

Relazione per il progetto di Business Intelligence per i Servizi Finanziari, gennaio 2023

Piacente Cristian 866020

Università degli Studi di Milano – Bicocca

Corso di Laurea in Informatica, A.A 2022-23

Sommario

Sommario dei dati utilizzati	3
Scelta stocks	3
Download dei dati da Yahoo! Finance	6
Fusione delle serie in un unico DataFrame	6
Presentazione dei dati con un grafico	8
Statistiche descrittive	11
Rendimento cumulato e composto annuo	11
Rendimenti semplici e logaritmici	12
Presentazione dei rendimenti con istogrammi	15
Grafici diagnostici a 4 sezioni	16
Statistiche descrittive univariate	20
Matrice di covarianze	22
Matrice di correlazione	23
Andamento correlazioni titoli	24
Grafici di dispersione correlazioni medie	26
Analisi di previsione.....	27
Test Adfuller	27
Trend, seasonality e resid	27
Grid search	31
SARIMAX.....	31
MSE e RMSE.....	34
Strategie di trading e backtesting	35
Costruzione strategia	35
Backtesting	36

Rendimenti e confronto con Buy & Hold	37
CAPM	41
Beta di ciascun titolo rispetto al mercato (S&P 500)	41
Esposizione di ciascun titolo ai fattori Fama-French	42
Utilizzo di beta per calcolare rendimento atteso annuo	45
Costruzione di portafoglio	46
Rendimenti passati: metodo di simulazione.....	46
Rendimenti attesi costruiti: metodo di simulazione	50
Rendimenti passati: metodo analitico.....	53
Rendimenti attesi costruiti: metodo analitico	55
Conclusioni	57

Sommario dei dati utilizzati

Scelta stocks

Per la scelta di 6 stocks del mercato americano appartenenti all'indice S&P500, ho deciso di utilizzare i seguenti settori (classificazione della Global Industry Classification Standard):

- **Health Care**
- Consumer Discretionary (sub-industry **Automobile Manufacturers**)
- Communication Services (sub-industry **Movies & Entertainment**).

Ho scelto di analizzare due titoli della sub-industry Automobile Manufacturers perché in questi anni diverse **case automobilistiche** stanno facendo notevoli **progressi** in diverse aree, sia per quanto riguarda tematiche come l'inquinamento dell'ambiente (nuove auto elettriche) sia l'evoluzione della tecnologia con la guida autonoma.

Ho preso in considerazione gli altri due settori (**Health Care** e **Communication Services**) principalmente per osservare cos'è successo ai titoli dal 2020 con la pandemia globale.

Per quanto riguarda le azioni scelte, di seguito una tabella:

Ticker	Compagnia	Settore/Sub-industry	Descrizione & notizie
UNH	UnitedHealth Group	Health Care	<p>UnitedHealth Group è una società multinazionale americana assicurativa e sanitaria (managed healthcare, sistema in cui le spese mediche sono controllate limitando i servizi offerti da parte di dottori ed ospedali), che si occupa di sanità e benessere.</p> <p>È un insieme di operatori sanitari, aziende farmaceutiche e servizi di prescrizione che fornisce benefit sanitari a clienti e mercati.</p> <p>Nel 2012 la società acquisisce XLHealth, che prestava particolare attenzione ai malati speciali e con malattie croniche. Inoltre, nello stesso anno, si ha la fusione con un'importante assicurazione sanitaria brasiliana.</p> <p>Nel 2017 si ha l'unione tra la divisione Optum di UnitedHealth e Rally Health (fondato da dirigenti di Audax Health).</p>

			<p>Nel 2019 la divisione Optum acquisisce Davita Medical Group per \$4.3 bilioni. Sempre nel 2019, acquisisce Equian per \$3.2 bilioni e il 19 giugno la piattaforma online PatientsLikeMe.</p> <p>IL 31 marzo 2021, UnitedHealth Group ha avuto una capitalizzazione di mercato di \$400.7 bilioni.</p>
JNJ	Johnson & Johnson	Health Care	<p>Johnson & Johnson è una società farmaceutica multinazionale statunitense, che produce farmaci, prodotti per la cura personale e apparecchiature mediche.</p> <p>È stata bersaglio di diverse cause legali, per i suoi prodotti a base di talco.</p> <p>Nel 2017, una giuria della California ha ordinato di pagare \$417 milioni a una donna che ha affermato di aver sviluppato un tumore alle ovaie dopo aver usato i prodotti a base di talco per l'igiene.</p> <p>Nel 2018, una giuria di St. Louis ha assegnato danni per quasi \$4.7 miliardi a 22 donne e alle loro famiglie, dopo aver dichiarato che l'amianto nella polvere di talco ha causato il cancro alle ovaie.</p> <p>Più di 11k persone hanno fatto causa per tumori causati da polvere di talco: Johnson & Johnson è costretta a rilasciare documenti che mostrano che erano a conoscenza della contaminazione da amianto almeno dal 1971.</p> <p>Il 19 dicembre 2018 il verdetto della giuria si pronuncia a favore degli accusatori ed è costretta a pagare \$4.14 miliardi di danni punitivi e \$550 milioni di danni compensativi.</p> <p>Durante la pandemia, è una delle prime case farmaceutiche a produrre un vaccino, approvato dall'Ema nel 2021, monodose ed efficace anche contro le varianti.</p> <p>Nell'agosto 2022, con un numero di 40k cause legali, Johnson & Johnson interrompe la produzione dei prodotti a base di talco.</p>
GM	General Motors	Automobile Manufacturers	<p>General Motors Corporation è un'azienda statunitense produttrice di autoveicoli, con marchi in tutto il mondo quali: Cadillac, Chevrolet, GM Korea, GMC, e Buick.</p> <p>Ha posseduto anche quote di Isuzu, Subaru e Suzuki in Giappone.</p> <p>È il più grande produttore automobilistico statunitense e uno dei più grandi al mondo.</p> <p>Nel 2017 Opel viene ceduta a PSA e General Motors esce dal mercato europeo auto.</p> <p>Nel 2018 chiude 5 stabilimenti nel Nord America e vi è il taglio di oltre 14k posti di lavoro. Trump minaccia di sopprimere ogni sovvenzione a General Motors, comprese quelle per l'auto elettrica.</p> <p>Nel 2020 esce quasi totalmente dai mercati australiano e neozelandese.</p>
TSLA	Tesla	Automobile Manufacturers	<p>Tesla, Inc. è una multinazionale statunitense specializzata nella produzione di auto elettriche, pannelli fotovoltaici e sistemi di stoccaggio energetico. Il nome è in onore a Nikola Tesla.</p> <p>Sta accelerando la transizione all'uso di fonti di energia rinnovabili: ciò include la produzione di veicoli elettrici ad alte prestazioni.</p> <p>Il 26 marzo 2009 la Model S è la prima berlina completamente elettrica e le consegne inizieranno nel 2012.</p>

			<p>Nel 2012, inoltre, viene presentata la Model X, il primo SUV a trazione elettrica, e inizia la costruzione dei Supercharger (rete di stazioni di ricarica rapida).</p> <p>IL 2013 è un anno difficile per Tesla: Elon Musk ha intenzione di vendere l'azienda a Google.</p> <p>Nel 2015 vi è l'acquisto di Riviera Tool e l'inizio delle consegne della Model X, con 100k vendite a dicembre.</p> <p>Il 31 marzo 2016 viene presentata la berlina Model 3, con 325k prenotazioni nella settimana successiva. Ad agosto vi è l'acquisto di SolarCity per \$2.6 miliardi, mentre a novembre Elon Musk ha acquisito la società tedesca Grohmann Automation.</p> <p>Nel 2017 supera Ford e General Motors in capitalizzazione di mercato, diventando l'azienda automobilistica statunitense di maggior valore.</p> <p>Nel 2019 acquista la produttrice di ultracondensatori Maxwell Technologies e la startup americana DeepScale, oltre a Hibar Systems.</p> <p>Il 25 ottobre 2021 Hertz effettua un maxi ordine di 100k Tesla Model 3 e la società raggiunge una capitalizzazione di mercato di \$1k miliardi.</p>
NFLX	Netflix	Movies & Entertainment	<p>Netflix è una società statunitense che si occupa di streaming via Internet di film, serie tv e altri contenuti d'intrattenimento a pagamento.</p> <p>È nata come azienda di noleggio di DVD e videogiochi.</p> <p>Da gennaio 2016 è accessibile il servizio di streaming in oltre 190 paesi, con 74 milioni di clienti.</p> <p>Nel 2016 è leader del settore on demand con 93.8 milioni di abbonati.</p> <p>Nel 2020, con l'avvenuta del COVID-19, concentra le proprie risorse in contenuti nuovi originali, ottenendo in tutto 182 milioni di abbonati.</p> <p>Nel 2022 perde 200k abbonati: potrebbe essere dovuto alla sospensione del servizio in Russia e allo stop dell'account sharing.</p>
DIS	Disney	Movies & Entertainment	<p>Disney è una multinazionale statunitense, in origine uno studio di animazione.</p> <p>Nel 2017 l'azienda decide di acquistare la 21st Century Fox per \$52.4 miliardi, approvato dall'Antitrust del Dipartimento di Giustizia degli Stati Uniti. L'acquisizione si è conclusa nel 2019.</p> <p>Ho scelto questo asset per osservare cos'è successo con il lancio del suo servizio Disney+ nel 2019: servizio di video on demand gestito da Disney Media and Entertainment Distribution.</p> <p>La piattaforma distribuisce film e serie tv prodotti da The Walt Disney Studios e da Walt Disney Television.</p> <p>Nel maggio 2021 raggiunge 103.6 milioni di abbonati, mentre in aprile 2022 137.7 milioni.</p>

Download dei dati da Yahoo! Finance

Per scaricare i dati da Yahoo! Finance ho utilizzato la libreria `yfinance`, che mi permette di ottenere dati OHLC (Open High Low Close) daily, indicando il ticker, la data di inizio periodo (30-11-2012) e la data di fine periodo (30-11-2022).

Per ogni asset scaricato, si ottiene un dataframe con il prezzo di apertura (Open), il prezzo massimo raggiunto (High), il prezzo minimo raggiunto (Low), l'ultimo prezzo a cui è stata scambiata l'azione (Close), il prezzo di chiusura aggiustato (Adj Close) e il numero di quote scambiate (Volume).

Per esempio, queste sono le prime 5 righe del dataframe corrispondente al titolo UnitedHealth Group:

Date	Open	High	Low	Close	Adj Close	Volume
2012-11-30	54.180000	54.880001	54.070000	54.389999	46.453583	7818600
2012-12-03	54.560001	54.950001	53.830002	53.889999	46.026543	3861300
2012-12-04	53.959999	54.240002	53.520000	53.540001	45.727623	5948900
2012-12-05	53.730000	54.230000	53.610001	53.840000	45.983860	4495100
2012-12-06	53.900002	53.900002	53.279999	53.650002	45.821571	3771900

Tuttavia, per ogni dataframe prendo in considerazione solo la colonna Adj Close.

Fusione delle serie in un unico DataFrame

Unisco le series, corrispondenti agli Adj Close, in un unico dataframe tramite la funzione concat di Pandas (in alternativa avrei potuto usare join): creo un nuovo dataframe vuoto, per ogni asset converto la series Adj Close in un dataframe “temporaneo” con `to_frame()` e concateno il dataframe totale con questo dataframe “temporaneo”.

Ottengo così un dataframe che contiene gli adjusted close dei titoli.

Prime righe del dataframe totale:

	Adj Close_UNH	Adj Close_JNJ	Adj Close_GM	Adj Close_TSLA	Adj Close_NFLX	Adj Close_DIS
Date						
2012-11-30	46.453583	53.113289	19.912642	2.254667	11.672857	44.462765
2012-12-03	46.026550	53.052326	19.627956	2.308000	10.857143	44.131496
2012-12-04	45.727619	53.212303	19.551014	2.260000	12.378571	44.140442
2012-12-05	45.983852	53.296066	19.235554	2.247333	11.910000	44.400085
2012-12-06	45.821575	53.357014	18.904699	2.260000	12.310000	44.600101

Ultime righe:

	Adj Close_UNH	Adj Close_JNJ	Adj Close_GM	Adj Close_TSLA	Adj Close_NFLX	Adj Close_DIS
Date						
2022-11-22	521.482483	176.820007	39.881310	169.910004	286.690002	96.209999
2022-11-23	528.082153	177.009995	40.021000	183.199997	291.500000	98.879997
2022-11-25	535.967834	177.240005	40.370220	182.860001	285.540009	98.870003
2022-11-28	530.634277	177.330002	39.292618	182.919998	281.170013	95.690002
2022-11-29	526.377380	176.089996	39.661797	180.830002	280.959991	94.690002

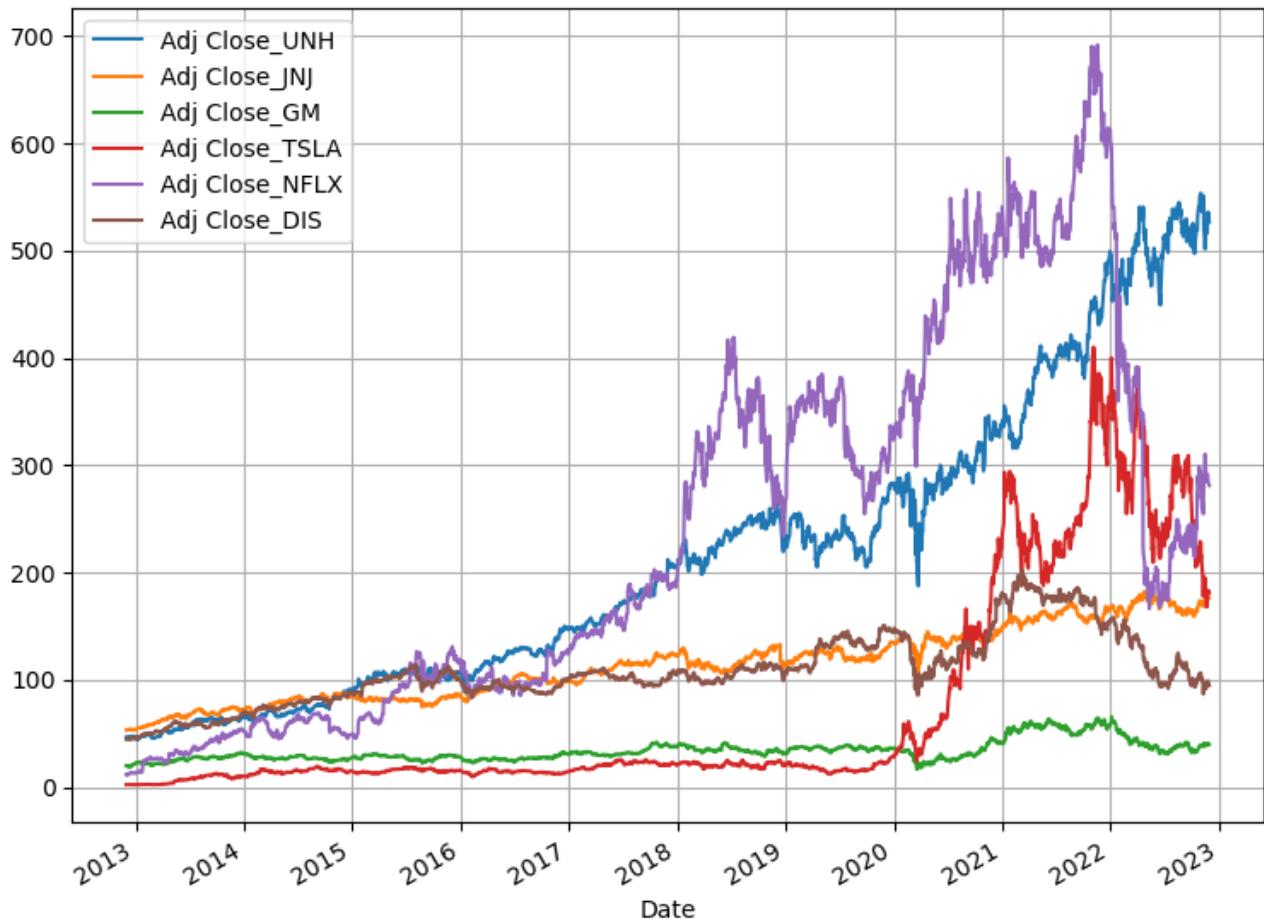
Dati mostrati con la funzione describe:

	Adj Close_UNH	Adj Close_JNJ	Adj Close_GM	Adj Close_TSLA	Adj Close_NFLX	Adj Close_DIS
count	2517.000000	2517.000000	2517.000000	2517.000000	2517.000000	2517.000000
mean	214.772238	113.268050	33.610196	71.796687	236.612867	107.899626
std	137.922804	33.036382	9.505687	101.868080	175.370038	32.942904
min	44.074871	52.922855	16.724537	2.194000	10.857143	44.131496
25%	102.878525	82.829933	26.974800	14.433333	87.879997	90.145126
50%	198.205887	114.352211	31.132090	18.648001	187.759995	102.866318
75%	285.904572	137.433502	37.119850	59.873333	361.809998	124.980003
max	553.443970	182.411407	65.444710	409.970001	691.690002	201.910004

Noto che ho a disposizione 2517 prezzi giornalieri e che il valore più alto tra tutti i titoli viene raggiunto da Netflix.

Presentazione dei dati con un grafico

Ho iniziato la visualizzazione dei dati plottando in un unico grafico il dataframe con tutti gli adjusted close.

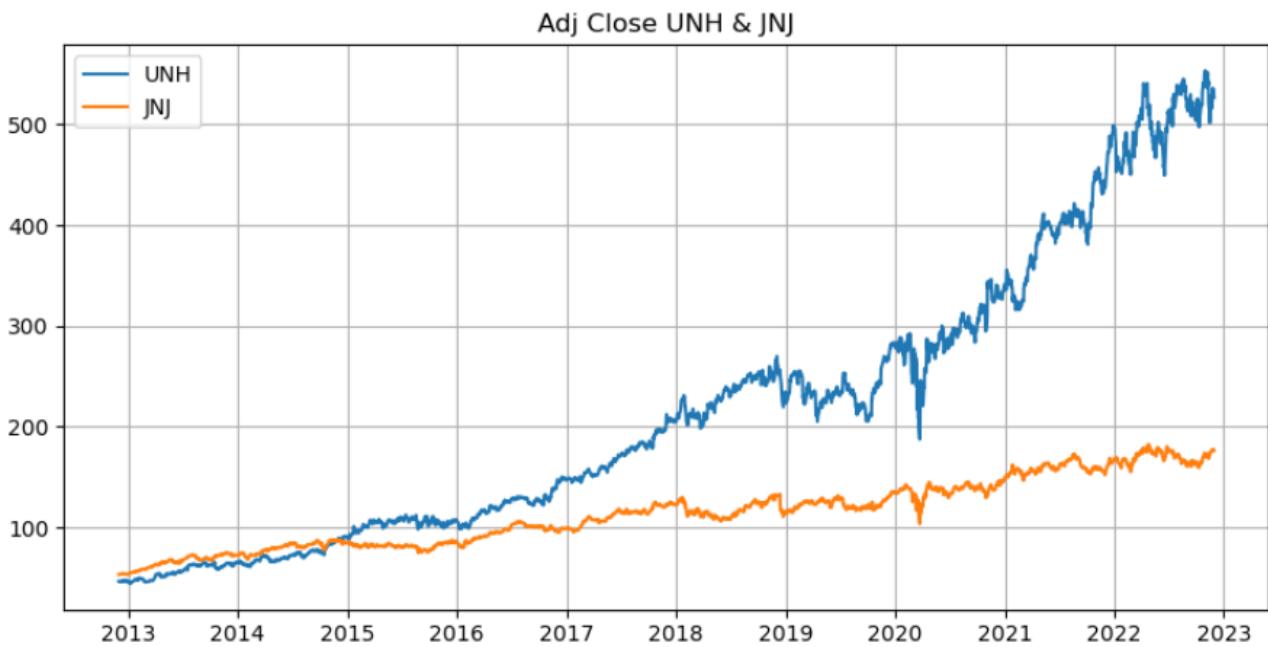


Da questo grafico posso confrontare l'andamento generale dei vari titoli: è notevole il prezzo massimo raggiunto da Netflix in tutto il periodo, arrivando quasi a \$700 alla fine del 2021.

Tuttavia, sono anche alti i prezzi di UnitedHealth Group, in continua crescita con trend bullish, e non si può non notare di quanto sia salito il valore azionario di Tesla dal 2020.

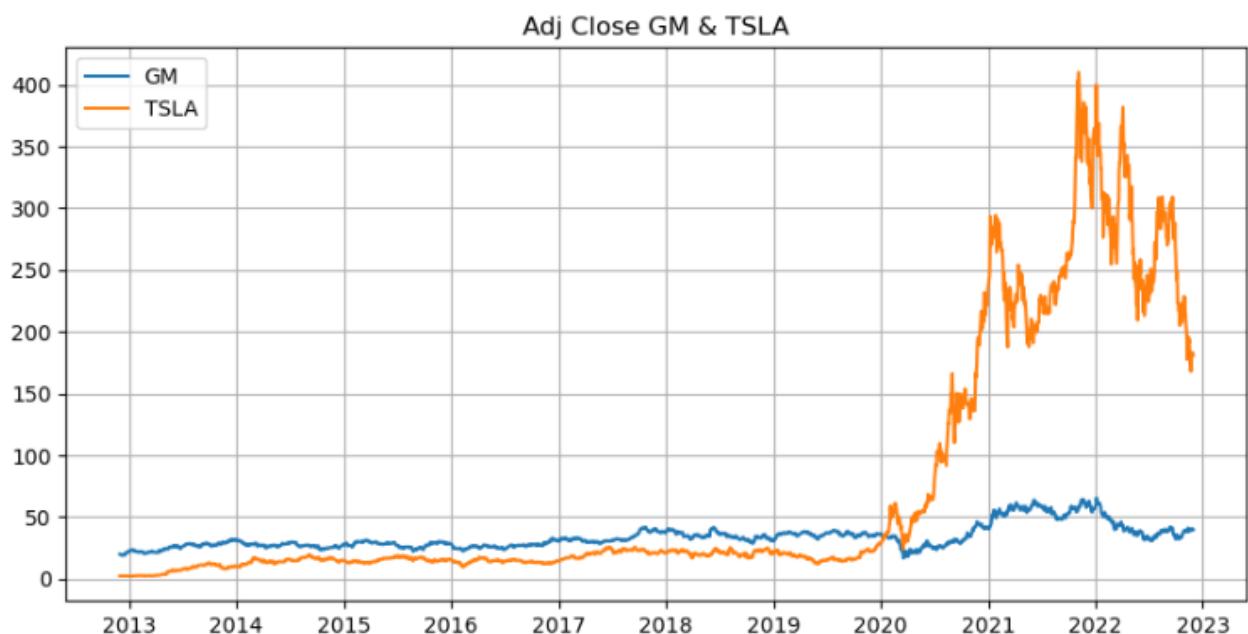
Posso confrontare l'andamento dei prezzi con le notizie storiche riportate nella sezione “Scelta stocks”.

Per fare questo lavoro in maniera più chiara, mostro separatamente per ogni settore la coppia di asset in un grafico.



Oltre al calo dei prezzi di entrambi a inizio 2020 probabilmente dovuto al COVID-19, si può notare come l'acquisto di diverse società da parte di UnitedHealth Group corrisponde sempre alla crescita di valore del titolo (in particolare nel 2017).

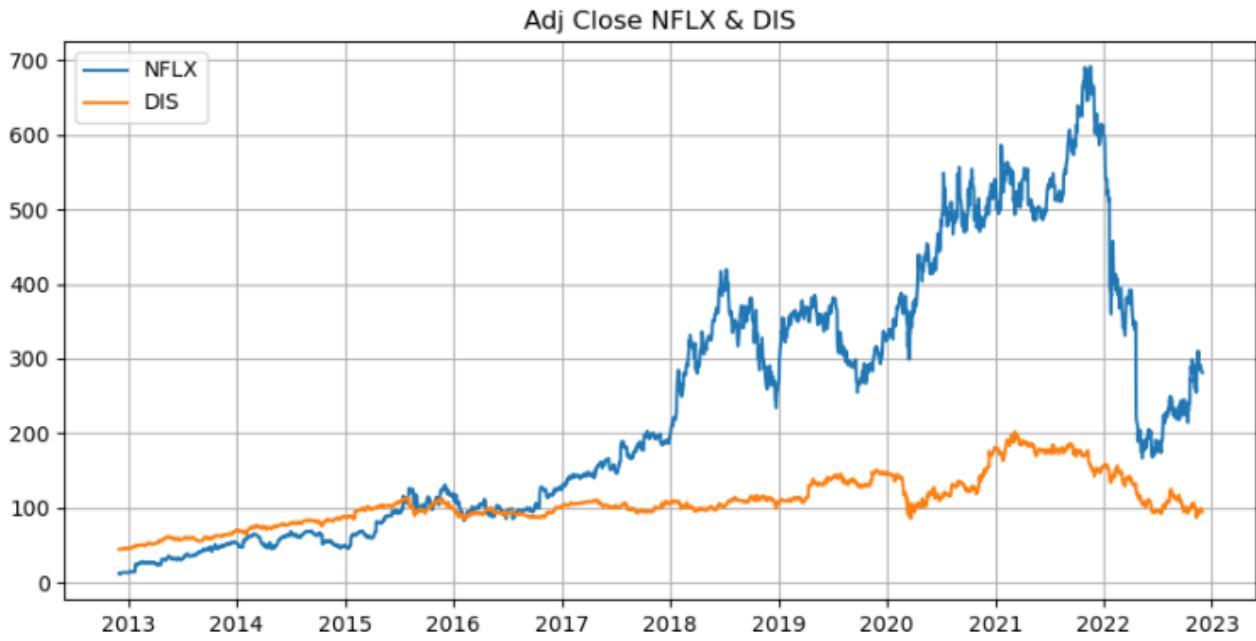
Johnson & Johnson non è cresciuta molto di valore nel tempo, poiché è stata bersaglio di molte cause legali dovute ai suoi prodotti a base di talco.



In questo grafico è impossibile non osservare di quanto sia “schizzato” il prezzo di Tesla a partire dal 2020 (dovuto all'aumento di interesse da parte dell'investitore), ma il prezzo massimo viene raggiunto durante la

fine del 2021: infatti, il 25 ottobre 2021 Hertz ha ordinato 100k Tesla Model 3.

Per quanto riguarda General Motors, facendo un confronto con le notizie, nel 2018 scende di valore, quando taglia oltre 14k posti di lavoro e Trump minaccia di sopprimere ogni sovvenzione.



Nel 2020 Netflix, con il COVID-19 e la fruizione di nuovi contenuti originali, aumenta di molto il suo valore, ma dal grafico si può osservare come l'apice si raggiunge a fine 2021.

Corrisponde anche la notizia sulla perdita di 200k abbonati nel 2022: vi è una discesa in forte pendenza, che può essere dovuta allo stop dell'account sharing e alla sospensione del servizio in Russia.

Per quanto riguarda Disney, l'acquisizione della 21st Century Fox si conclude nel 2019 e infatti vi è un aumento dei prezzi, ma questo è dovuto anche al lancio del servizio Disney+ e a inizio 2021 si raggiunge il prezzo più alto.

Statistiche descrittive

Rendimento cumulato e composto annuo

Di seguito il rendimento cumulato di ciascun titolo nell'intero periodo:

cum_rtn_UNH	cum_rtn_JNJ	cum_rtn_GM	cum_rtn_TSLA	cum_rtn_NFLX	cum_rtn_DIS
10.331252	2.315368	0.99179	79.20253	23.069513	1.129647

Il rendimento cumulato è uguale alla quantità aggregata che l'investimento ha guadagnato o perso col tempo, indipendente da quanto tempo è trascorso.

Si tratta di una percentuale, quindi, investendo il 30-11-2012 per 10 anni, si hanno i seguenti rendimenti cumulati:

- UnitedHealth Group ≈ 1033%
- Johnson & Johnson ≈ 231%
- General Motors ≈ 99%
- Tesla ≈ 7920%
- Netflix ≈ 2306%
- Disney ≈ 112%

Per quanto riguarda il rendimento composto annuo, a differenza del rendimento cumulato questo è annualizzato.

y_comp_rtn_UNH	y_comp_rtn_JNJ	y_comp_rtn_GM	y_comp_rtn_TSLA	y_comp_rtn_NFLX	y_comp_rtn_DIS
0.274758	0.127335	0.071333	0.550311	0.374506	0.078526

Questi numeri si possono interpretare come i tassi di ritorno annualizzati.

Rendimenti semplici e logaritmici

I rendimenti semplici (e poi logaritmici) sono calcolati con frequenza giornaliera, quindi l'holding period è un giorno: calcolo il ritorno netto semplice (= rendimento semplice) su un giorno di investimento.

Queste sono le prime righe del dataframe dei rendimenti semplici giornalieri:

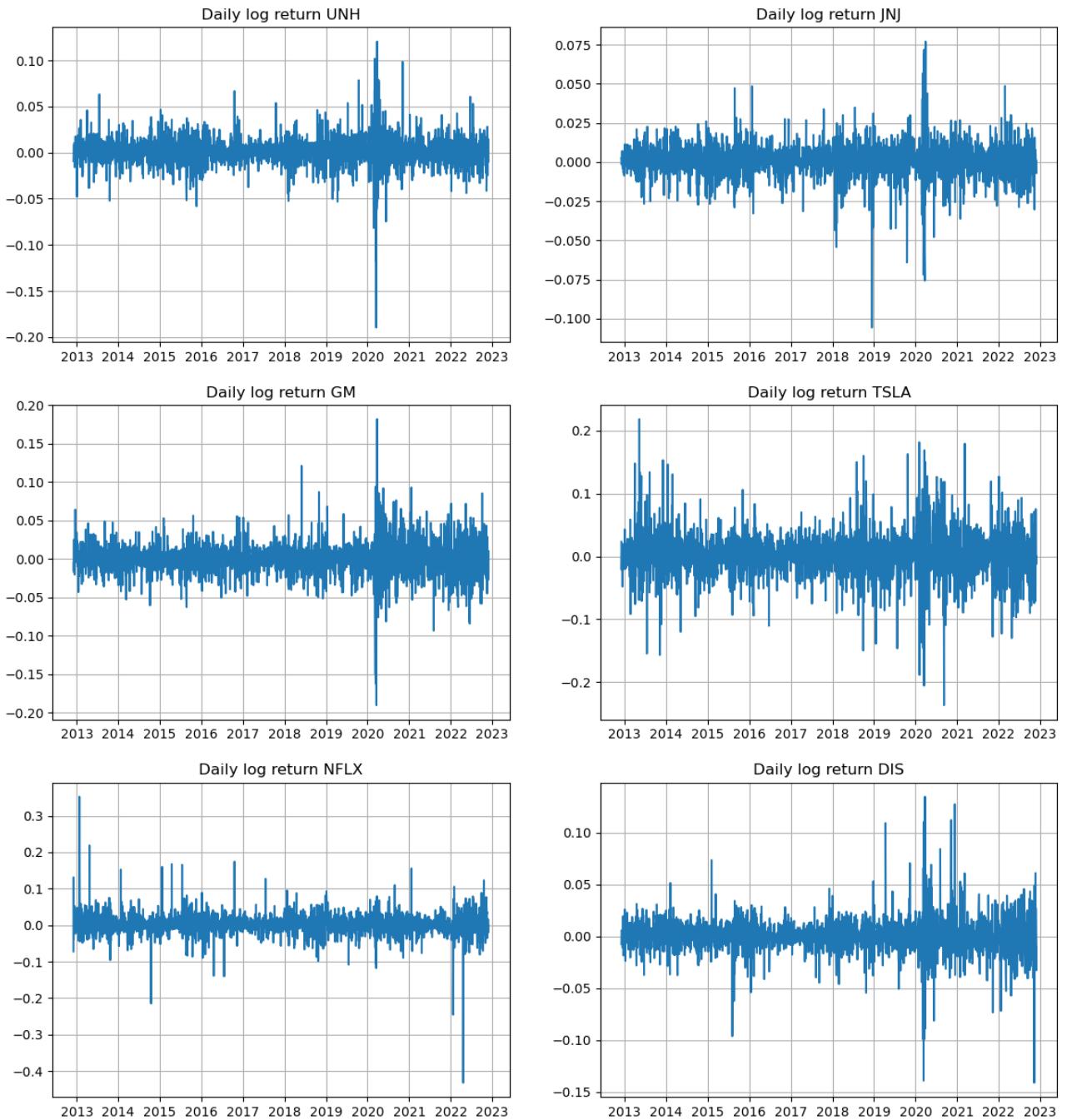
Date	[D] Simple ret. UNH	[D] Simple ret. JNJ	[D] Simple ret. GM	[D] Simple ret. TSLA	[D] Simple ret. NFLX	[D] Simple ret. DIS
2012-12-03	-0.009193	-0.001147	-0.014297	0.023655	-0.069881	-0.007451
2012-12-04	-0.006495	0.003015	-0.003920	-0.020797	0.140131	0.000203
2012-12-05	0.005603	0.001574	-0.016135	-0.005605	-0.037853	0.005882
2012-12-06	-0.003529	0.001143	-0.017200	0.005636	0.033585	0.004505
2012-12-07	0.008103	0.005710	0.025234	0.007965	-0.002205	0.003669

I rendimenti logaritmici non sono altro che il logaritmo di $1 +$ rendimento semplice:

Date	[D] Log ret. UNH	[D] Log ret. JNJ	[D] Log ret. GM	[D] Log ret. TSLA	[D] Log ret. NFLX	[D] Log ret. DIS
2012-12-03	-0.009235	-0.001148	-0.014400	0.023379	-0.072443	-0.007479
2012-12-04	-0.006516	0.003010	-0.003928	-0.021017	0.131144	0.000203
2012-12-05	0.005587	0.001573	-0.016267	-0.005621	-0.038588	0.005865
2012-12-06	-0.003535	0.001143	-0.017350	0.005621	0.033034	0.004494
2012-12-07	0.008070	0.005694	0.024921	0.007933	-0.002207	0.003662

Ho deciso di mostrare nei grafici solamente i rendimenti logaritmici, proprio perché, come già detto, si calcolano a partire dai rendimenti semplici e mi sembrava ripetitivo mettere entrambi.

Di seguito mostro i rendimenti logaritmici giornalieri per ogni titolo:



Possiamo osservare come si tratta di serie la cui media si aggira intorno allo zero: questo lo posso confermare eseguendo la funzione describe()

	[D] Log ret. UNH	[D] Log ret. JNJ	[D] Log ret. GM	[D] Log ret. TSLA	[D] Log ret. NFLX	[D] Log ret. DIS
count	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000
mean	0.000965	0.000476	0.000274	0.001743	0.001264	0.000300
std	0.015964	0.011146	0.021475	0.035570	0.030039	0.016569
min	-0.189671	-0.105782	-0.190227	-0.236518	-0.432578	-0.141139
25%	-0.006775	-0.004501	-0.009934	-0.014888	-0.012155	-0.006737
50%	0.001054	0.000530	0.000599	0.001427	0.000434	0.000422
75%	0.008492	0.006011	0.010333	0.018793	0.014880	0.007700
max	0.120437	0.076940	0.181848	0.218292	0.352230	0.134639

Potrei quindi dire che i rendimenti logaritmici hanno in comune il fatto di essere serie stazionarie.

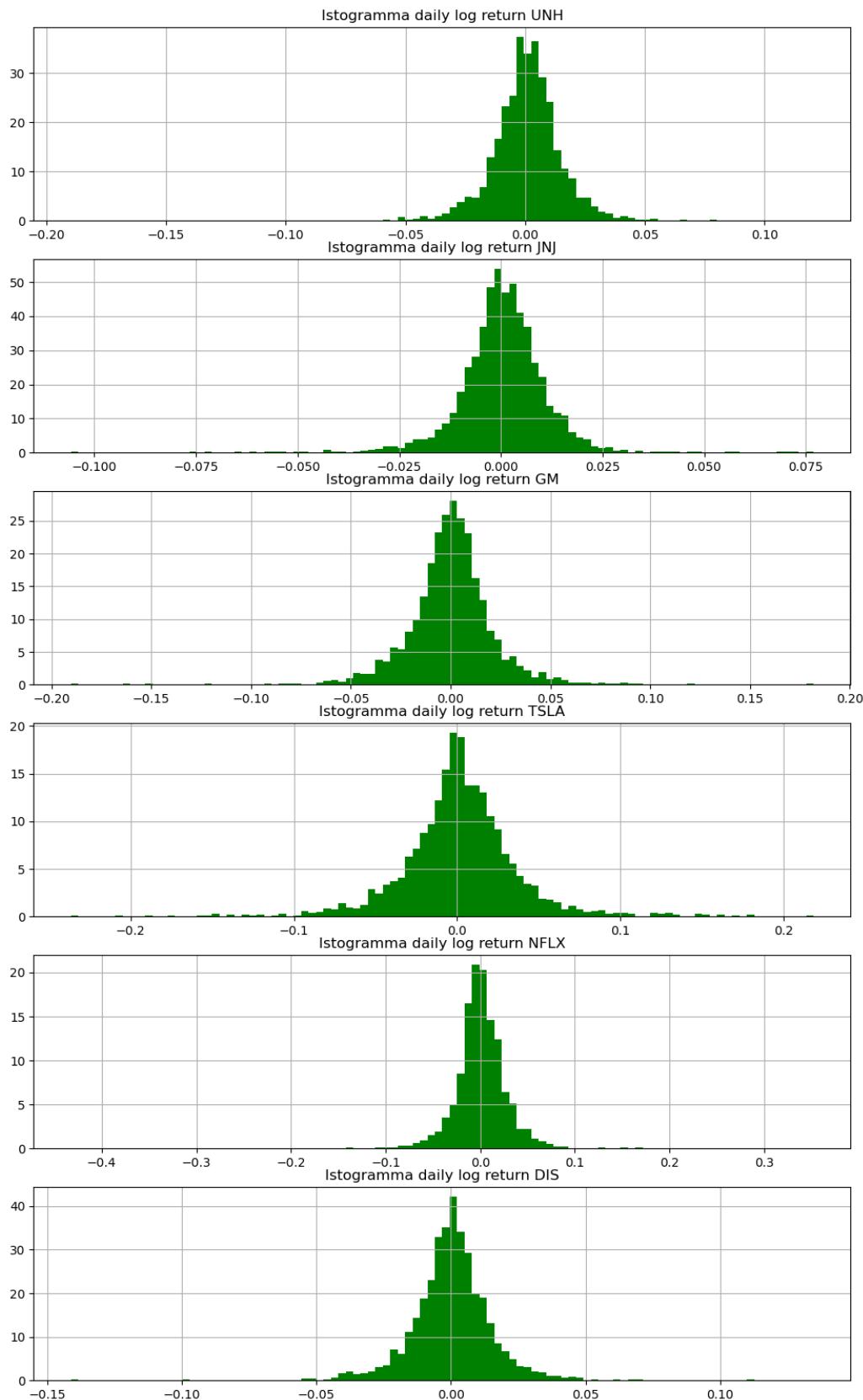
Tornando ai grafici, possiamo osservare come la maggior parte dei rendimenti abbia avuto grandi variazioni a inizio 2020 (COVID-19), ma il minimo è stato raggiunto da Netflix durante l'inizio del 2022, per la notizia già discussa riguardo alla perdita di 200k abbonati.

Basandomi solo sul grafico, non posso saltare a conclusioni sull'eventuale correlazione positiva fra società dello stesso settore.

Posso invece osservare che ci sono stati momenti molto lontani dalla media. Alcuni momenti che evidenzio:

- 2013: Netflix vede una salita molto alta, questo è probabilmente dovuto alla sua entrata nel settore della produzione, oltre all'espansione territoriale in Austria, Belgio, Francia, Germania, Lussemburgo e Svizzera e all'accordo con la Marvel Entertainment per la trasmissione di serie tv
- 2018: Johnson & Johnson, a dicembre, come riportato tra le notizie nella sezione "Scelta stocks", è costretta a pagare \$4.14 miliardi di danni punitivi e \$550 milioni di danni compensativi per la contaminazione di prodotti a base di talco; ciò è coerente con la discesa evidenziata nel grafico
- 2019: Disney lancia il servizio on demand Disney+ negli USA (e nel 2020 nel resto del mondo) e inoltre viene annunciato il controllo di Hulu (piattaforma di streaming e video on demand), in concorrenza con Netflix.

Presentazione dei rendimenti con istogrammi



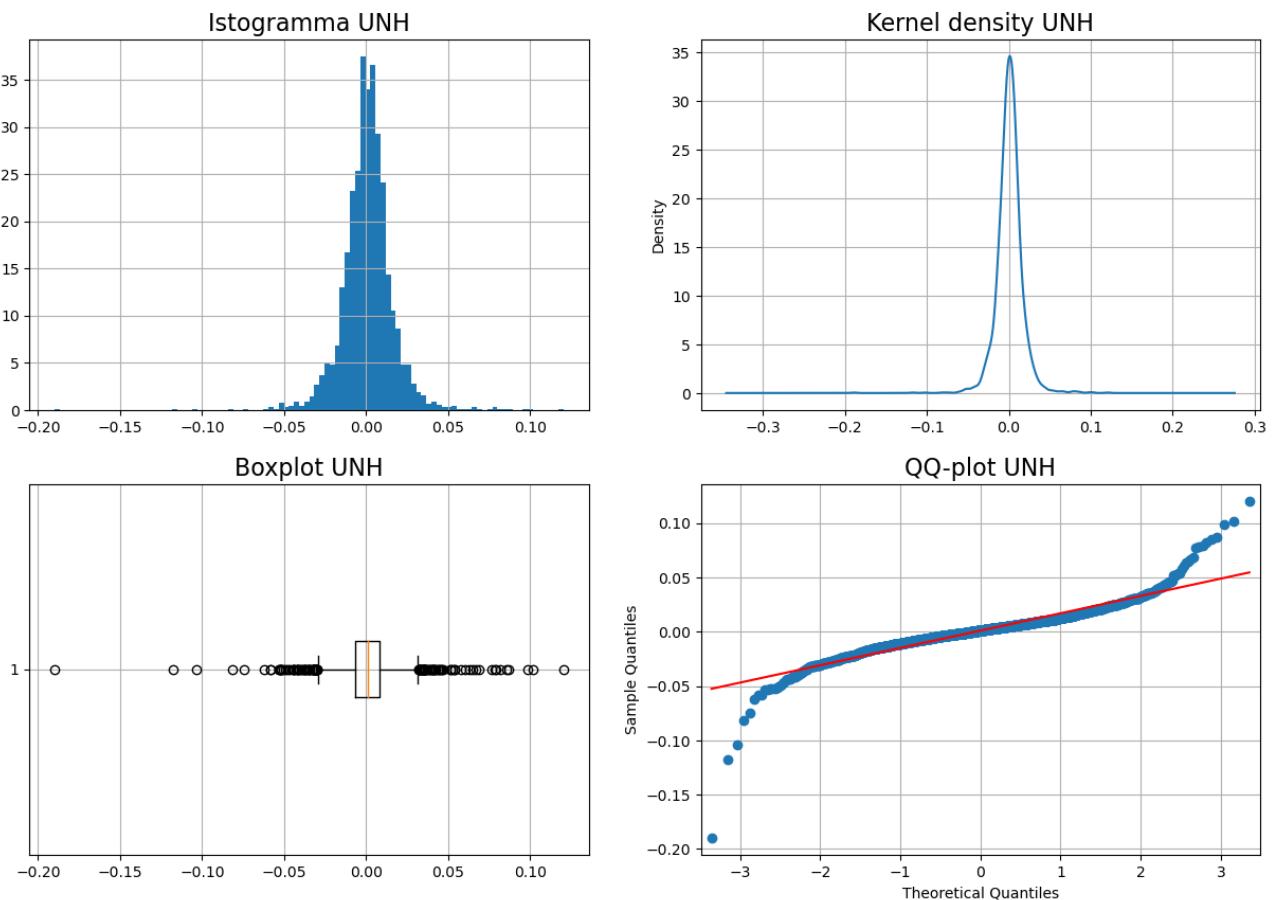
Da questi grafici posso osservare che le distribuzioni sono più “appuntite” rispetto a quella di una normale.

Ne avrò la conferma più avanti una volta visualizzato il valore di curtosi, che mi dice lo “spessore” delle code.

Da una prima osservazione, l’asset con una distribuzione più “appuntita” sembra essere Netflix.

Grafici diagnostici a 4 sezioni

Grafici diagnostici a 4 sezioni - UNH



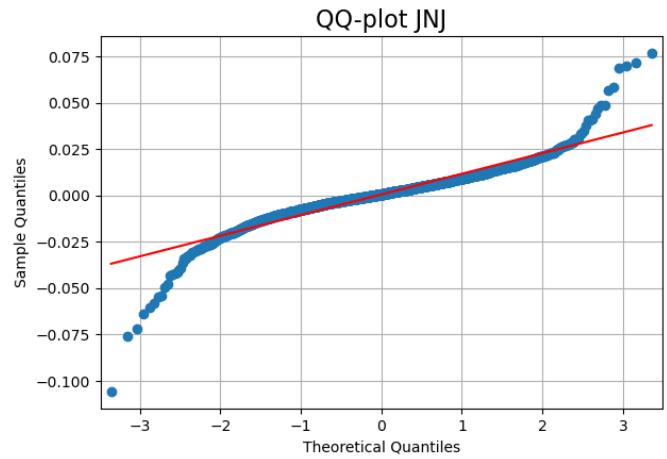
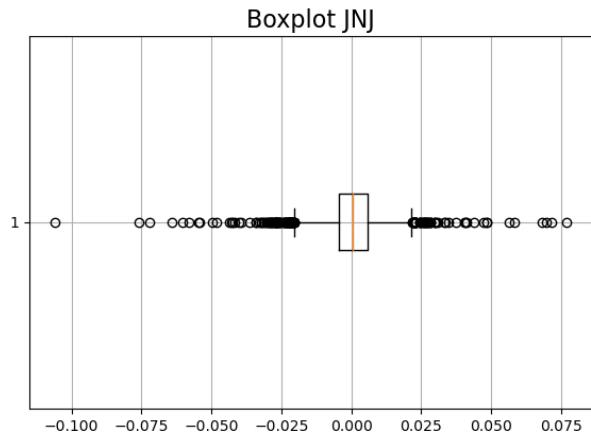
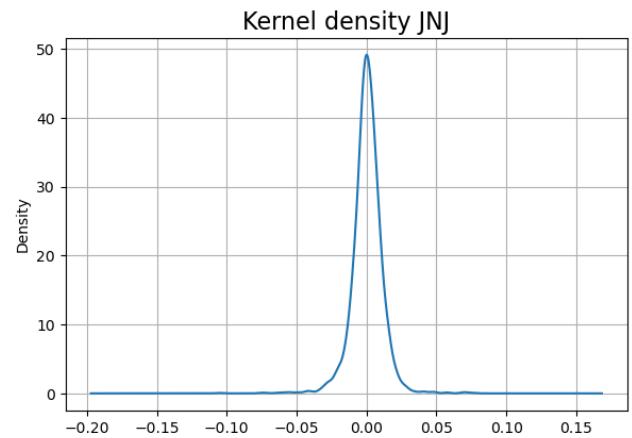
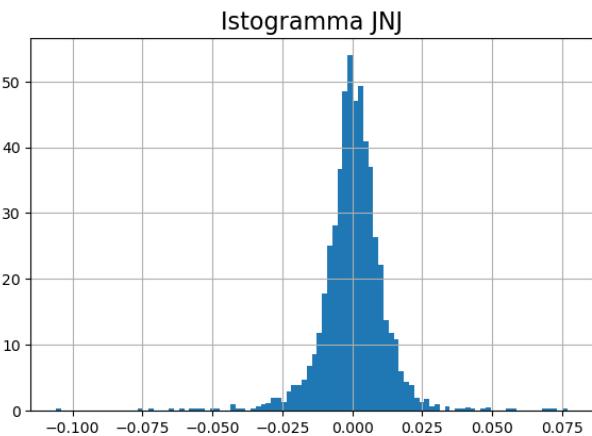
Come già detto, la **distribuzione** dei rendimenti (logaritmici) è molto più **appuntita** rispetto alla distribuzione di una normale: è sufficiente guardare il grafico kernel density.

Dal boxplot si può notare come ci siano molti **outliers**.

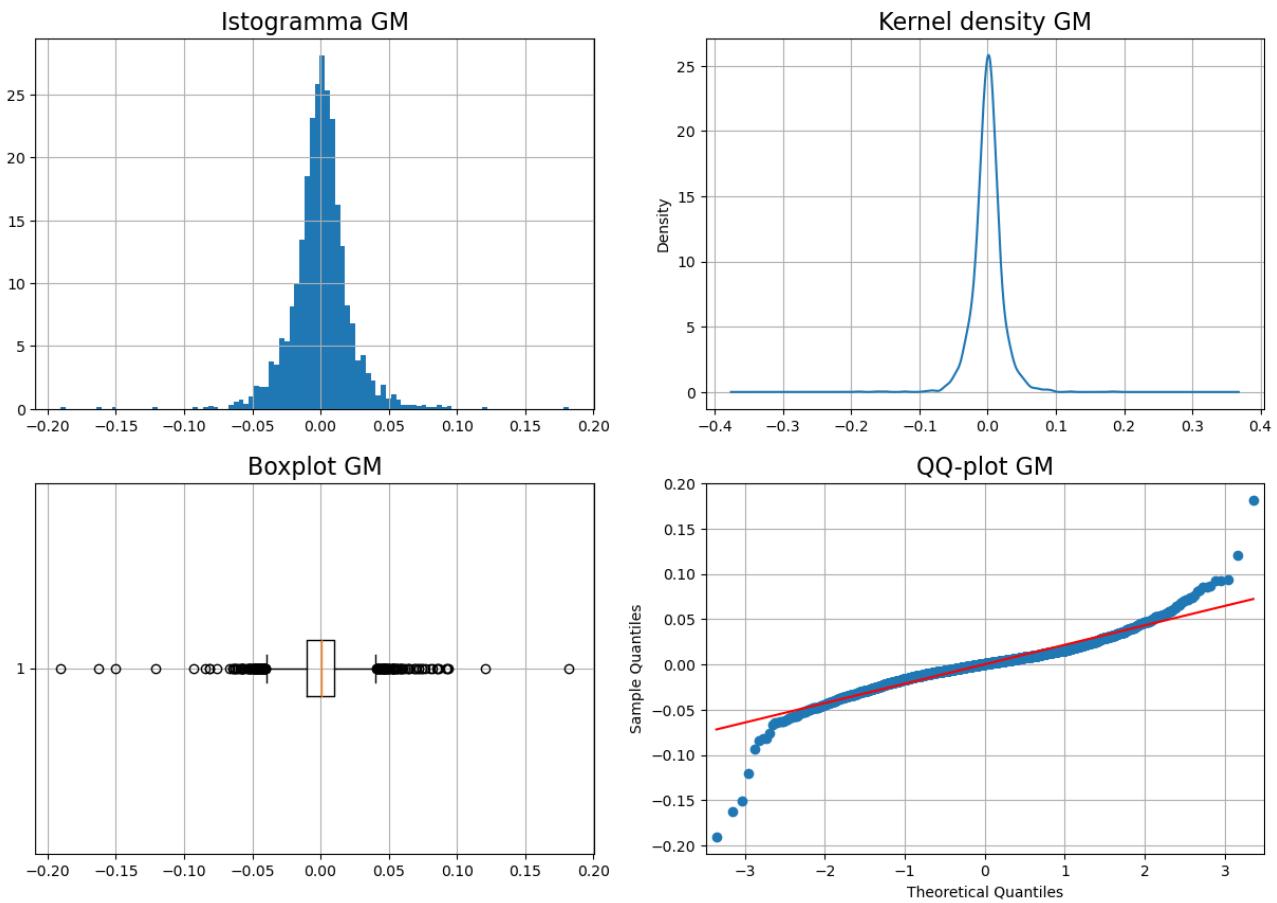
Il grafico Q-Q plot mi permette di confrontare la distribuzione con la distribuzione della normale. Noto che i punti non si addensano sulla diagonale. Ho molti punti condensati che crescono lentamente.

Queste osservazioni valgono anche per i rendimenti degli altri asset, di cui riporto le immagini qui sotto.

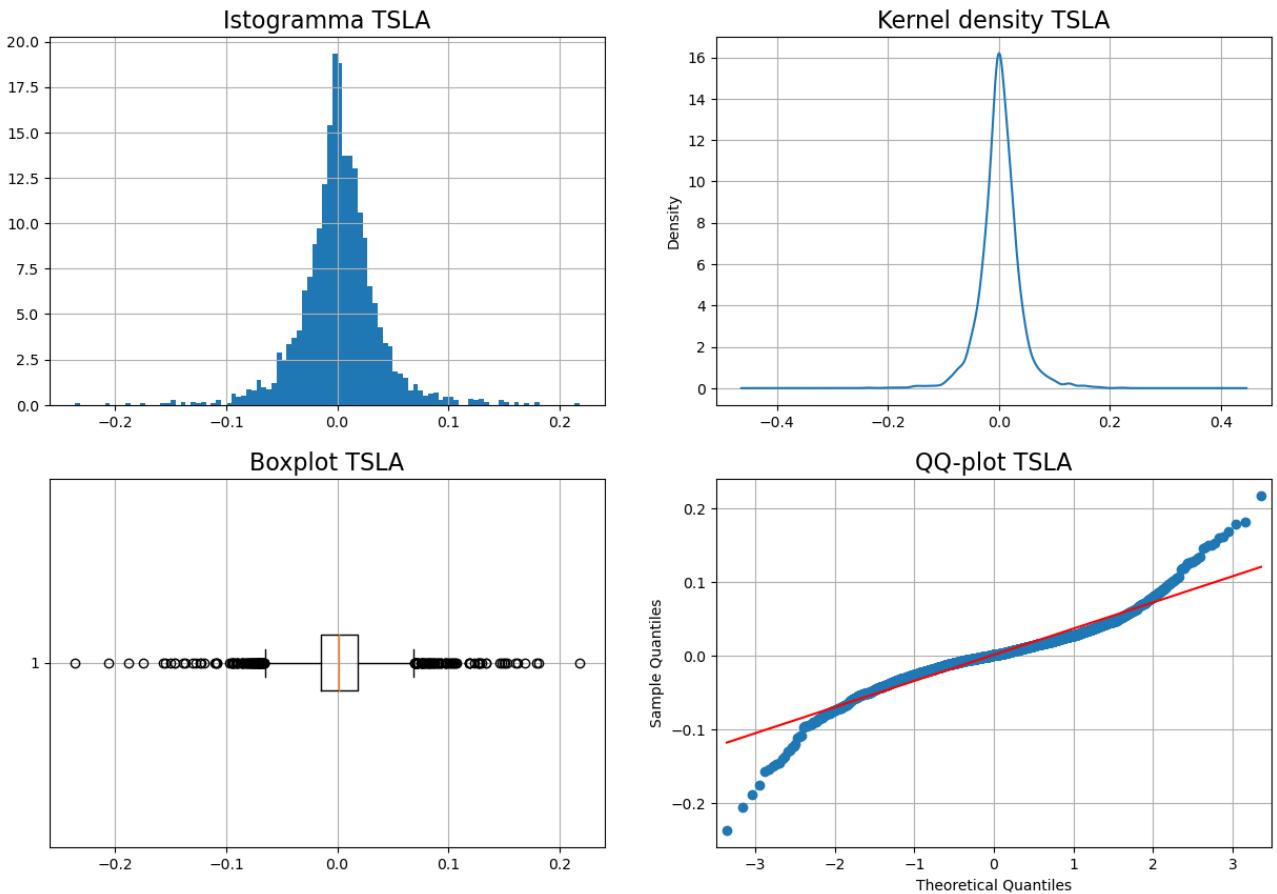
Grafici diagnostici a 4 sezioni - JNJ



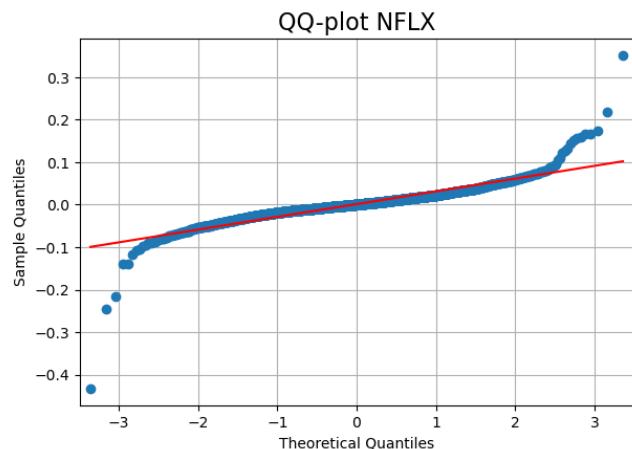
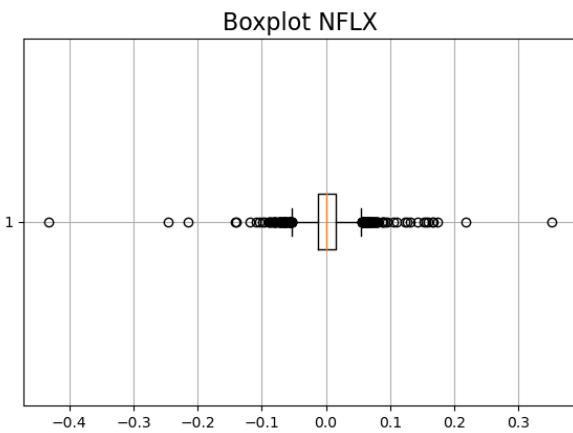
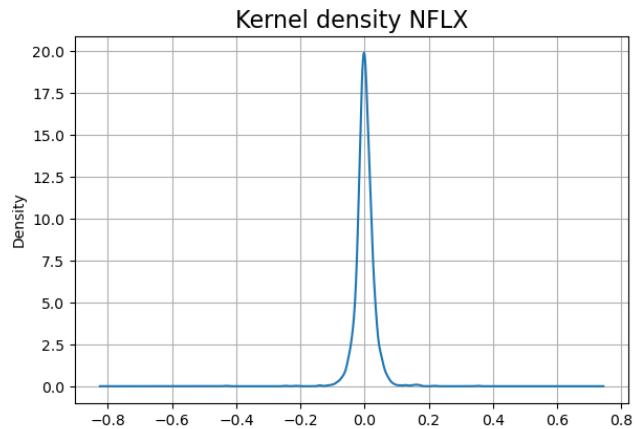
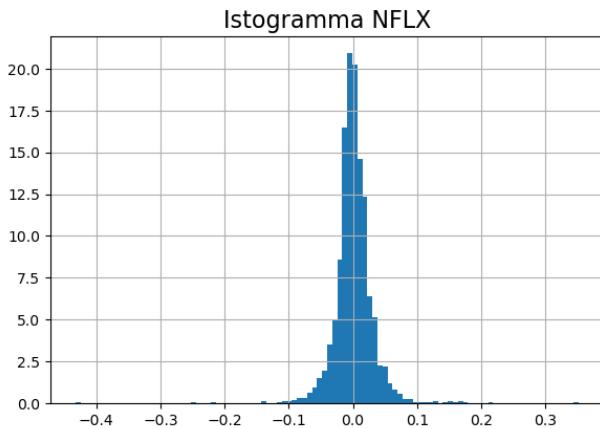
Grafici diagnostici a 4 sezioni - GM



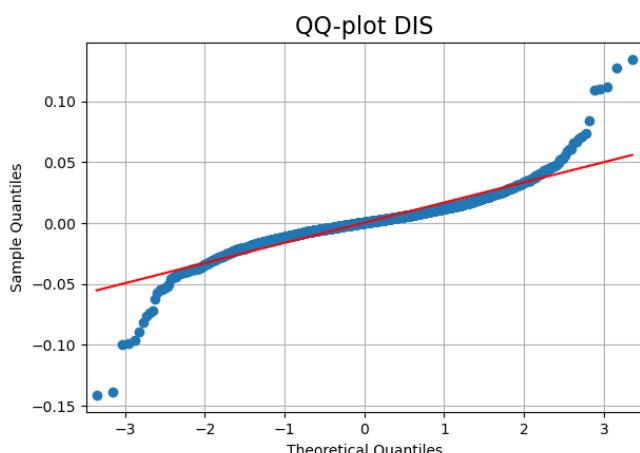
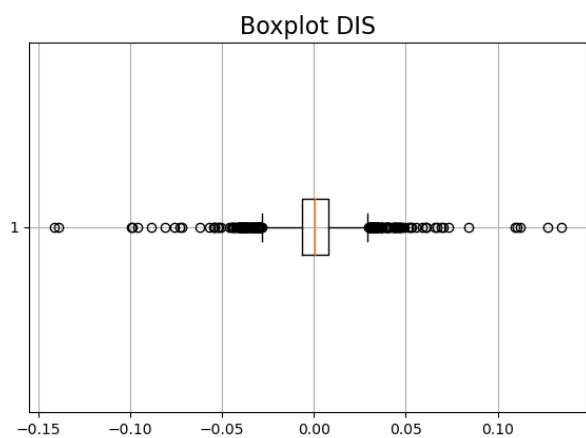
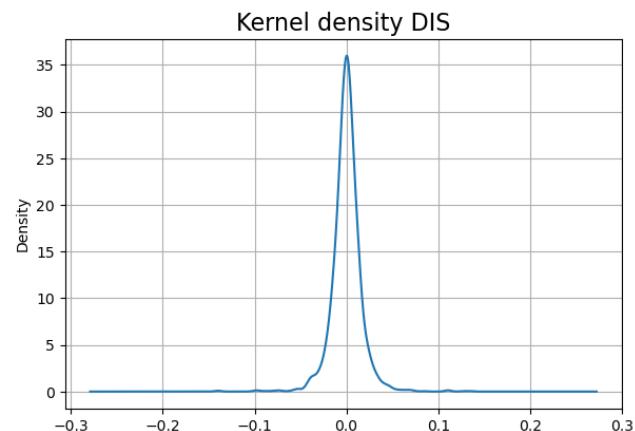
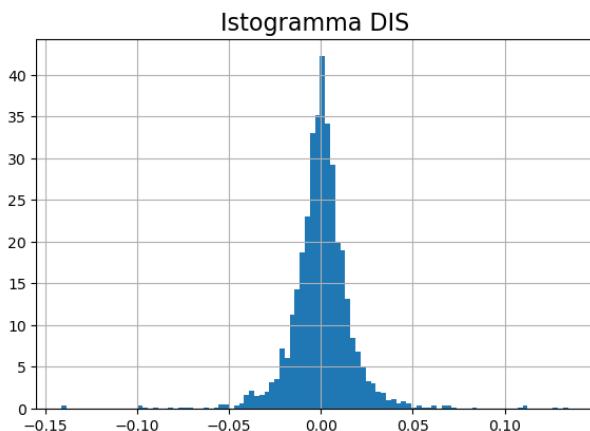
Grafici diagnostici a 4 sezioni - TSLA



Grafici diagnostici a 4 sezioni - NFLX



Grafici diagnostici a 4 sezioni - DIS



Statistiche descrittive univariate

Per ogni asset ho calcolato media, varianza, deviazione standard, asimmetria e curtosi:

```
*** Statistiche descrittive univariate UNH ***
    Media:      0.000965
    Varianza:   0.000255
    Std. dev.:  0.015964
    Asimmetria: -0.39742
    Curtosi:    14.773468

*** Statistiche descrittive univariate JNJ ***
    Media:      0.000476
    Varianza:   0.000124
    Std. dev.:  0.011146
    Asimmetria: -0.474642
    Curtosi:    10.249184

*** Statistiche descrittive univariate GM ***
    Media:      0.000274
    Varianza:   0.000461
    Std. dev.:  0.021475
    Asimmetria: -0.196397
    Curtosi:    8.716005

*** Statistiche descrittive univariate TSLA ***
    Media:      0.001743
    Varianza:   0.001265
    Std. dev.:  0.03557
    Asimmetria: 0.020611
    Curtosi:    5.181715

*** Statistiche descrittive univariate NFLX ***
    Media:      0.001264
    Varianza:   0.000902
    Std. dev.:  0.030039
    Asimmetria: -0.41181
    Curtosi:    31.01424

*** Statistiche descrittive univariate DIS ***
    Media:      0.0003
    Varianza:   0.000275
    Std. dev.:  0.016569
    Asimmetria: -0.029956
    Curtosi:    12.546346
```

Il valore di **curtosi** è positivo: mi conferma le precedenti osservazioni.

La curva viene detta “**leptocurtica**”: più “appuntita” di una normale.

Il valore di **skewness** (asimmetria) mi dice se una delle due code è più lunga rispetto a quella della distribuzione normale: i valori negativi mi fanno osservare che la coda sinistra è più lunga e questo avviene per tutti gli asset tranne Tesla, la cui coda destra è più lunga.

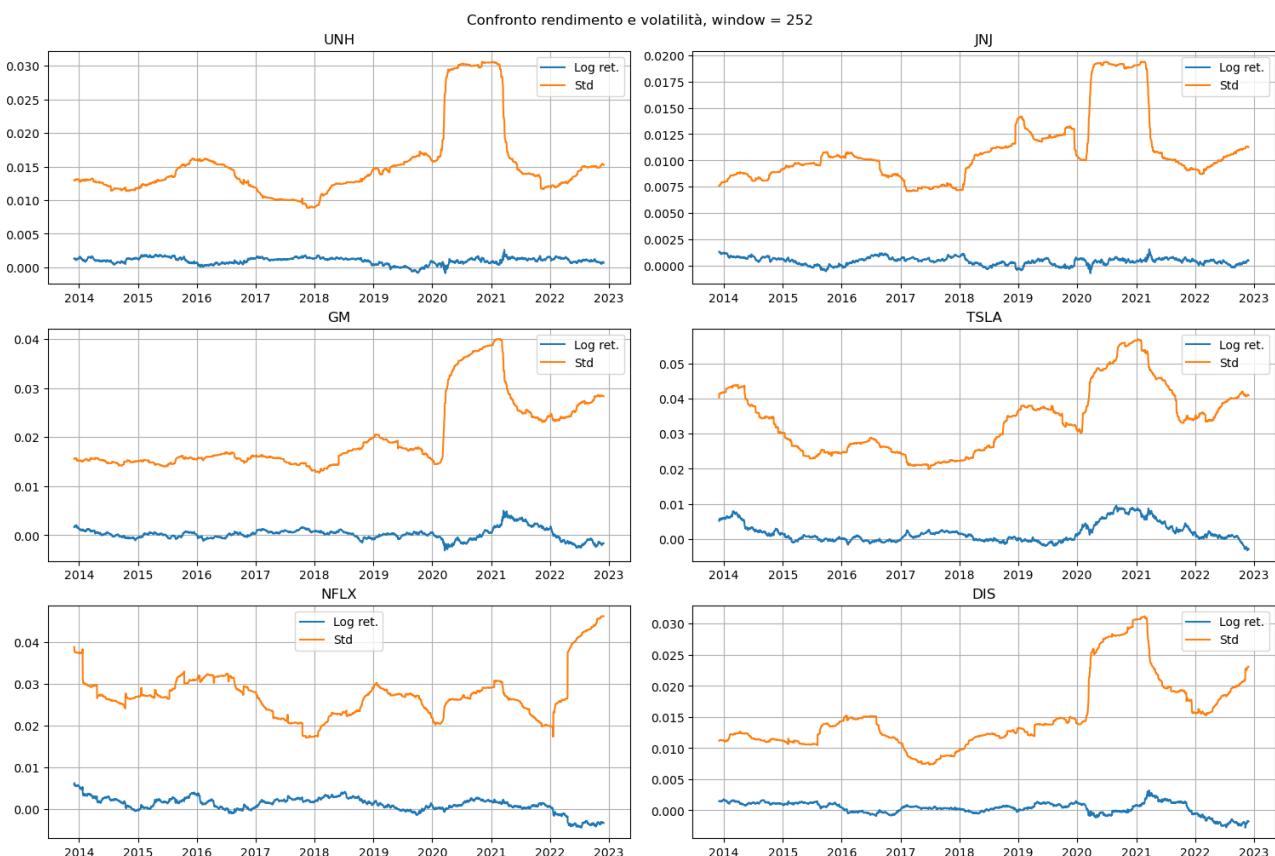
L'azione con la distribuzione di rendimenti più vicina alla normale è Tesla, mentre quella più lontana è Netflix, con curtosi molto alta.

Analizzando la media, il rendimento più alto è quello di Tesla, ma è anche vero che ha una deviazione standard alta: in particolare **Tesla è il titolo con deviazione standard più alta**.

Sempre facendo riferimento alla media, il rendimento più basso è quello del titolo General Motors.

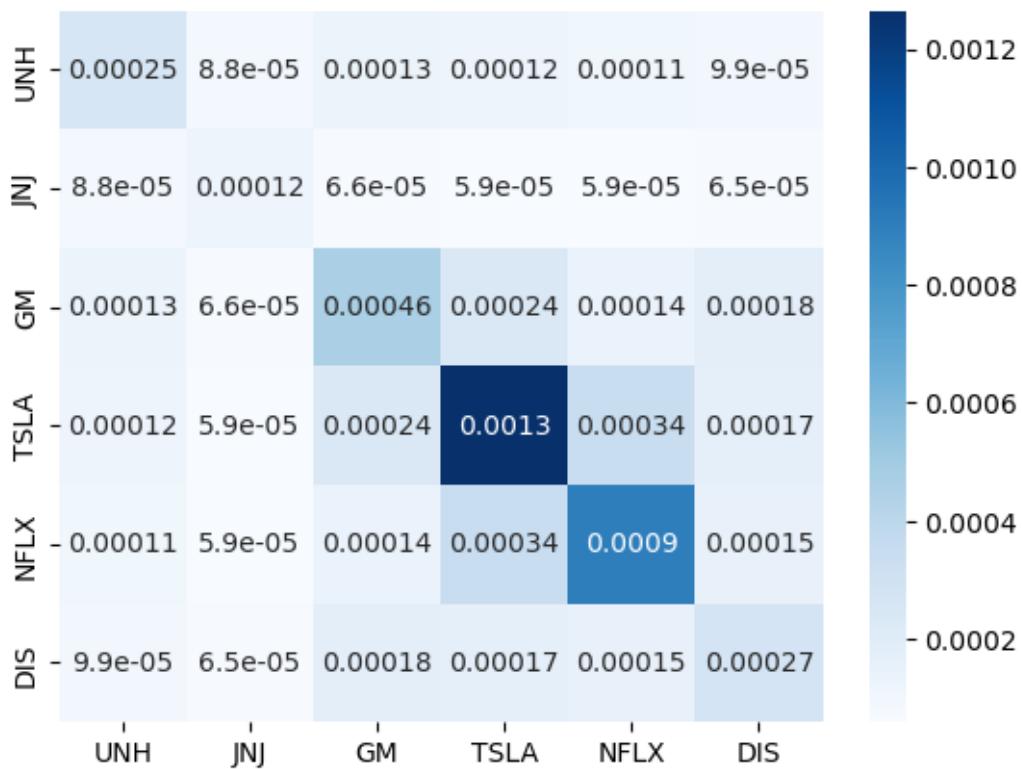
Per quanto riguarda la deviazione standard più bassa, è il caso di Johnson & Johnson: potrei pensare che si tratti del titolo meno volatile.

Confronto nel tempo rendimento e volatilità piazzando altri grafici, che ne mostrano l'andamento con una time window di 252 giorni (= numero di trading days in un anno).



Matrice di covarianze

La matrice di covarianze ci dà la covarianza tra qualsiasi coppia di asset; inoltre ci permette di risalire alle **varianze** dei singoli asset, leggendo la diagonale principale.



Noto che Tesla è il titolo con varianza più alta, come confermato anche nel punto precedente.

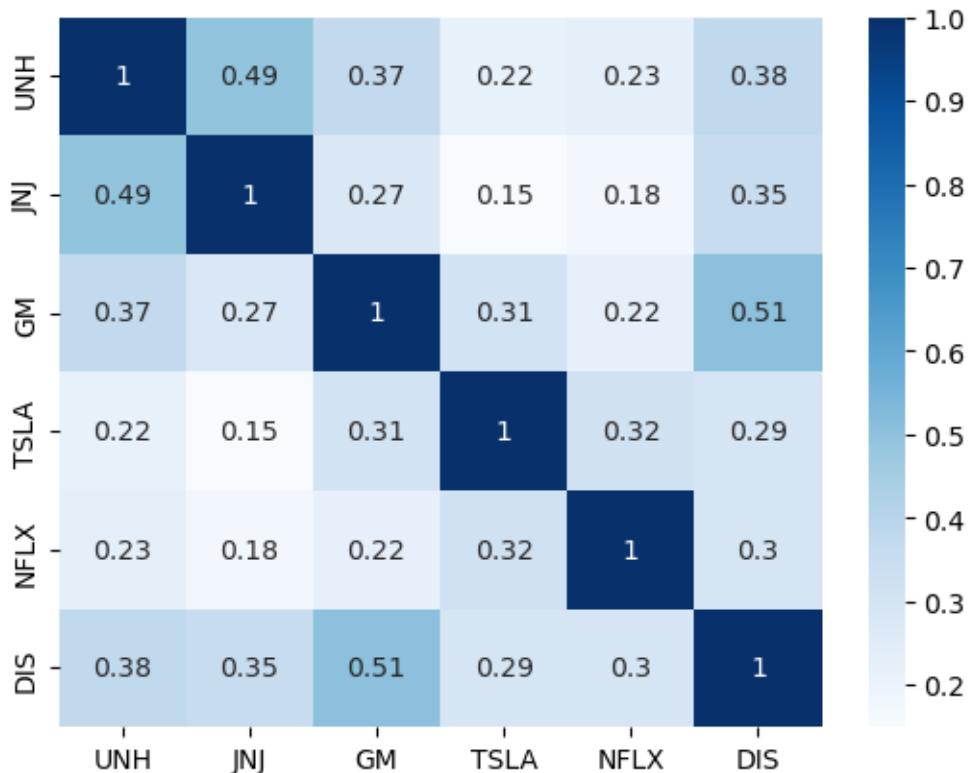
Di particolare importanza è anche la varianza di Netflix.

Guardando il resto della matrice (cioè le **covarianze**), noto come alcuni asset siano più variabili di altri: Tesla e Netflix, nonostante siano di due settori diversi, è la coppia con variabilità più alta.

Al secondo posto abbiamo General Motors e Tesla.

Matrice di correlazione

La matrice di correlazione ci dà invece una misura di correlazione tra coppie di asset.

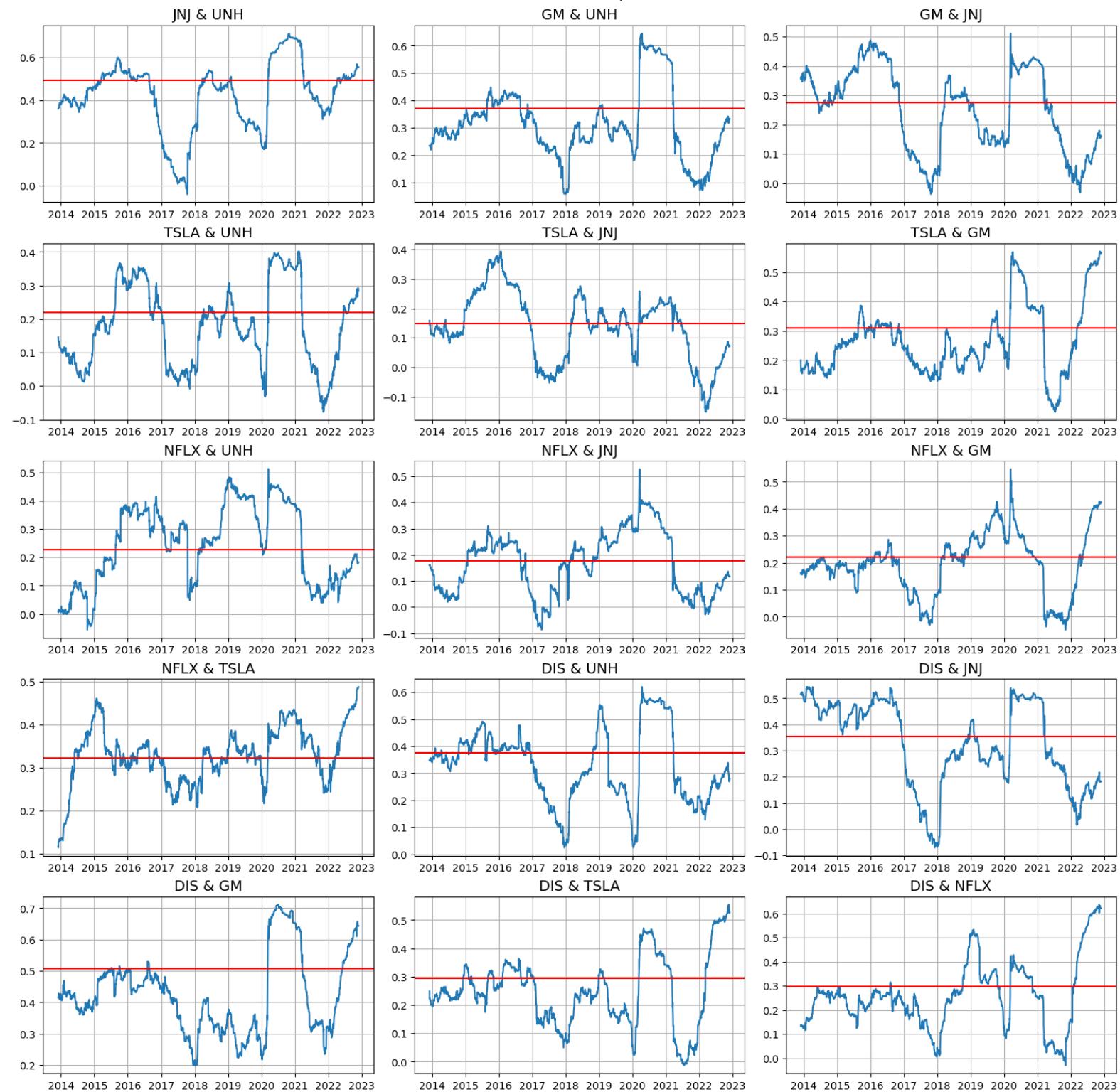


Inaspettatamente, osservo che i titoli **più correlati** sono Disney e General Motors, mentre la seconda coppia è Johnson & Johnson e UnitedHealth Group, che mi aspettavo in quanto dello stesso settore.

I titoli **meno correlati** consistono nelle coppie Tesla e Johnson & Johnson, Netflix e Johnson & Johnson.

Andamento correlazioni titoli

Andamento correlazioni titoli, window = 252



Questi grafici evidenziano la correlazione nel tempo (window sempre di 252 giorni) di tutte le 15 coppie distinte tra i titoli.

La riga in rosso corrisponde al valore della correlazione media, estratta dalla matrice di correlazione.

Dalle immagini posso avere la conferma di quanto osservato prima attraverso la matrice e inoltre posso vedere nel tempo quali sono i periodi in cui è stata superata la correlazione media.

Una cosa che posso osservare è che nel 2020 tutte le coppie di titoli hanno registrato un'alta correlazione.

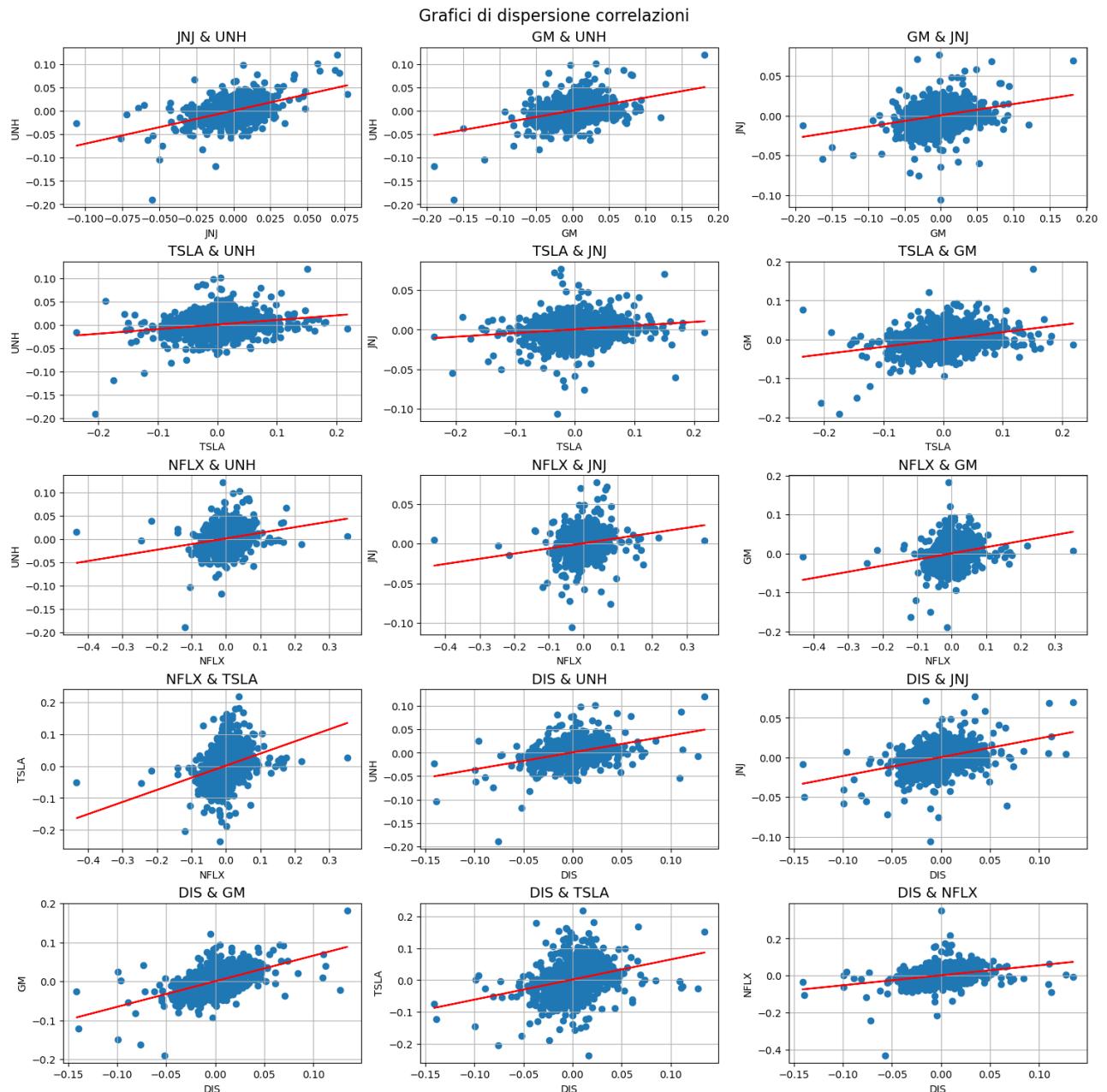
Nel tempo le correlazioni hanno avuto sia discese sia salite, ma è importante notare come guardando nell'ultimo periodo ci sia una tendenza rialzista per coppie di asset dello stesso settore, che può essere poco evidente rispetto alla correlazione media (coppia Johnson & Johnson e UnitedHealth Group) oppure molto evidente (coppie Tesla e General Motors, Disney e Netflix).

Un'altra cosa da sottolineare è il fatto che ci siano coppie la cui correlazione è stata negativa per certi archi temporali. Queste riguardano anche titoli dello stesso settore.

Non mi sembra significativo mostrare un confronto tra correlazioni e rendimenti, in quanto le correlazioni riguardano coppie di asset mentre i rendimenti un singolo asset.

Invece, può essere significativo mostrare tramite grafici di dispersione (scatterplot) la relazione lineare per ogni coppia di rendimenti.

Grafici di dispersione correlazioni medie



La retta rossa è la retta di regressione che meglio approssima la relazione lineare tra i punti. Osservando sia la dispersione dei punti sia la pendenza della retta di regressione, possiamo dire di avere una conferma della relazione tra rendimenti discussa precedentemente.

Analisi di previsione

Per l'analisi di previsione, innanzitutto preparo un dataframe di adjusted close mensili. Sul resample utilizzo last() per non lavorare su dati stimati come la media, ma dati di fine mese.

Ciò che andrò a prevedere sono i prezzi e non i rendimenti.

Test Adfuller

A scopo didattico, verifico se le serie dei prezzi sono stazionarie o meno.

Eseguendo il test statistico Adfuller (con ipotesi nulla: serie non-stazionaria, ipotesi alternativa: serie stazionaria), ottengo i seguenti p-value:

- UnitedHealth Group ≈ 0.99746
- Johnson & Johnson ≈ 0.96641
- General Motors ≈ 0.26295
- Tesla ≈ 0.00079
- Netflix ≈ 0.58979
- Disney ≈ 0.32969

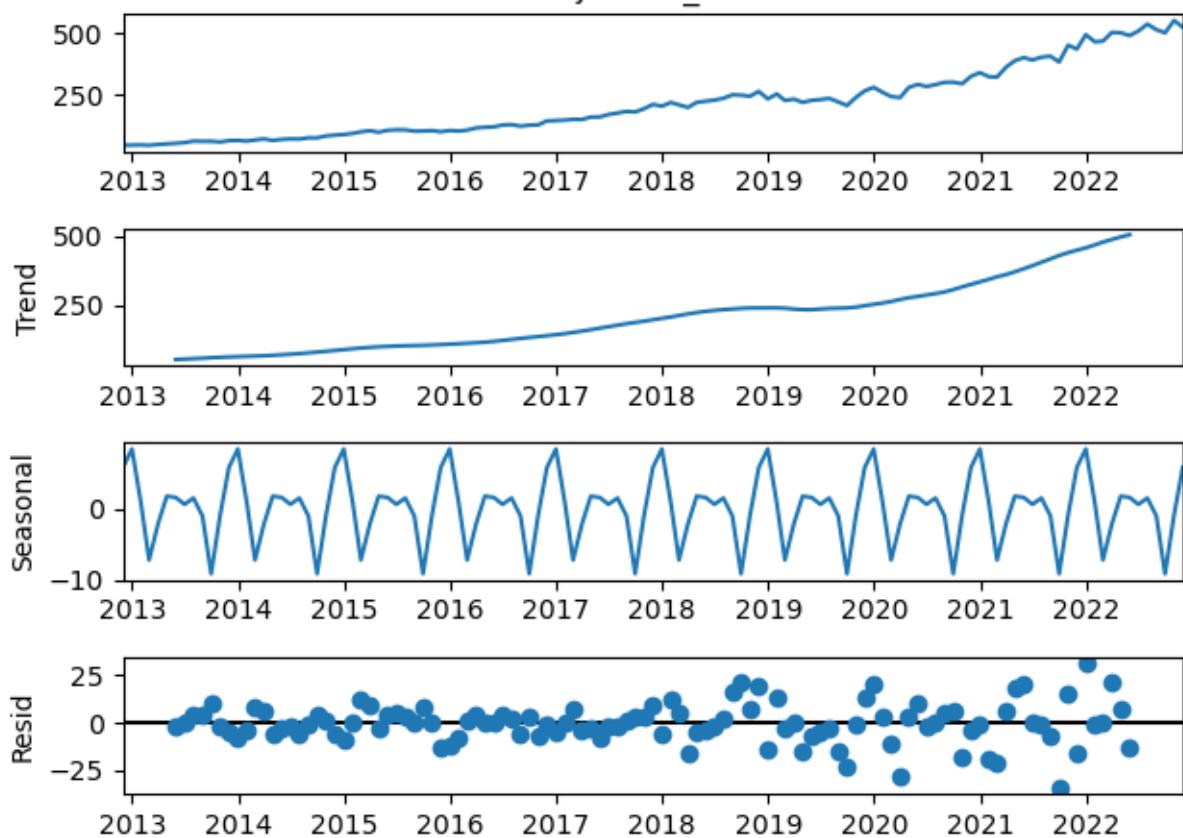
Stabilendo un livello di significatività pari a 0.05, rifiuto l'ipotesi nulla se il p-value è ≤ 0.05 : l'unica serie che risulta stazionaria è Tesla, ma di pochi ordini di grandezza. Vedremo come non sarà influente.

La non-stazionarietà non è un problema in quanto andrò ad utilizzare SARIMAX come modello di previsione, che utilizza la differenziazione per trasformare una serie non-stazionaria in una serie stazionaria.

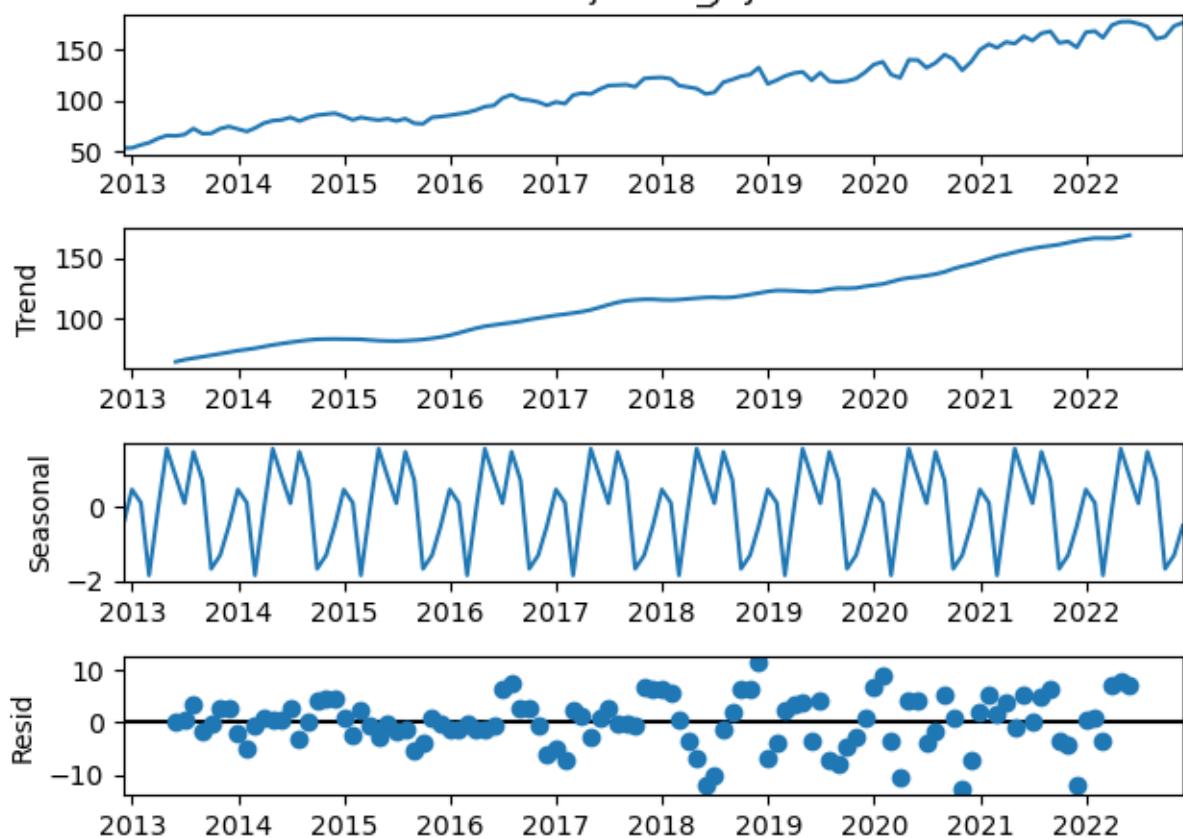
Trend, seasonality e resid

A questo punto, controllo se esiste un trend e una stagionalità, tramite la decomposizione delle serie. Essa mi fornisce la componente di trend, la componente stagionale e la componente "random" (rumore, residui).

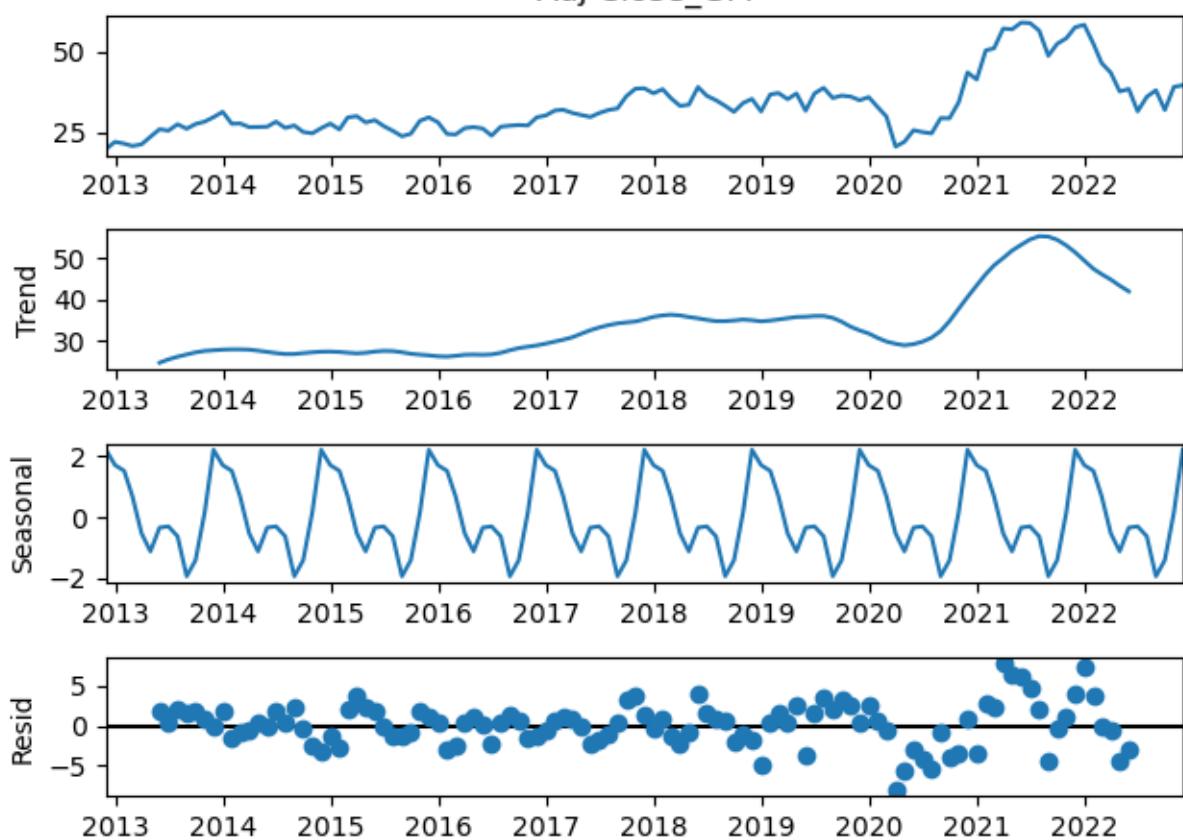
Adj Close_UNH



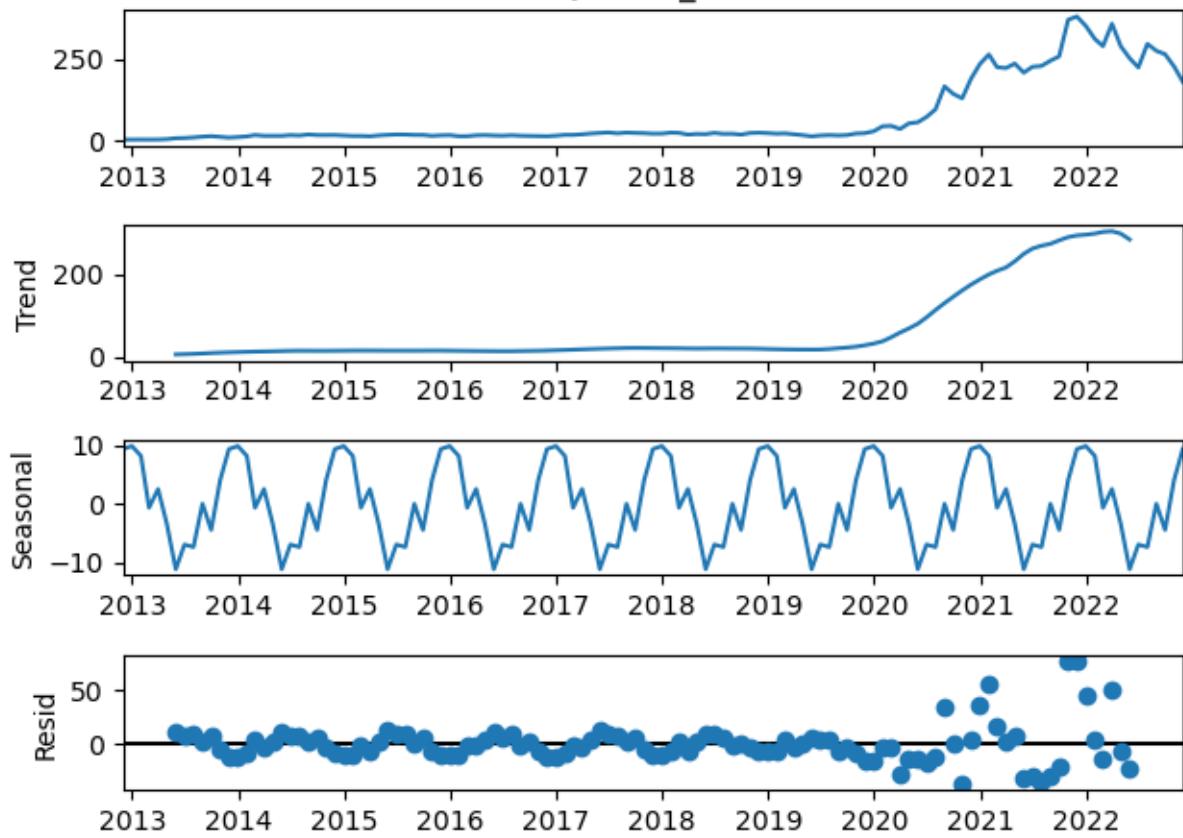
Adj Close_JNJ



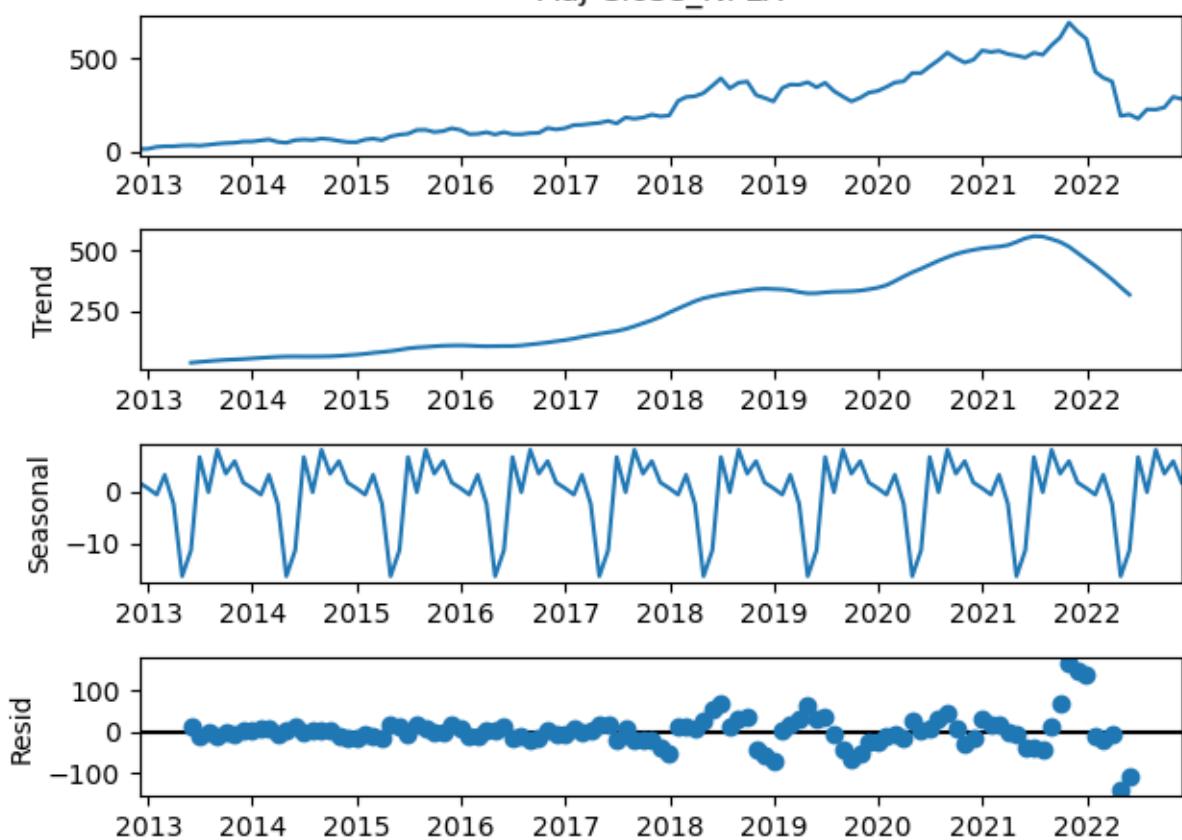
Adj Close_GM



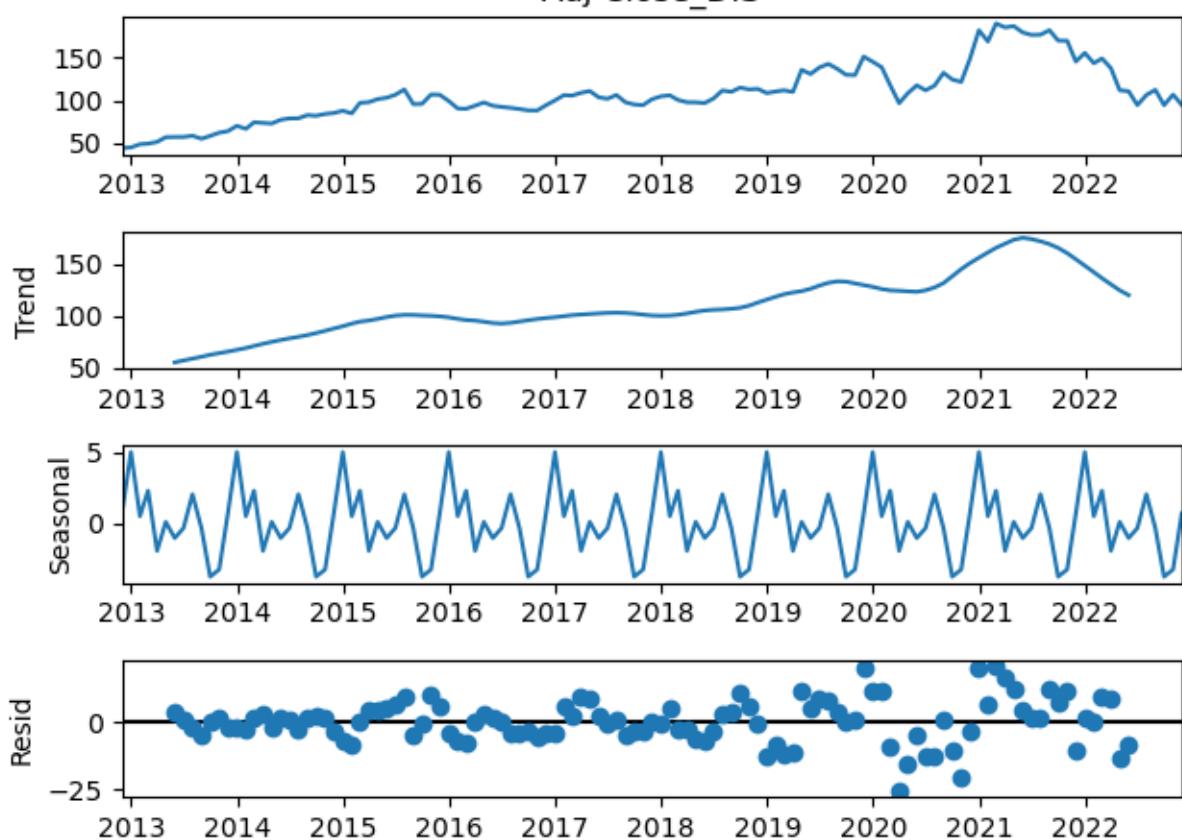
Adj Close_TSLA



Adj Close_NFLX



Adj Close_DIS



Osservo che in tutte le serie sono presenti una componente stagionale e una componente di trend.

Grid search

Utilizzo l'algoritmo grid search per cercare gli iperparametri ottimali da assegnare al modello, ottenendo il seguente assegnamento:

```
{'UNH': (797.8604888313384, (1, 1, 0), (1, 1, 0, 12)),  
 'JNJ': (625.1275197637207, (1, 1, 1), (0, 1, 1, 12)),  
 'GM': (493.28281983229346, (0, 1, 0), (1, 1, 1, 12)),  
 'TSLA': (844.2915014692462, (1, 1, 1), (1, 1, 0, 12)),  
 'NFLX': (923.686712000156, (0, 1, 0), (0, 1, 1, 12)),  
 'DIS': (717.5213637872664, (0, 1, 0), (0, 1, 1, 12))}
```

Il primo valore mostrato è l'AIC (Akaike Information Criterion): stimatore dell'errore di predizione che la grid search tende a minimizzare.

Vi sono poi gli iperparametri trend (p, d, q) e seasonal (P, D, Q, m), con $m=12$ che ho passato perché ho 12 mesi come stagionalità.

Noto che nella parte Integrativa di ARIMA ho come iperparametro $d=1$ (così come anche l'iperparametro D nella componente seasonal) per ogni serie, perciò SARIMAX utilizzerà la **differenziazione** per rendere **tutte** le serie stazionarie, in fase di fitting.

Per effettuare la grid search ho utilizzato i primi 110 mesi, cioè training set e validation set.

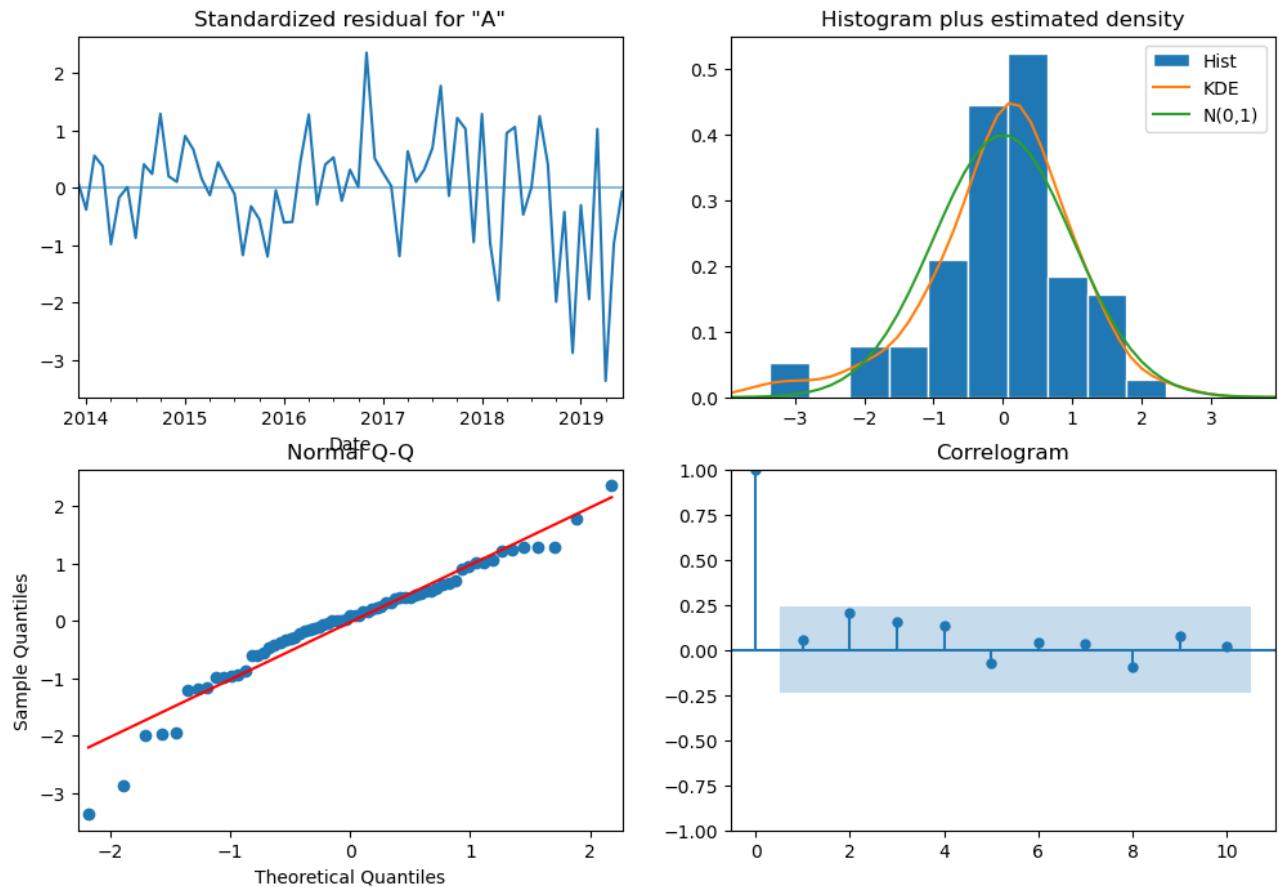
SARIMAX

Per ogni asset, creo il relativo modello SARIMAX effettuando fit() sul training set (primi 80 mesi).

Per ogni modello creato, ho osservato il summary ma non mi sembra significativo e quindi non lo riporto qui. Questo perché nonostante ci siano modelli con AIC minori rispetto ad altri, in seguito si evidenzierà come la previsione dei movimenti nel prezzo di mercato sia sempre scarsa: non tutto è prevedibile.

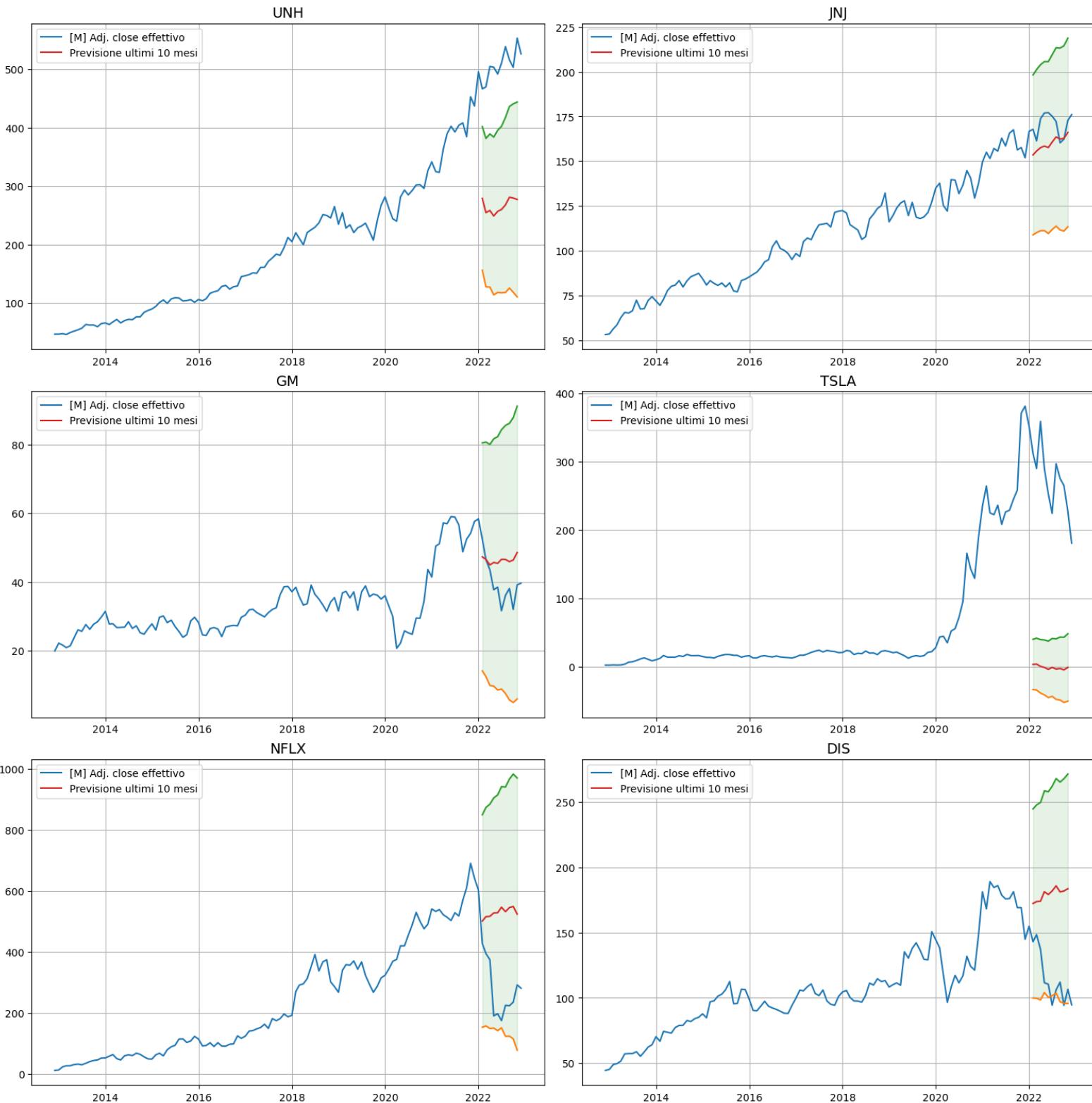
Una cosa che si può notare dai grafici è che la distribuzione dei residui risulta molto simile a quella della normale.

Ad esempio, questi sono i grafici generati dal modello associato a UnitedHealth Group:



Sia dall'istogramma sia dal Q-Q plot vediamo quanto detto sopra.

Fatte queste diverse osservazioni, per ogni titolo effettuo la previsione degli ultimi 10 mesi, mostrando anche l'intervallo di confidenza.



Come accennato prima, la previsione rispetto all'adjusted close effettivo degli ultimi 10 mesi risulta scarsa, soprattutto per UnitedHealth Group e per Tesla perché l'intervallo di confidenza non comprende il valore effettivo.

Questo fa capire come l'andamento dei prezzi **non sia sempre prevedibile**.

MSE e RMSE

Per misurare l'errore delle previsioni degli ultimi 10 mesi rispetto ai valori effettivi, calcolo l'errore quadratico medio (Mean Squared Error) e anche la sua radice (Root Mean Square Error).

----- MSE e RMSE UNH -----

MSE: 60901.21301010971
RMSE: 246.7817112553313

----- MSE e RMSE TSLA -----

MSE: 73501.06901435158
RMSE: 271.11080578676973

----- MSE e RMSE JNJ -----

MSE: 176.79885125442996
RMSE: 13.296572913891382

----- MSE e RMSE NFLX -----

MSE: 79061.61716192737
RMSE: 281.17897709808847

----- MSE e RMSE GM -----

MSE: 81.06617629977988
RMSE: 9.003675710496235

----- MSE e RMSE DIS -----

MSE: 5028.236784853556
RMSE: 70.91006123854045

Da questi valori posso confermare quanto osservato prima dal grafico: l'errore nelle previsioni di Tesla e UnitedHealth Group sono alti.

Noto anche come l'errore più alto è quello di Netflix, nonostante il prezzo effettivo sia nell'intervallo di confidenza (in quanto ha un'ampiezza molto grande): infatti, guardando il prezzo forecasted si nota questa grande differenza rispetto al prezzo effettivo.

Strategie di trading e backtesting

Costruzione strategia

Come strategia ho scelto di utilizzare una strategia combo basata sulle bande di Bollinger e sull'indicatore RSI.

Come granularità dei dati scelgo di utilizzare dati giornalieri, per avere potenzialmente più segnali in tutto il periodo e per avere grafici più “smooth” rispetto ad avere salti da un mese all’altro.

Per le **bande di Bollinger**, uso una media mobile a $G = 20$ giorni a cui aggiungo (per determinare la banda superiore) e sottraggo (banda inferiore) la deviazione standard moltiplicata per $F = 2$. Queste due bande mi generano un segnale di **buy** se il prezzo esce dalla banda inferiore e successivamente rientra, mentre ho un segnale di **sell** nel caso in cui il prezzo esce dalla banda superiore e poi rientra.

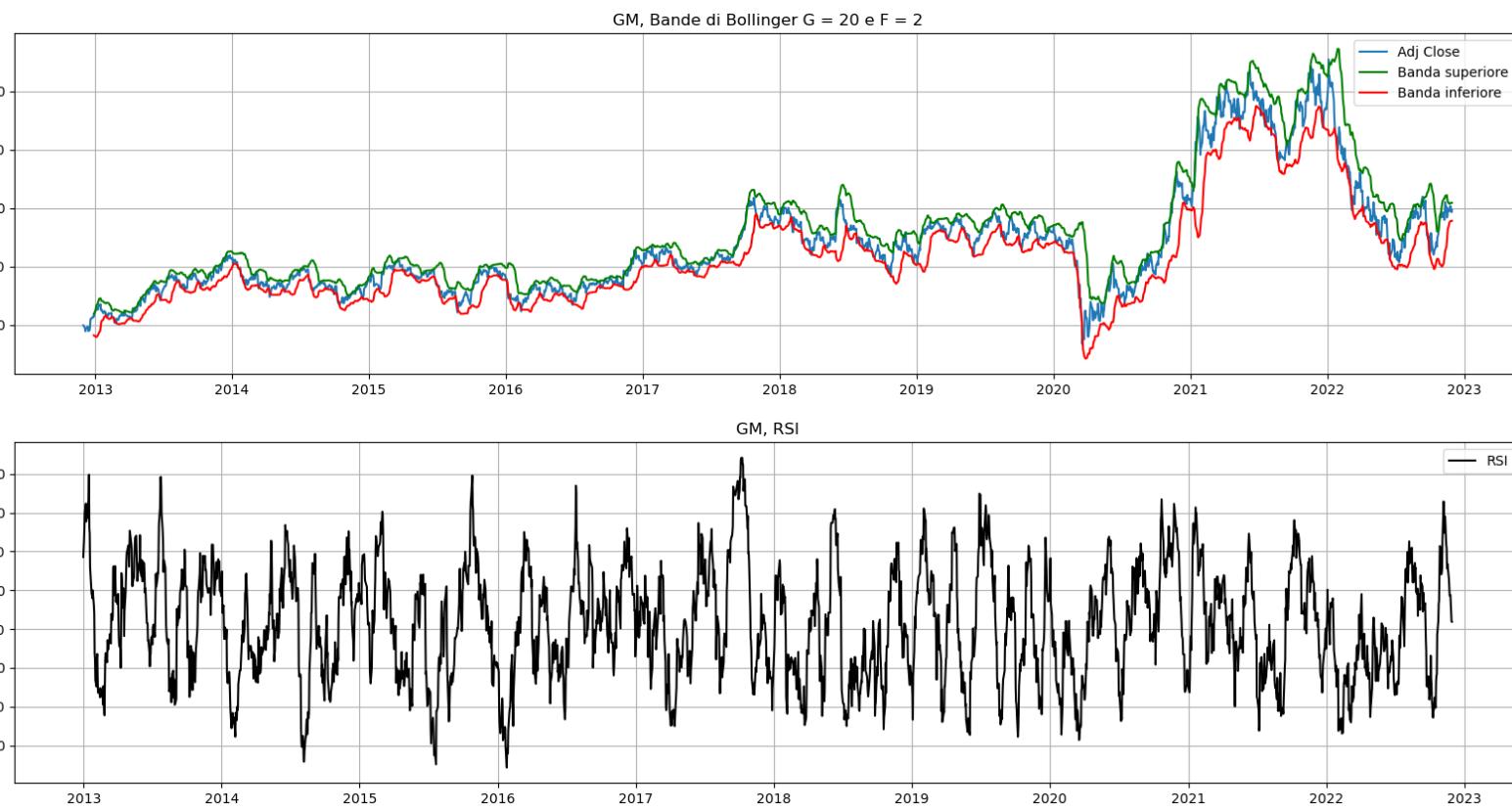
Invece, l'indicatore **RSI** (indice di forza relativa, Relative Strength Index) è un indicatore di momentum che può assumere valore da 0 a 100: se ho un $RSI < 30$ ho un segnale di **acquisto** (situazione di ipervenduto), mentre se $RSI > 70$ ho un segnale di **vendita** (ipercomprato).

Combino i due segnali delle bande di Bollinger e di RSI, in modo tale che considero un segnale solo se viene generato da entrambe le componenti.

Come titolo per la mia strategia ho scelto di utilizzare General Motors perché altrimenti in caso di un forte mercato l'indicatore RSI darà prematuramente segnali di ipercomprato o ipervenduto.

Di seguito mostro gli ultimi dati del dataframe che ho generato e poi plotto sia le bande di Bollinger sia l'oscillatore RSI.

Date	Adj Close	upper_band	lower_band	rsi
2022-11-22	39.881310	40.721173	37.444993	60.787180
2022-11-23	40.021000	40.762433	37.628235	58.647205
2022-11-25	40.370220	40.867922	37.752235	58.805510
2022-11-28	39.292618	40.873129	37.799911	51.970274
2022-11-29	39.661797	40.902398	37.820530	51.862890



Backtesting

A questo punto, utilizzando gli adjusted close giornalieri di General Motors, genero segnali di buy e sell utilizzando le regole sopra descritte, cioè quando il prezzo esce dalla banda superiore e poi vi rientra E RSI > 70 ho un segnale di vendita, mentre quando il prezzo esce dalla banda inferiore e poi vi rientra E RSI < 30 ho un segnale di acquisto.

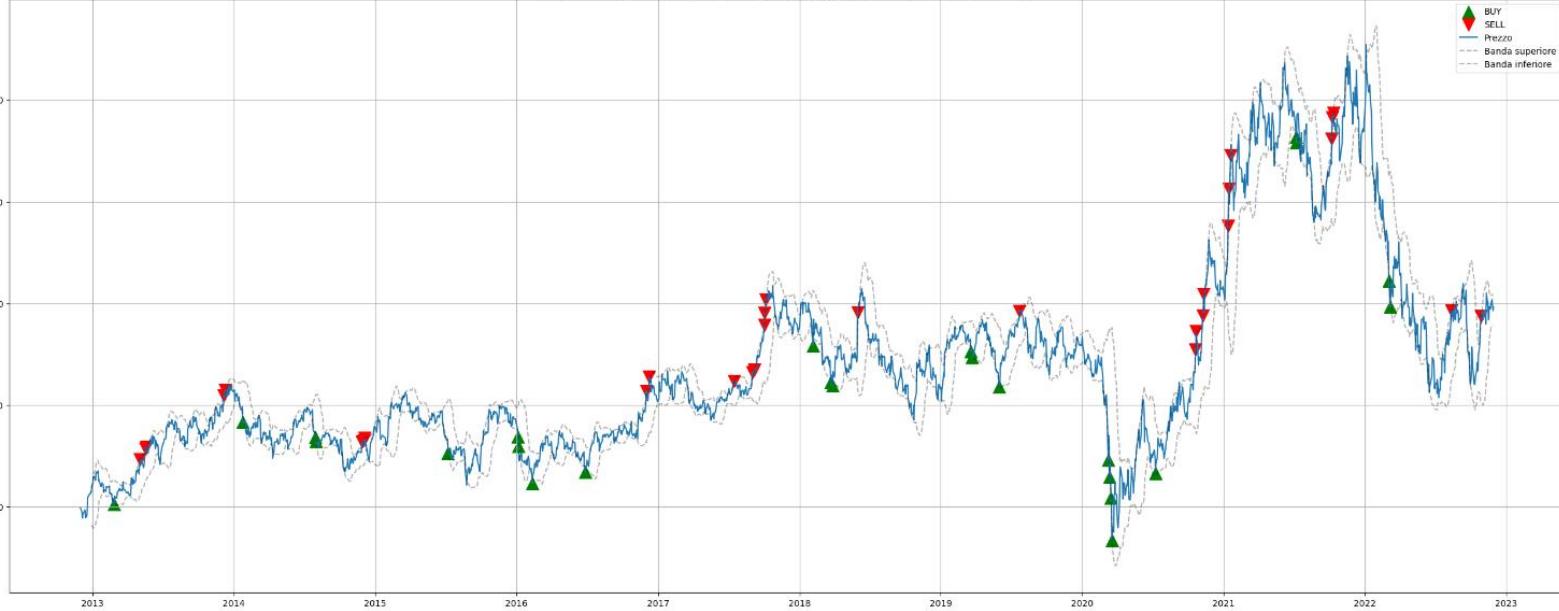
Distinguo tra 3 possibili segnali:

- 1: segnale di vendita
- 0: nessun segnale (rimango con la posizione attuale)
- 1: segnale di acquisto

Contando i valori assunti, ho 30 segnali di vendita, 24 segnali di acquisto e 2463 giorni senza alcun segnale.

Plotto i segnali generati in un grafico.

Strategia combo RSI + bande di Bollinger, GM, G = 20 e F = 2



Oltre ai segnali ho memorizzato nel dataframe anche le posizioni, in cui vado a rimpiazzare i segnali che hanno valore 0 con il precedente valore diverso da 0 (se possibile, ovviamente).

Le posizioni mi torneranno utile per calcolare i rendimenti.

Rendimenti e confronto con Buy & Hold

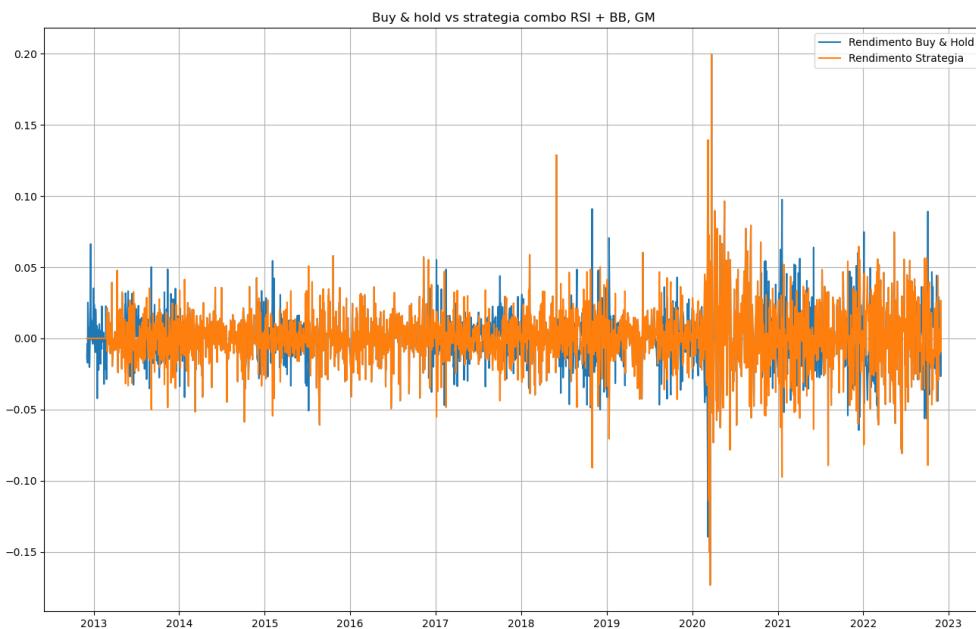
Per valutare la performance della strategia occorre confrontare con la strategia Buy & Hold, strategia usata come benchmark perché è la più semplice in quanto dopo aver investito non si fa più trading.

Calcolo i rendimenti **cumulati** non sull'intero periodo ma con il passare dei giorni (quelli che a lezione di teoria abbiamo definito come **ritorno lordo semplice a k-periodi**).

Sfrutto le posizioni che ho calcolato prima per ottenere i rendimenti semplici della strategia (moltiplicando le posizioni con i rendimenti semplici dell'asset) ed effettuo `cumprod()` su 1 + rendimenti semplici, per calcolare i rendimenti cumulati ogni giorno (sottolineo, ancora una volta, che non si tratta dei rendimenti cumulati sull'intero periodo).

Per la Buy & Hold è sufficiente ripetere i calcoli partendo dai rendimenti semplici dell'asset.

Prima di fare il confronto con i dati ottenuti, volevo motivare la scelta di non utilizzare direttamente i rendimenti semplici delle due strategie. Se non utilizzassi i rendimenti cumulati ma i rendimenti semplici, avrei il seguente grafico:

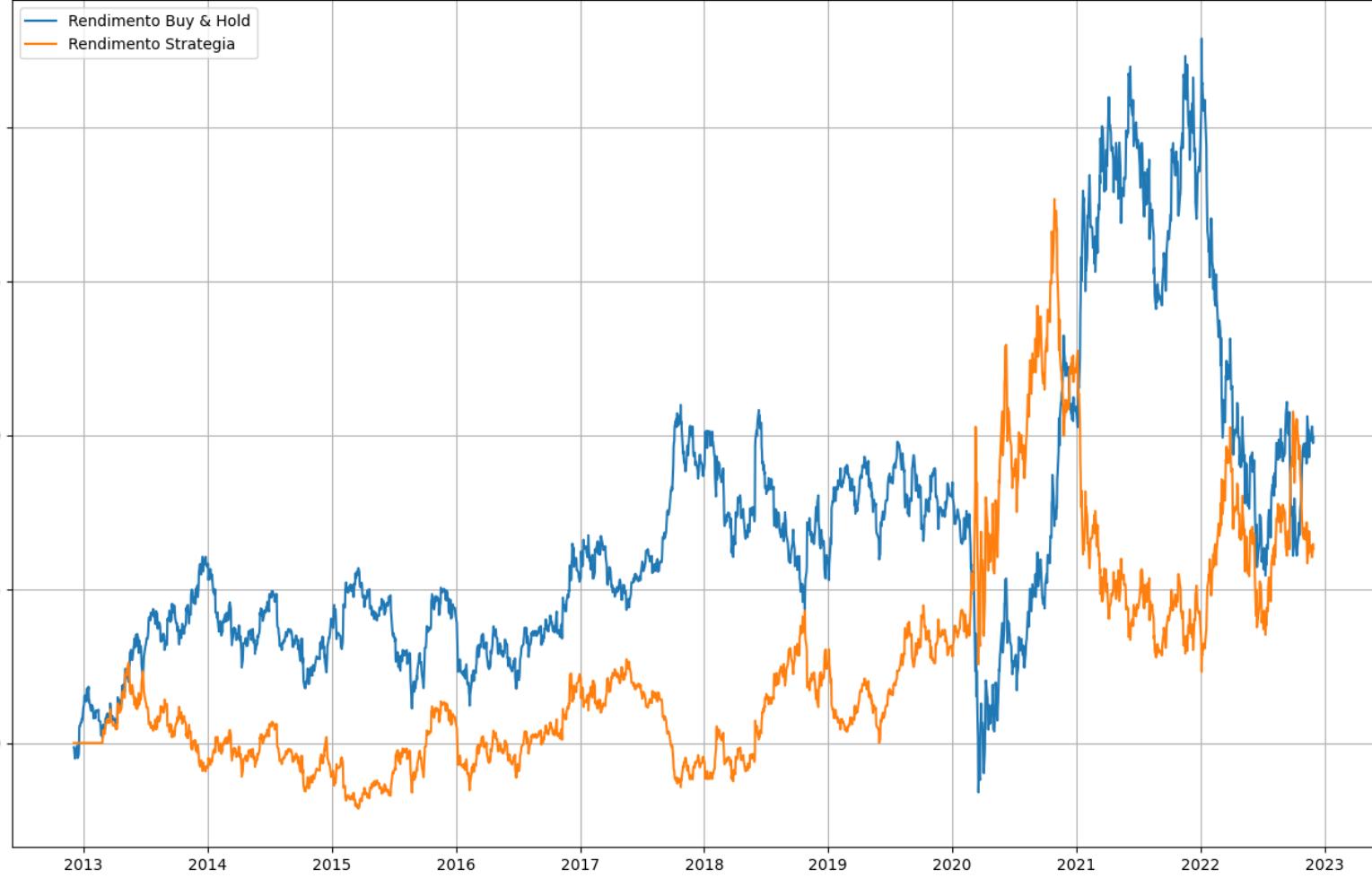


Questo grafico non è per niente significativo (scherzando sembrano spaghetti), quindi mi sembra particolarmente appropriato non utilizzare i rendimenti semplici ma i rendimenti cumulati.

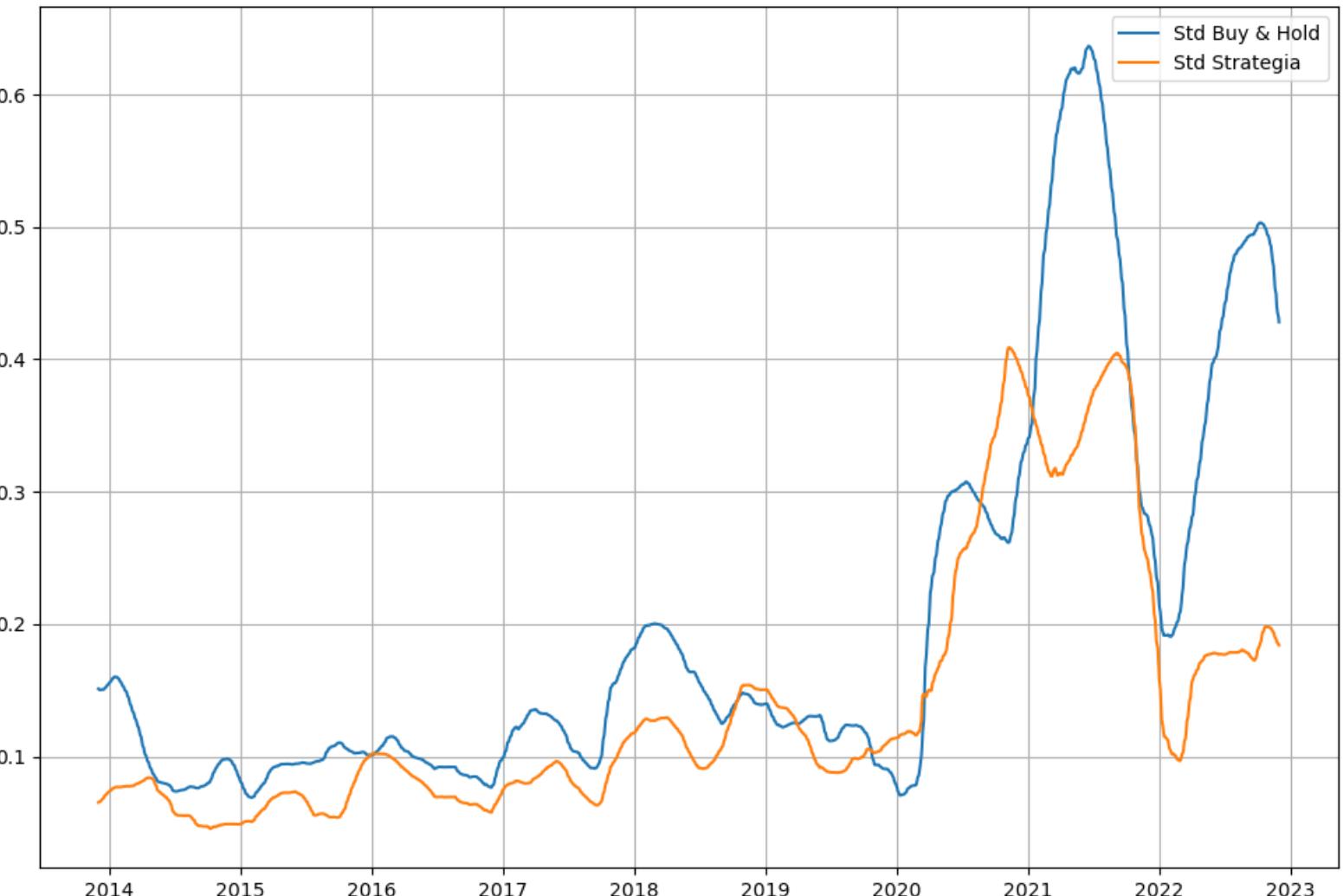
Ora che ho motivato la scelta di utilizzare i rendimenti cumulati, mostro le ultime righe del nuovo dataframe ottenuto, poi plotto in un grafico il confronto tra le due strategie e anche la volatilità con il passare del tempo (come al solito utilizzo window di 252 giorni).

	buy_and_hold	strategy
Date		
2022-11-22	2.002813	1.622656
2022-11-23	2.009829	1.616972
2022-11-25	2.027366	1.602863
2022-11-28	1.973250	1.645648
2022-11-29	1.991790	1.630186

Buy & hold vs strategia combo RSI + BB, GM



Volatilità



Posso fare diverse osservazioni guardando il confronto tra le due strategie.

La strategia dal 2013 al 2020 non ha mai superato la Buy & Hold, poi nel 2020 è riuscita ad avere rendimenti più alti ma questa situazione si rivelerà temporanea, in quanto con l'inizio del 2021 è tornata a rendere meno rispetto alla strategia di benchmark.

Potenzialmente nel futuro potrebbe avere rendimenti migliori, in quanto dal 2022 continuano a incrociarsi i rendimenti della Buy & Hold e della strategia, e una cosa importante da sottolineare è che la volatilità della strategia non è alta quanto quella della Buy & Hold.

Infatti, nel complesso la volatilità della strategia è più bassa rispetto alla volatilità della Buy & Hold, quindi la Buy & Hold per questo asset mi permette di avere rendimenti migliori rispetto alla mia strategia ma con volatilità più alta.

Come ulteriore conclusione, posso dire che la mia strategia, nonostante sia già una combo, per risultare efficiente ha bisogno di essere utilizzata assieme ad altre strategie che forniscono segnali, poiché ad esempio le bande di Bollinger possono dare falsi segnali in generale.

CAPM

Beta di ciascun titolo rispetto al mercato (S&P 500)

Dopo aver scaricato i dati dell'indice azionario S&P 500 (pesato per capitalizzazione) da yfinance (ticker ^GSPC), inizio calcolando i suoi rendimenti semplici mensili.

Per ogni titolo, creo un nuovo dataframe unendo i rendimenti mensili dell'asset con quelli del mercato (S&P 500), poi calcolo la matrice di covarianze perché mi serve per il calcolo del coefficiente beta.

Per calcolare beta basta dividere la covarianza tra asset e mercato per la varianza del mercato.

Di seguito riporto i valori di **beta** per ogni asset:

```
{ 'UNH': 0.72517717579179,  
  'JNJ': 0.6238125839622325,  
  'GM': 1.3606001482326306,  
  'TSLA': 1.7465926791682234,  
  'NFLX': 1.3127338236947337,  
  'DIS': 1.2678714477433939}
```

Posso osservare che tutti sono correlati rispetto al movimento del mercato, essendo positivi, ma solo gli ultimi 4 titoli (**General Motors, Tesla, Netflix e Disney**) hanno **beta > 1**.

Questi ultimi 4 asset si muovono nella stessa direzione del mercato con volatilità più alta: potrebbero battere il mercato.

Invece, per quanto riguarda la volatilità degli altri titoli non posso affermare nulla.

Esposizione di ciascun titolo ai fattori Fama-French

Il modello Fama-French a tre fattori consiste in:

- fattore mercato (**MKT**): dipendenza dall'andamento del mercato
- fattore dimensione (**Small Minus Big**): rendimento in eccesso delle azioni a piccola capitalizzazione rispetto alle grandi
- fattore valore (**High Minus Low**): rendimento in eccesso delle azioni con un rapporto alto fra patrimonio e prezzo (value stocks) e quelle con un rapporto basso (growth stocks).

Dopo aver [scaricato i dati](#), tolgo le prime 3 righe e la parte annuale: prendo in considerazione i dati mensili dei 3 fattori (MKT, SMB e HML) e del tasso risk-free (RF).

Creo un dataframe in cui sistemo i dati dal 30-11-2012 al 30-11-2022 e riporto di seguito sia le prime righe sia le ultime righe:

	mkt	smb	hml	rf		mkt	smb	hml	rf
date					date				
2012-12	0.0118	0.0150	0.0351	0.0001	2022-06	-0.0843	0.0209	-0.0597	0.0006
2013-01	0.0557	0.0033	0.0096	0.0000	2022-07	0.0957	0.0281	-0.0410	0.0008
2013-02	0.0129	-0.0028	0.0011	0.0000	2022-08	-0.0377	0.0139	0.0031	0.0019
2013-03	0.0403	0.0081	-0.0019	0.0000	2022-09	-0.0935	-0.0082	0.0003	0.0019
2013-04	0.0155	-0.0236	0.0045	0.0000	2022-10	0.0783	0.0010	0.0806	0.0023

Ora ho a disposizione tutti i dati per generare il modello a tre fattori.

Per ogni titolo, calcolo l'eccesso di ritorno dell'asset rispetto al risk free, poi creo un modello di regressione lineare OLS (Ordinary Least Squares), in cui si va ad approssimare questo valore come somma dei tre fattori Fama-French, cioè eccesso di ritorno dell'asset $\sim \text{mkt} + \text{smb} + \text{hml}$.

Effettuato il fit di ogni modello, riporto di seguito la parte fondamentale di ogni summary:

UnitedHealth Group

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0148	0.005	3.208	0.002	0.006	0.024
mkt	0.6956	0.107	6.478	0.000	0.483	0.908
smb	0.0118	0.185	0.064	0.949	-0.354	0.377
hml	-0.0227	0.127	-0.178	0.859	-0.275	0.230

Johnson & Johnson

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0037	0.003	1.112	0.268	-0.003	0.010
mkt	0.6388	0.078	8.186	0.000	0.484	0.793
smb	-0.4427	0.134	-3.300	0.001	-0.708	-0.177
hml	0.0649	0.093	0.701	0.485	-0.118	0.248

General Motors

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.0031	0.006	-0.541	0.589	-0.014	0.008
mkt	1.1966	0.132	9.048	0.000	0.935	1.459
smb	0.7925	0.227	3.486	0.001	0.342	1.243
hml	0.6799	0.157	4.333	0.000	0.369	0.991

Tesla

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0349	0.016	2.187	0.031	0.003	0.066
mkt	1.8000	0.370	4.867	0.000	1.067	2.532
smb	0.2265	0.636	0.356	0.722	-1.033	1.486
hml	-0.9744	0.439	-2.221	0.028	-1.844	-0.105

Netflix

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0221	0.012	1.824	0.071	-0.002	0.046
mkt	1.3657	0.281	4.858	0.000	0.809	1.923
smb	-0.0198	0.483	-0.041	0.967	-0.977	0.937
hml	-0.8983	0.334	-2.693	0.008	-1.559	-0.238

Disney

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.0034	0.005	-0.685	0.495	-0.013	0.006
mkt	1.2480	0.114	10.972	0.000	1.023	1.473
smb	-0.0146	0.196	-0.075	0.941	-0.402	0.373
hml	0.1272	0.135	0.943	0.348	-0.140	0.395

Guardando la colonna **coef**, si ottiene l'esposizione di ciascun titolo rispetto ai 3 fattori Fama-French.

Valutando singolarmente i fattori, risultano le seguenti esposizioni:

- **mkt**: Tesla ha la dipendenza dall'andamento del mercato più alta (1.8), così come avevamo già osservato dai coefficienti beta. La dipendenza più bassa dall'andamento del mercato è data da Johnson & Johnson (0.6388): infatti, questo ha il beta minore.
- **smb**: il fattore Small Minus Big più alto è dato da General Motors (0.7925), mentre quello più basso Johnson & Johnson (-0.4427).
- **hml**: il fattore High Minus Low più alto è quello corrispondente a General Motors (0.6799) e quello più basso è di Tesla (-0.9744).

Utilizzo di beta per calcolare rendimento atteso annuo

È possibile utilizzare i valori di beta calcolati precedentemente per calcolare il rendimento atteso annuo di ogni asset.

Si deve utilizzare la formula $E(R_i) = r_f + \beta_i(E(R_M) - r_f)$.

Ho bisogno di stimare due dati: il **tasso risk-free annuo** e il **rendimento atteso annuo di mercato**.

Per effettuare le stime, utilizzo il dataframe precedentemente creato per i fattori Fama-French, in quanto abbiamo la colonna **rf** e la colonna **mkt**, che fanno riferimento a dati mensili.

Utilizzo gli ultimi 12 mesi per entrambi e ne faccio la media, ottenendo stime mensili, poi moltiplico per 12 e ottengo le seguenti stime annue:

```
Tasso risk-free annuo: 0.0081
Rendimento atteso annuo di mercato: -0.1789
```

Uso queste stime e i beta calcolati in precedenza per calcolare il rendimento atteso annuo di ogni asset (secondo la formula riportata sopra):

```
Rendimento atteso annuo di UNH: -0.12750809837020283
Rendimento atteso annuo di JNJ: -0.10855307488422956
Rendimento atteso annuo di GM: -0.24633218948019628
Rendimento atteso annuo di TSLA: -0.3185128310044578
Rendimento atteso annuo di NFLX: -0.23738122503091524
Rendimento atteso annuo di DIS: -0.2289919635028822
```

È importante osservare che tutti i rendimenti attesi annui sono risultati negativi. La motivazione consiste nel fatto che la mia stima del rendimento atteso annuo di mercato è negativo.

Costruzione di portafoglio

Vi sono due modi per scegliere i rendimenti da utilizzare:

- rendimenti passati: utilizzo i primi 108 mesi
- rendimenti attesi costruiti nel punto precedente.

Per ognuno di essi, devo utilizzare sia il metodo di simulazione (Monte Carlo) sia il metodo analitico (SciPy).

Mi salvo in un dataframe i rendimenti mensili di ogni asset dei primi 108 mesi, in quanto mi servono ogni volta che utilizzo i rendimenti passati.

Questa è la coda del dataframe:

Date	rtn_UNH	rtn_JNJ	rtn_GM	rtn_TSLA	rtn_NFLX	rtn_DIS
2021-07-31	0.029418	0.045283	-0.039378	0.011034	-0.020144	0.001422
2021-08-31	0.009825	0.011375	-0.137755	0.070605	0.099735	0.029997
2021-09-30	-0.058036	-0.067175	0.075495	0.054042	0.072296	-0.066906
2021-10-31	0.178456	0.008545	0.032631	0.436530	0.131025	-0.000591
2021-11-30	-0.035290	-0.036399	0.063200	0.027612	-0.070128	-0.142959

Inoltre, inizializzo alcune variabili globali: il numero di portfolio da generare (per il metodo di simulazione), il numero di mesi da considerare (12), il numero degli asset (ovviamente 6) e la matrice di covarianze già moltiplicata per 12.

Rendimenti passati: metodo di simulazione

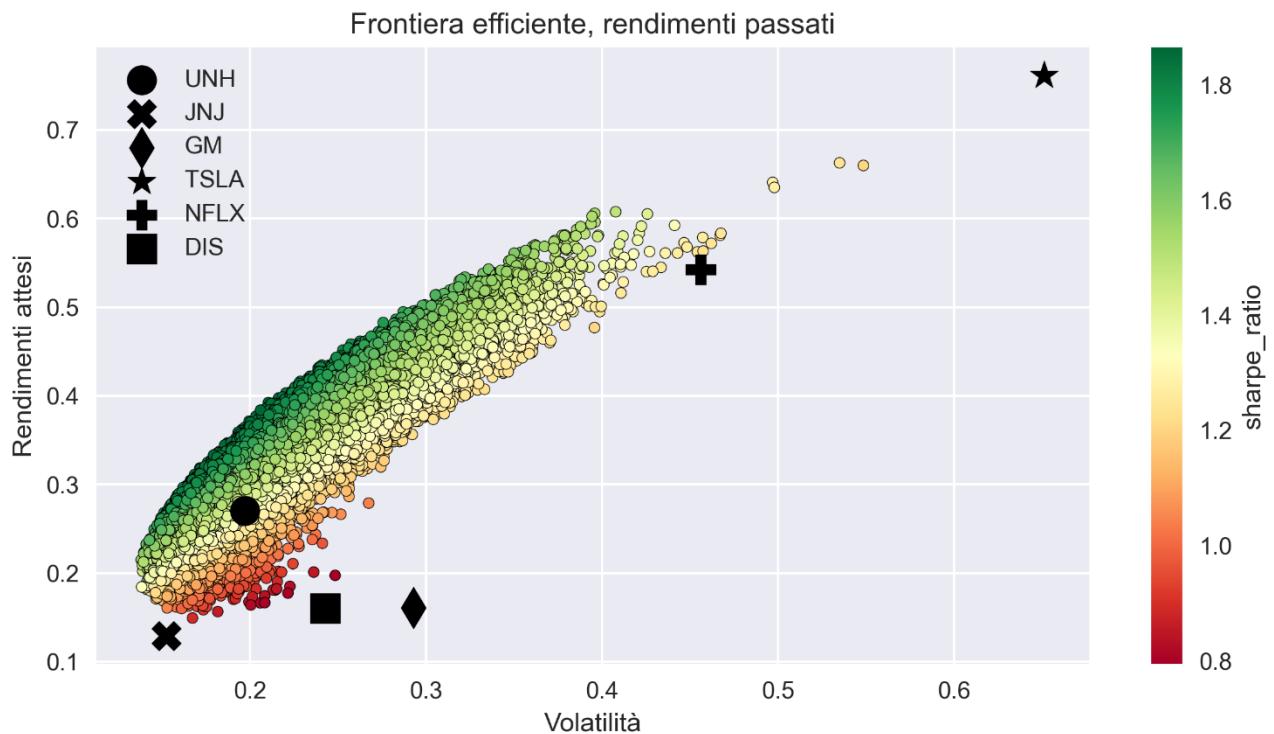
In questo caso, come rendimenti medi (annui) si utilizza il vettore con la media di ciascun asset del dataframe, a cui moltiplico 12.

Il metodo di simulazione che utilizzo è il metodo Monte Carlo, che si basa sul campionamento casuale per ottenere risultati.

Genero una matrice di pesi casuali, poi i rendimenti di portfolio, definiti come il prodotto tra i pesi e i rendimenti medi, e infine creo un dataframe in cui inserisco i **rendimenti**, la **volatilità** (per ogni asset calcolata come la radice quadrata del prodotto tra il vettore dei pesi

trasposto e il prodotto tra matrice di covarianze e il vettore dei pesi) e lo **Sharpe Ratio** (rapporto tra rendimenti e volatilità, in quanto non abbiamo un risk-free).

A questo punto, finalmente posso generare la **frontiera efficiente di Pareto** e poi plottarla:



Ho rappresentato con dei marker i portfolio composti solo da un asset, ad esempio la stella indica il portfolio in cui diamo peso 1 a Tesla e peso 0 a tutti gli altri stock.

Si può notare come il portfolio con solo **Tesla** sia quello con il rendimento atteso più alto, ma anche quello con la volatilità più alta. Il portfolio con volatilità più bassa è il portfolio di **Johnson & Johnson**.

A questo punto cerco i **portfolio ottimali**, che sono 3:

- **portfolio con massimo sharpe ratio**
- **portfolio con volatilità minima**
- **portfolio con rendimento atteso massimo**

Per ognuno di essi stampo il rendimento atteso, la volatilità, lo Sharpe Ratio, i pesi e il beta del portfolio rispetto al mercato.

Portfolio con max Sharpe Ratio:

```
==> Portfolio con sharpe ratio massimo ==>
Rendimento atteso: 0.3923710371750611
Volatilità: 0.21042177736832982
Sharpe Ratio: 1.8646883515684824
```

```
==> Pesi portfolio con max sr ==>
UNH: 0.5473033331338616
JNJ: 0.057464589618536316
GM: 0.02050637450822323
TSLA: 0.14789629978272545
NFLX: 0.22138635643069282
DIS: 0.005443046525960732
```

```
==> Beta portfolio con max sr rispetto al mercato ==>
Beta: 1.016477031702008
```

Portfolio con volatilità minima:

```
==> Portfolio con volatilità minima ==>
Rendimento atteso: 0.21478800277648977
Volatilità: 0.13816717614966842
Sharpe Ratio: 1.5545515857096368
```

```
==> Pesi portfolio con min volatility ==>
UNH: 0.24875712553015525
JNJ: 0.4963599421654836
GM: 0.09560040979242178
TSLA: 0.005935251819709933
NFLX: 0.10212642866696281
DIS: 0.05122084202526655
```

```
==> Beta portfolio con min volatility rispetto al mercato ==>
Beta: 0.8294752272971109
```

Portfolio con rendimento massimo:

```
==> Portfolio con rendimento massimo ==>
Rendimento atteso: 0.662898418895985
Volatilità: 0.53501267364025
Sharpe Ratio: 1.239033113712232
```

```
==> Pesi portfolio con max returns ==>
UNH: 0.0842464771958401
JNJ: 0.00541429757290487
GM: 0.01021605396530504
TSLA: 0.8018063410558482
NFLX: 0.030174336451198228
DIS: 0.06814249375890366
```

```
==> Beta portfolio con max returns rispetto al mercato ==>
Beta: 1.6048069735829111
```

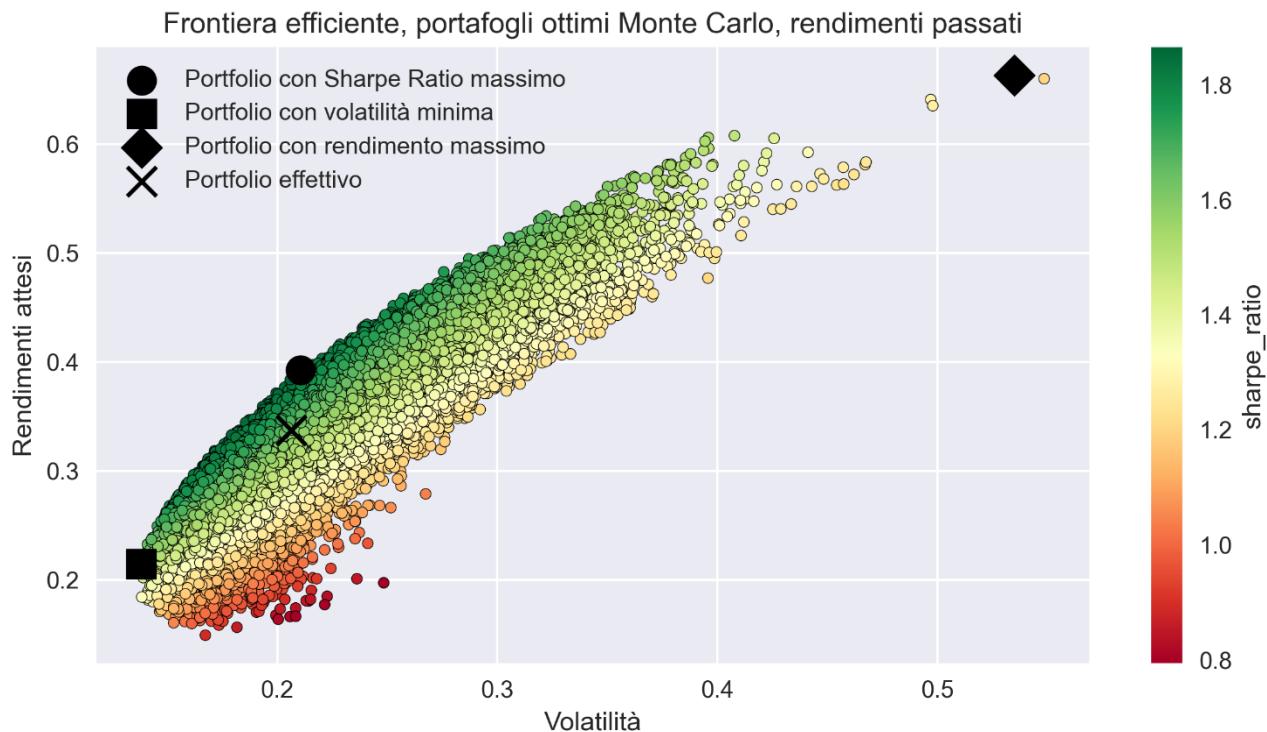
A questo punto manca da considerare un portfolio importante: il **portfolio effettivo**, in cui tutti i pesi sono uguali tra di loro, quindi avendo 6 asset ho che i pesi sono tutti 1/6.

Di seguito il portfolio effettivo:

```
== Portfolio effettivo rendimenti passati ==
Rendimento atteso: 0.337493427581276
Volatilità: 0.20633644594816208
Sharpe Ratio: 1.6356462186329626

Beta: 1.1727979764321672
```

Infine, mostro questi 4 (3 ottimali e quello effettivo) portafogli in un grafico:



Si noti come tutti e 3 i portfolio ottimi siano sulla frontiera di Pareto.

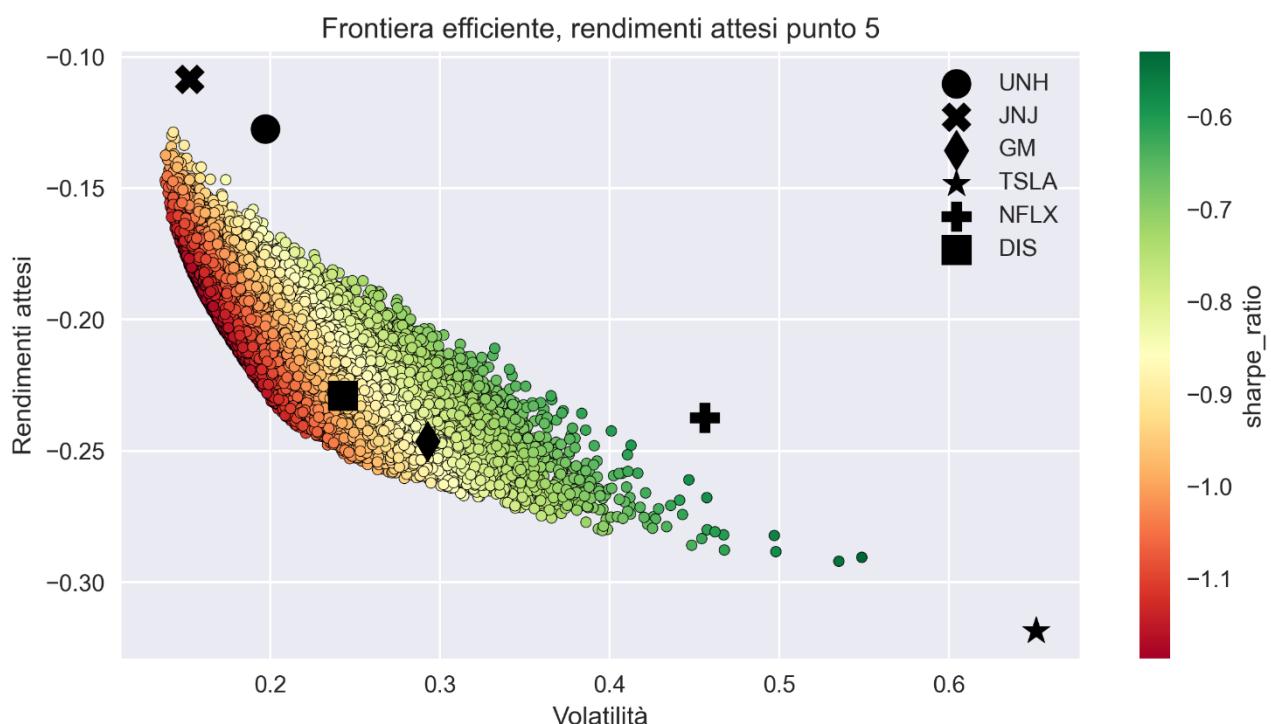
Rendimenti attesi costruiti: metodo di simulazione

In questo caso come **rendimenti medi annui** si utilizzano quelli **stimati** nell'ultimo punto del CAPM.

Siccome sono già annui, non devo moltiplicare per 12.

Ora, ripeto tutto il procedimento svolto nel punto precedente (metodo di Monte Carlo), in quanto l'unica differenza consiste nell'utilizzo di rendimenti diversi.

Ottengo la seguente **frontiera efficiente di Pareto**:



È evidente come il ramo di iperbole sia “specchiato”: la motivazione consiste nel fatto che ho utilizzato rendimenti attesi annui tutti negativi, a causa delle stime.

Ora che i rendimenti sono tutti negativi, si può notare come si sia ribaltata la situazione: il portfolio di Tesla è il portfolio con rendimento atteso minimo.

A questo punto calcolo i 3 portafogli ottimi e il portfolio effettivo.

Portfolio con max Sharpe Ratio:

```
==> Portfolio con sharpe ratio massimo ==>
Rendimento atteso: -0.2905039103131825
Volatilità: 0.5485356472898564
Sharpe Ratio: -0.5295989636197243

==> Pesi portfolio con max sr ==>
UNH: 0.0015573575497629072
JNJ: 0.11061749348275861
GM: 0.045939617394172044
TSLA: 0.8274675383306458
NFLX: 0.014354808537916167
DIS: 6.318470474443617e-05

==> Beta portfolio con max sr rispetto al mercato ==>
Beta: 1.5968123546159494
```

Portfolio con volatilità minima:

```
==> Portfolio con volatilità minima ==>
Rendimento atteso: -0.14701191605563368
Volatilità: 0.13816726554234995
Sharpe Ratio: -1.0640140808936596

==> Pesi portfolio con min volatility ==>
UNH: 0.24875712553015525
JNJ: 0.4963599421654836
GM: 0.09560040979242178
TSLA: 0.005935251819709933
NFLX: 0.10212642866696281
DIS: 0.05122084202526655

==> Beta portfolio con min volatility rispetto al mercato ==>
Beta: 0.8294754869285224
```

Portfolio con rendimento massimo:

```
==> Portfolio con rendimento massimo ==>
Rendimento atteso: -0.12868778839348657
Volatilità: 0.14296499817517092
Sharpe Ratio: -0.9001349283816247

==> Pesi portfolio con max returns ==>
UNH: 0.31636975345842916
JNJ: 0.5935300767053079
GM: 0.03421518519663267
TSLA: 0.029982391473708497
NFLX: 0.0010733179644009097
DIS: 0.024829275201520864

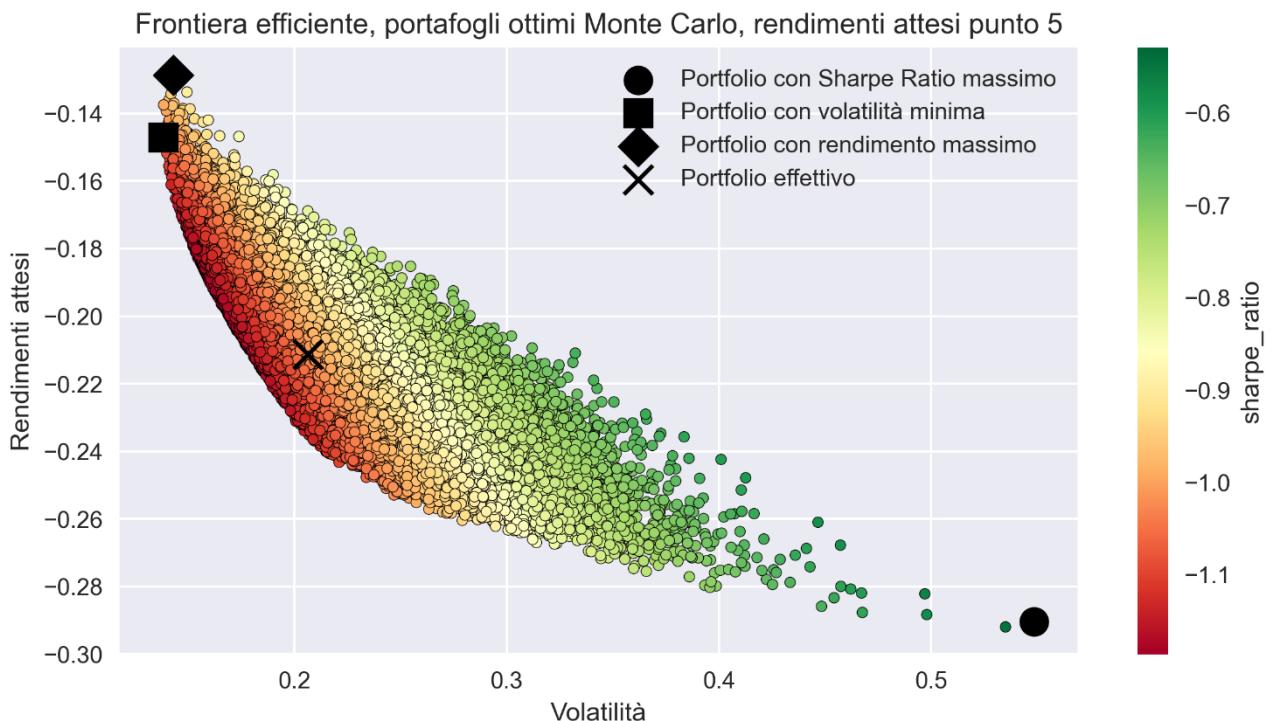
==> Beta portfolio con max returns rispetto al mercato ==>
Beta: 0.7314854994304096
```

Portfolio effettivo:

```
==== Portfolio effettivo ====
Rendimento atteso: -0.21121323037881395
Volatilità: 0.20633647807532557
Sharpe Ratio: -1.0236349498109978

Beta: 1.1727980234161173
```

Come prima, plotto i 4 portafogli:



Rendimenti passati: metodo analitico

Il metodo analitico consiste nell'utilizzare la libreria SciPy per risolvere un problema di ottimizzazione, in cui la funzione obiettivo è minimizzare la volatilità di un portfolio, imponendo vincoli sul rendimento del portfolio (che dev'essere uguale alla somma del prodotto tra pesi e rendimenti medi) e sulla somma dei pesi (che dev'essere 1).

Questo mi permette, per ogni rendimento nel range generato, di trovare un portfolio la cui volatilità è minima e che rispetta i vincoli.

Come range di rendimenti utilizzo [minimo rendimento Monte Carlo, massimo rendimento Monte Carlo], ovviamente associando correttamente i casi rendimenti passati e rendimenti attesi costruiti.

La differenza è che con il metodo analitico non abbiamo, ad esempio, 10^5 (valore della variabile definita all'inizio, prima del metodo Monte Carlo) portafogli: non si utilizza alcuna matrice di pesi. La matrice dei pesi serviva solo per il metodo Monte Carlo, qui calcolo 200 portafogli nel range dei rendimenti.

Perciò, non mi sembra significativo plottare lo stesso grafico dei portfolio generati con metodo Monte Carlo, con l'aggiunta di trattini colorati per evidenziare il range di cui si hanno i portafogli trovati col metodo analitico.

Detto questo, salvo in un dataframe tutti i 200 portfolio calcolati dall'ottimizzazione di SciPy, memorizzando per ognuno di essi rendimento, volatilità e Sharpe Ratio. Questa è la coda del dataframe:

	returns	volatility	sharpe_ratio
195	0.652573	0.436026	1.496636
196	0.655154	0.439055	1.492191
197	0.657735	0.442223	1.487340
198	0.660317	0.445526	1.482106
199	0.662898	0.448962	1.476512

A questo punto calcolo i 3 portafogli ottimi.

Portfolio con max Sharpe Ratio:

```
== (SciPy) Portfolio con sharpe ratio massimo ==
Rendimento atteso: 0.3608670250527365
Volatilità: 0.19196903835115886
Sharpe Ratio: 1.8798188924227532

== (SciPy) Pesi portfolio con max sr ==
UNH: 0.5445497968334795
JNJ: 0.13696544767355393
GM: 0.0
TSLA: 0.10557448048075452
NFLX: 0.21291027501221205
DIS: 5.209252625613947e-18

== (SciPy) Beta portfolio con max sr rispetto al mercato ==
Beta: 0.9442259792716284
```

Portfolio con volatilità minima:

```
== (SciPy) Portfolio con volatilità minima ==
Rendimento atteso: 0.19049034075191254
Volatilità: 0.1354771615728679
Sharpe Ratio: 1.406069765120191

== (SciPy) Pesi portfolio con min volatility ==
UNH: 0.21947433960488427
JNJ: 0.6130330335464607
GM: 0.09400748849376026
TSLA: 0.0
NFLX: 0.0659837202478628
DIS: 0.007501418107032014

== (SciPy) Beta portfolio con min volatility rispetto al mercato ==
Beta: 0.7656123409372876
```

Portfolio con rendimento massimo:

```
== (SciPy) Portfolio con rendimento massimo ==
Rendimento atteso: 0.6628984199496517
Volatilità: 0.4489624572243001
Sharpe Ratio: 1.4765119205022301

== Pesi portfolio con max returns ==
UNH: 0.0
JNJ: 0.0
GM: 0.0
TSLA: 0.5494979328211422
NFLX: 0.4505020671788578
DIS: 0.0

== (SciPy) Beta portfolio con max returns rispetto al mercato ==
Beta: 1.551138367913563
```

Perché non c'è il portafoglio effettivo? Perché è lo stesso del metodo Monte Carlo con rendimenti passati, avendo pesi uguali e rendimenti medi uguali.

Rendimenti attesi costruiti: metodo analitico

Si tratta banalmente di ripetere il procedimento ma utilizzando i rendimenti attesi costruiti nell'ultimo punto del CAPM, senza moltiplicare per 12 essendo già annui.

Si deve inoltre far attenzione a utilizzare il range corretto di rendimenti, ossia lo stesso per Monte Carlo con rendimenti attesi costruiti e non rendimenti passati, in quanto si è mostrato come i rendimenti attesi costruiti erano tutti negativi e quindi cambiava la frontiera di Pareto.

Calcolando i 3 portafogli ottimi ottengo:

Portfolio con max Sharpe Ratio:

```
==== (SciPy) Portfolio con sharpe ratio massimo ===
Rendimento atteso: -0.2919989016947674
Volatilità: 0.4556364490214417
Sharpe Ratio: -0.6408594007829832

==== (SciPy) Pesi portfolio con max sr ===
UNH: 2.2233664101767493e-16
JNJ: 0.0
GM: 0.36732742679219216
TSLA: 0.6326725732078076
NFLX: 8.218252467440124e-17
DIS: 5.4357511988096734e-17

==== (SciPy) Beta portfolio con max sr rispetto al mercato ===
Beta: 1.6048069609044977
```

Portfolio con volatilità minima:

```
==== (SciPy) Portfolio con volatilità minima ===
Rendimento atteso: -0.13525305927997022
Volatilità: 0.13547655869574418
Sharpe Ratio: -0.9983502724166776

==== (SciPy) Pesi portfolio con min volatility ===
UNH: 0.21897893137415253
JNJ: 0.6120056774786282
GM: 0.09416849424265297
TSLA: 4.7704895589362195e-18
NFLX: 0.06678833055314468
DIS: 0.008058566351421543

==== (SciPy) Beta portfolio con min volatility rispetto al mercato ===
Beta: 0.7665938999075932
```

Portfolio con rendimento massimo:

```
==== (SciPy) Portfolio con rendimento massimo ===
Rendimento atteso: -0.12868778839348657
Volatilità: 0.1359588463380743
Sharpe Ratio: -0.9465201556174758

==== Pesi portfolio con max returns ===
UNH: 0.22783180460714916
JNJ: 0.6536917535237717
GM: 0.061787247946463814
TSLA: 7.37257477290143e-18
NFLX: 0.056689193922615275
DIS: 3.686287386450715e-18

==== (SciPy) Beta portfolio con max returns rispetto al mercato ===
Beta: 0.7314854994286216
```

Perché non c'è il portafoglio effettivo? Stessa motivazione di prima, cioè è lo stesso del metodo Monte Carlo con rendimenti attesi costruiti.

Conclusioni

Nel corso di questo progetto sono emerse diverse modalità per effettuare analisi finanziarie sia per quanto riguarda i singoli asset, sia per quanto riguarda la costruzione ed ottimizzazione di portafoglio.

La parte di statistiche descrittive ha evidenziato come le distribuzioni dei rendimenti logaritmici degli asset scelti siano più “appuntite” rispetto a una normale e quindi si parla di leptocurtica, oltre ad evidenziare che ci possono tranquillamente essere diversi outliers a prescindere dal tipo di distribuzione. Si è evidenziato anche come non è detto che le correlazioni più alte, nella matrice di correlazione, si trovino nelle coppie tra titoli appartenenti allo stesso settore (si osservi Disney e General Motors).

La parte di analisi di previsione pone delle basi per l'utilizzo di algoritmi di Machine Learning in ambito finanziario, tramite il forecasting su serie di prezzi. Nonostante ciò, è emerso che non è facile prevedere l'andamento dei mercati: infatti, la maggior parte delle previsioni era fuori strada.

È stata particolarmente interessante la parte di strategie di trading, poiché ci permette di definire una nostra strategia e di effettuare il backtesting generando dei segnali nell'intero periodo e calcolando i rendimenti. Anche qui è emerso il fatto che non è facile prevedere l'andamento dei prezzi a proprio vantaggio.

La parte di CAPM ci permette di calcolare l'andamento di ogni asset rispetto al mercato, che è di particolare rilevanza nel momento in cui il titolo segue il movimento del mercato con volatilità più alta ($\beta > 1$).

Sinceramente non mi ha colpito l'esposizione ai fattori Fama-French.

Invece, la costruzione di portafoglio è stata interessante perché permette di trovare i portafogli ottimali con diversi metodi e con l'utilizzo di diversi rendimenti, riuscendo a fornire in output la combinazione ottimale dei pesi che ci garantisce massimo Sharpe Ratio o volatilità minima o massimo rendimento atteso.

In tutto, questo progetto pone delle basi per analisi in ambito finanziario, ma ci si può soffermare su alcuni punti (ad esempio per la parte di Machine Learning servirebbe un intero corso) ed approfondirli rispetto ad altri. Si tratta di uno schema molto generale che può dare spunto per un'analisi nel dettaglio rispetto a particolari aspetti.