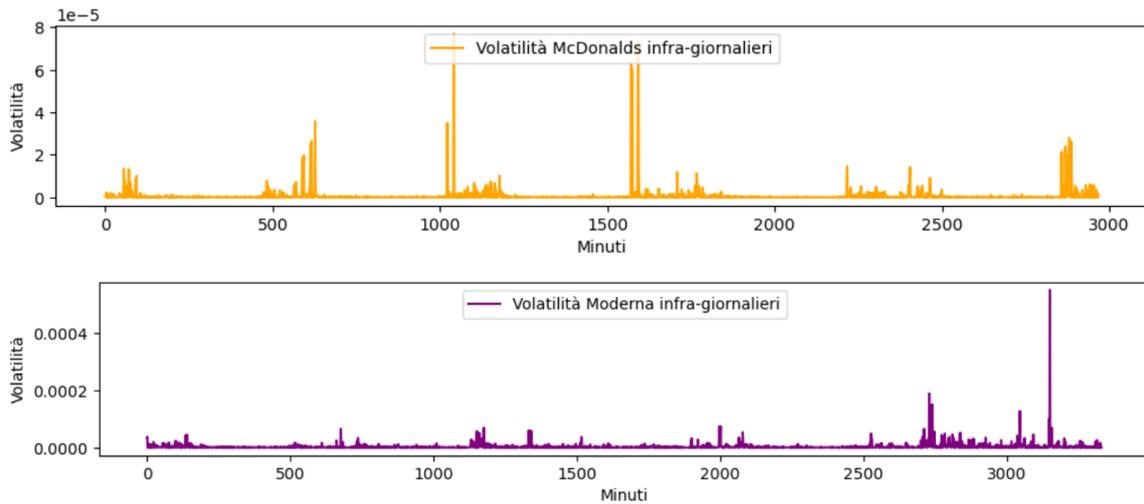
Numero di previsioni corrette: 20  
 Numero totale di previsioni: 40  
 Accuratezza modello GRU : 50.00%

Il modello GRU sembra raggiungere livelli leggermente più bassi rispetto al modello basato sulla stessa tipologia di rete ricorrente precedente, nonostante i risultati dei primi due titoli siano gli stessi dei due modelli precedenti. Questo fenomeno può essere dovuto all'assenza di una specifica cella deputata a contenere la memoria di lungo termine della rete, anche se non sono presenti grosse differenze rispetto al modello LSTM.

Da evidenziare la difficoltà del modello a identificare i pattern di rendimento per i titoli di Procter & Gamble e Moderna, come successo nel caso della rete CNN.

Grafico 32. Volatilità dei cinque titoli

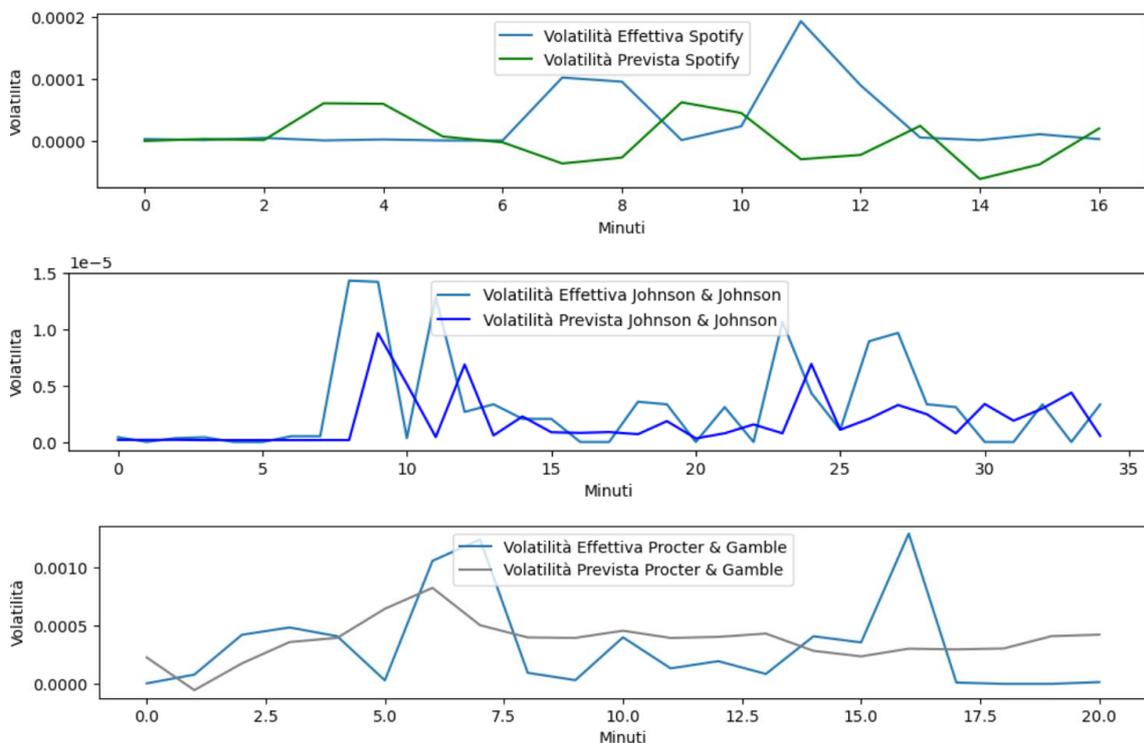


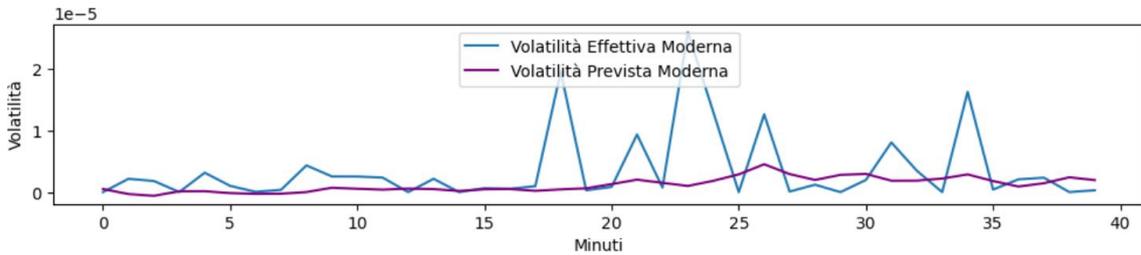
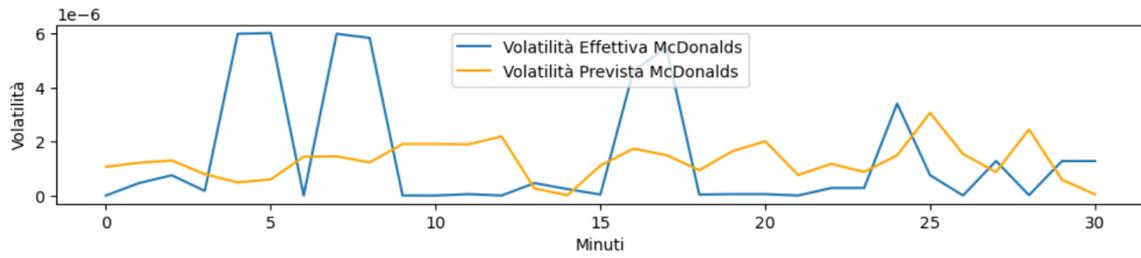


Dal grafico 32 la variabilità dei rendimenti emerge maggiormente rispetto al grafico 26; questa avviene con una ripetitività coerente all'apertura e chiusura dei mercati, come detto in precedenza.

Di seguito vengono riportati i risultati sull'analisi della volatilità.

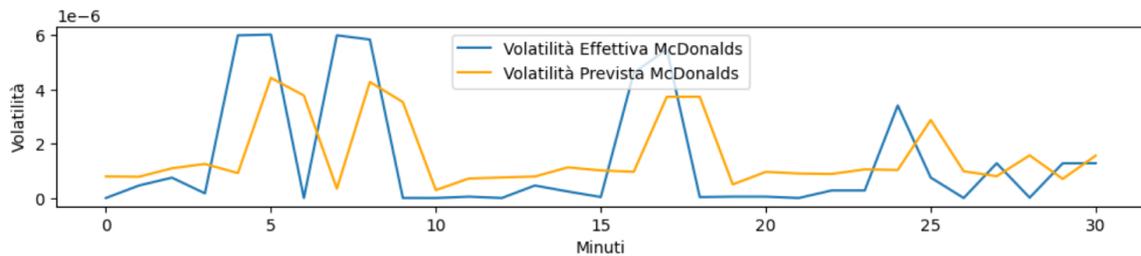
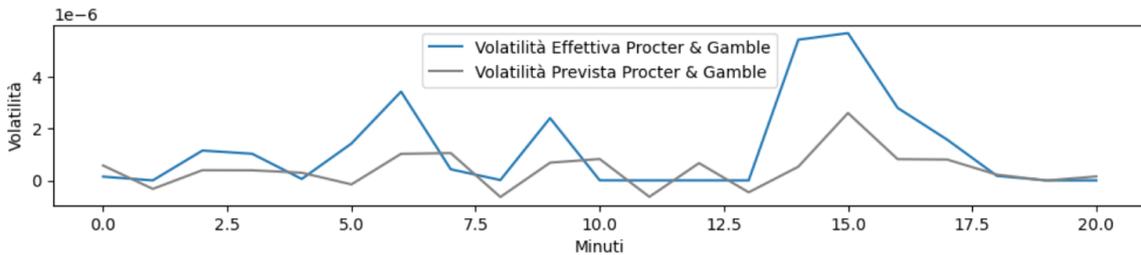
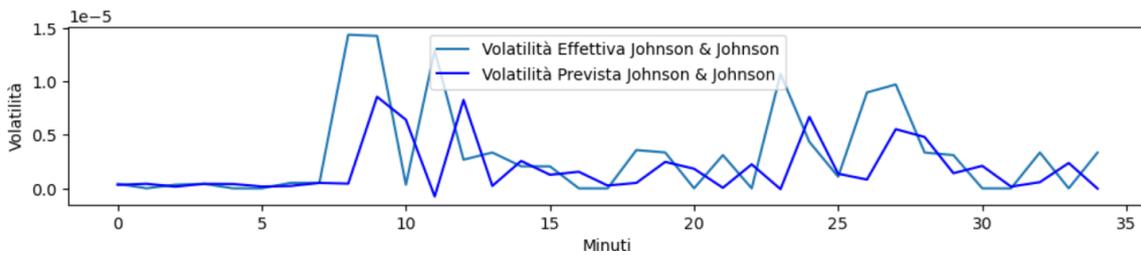
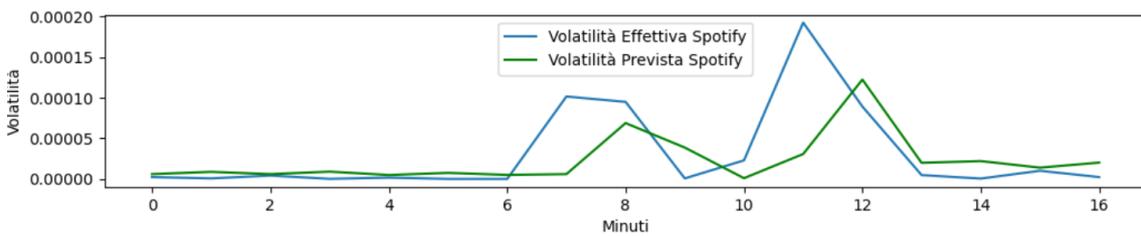
Grafico 33. Confronto tra volatilità effettiva e predetta dalla rete neurale con CNN

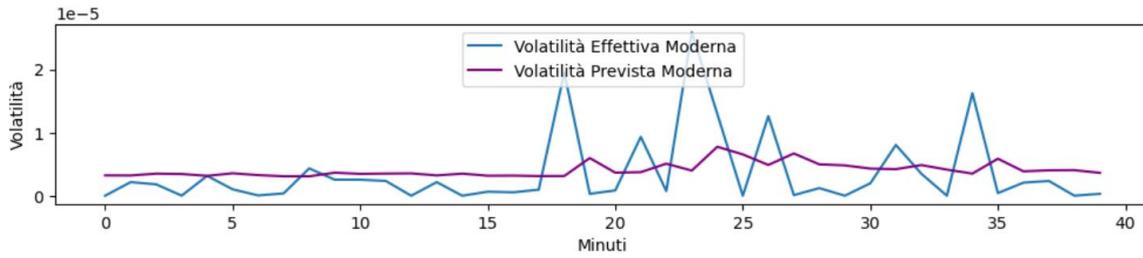




Per quanto concerne la previsione sulla volatilità utilizzando il modello convoluzionale, il modello sembra captare i picchi di volatilità nei minuti successivi rispetto alla realizzazione empirica, come si può ben vedere nel caso di P&G e McDonalds.

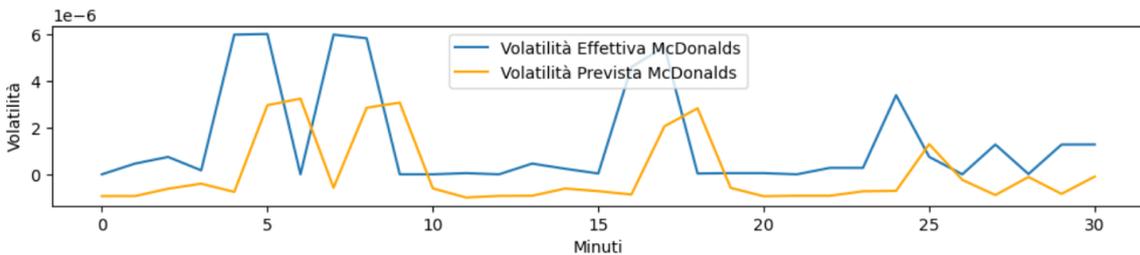
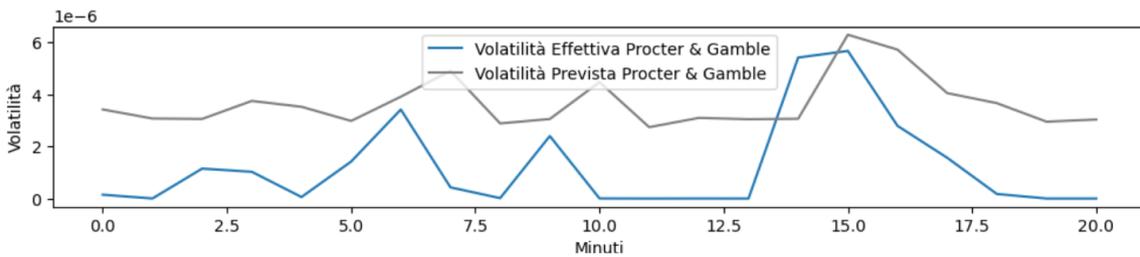
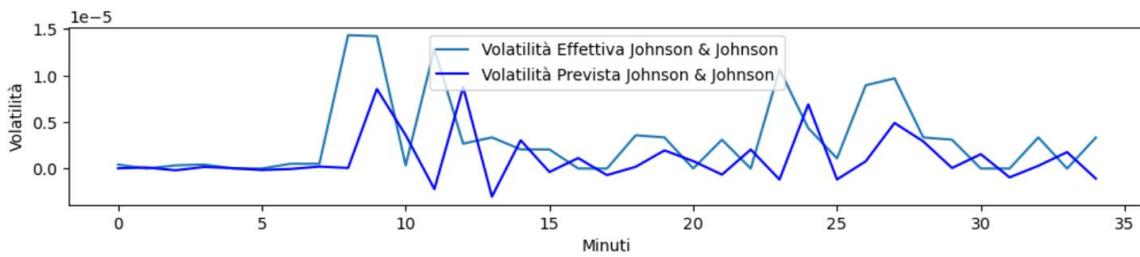
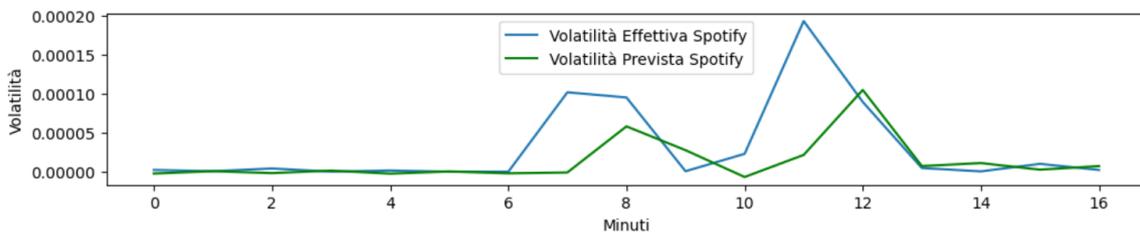
Grafico 34. Confronto tra volatilità effettiva e predetta dalla rete neurale con LSTM

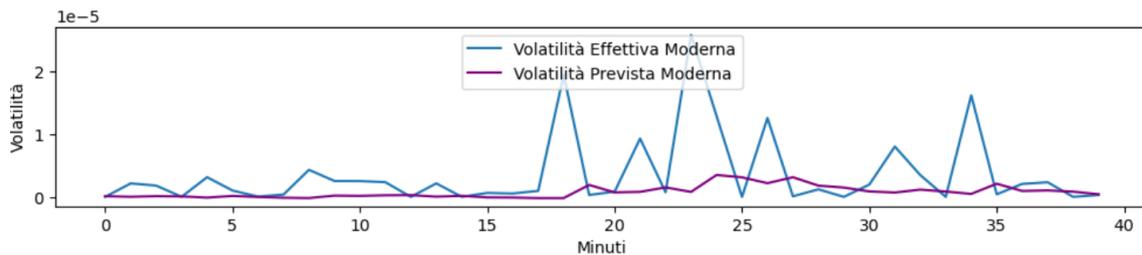




In ultima istanza l'analisi sulla volatilità sembra essere più accurata rispetto al modello precedente (seppur il modello aggiusti le volatilità nei minuti successivi rispetto a ciò che avviene empiricamente).

Grafico 35. Confronto tra volatilità effettiva e predetta dalla rete neurale con GRU



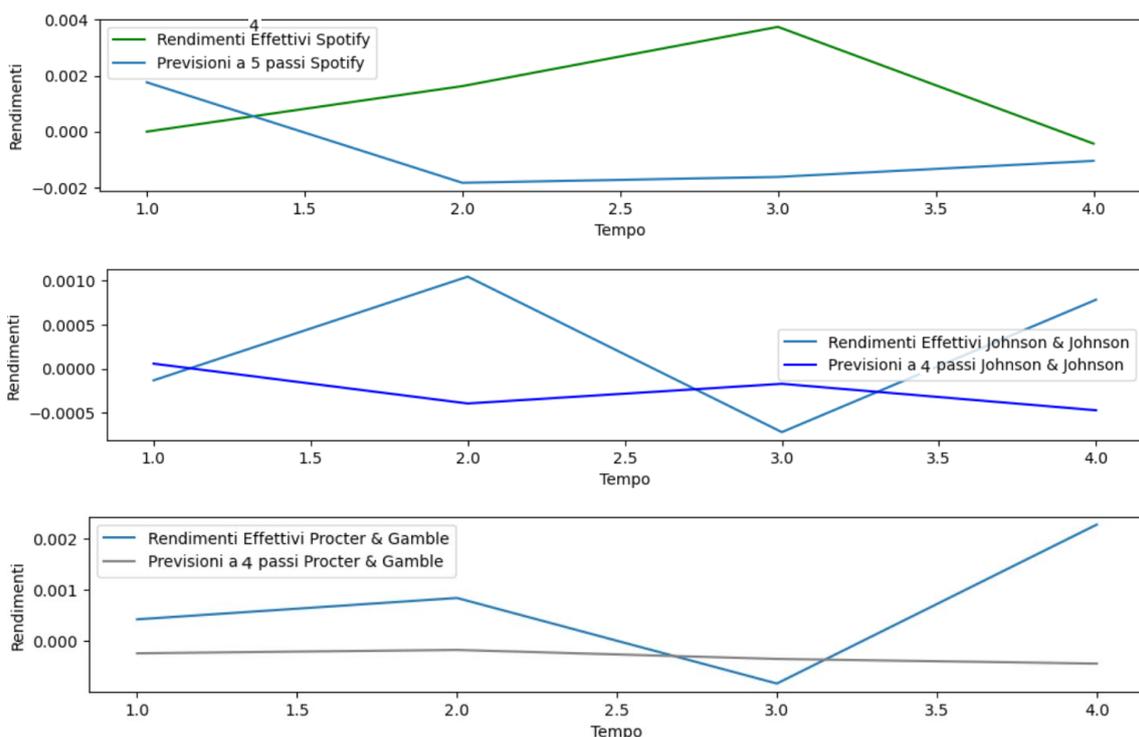


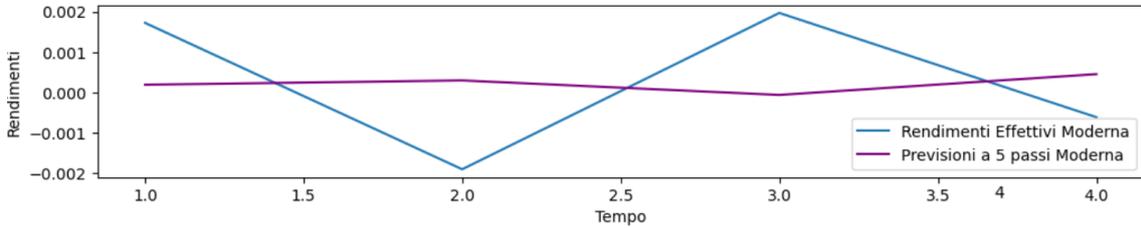
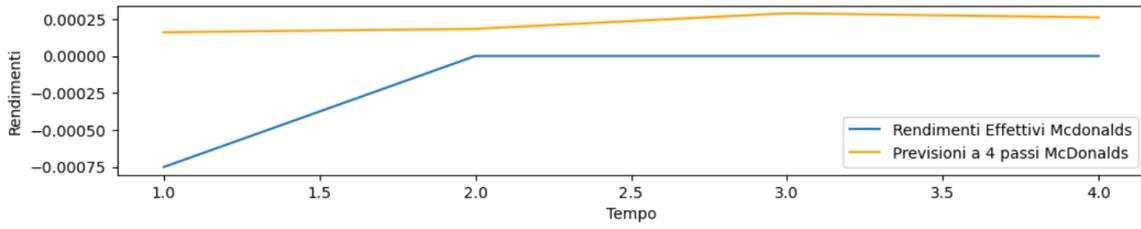
Il modello, per ciò che riguarda la volatilità, segue in linea di massima ciò che è stato detto in precedenza, con le previsioni sulla volatilità che faticano a percepire queste variazioni improvvise, che rappresentano degli outliers. Le previsioni utilizzando questo modello sono simili a quelle ottenute facendo uso del modello Long Short Memory.

### 3.2.1. Previsione a n passi

Come realizzato per i pattern di rendimenti giornalieri, può essere utile comprendere come si comporta il modello sulla base dell'idea di realizzare una serie di previsioni "a scatola chiusa"; per fare ciò il modello utilizzerà anche i nuovi rendimenti predetti, realizzando quindi una previsione completamente fuori dal sample di dati utilizzato per il training, la validazione e il test.

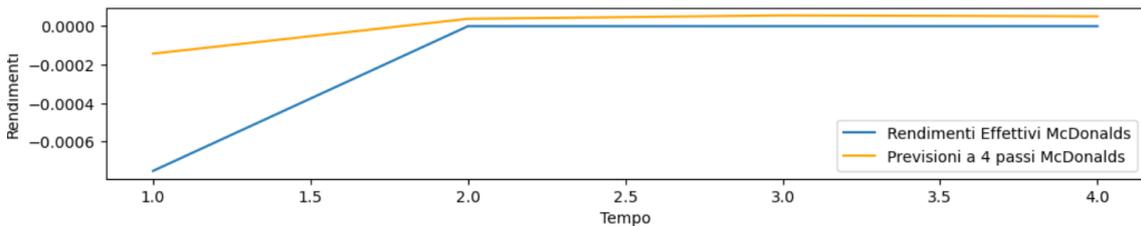
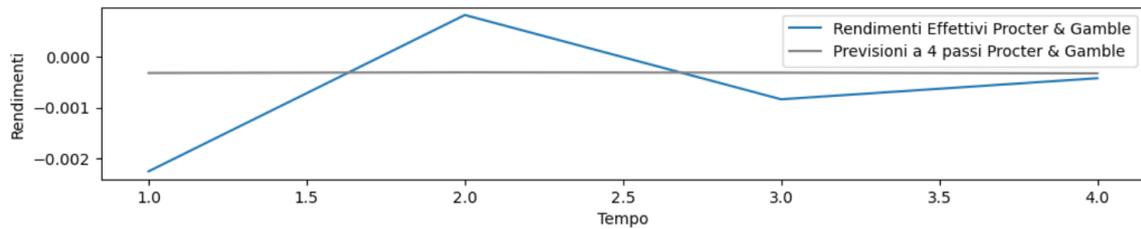
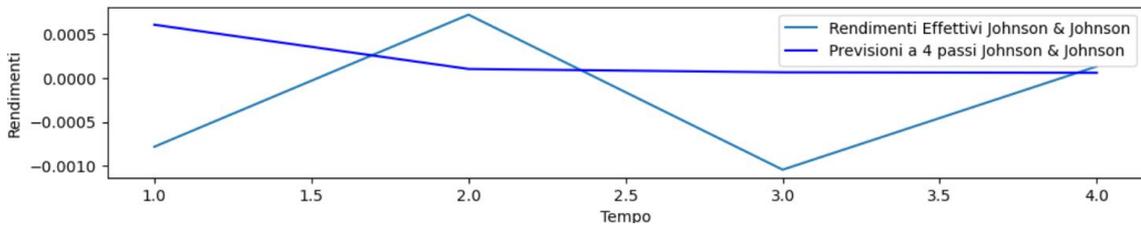
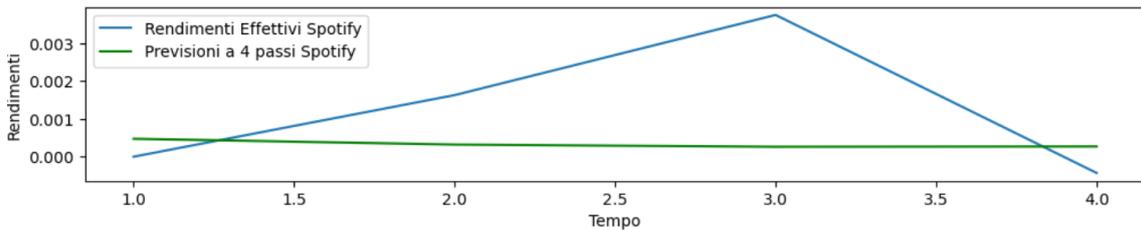
Grafico 36. Previsione a quattro passi per i rendimenti con modello CNN

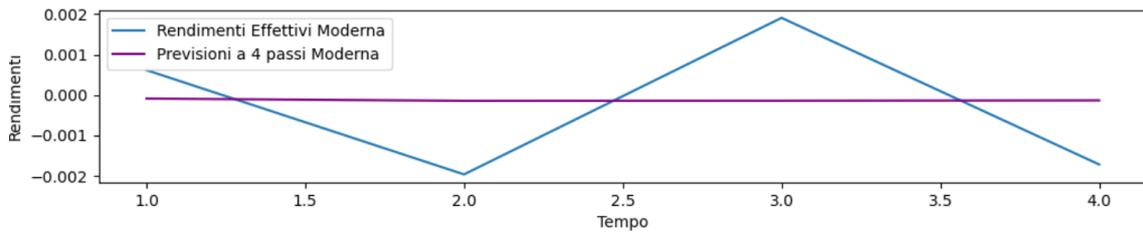




Dal grafico 36 emerge una generalizzata difficoltà a comprendere i futuri movimenti di prezzo dei diversi titoli, ad eccezione del titolo McDonald, la cui previsione sembra seguire con logica l'andamento futuro dei rendimenti.

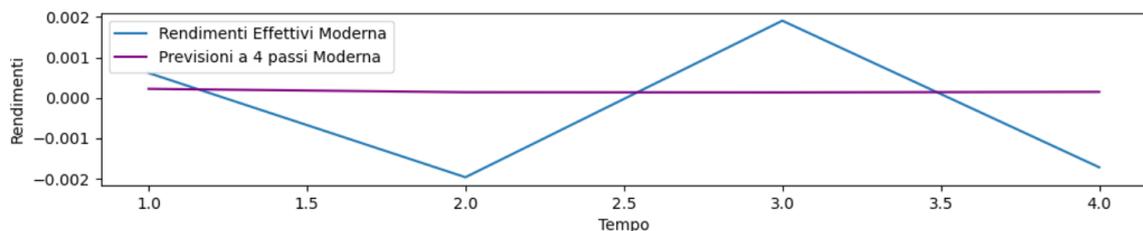
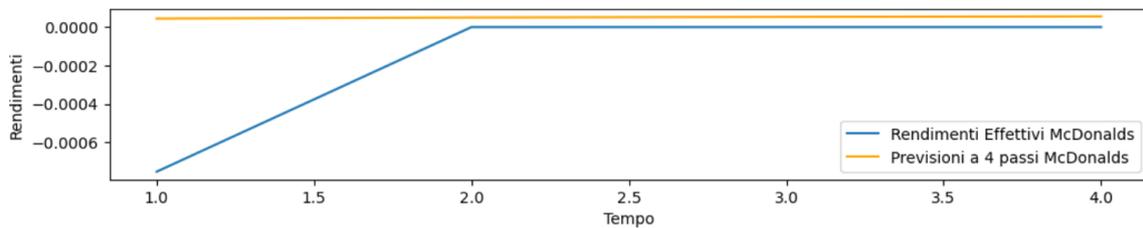
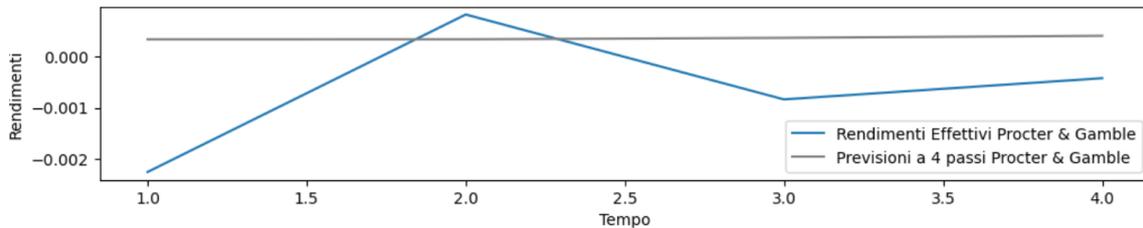
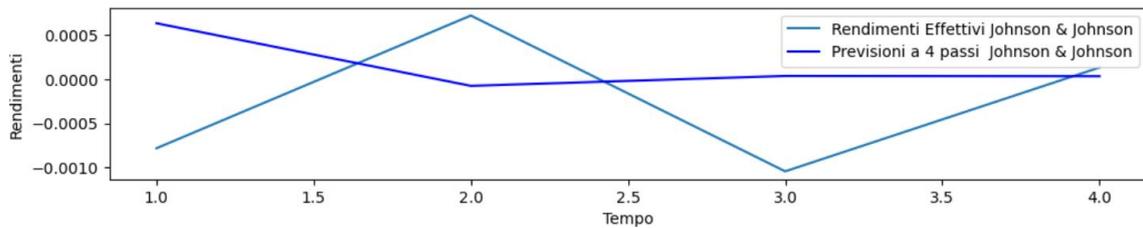
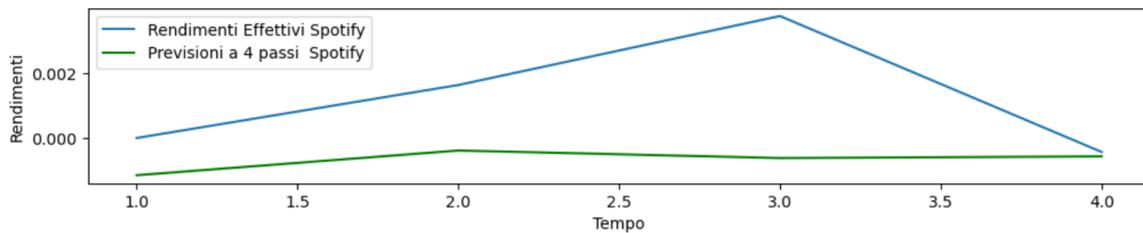
Grafico 37. Previsione a quattro passi dei rendimenti con modello LSTM





La tendenza delle previsioni ad n passi anche in questo caso non dà risposte completamente soddisfacenti, ad eccezione del titolo McDonalds, perché non segue perfettamente l'andamento dei rendimenti, benché in diversi casi venga individuato in modo corretto la natura del rendimento (positivo o negativo).

Grafico 38. Previsione a quattro passi dei rendimenti con modello GRU



Anche questa rete non sembra implementabile per formulare previsioni a breve termine su campioni out-of-sample.

## Capitolo IV – Strategia di trading

In questo capitolo si presenta una strategia di trading che utilizza i segnali di rialzo o ribasso come indicazione di apertura o chiusura di una posizione sui titoli utilizzati nell'analisi, andando a considerare le previsioni sul testing set ottenute nel capitolo III.

Le motivazioni dell'implementazione di questa applicazione sono di voler verificare la fattibilità di una strategia d'investimento in un titolo finanziario seguendo le previsioni della porzione di testing set fatte dai modelli a reti neurali presentati nel capitolo II e implementati nel capitolo III, e verificare in quanti scenari la trading strategy avrebbe portato dei risultati positivi qualora fosse stata utilizzata.

Per realizzare la strategia di trading verranno utilizzati gli stessi titoli del capitolo III, con la distinzione in base alla frequenza con cui vengono registrati i dati: i titoli Ferrari, Moncler, Unicredit, De Longhi e Poste Italiane con frequenza di dati giornaliera mentre Spotify, Johnson & Johnson, Procter & Gamble, McDonalds e Moderna con frequenza di aggiornamento di un minuto. Nel caso di quest'ultima categoria la finestra temporale della strategia non sarà la stessa per tutti i titoli, poiché è possibile che non vengano scambiati ogni minuto e può trascorrere più tempo tra una transazione e l'altra, come spiegato nel capitolo III.

Le ipotesi su cui si basa l'applicazione sono di operare in presenza di un mercato liquido e privo di frizioni, dove si possa acquistare e vendere uno stock nel momento in cui lo si desidera; si assume inoltre l'assenza di tasse e di costi di transazione sulle operazioni effettuate.

È stato opportuno impostare un ciclo for che andasse a reiterare la procedura di acquisto o vendita del titolo controllando che la previsione del testing set fosse di un rendimento positivo o negativo.

Nel caso in cui la previsione per il giorno/minuto successivo risulti positiva, il segnale trasmesso per il giorno corrente sarà di acquisto, mentre viceversa sarà venduto lo stock in previsione di rendimenti negativi.

Input Python 2. Istruzione di un ciclo per sviluppare la strategia di trading

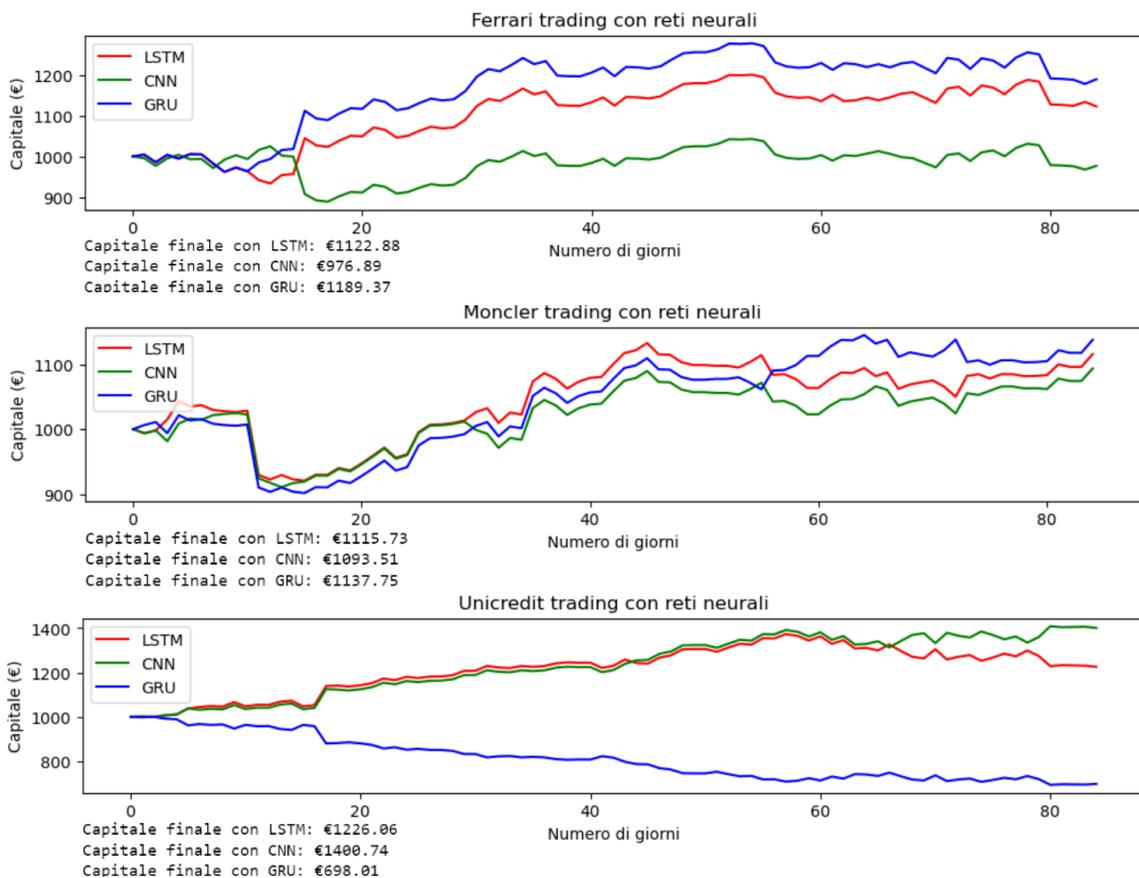
```
# Strategia di trading basata sulla previsione
capitale_iniziale = 1000
capitale = [capitale_iniziale]

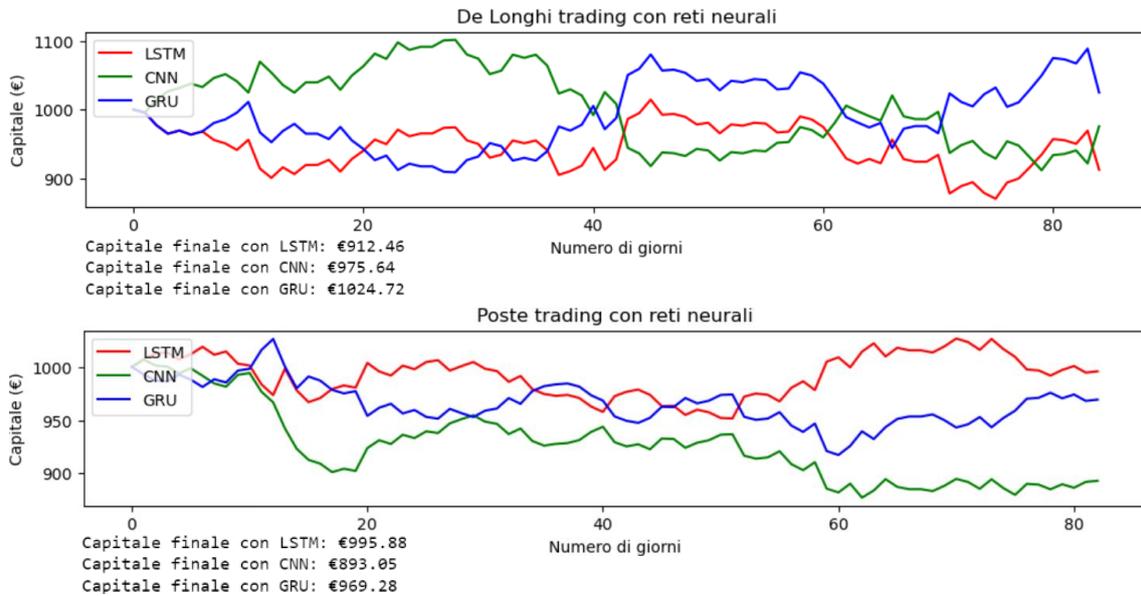
for i in range(0, len(y_test_flat)-1):
    if predictions[i+1] > 0: # Se La previsione è positiva, acquista
        operazione = 1
    elif predictions[i+1] < 0: # Se La previsione è negativa, vendi
        operazione = -1
    else:
        operazione = 0 # Nessuna operazione se il segnale è debole

# Calcolo del nuovo capitale basato sul rendimento reale e sull'operazione
nuovo_capitale = capitale[-1] * (1 + operazione * y_test_flat[i])
capitale.append(nuovo_capitale2)
```

A questo punto, per ogni titolo, vengono istruiti e allenati i tre modelli di reti neurali LSTM, GRU e CNN definiti in precedenza, e successivamente utilizzati per prevedere il movimento del prezzo dei titoli.

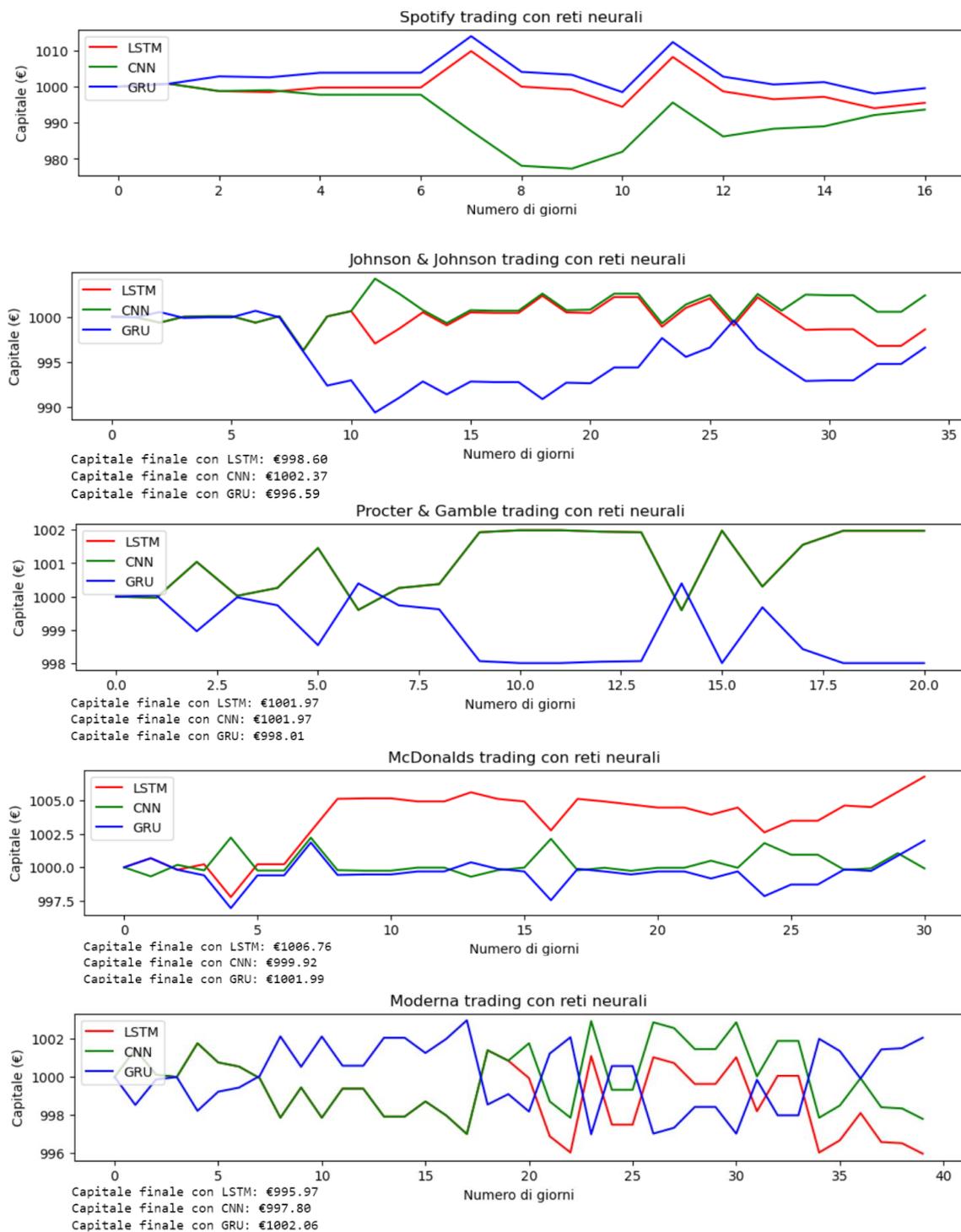
Grafico 39. Equity line delle reti neurali applicata ai dati giornalieri





Analizzando nel dettaglio i grafici, possiamo notare come in tre casi su cinque il modello GRU permetta di raggiungere dei livelli di trading capital più elevati rispetto alle reti LSTM e CNN. Questa rete non si rivela tuttavia affidabile al 100%: esaminando, infatti, i risultati per il titolo Unicredit, emerge una ingente perdita di capitale totalmente sproporzionata rispetto agli altri due modelli di rete neurale; questo dato è comprensibile guardando ai risultati delle previsioni dei rendimenti del titolo utilizzando la rete GRU (grafico 21 del capitolo II), che arriva ad un livello di accuratezza più basso (44,71%) rispetto a ciò che viene raggiunto dai modelli basati sulle reti LSTM e CNN per il medesimo titolo (rispettivamente 61,18% e 51,76%). In otto casi su quindici (tre i modelli utilizzati per cinque titoli) utilizzare questa strategia si rivela redditizio nella finestra temporale del testing set, un numero buono ma non accettabile per il rischio intrinseco che coinvolge questo tipo di operazione, con possibilità di ingenti perdite nel breve periodo. I titoli di cui è stato compreso maggiormente l'andamento sono Ferrari e Moncler, che chiudono entrambi in positivo a prescindere da modello utilizzato, sintomo di una minore variabilità dei dati e di una maggiore coerenza con i dati della fase di addestramento delle reti.

Grafico 40. Equity line delle reti neurali applicata ai dati infragiornalieri



Una strategia di trading sul brevissimo termine non può portare a risultati completamente soddisfacenti in termini di profitto, poiché le variazioni di prezzo minuto su minuto sono spesso minime; può però far comprendere se la direzione di inferenza intrapresa è quella corretta. Dai grafici riportati le variazioni di capitale sono minime, e in otto casi su quindici

il capitale si aggira attorno al capitale inizialmente investito. Non sono presenti forti squilibri di capitale in una strategia che utilizzi i dati infra-giornalieri.

Questa strategia non ha nessun tipo di effetto sul futuro, ma è stata pensata per verificare gli effetti “reali” dell’applicazione di una rete neurale sulla previsione dei rendimenti di alcuni titoli.

## Conclusioni

Lo scopo della tesi era quello di descrivere e verificare il comportamento delle reti neurali come metodo di machine learning nei confronti di serie storiche finanziarie, utilizzando due set di dati: uno basato su dati giornalieri e uno basato su dati infra-giornalieri.

Se i modelli a liscio esponenziale e i modelli auto-regressivi e a media mobile si sono rivelati insufficienti per trattare le serie dei prezzi e rendimenti a causa di una incapacità di adattamento ai cambiamenti del mercato, i modelli a reti neurali rappresentano una delle opzioni messe a disposizione dal machine learning che hanno la capacità di formulare previsioni grazie ad un corposo set di dati di addestramento.

L'efficienza del mercato, argomento trattato nel primo capitolo dell'elaborato, è un concetto che ha trovato riscontro nella prima fase di studio, quella che trattava i pattern giornalieri; in quel caso, infatti, durante la fase di ricerca degli iper-parametri come la batch size e le epoche di addestramento è stato complesso trovare i valori adeguati per una efficace fase di training e di fitting rispetto al testing set. Questo può indicare che anche un metodo di riconoscimento non lineare di pattern finanziari come possono essere le reti neurali fa fatica a raggiungere livelli di accuratezza e fitting soddisfacenti a causa di un livello di efficienza del mercato che implica la presenza di arbitraggisti e/o operatori che riescono ad assorbire in pochissimo tempo le inefficienze; è pacifico infatti che, a fronte ad esempio di un aumento improvviso ed eccessivo dei prezzi consegua una successiva azione da parte degli operatori che assumeranno posizioni short sugli stessi stock, andando a sgonfiare man mano il prezzo perché la pressione dei venditori è maggiore di quella dei compratori, e viceversa nel caso di una forte ed improvvisa diminuzione dei prezzi.

Per ciò che concerne le rilevazioni ad alta frequenza dei dati finanziari, i risultati emersi sono di pattern i cui andamenti rimangono impossibili da prevedere con esattezza, ma con meno rumore e variabilità; è invece presente una ripetitività periodica dei rendimenti che potrebbe aiutare i modelli a reti neurali a migliorare lo score precedente.

Il processo di sviluppo delle reti neurali artificiali è ancora agli albori e le potenzialità che possono raggiungere, acquisendo una sempre maggiore quantità di dati ed informazioni, potrebbe permettere di prevedere con maggiore accuratezza i movimenti del mercato;

ipotizzando che in futuro questo metodo di machine learning riuscisse a garantire un'affidabilità tale da poter strutturare una strategia di trading redditizia, sarebbe possibile conseguire profitti considerevoli minimizzando il rischio.

Ma per quanto potrebbe durare questa strategia? Considerando che l'evoluzione tecnologica e la standardizzazione dei processi digitali porterà un'ampia accessibilità di nuovi strumenti di intelligenza artificiale, il beneficio che porterebbe l'utilizzo di modelli di machine learning che possono sfruttare tutti gli investitori alla ricerca di extra-profitti sarebbe minimo, anche in quel caso infatti tornerebbe in gioco il concetto di efficienza del mercato; uno scenario in cui tutti hanno la possibilità di effettuare le stesse operazioni nello stesso momento, basate su un algoritmo predittivo strutturato su una rete neurale artificiale, diverrebbe meno remunerativo di ciò che ci si aspetterebbe in partenza. Ipotizziamo che per assurdo esistesse un modello predittivo basato su una rete neurale con un alto grado di accuratezza che prevede rendimenti positivi di un determinato titolo nei prossimi giorni, il segnale nella strategia di trading sarebbe quello di assumere una posizione di acquisto del titolo; se questo metodo fosse utilizzato da tutti gli investitori il prezzo aumenterebbe e la possibilità di ottenere rendimenti positivi diminuirebbe poiché andrebbe a scomparire l'occasione di acquistare uno stock ad un prezzo inferiore al suo reale valore futuro.

La possibilità di trarre profitti diminuirebbe drasticamente e le inefficienze verrebbero intercettate e neutralizzate in un minor intervallo di tempo rispetto agli standard attuali, e la reale differenza la farebbe chi è in possesso della capacità computazionale più elevata per processare dati utilizzando lo stesso modello basato su una rete neurale.

## Bibliografia e Sitografia

### Introduzione

1. Fama, E. F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *The Journal of Finance*, 25(2), pp. 383–417.  
<https://doi.org/10.2307/2325486>
2. Fama, E. F. (1998), "Market efficiency, long-term returns, and behavioral finance", *Journal of Financial Economics*, Volume 49(3), pp. 283-306  
[https://doi.org/10.1016/S0304-405X\(98\)00026-9](https://doi.org/10.1016/S0304-405X(98)00026-9)
3. Tobin, J. (1984), "On the efficiency of the financial system", New York, *Lloyds Bank Review*, No. 153.
4. Blake, C. R., Elton, E. J., & Gruber, M. J. (1993), "The Performance of Bond Mutual Funds", *The Journal of Business*, 66(3), pp. 371–403.  
<http://www.jstor.org/stable/2353206>
5. Keown, A. J. & Pinkerton, J. M. (1981). "Merger announcements and insider trading Activity: An Empirical investigation". *The Journal of Finance*, 36(4), 855–869. <https://doi.org/10.1111/j.1540-6261.1981.tb04888.x>
6. Werner, F. M. De Bondt & Thaler, R. (1985), "Does the Stock Market Overreact? ", *The Journal of Finance*, 40(3), pp. 793–805.  
<https://doi.org/10.2307/2327804>
7. Kahneman, D. & Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk. *Econometrica*", 47(2), p. 12.  
<https://doi.org/10.2307/1914185>
8. Kampman, T. (2012), "Behavioral Finance: The January effect", Tilburg University
9. Rozeff, M. & Kinney, W. (1976), "Capital market seasonality: The case of stock returns", *Journal of Financial Economics*, 3(4), pp. 379-402.  
[https://doi.org/10.1016/0304-405X\(76\)90028-3](https://doi.org/10.1016/0304-405X(76)90028-3)
10. Hiller, D. & Marshall, A. (2002), "Insider trading, tax-loss selling, and the turn-of the-year effect", *International Review of Financial Analysis*, 11(1), pp. 73-84. [https://doi.org/10.1016/S1057-5219\(01\)00065-5](https://doi.org/10.1016/S1057-5219(01)00065-5)

11. Basu, S. (1977), "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis", *The Journal of Finance*, 32(3), 663–682.  
<https://doi.org/10.2307/2326304>
12. Banz, R. W. (1981). "The relationship between return and market value of common stocks". *Journal of Financial Economics*, 9(1), 3–18.  
[https://doi.org/10.1016/0304-405x\(81\)90018-0](https://doi.org/10.1016/0304-405x(81)90018-0)
13. Keim, D. (1983), "Size-related anomalies and stock return seasonality: Further empirical evidence", *Journal of Financial Economics*, Volume 12(1), pp. 13-32. [https://doi.org/10.1016/0304-405X\(83\)90025-9](https://doi.org/10.1016/0304-405X(83)90025-9)

### Capitolo I

1. Treccani, *Enciclopedia della matematica* (2013).  
[https://www.treccani.it/enciclopedia/processo-stocastico\\_\(Enciclopedia-della-Matematica\)/](https://www.treccani.it/enciclopedia/processo-stocastico_(Enciclopedia-della-Matematica)/)
2. Holt, C. C. (2004). "Forecasting seasonals and trends by exponentially weighted moving averages". *International Journal of Forecasting*, 20(1), 5–10. <https://doi.org/10.1016/j.ijforecast.2003.09.015>
3. Little, A. & Brown, R. (1956), "Exponential smoothing for predicting demand", Philip Morris Records, Master Settlement Agreement.  
<https://www.industrydocuments.ucsf.edu/docs/jzlc0130>
4. Treccani, *Dizionario di Economia e Finanza* (2012).  
[https://www.treccani.it/enciclopedia/stagionalita\\_\(Dizionario-di-Economia-e-Finanza\)/](https://www.treccani.it/enciclopedia/stagionalita_(Dizionario-di-Economia-e-Finanza)/)
5. Pegels, C. C. (1969). "Exponential Forecasting: Some New Variations". *Management Science*, 15(5), 311–315.  
<http://www.jstor.org/stable/2628137>
6. Box, G. & Jenkins, G. (1970) "Time Series Analysis: Forecasting and Control". Holden-Day, San Francisco.

7. Mandelbrot, B. (1963). "The variation of certain speculative prices". *The Journal of Business*, 36(4), 394. <https://doi.org/10.1086/294632>
8. Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
9. Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
10. Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization in the brain". *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
11. Swanson, N. R. & White, H. (1997). "A model selection approach to Real-Time macroeconomic forecasting using linear models and artificial neural networks". *The Review of Economics and Statistics*, 79(4), 540–550. <https://doi.org/10.1162/003465397557123>

## Capitolo II

1. Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit". *Nature*, 405(6789), 947–951. <https://doi.org/10.1038/35016072>
2. Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2017). "Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks". 2017 IEEE 19th Conference on Business Informatics. <https://doi.org/10.1109/cbi.2017.23>
3. Song, Y., Cai, C., Ma, D., & Li, C. (2024). "Modelling and forecasting high-frequency data with jumps based on a hybrid nonparametric regression and LSTM model". *Expert Systems With Applications*, 237, 121527. <https://doi.org/10.1016/j.eswa.2023.121527>
4. Ghadimpour, M. & Ebrahimi, S. B. (2022). "Forecasting Financial Time Series Using Deep Learning Networks: Evidence from Long-Short Term

Memory and Gated Recurrent Unit”. Iranian Journal of Finance, 6(4), 81–  
94. <https://doi.org/10.30699/ijf.2022.313164.1286>