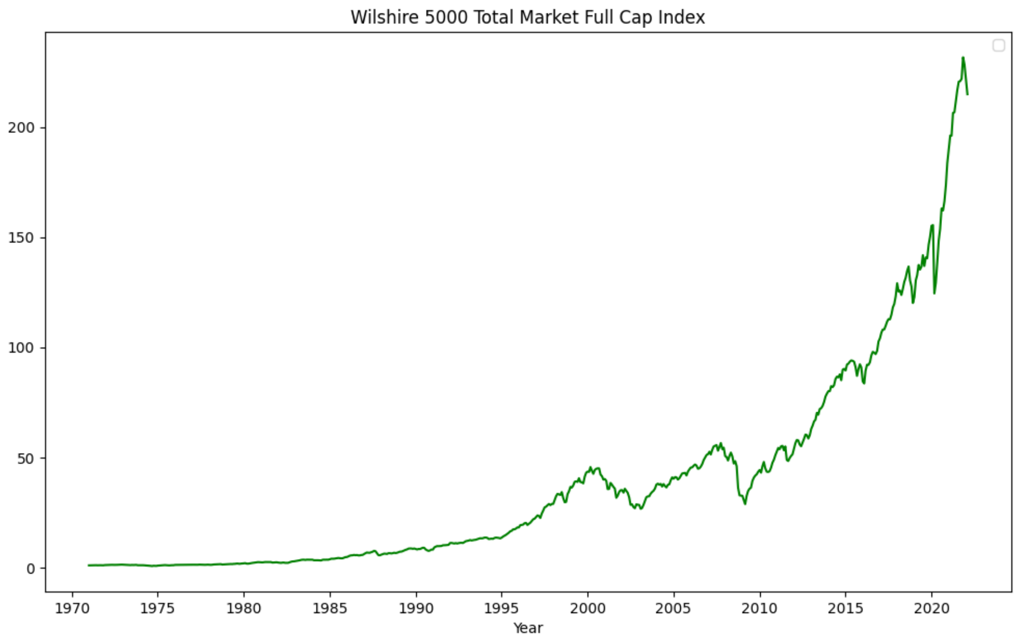
Andrea Ferrari – Time Series – MDS 2021-2022

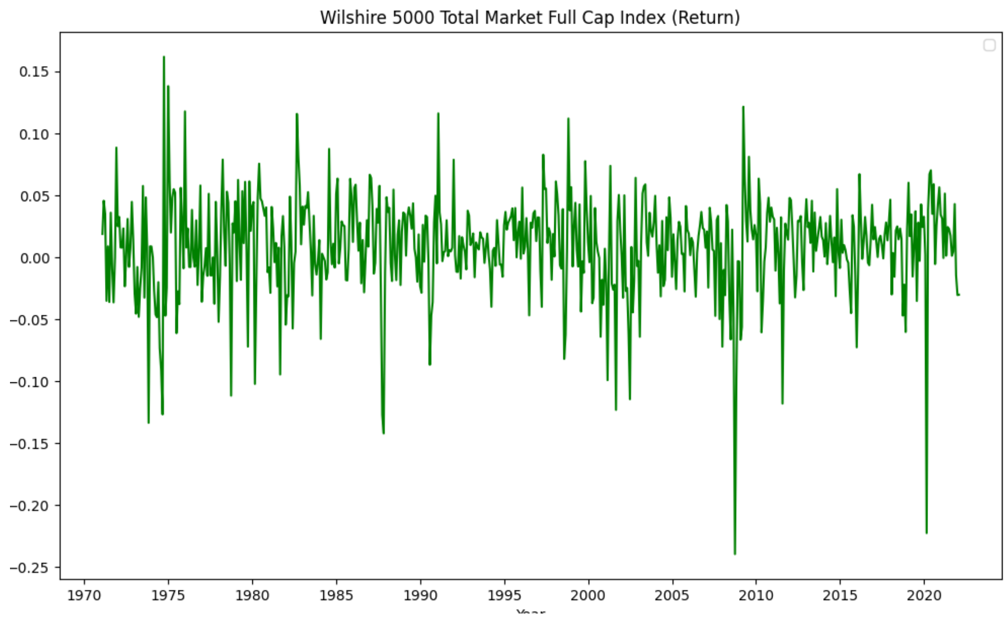
La serie storica

La serie storica utilizzata è il 'Wilshire 5000 Total Market Full Cap Index', o più semplicemente il 'Wilshire 5000', che è un indice di mercato, costruito pesando le azioni delle aziende quotate per la loro capitalizzazione di mercato, di tutto il mercato azionario statunitense. A gennaio 2020, l'indice era composto da 3473 azioni.

I dati sono stati raccolti dalla FRED con cadenza mensile, prendendo come valore l'ultimo dato disponibile nel mese. La serie parte nel 1971 e finisce nel 2022. Il link alla pagina FRED: https://fred.stlouisfed.org/series/WILL5000INDFC

Una rappresentazione grafica dei dati, e poi dei ritorni:





È evidente dal primo grafico che la serie storica abbia un trend rialzista, perciò andrà sicuramente applicato il modello ARMA sui ritorni, quindi un modello ARIMA.

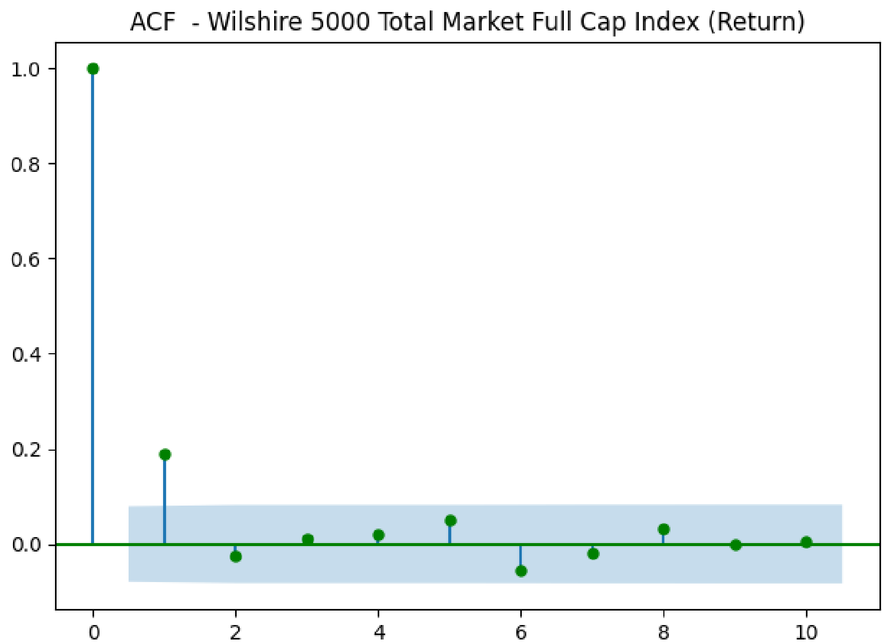
Visivamente i ritorni mensili sembrano essere stazionari, con media e varianza abbastanza costanti nel tempo. Vista la tendenza rialzista la media sarà sicuramente maggiore di zero; per quanto riguarda la varianza sono presenti due ‘outliers’ nella seconda parte della storia: uno nel 2008 (crisi Lehman-Brothers) e l’altro nel 2020 (pandemia), che però non sono eccessivi e non dovrebbero impedire alla varianza di rimanere costante nel tempo.

Ho quindi calcolato la differenza prima della serie storica: ho applicato il logaritmo ai dati e calcolato la variazione al tempo ‘t’ rispetto a ‘t-1’.

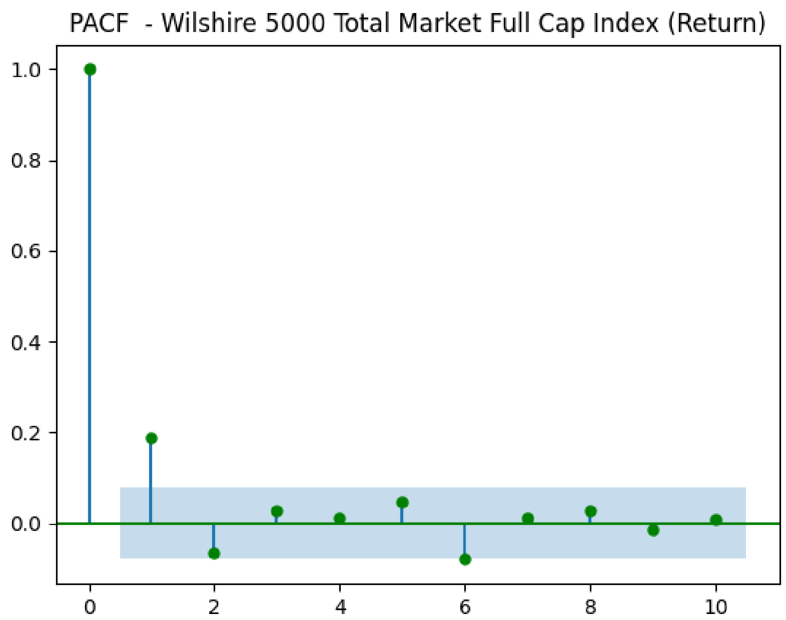
Autocorrelazione della serie storica

La funzione di autocorrelazione cattura le somiglianze interne alla serie storica, vede cioè se la serie traslata opportunamente somiglia a sé stessa.

Il primo grafico rappresenta l’autocorrelazione della serie storica, trasformata nella sua differenza prima, con sé stessa fino a 10 periodi precedenti (lags). Escludendo ovviamente il lag al tempo 0, sempre uguale ad 1, soltanto il lag ‘1’, che corrisponde al tempo il ‘t-1, è oltre al valore soglia e quindi significativo, mentre tutti gli altri sono sotto alla soglia di significatività.



Questo significa che soltanto il ritorno al tempo ‘t-1’ può essere utile per predire il ritorno al tempo ‘t’, e dunque solo il ritorno al tempo ‘t’ può essere utile per predire il ritorno al tempo ‘t+1’.



L’autocorrelazione parziale, che misura la correlazione di ciascun lag con il ritorno al tempo ‘t’ escludendo la correlazione dei tempi precedenti, conferma il fatto che l’unico lag che può esserci utile è quello più recente.

Test di stazionarietà

Nell'analisi delle serie storiche in campo econometrico è possibile che si manifestino delle tendenze che potrebbero rendere le regressioni spurie. Questi trend possono essere stocastici, nel caso ci sia non stazionarietà in varianza, o deterministico, nel caso la non stazionarietà sia in media.

Test di Dickey-Fuller (DF)

Il test di Dickey-Fuller, che appartiene a una categoria di test chiamata "Unit Root Test", permette di valutare se esiste una tendenza nelle variabili che renda la regressione spuria, e quindi per testare la stazionarietà di una serie temporale.

La radice unitaria (“Unit Root”) è una caratteristica di una serie temporale che la rende non stazionaria. Si dice che una radice unitaria esiste in una serie temporale del valore di alfa = 1 nell'equazione seguente.

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

Dove:

- Y_t è il valore della serie temporale al tempo 't'
- X_e è una variabile esogena (una variabile esplicativa separata, che è anche una serie storica).

Un test di Dickey-Fuller è un test radice unitaria che verifica l'ipotesi nulla che α=1 nella seguente equazione del modello. Alfa è il coefficiente del primo lag su Y.

Se c’è la presenza di una radice unitaria, allora la serie storica non è stazionaria. Inoltre, il numero di radici unitarie contenute nella serie corrisponde al numero di operazioni di differenziazione necessarie per rendere stazionaria la serie.

Il test è ‘one-sided’, a una sola coda, e la distribuzione dipende dal fatto che il test includa a una costante, un'intercetta e una tendenza, o nessuna di esse.

Test di Dickey-Fuller aumentato (ADF)

Come suggerisce il nome, il test “Agumented Dickey-Fuller” (ADF) è una versione "aumentata" del test Dickey Fuller.

Il test ADF espande l'equazione del test Dickey-Fuller aggiungendo lags di Δyt, per catturare correlazioni di ordine superiore. Questo aggiunge più completezza al test.

L'ipotesi nulla, tuttavia, è sempre la stessa del test di Dickey Fuller.

L'ipotesi nulla presuppone la presenza di una radice unitaria, cioè $\alpha=1$; se il p-value ottenuto è inferiore al livello di significatività (solitamente 0.05) si rifiuta l'ipotesi nulla, deducendo quindi che la serie è stazionaria. Viceversa, se il p-value è maggiore di 0.05, la serie non è stazionaria.

L’ADF test sulla serie storica ha prodotto i seguenti risultati:

- T-test: -16.70
- P-value: 0.0

Dunque, con un p-value uguale a zero rifiutiamo H0, deducendo che i ritorni della serie storica sono stazionari.

Il miglior modello

Ho calcolato vari modelli ARMA sui ritorni, quindi di fatto ARIMA, variando ‘p’ e ‘q’ tra 0 e 4:

- ‘p’: è l’ordine del processo autoregressivo
- ‘q’: è l’ordine del processo a media mobile
- ‘d’: (in ARIMA) è uguale ad 1, in quanto calcoliamo la differenza prima dei dati, chiamati in gergo i ‘ritorni azionari’

Di seguito i migliori modelli per entrambi i criteri:

p	q	AIC
0	1	-2223.67
1	1	-2222.33
0	2	-2222.25
2	0	-2221.77
1	0	-2221.32
2	1	-2220.52
0	3	-2220.41
3	0	-2220.25
1	2	-2219.78
2	2	-2218.66
1	3	-2218.6

p	q	BIC
0	1	-2210.41
1	0	-2208.07
1	1	-2204.66
0	2	-2204.57
2	0	-2204.09
2	1	-2198.43
0	3	-2198.31
3	0	-2198.16
1	2	-2197.69
2	2	-2192.15
1	3	-2192.09

Senza ombra di dubbio, in quanto lo è per entrambi i criteri, il miglior modello è lo 0,1. Quindi di fatto un modello MA(1). Oltre ai criteri informativi, abbiamo pochi dubbi sulla scelta in quanto l’ordine 1 è in linea con ciò che abbiamo visto nei grafici ACF, PACF, ed è il anche tra i modelli più semplici: con meno parametri da stimare.

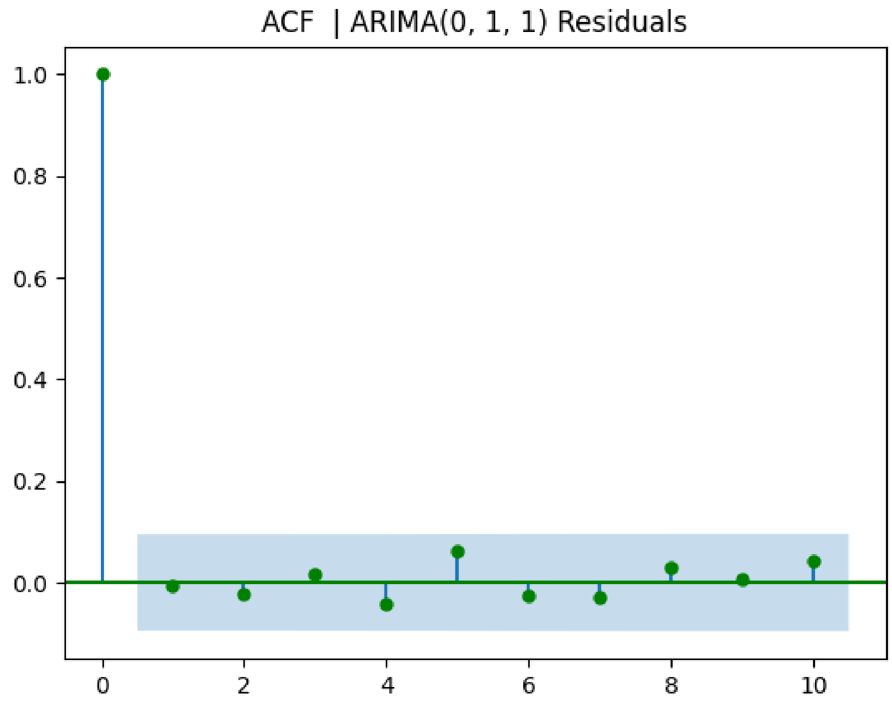
Autocorrelazione dei residui

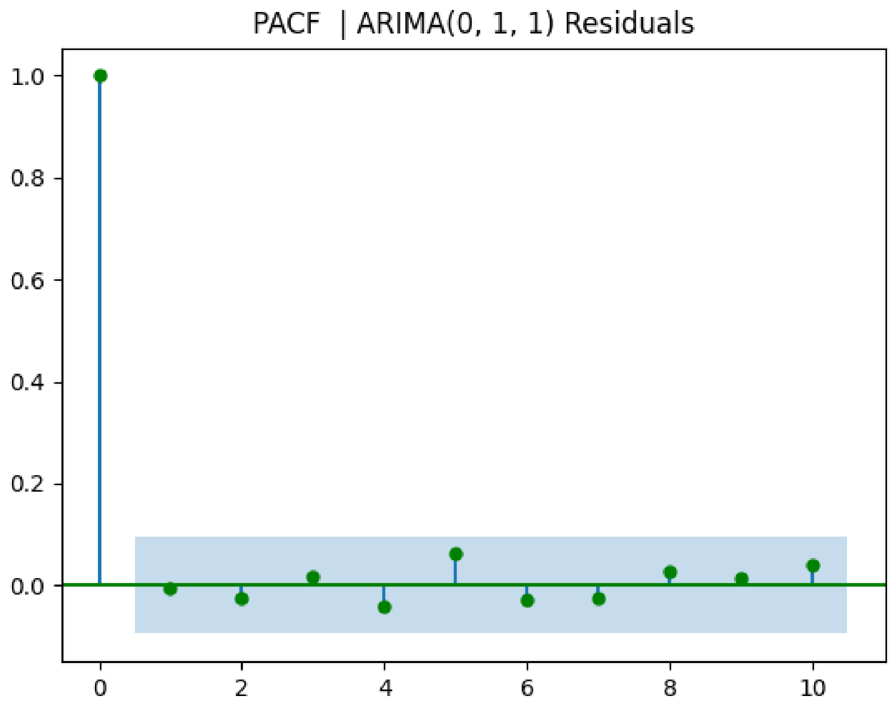
Se abbiamo scelto un buon modello, analizzando l'autocorrelazione dei residui dovremmo osservare che per nessun lag il livello di significatività superi la soglia.

i residui al tempo t non devono essere correlati coi residui al tempo $t - k$ (con $k > 0$), cioè il passato non deve in alcun modo essere utilizzato per prevedere il presente. In termini euristici, i residui devono configurarsi in una serie storica composta da numeri non legati da alcuna relazione (o privi di memoria).

Qualora i residui mostrino un qualche grado di autocorrelazione, questo significa che il modello stimato è "mispecificato", cioè non è stato in grado di catturare tutte le relazioni dinamiche contenute nella variabile dipendente. Ciò può accadere soprattutto quando ci sono variabili esplicative omesse oppure quando le stesse variabili esplicative disponibili hanno una limitata capacità di spiegazione del fenomeno oggetto dell'analisi. In questo caso, qualsiasi dinamica che non viene catturata dal modello si scarica automaticamente all'interno dei residui "sporcandoli";

In questo caso nessun residuo di nessun lag è autocorrelato con i residui al tempo 't', e questo è vero sia osservando il grafico ACF, sia il grafico PACF.





Possiamo concludere questa parte dicendo che il modello ARMA(0,1) è stato in grado di catturare tutte le relazioni dinamiche contenute nella variabile dipendente.

Previsioni dei Modelli

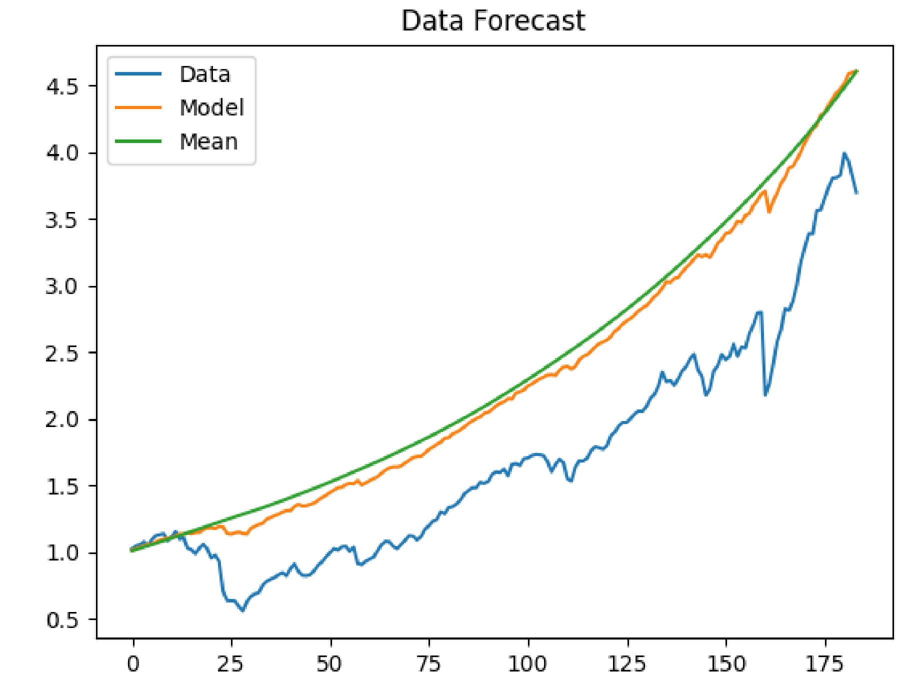
La serie storica, composta da 613 osservazioni, è stata divisa in 2 parti:

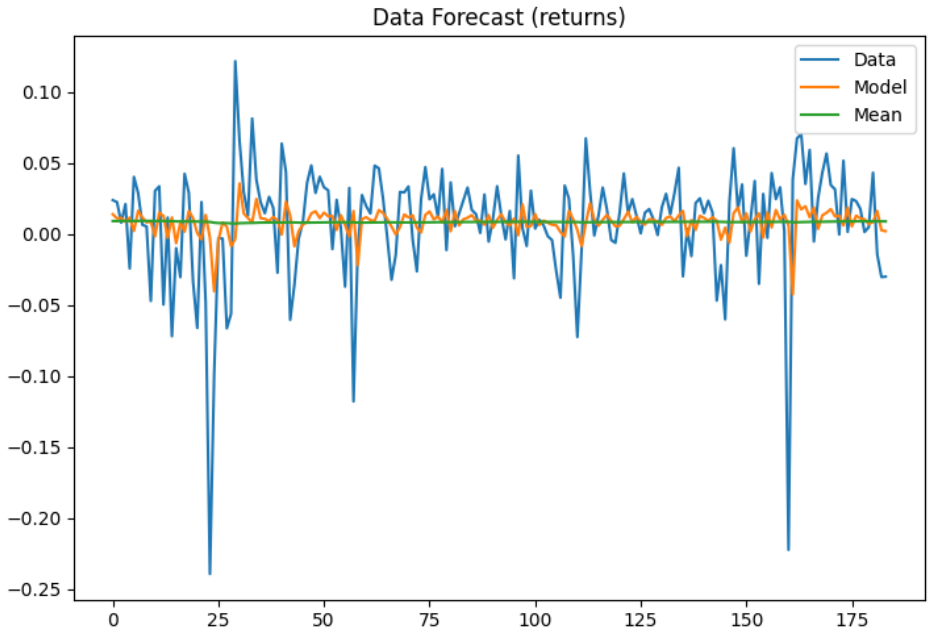
- In-Sample: dal 1971 al 2006 (429 osservazioni, il 70%)
- Out-of-Sample: dal 2007 al 2022 (184 osservazioni, il 30%)

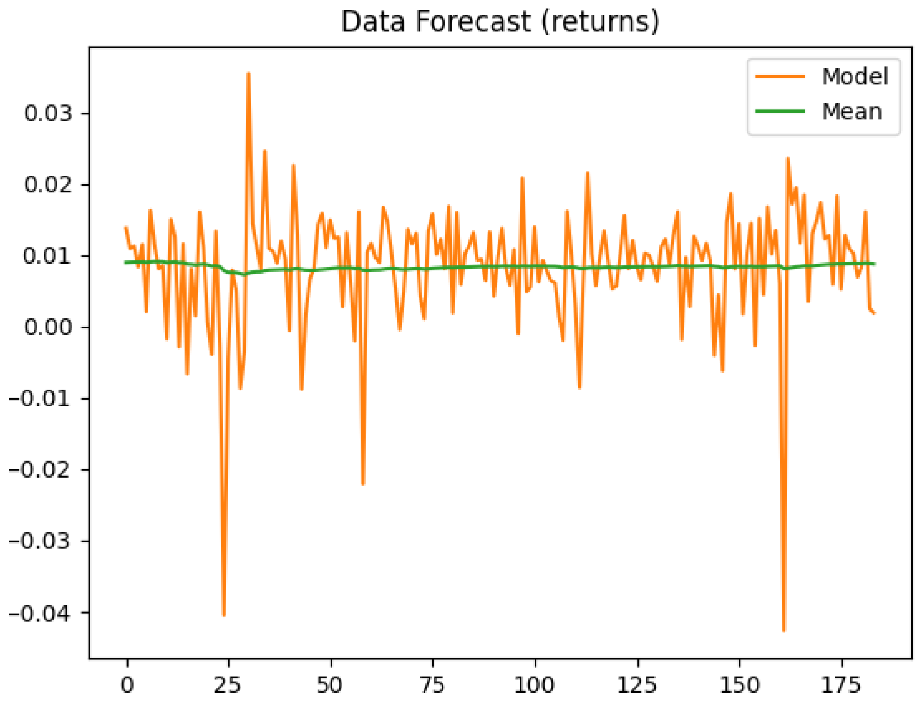
Sono stati implementati 2 tipi di modelli per prevedere il ritorno al tempo ‘t+1’:

- Modello ARMA (sui ritorni, quindi ARIMA)
    - Di ordine 0,1
    - Nei grafici sotto la sua etichetta è ‘Model’
- Modello che utilizza la media
    - La previsione di questo modello per ‘t+1’ sarà la media di tutti i valori della serie storica dal primo dato disponibile fino a ‘t’.
    - Nei grafici sotto la sua etichetta è ‘Mean’

Le previsioni sono state fatte con il metodo ‘rolling’, e sono state fatte 184 previsioni, dal 2007 sino al 2022.

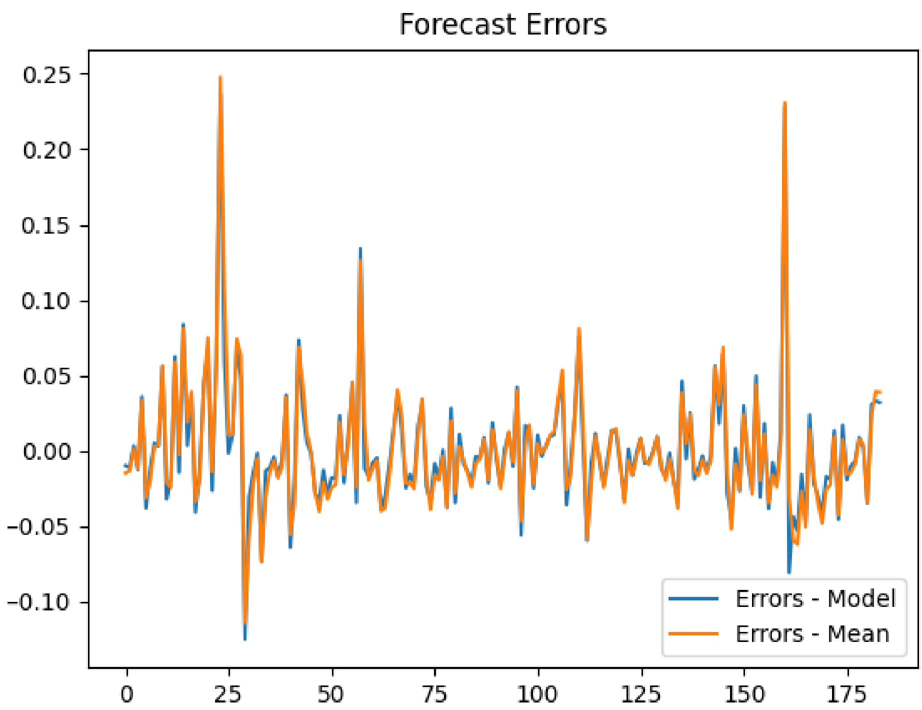

Data Forecast
Data
Model
Mean
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
25
50
75
100
125
150
175


Data Forecast (returns)
Data
Model
Mean
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
0
25
50
75
100
125
150
175



Da questi due grafici si osserva che entrambi i forecast hanno una variabilità nettamente inferiore a quella poi realizzatasi dalla serie storica. Questo è ancora più vero, ovviamente per il forecast creato dal modello 'Mean'.

Nel grafico di seguito, gli errori di previsione:



Gli errori di previsione si somigliano molto in quanto entrambi i modelli hanno il difetto di sottostimare la variazione futura dell'indice azionario.

Mean Squared Forecast Error

La radice del ‘Mean Squared Forecast Error’ dei due modelli è:

- ARMA(0,1): 0.0405
- Mean: 0.0415

L’errore di previsione è inferiore per il modello ARMA, anche se di poco.

Diebold-Mariano Test

Vogliamo vedere quale modello ci offre le previsioni migliori: una migliore accuratezza predittiva. L'approccio più ovvio è selezionare il modello con l'errore di previsione inferiore, ma facciamo un ulteriore passo avanti e determiniamo se questa differenza è significativa, o è semplicemente dovuta alla scelta specifica dei valori dei dati nel campione.

Il Diebold-Mariano Test ci dice se la media della differenza del quadrato degli errori di previsione dei due modelli è statisticamente diversa da zero. In altre parole, se, sulla base degli errori di previsione, il modello A è statisticamente migliore del modello B.

L’ipotesi nulla del test è che il modello A non sia significativamente migliore del modello B (E[d]=0).

Il risultato del “Diebold-Mariano Test”:

- T-stat: -1.156
- P-value: 0.2492

Rifiutiamo l’ipotesi nulla in quanto il p-value è maggiore di 0.05, quindi il modello A (ARMA) è significativamente migliore nelle previsioni del modello B (Media).