

ALMA MATER STUDIORUM · UNIVERSITÀ DI  
BOLOGNA

---

SCUOLA DI ECONOMIA

Corso di Laurea in Finanza, Mercati e Istituzioni

Ottimizzazione della gestione del  
portafoglio tramite tecniche di  
clustering gerarchico

Relatore:  
Stefano Ferretti

Presentata da:  
Gian Maria Lunghi

II Sessione: 17/12/2020  
Anno Accademico 2019/2020

*A mio fratello che ha investito a fondo perduto sulla mia istruzione e alla  
potenza della sua macchina su cui posso sempre contare.*

*A mago data la continua sfida per dominare la black side, a fonzies su cui  
posso contare per una visione delle cose inaspettata e a cavo\_25 per  
ricordare i pilastri da cui sono partito.*

*A fabio con cui condivido l'apprezzamento degli FPS, mi auguro di poter  
lavorare assieme un giorno.*

*Al buon lombardo per la quale mai ripeterei le sessioni nella saletta a  
Borgonuovo.*

*A Juan e Gigi quale auguro di non diventare mai tutto ciò che hanno  
sempre odiato.*



# Indice

|  |           |
|--|-----------|
| <b>Introduzione</b>  | <b>5</b>  |
| <b>1 Il portafoglio</b>  | <b>9</b>  |
| 1.1 Rischio e rendimento . . . . .                                   | 11        |
| 1.2 Costruzione del portafoglio . . . . .                            | 17        |
| <b>2 HRP e algoritmi per la creazione di portafogli</b>              | <b>23</b> |
| 2.1 Costruzione delle matrici di covarianza e correlazione . . . . . | 25        |
| 2.2 Metodi di clustering . . . . .                                   | 30        |
| 2.3 Quasi diagonalizzazione . . . . .                                | 34        |
| 2.4 Bipartizione ricorsiva . . . . .                                 | 36        |
| <b>3 Hrp in approccio stock market</b>                               | <b>41</b> |
| 3.1 Algoritmo iterativo per massimizzazione SR . . . . .             | 45        |
| <b>4 Dati e metodologia</b>  | <b>51</b> |
| 4.1 Out of Sample Analysis . . . . .                                 | 54        |
| 4.2 In Sample Analysis . . . . .                                     | 57        |
| <b>5 Spunti per lavori futuri</b>                                    | <b>61</b> |
| <b>Conclusioni</b>   | <b>63</b> |



# Introduzione

La costruzione di un portafoglio efficiente è da sempre uno degli obiettivi primari nell'ambito finanziario, che ci si interessi in ottica personale o da un punto di vista professionale è ricorrente l'investimento di risorse su base periodale per ottenere in un futuro un guadagno pecuniario.

Il problema viene affrontato nel tempo da diversi economisti quale Markowitz 1952, costruendo la frontiera dei portafogli efficienti secondo una relazione data dalla media dei rendimenti e dalla varianza e poi Black 1972, portando il modello di equilibrio per i mercati di capitali ovvero il CAPM.

L'approccio verso la costruzione del portafoglio subisce una notevole spinta dall'avvento dei computer e successivamente da internet, nel primo caso si ha avuto una migliore maneggevolezza di mole di dati importanti, si pensi al solo foglio di calcolo quanto riduca i tempi rispetto a metodi non digitali, nel secondo caso invece l'accesso a dati provenienti da diversi punti del mondo, a qualsiasi ora e sotto molteplici formati. Grazie allo sviluppo di entrambe le tecnologie si è arrivati a metodi più complessi di computazione permettendo così l'incrocio di *big data* e lo sviluppo di algoritmi di *machine learning* per lo sviluppo dell'obiettivo primario, il portafoglio.

Il *Hierarchical Risk Parity* è un approccio portato alla luce da De Prado 2016 [1], attraverso il quale la costruzione di portafoglio avviene tramite tre fasi principali, il *clustering*, la quasi-diagonalizzazione e in fine la bipartizione ricorsiva ottenendo come risultato la composizione di portafoglio ottimale per una riduzione del rischio.

I primi problemi si riscontrano nella costruzione della matrice di covarianza

e correlazione tra asset, infatti per stimare la covarianza dei rendimenti sarebbero necessarie  $N(N+1)/2$  indipendenti e identicamente distribuite (*iid*) osservazioni di rendimenti. La finestra ottimale sarebbe composta da almeno 1 anno di osservazioni giornaliere, ricordando che nel mercato si riscontra la caratteristica di non stazionarietà.

La stazionarietà è un presupposto nell'utilizzo di molte procedure statistiche per l'analisi di serie storiche, una delle cause principali che non permettono l'utilizzo dei dati *naive* è la presenza di trend nella serie, rimuovendo quindi l'assunzione di dati stazionari.

In secondo luogo possono verificarsi eventi di *cross-correlation* tra *time series* dal momento che si analizzano potenzialmente un numero elevato di osservazioni e un numero importante di asset, quindi la probabilità che questo fenomeno si presenti aumenta.

Per ovviare a questi due problemi si è deciso di utilizzare un approccio innovativo per la stima della matrice di correlazione, ovvero la *Detrended Cross Correlation Analysis (DCCA)*. Tramite questo strumento di analisi la serie temporale viene divisa in sottogruppi  $(N-n)$  di uguale lunghezza  $(n)$  costruendo quindi una *sliding window* tra la serie storica.

Una volta ripuliti i dati dal trend in background si passa alla prima fase dell'algoritmo di HRP.

Vengono calcolate le distanze in base alla matrice di correlazione costruendo una matrice con le distanze si viene a creare un presupposto attraverso il quale l'algoritmo di linkage e quindi il clustering sarà in grado di convogliare gli investimenti simili nello stesso gruppo gerarchico, il risultato finale è un cluster che contiene tutti gli altri cluster a cascata. L'algoritmo di clustering può essere computato seguendo diverse procedure che verranno analizzate successivamente.

Il passo successivo riguarda la quasi-diagonalizzazione della matrice di covarianza, si ha una modifica delle colonne e delle righe basandosi sui risultati ottenuti dalla fase precedente restituendo come output una matrice nella quale gli asset simili sono raggruppati e gli asset diversi sono lontani.

A questo punto gli asset vengono sezionati ricorsivamente in modo tale da creare sottogruppi per l'assegnazione dei pesi di portafoglio. Ovviamente questa procedura si basa sulle matrici ottenute in precedenza, quindi, i pesi verranno ponderati per il rischio di ogni gruppo di asset. La procedura viene ripetuta finché non vi saranno più gruppi di asset, e quindi tutti i pesi saranno distribuiti ottenendo un portafoglio con  $w = 1$ .

L'elaborato è così strutturato, nel capitolo 1 vi sono le nozioni di base per la comprensione delle grandezze statistiche e degli indici di performance mentre nel capitolo 2 vi sarà la spiegazione dell'algoritmo di cluster gerarchici attraverso tutte le sue fasi. Il capitolo 3 prende in considerazione un approccio HRP orientato al mercato azionario e in fine, nel capitolo 4 si presenta l'approccio numerico e metodologico degli algoritmi gerarchici attraverso l'analisi di serie storiche e di simulazioni di Monte Carlo.





# Capitolo 1

## Il portafoglio

Il portafoglio riguarda la situazione nella quale una persona fisica o giuridica detiene strumenti di diverse categorie con un fine economico o di controllo su una determinata società entrando a far parte della compagine sociale. In questa sede ci si occuperà del fine economico tralasciando i diritti e obblighi derivanti dalla partecipazione in una società.

Gli strumenti facenti parte il mercato sono molto numerosi e si dividono sotto il gruppo dei titoli rappresentativi il capitale di rischio, obbligazioni, derivati, titoli negoziati sul mercato monetario e quote di fondi. In questo elaborato sono stati considerati solo i titoli azionari malgrado l'applicativo che verrà esposto successivamente possa essere ampliato prevedendo buona parte degli strumenti appena descritti. I titoli azionari vengono negoziati in molteplici indici tra i quali, DowJ, Nasdaq, FTSEMIB, LSE ecc. Le negoziazioni permettono agli operatori di acquistare e vendere titoli dietro un prezzo prefissato che si sviluppa nel corso della giornata in base a tanti fattori, quali, volatilità del mercato, scenari politici, economici e sociali. L'investitore acquista o vende titoli tramite il book di negoziazione che permette quindi l'incrocio tra la domanda e l'offerta. In un contesto del genere è facile capire quale ruolo importante hanno le informazioni, in questo senso si parla di efficienza di mercato sotto tre aspetti [2]:

- Efficienza in forma debole;

- Efficienza in forma semi-forte;
- Efficienza in forma forte.

La prima si sostanzia nel fatto che i prezzi osservati sul mercato riflettono tutta l'informazione disponibile nella serie storica del titolo. Essenzialmente viene proposta un'analisi della serie storica per la quale le informazioni sono implicite nelle giornate di apertura delle contrattazioni producendo quindi delle variazioni nel prezzo del titolo.

La seconda prevede che i prezzi riflettano tutta l'informazione relativa alla serie storica più qualunque altra informazione pubblica. In questo caso qualsiasi conferenza stampa, documento e articolo di pubblico dominio viene riflesso nel prezzo dell'azione.

L'ultima prevede che i prezzi riflettano tutta l'informazione relativa alla serie storica, qualunque altra informazione pubblica e qualunque informazione privata. In questo caso vi è l'inclusione delle informazioni cosiddette "privilegiate". L'efficienza in forma forte contiene le altre due, e così via a cascata, ovviamente non è vero il contrario.

É chiaro come situazioni del mercato come quelle appena descritte risultino improbabili dal punto di vista concreto del mercato e allo stato attuale dell'economia, in aggiunta non va sottovalutato l'impatto che la tecnologia ha sul mondo delle informazioni. Considerando la pubblicazione di Fama del 1970 appunto dove formalizzava questi tre tipi di efficienza vi è stato da allora un notevole sviluppo sia come già annunciato sotto il profilo tecnologico, sia comunicativo e normativo.

Il mercato azionario prevede la negoziazione di titoli che rappresentino diritti sul capitale della società quotata. Vi possono essere anche diverse categorie di azioni con diversi diritti nonostante non vengano analizzate in questo elaborato. Il processo che riguarda la rigerosità tecnica di trading non verrà presa in considerazione, così come gli oneri annessi a tale procedura, facendo riferimento ad esempio ai costi di transazione, ribilanciamento e annessi.

Il prezzo dell'azione rispecchia il valore che il mercato attribuisce alla società nella sua interezza. Il numero di azioni della società moltiplicato per il prez-

zo dell'azione singola restituisce la capitalizzazione di mercato della società stessa. Una precisazione in tale senso, viene considerato in questo calcolo solo una tipologia di azione ma il medesimo calcolo potrebbe essere esteso a tutte le categorie.

Il numero di azioni rappresentante un capitale sociale può cambiare nel tempo con l'introduzione di nuove azioni nel mercato da sottoscrivere, tramite conversioni di altre tipologie di strumenti, come le obbligazioni convertibili, ancora, tramite *split* o *merge* di azioni o per riacquisto di quelle già in circolazione. Le operazioni sui titoli azionari sono svariate e per portare al lettore un chiaro esempio riguardante il mercato, l'annuncio di una delle modalità sopra descritte provoca immediatamente un cambiamento sulle aspettative del titolo provocandone una variazione di prezzo, questo succede anche nella stessa società quando vi sono annunci che favoriscono una tipologia di azione rispetto ad un'altra, si pensi ad annunci di *Dual Class Unification* o *Dual Class Creation*.

## 1.1 Rischio e rendimento

Il principale mezzo di valutazione di uno strumento finanziario come il titolo azionario è il calcolo dei rendimenti, questi vengono calcolati su un orizzonte arbitrario di tempo  $\frac{1}{T}$ , dove  $T$  si riferisce alla serie storica completa. Il motivo per la quale non viene considerata l'intera serie storica del titolo ma solo una frazione di questa corrispondente a  $\frac{1}{T}$  riguarda soprattutto l'analisi che deve essere svolta, infatti nel caso di un orizzonte temporale di investimento su base annuale è controproducente tenere conto della serie storica degli ultimi 20 anni del titolo, i dati così analizzati rappresenterebbero circa 240 osservazioni mensili che in altri contesti sarebbero una buona base di partenza come quantitativo di osservazioni ma il problema riguarda appunto la varietà di scenari verificatisi in tale periodo. In ottica di *business forecasting* l'orizzonte temporale si aggira dai 3+ anni evidenziando già grosse lacune per quanto riguarda gli scenari possibili.

Scelta quindi la finestra temporale si procede con il calcolo dei rendimenti tramite l'analisi del prezzo  $P$  riguardante il sottostante<sup>1</sup> in diversi periodi di tempo:

$$r_i = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Ottenuti i rendimenti si può calcolare la media aritmetica di questi ricavandone il valore atteso, ovvero il valore che ci si aspetta di ottenere in tale orizzonte temporale. La media è il migliore strumento per la previsione delle aspettative sui rendimenti malgrado sia sensibile agli *outlier*<sup>2</sup>, per questo è possibile affiancarla a strumenti come la mediana, oppure a distribuire tali valori estremi tra tante osservazioni considerando una finestra temporale imponente.

È necessaria una precisazione sull'interpretazione dei dati ricavati dal mercato in ottica futura, infatti vi è un'assunzione molto forte per la quale ci si attende che l'informazione ricavata dalla serie storica possa ripresentarsi in futuro, o meglio utilizzare i dati storici per effettuare previsioni sul futuro. Ovviamente è facile capire come non sia possibile predire il futuro con certezza, perciò ci si affiderà alla statistica per ottenere risultati probabilistici rappresentanti gli scenari più comuni, in questo caso ottenendo il valore atteso.

I rendimenti ottenuti per ogni singolo asset verranno analizzati attraverso una misura di dispersione qual è la varianza e la deviazione standard, queste due misure permettono di dare all'investitore un'idea chiara di come il titolo si muova nel tempo rispetto alla propria media. Per dare al lettore una chiara idea delle misure di dispersione si immagina di tracciare sul pavimento una linea e di far cadere dei fogli perpendicolarmente a questa, i fogli per via di determinate variabili non cadranno tutti nello stesso punto, alcuni lontani e altri vicini a questa linea. La linea in questione rappresenta la media dei

---

<sup>1</sup>Con sottostante si intende lo strumento finanziario sulla quale fa riferimento il derivato. In questo caso, per analogia, si intende il titolo di riferimento per l'analisi.

<sup>2</sup>Con outlier si intende un valore anomalo, ovvero un valore distante dalle altre osservazioni disponibili. Potrebbe causare problemi quando si usano strumenti come la media aritmetica per l'osservazione di una popolazione statistica.

rendimenti del titolo mentre tutti i fogli rappresentano i valori assunti dal titolo in ogni giornata di contrattazione. Più i fogli variano nella distanza dalla linea, per analogia, il titolo sarà più volatile, viceversa, il titolo tenderà ad essere costante nella realizzazione dei rendimenti. Questa misura di dispersione può essere in alcuni casi un bene per l'investitore e per altri una situazione poco ottimista, il concetto viene riportato in equilibrio dal trade-off più comune, ovvero dalla relazione rendimento e rischio del titolo per la quale l'investitore determinerà una strategia in base alla propria funzione di utilità.

Il calcolo di questa variabilità del titolo rispetto alla propria media viene calcolata come segue:

$$\text{Varianza} = \sigma_x^2 = m_2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

Nel calcolo di portafoglio e per il calcolo di indici di performance viene usata spesso anche la radice quadrata della varianza ovvero:

$$\text{Deviazione standard} = \sigma_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

Queste due misure saranno utilizzate come parametro per valutare il rischio di un titolo e del portafoglio. Si porta il lettore a pensare ad una situazione nella quale il rendimento previsto, quindi il valore atteso di un investimento sia un certo ammontare e che la probabilità che questo si discosti da quanto valutato sia molto bassa. Nella situazione opposta si ha invece un investimento con le stesse caratteristiche del primo ma con una variabilità molto più accentuata, questa porterà inevitabilmente situazioni di incertezza in riguardo all'effettivo rendimento a scadenza.

Il punto focale non è tanto selezionare titoli che abbiano bassa volatilità ma costruire un portafoglio ben bilanciato che tenga conto di queste varianze distribuendole in modo ponderato per i propri vincoli.

Una panoramica più approfondita si evidenzia nell'analisi della distribuzione dei rendimenti, questi si assume che abbiano una distribuzione normale

studiabile attraverso momento terzo e momento quarto.

$$Skewness = m_3 = \frac{1}{T} \sum_i^T \left( \frac{(x_i - \bar{x})}{\sigma} \right)^3$$

$$Kurtosis = m_4 = \frac{1}{T} \sum_i^T \left( \frac{(x_i - \bar{x})}{\sigma} \right)^4$$

Ottenuti il momento terzo e momento quarto è possibile calcolarne i coefficienti indicativi per l'analisi della distribuzione così facendo:

$$C_{Skewness} = g_1 = \frac{m_3}{m_2^{3/2}}$$

$$C_{Kurtosis} = \frac{m_4}{m_2^2} - 3$$

Il coefficiente  $C_{Skewness}$  assume valore in base alla simmetria della distribuzione, ovvero:

- $g_1 = 0$ , se la distribuzione è simmetrica;  
La distribuzione è centrata sulla media con uguali osservazioni a sinistra e a destra.
- $g_1 > 0$ , se la distribuzione è asimmetrica positiva;  
Ci sono più valori nella parte destra della distribuzione e in termini di rendimenti questo è un vantaggio per la distribuzione poiché significa che vi sono più rendimenti positivi che negativi dunque il titolo tende a performare tralasciando l'entità di tali valori.
- $g_1 < 0$ , se la distribuzione è asimmetrica negativa.  
In questo caso vi sono più rendimenti negativi che positivi, parametro negativo nel caso generale e non tenendo conto dell'entità degli stessi nella distribuzione.

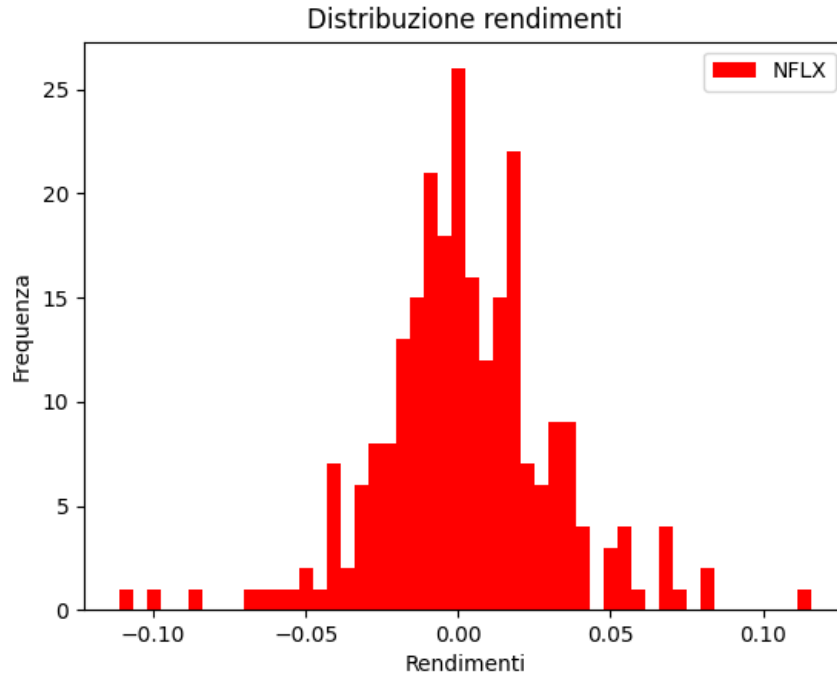


Figura 1.1: Esempio di distribuzione dei rendimenti del titolo NFLX, presenta una skewness pari a  $-0.005563$

La curtosi riguarda l'allontanamento dalla normalità distributiva osservando quindi una distribuzione modificata per quanto riguarda la classica normale standard. Il coefficiente viene così interpretato:

- $C_{Kurtosis} > 0$ , la curva si definisce leptocurtica;  
In questo caso si ha una maggiore frequenza di distribuzione attorno al valore centrale, le code saranno meno spesse.
- $C_{Kurtosis} < 0$ , la curva si definisce platicurtica;  
In questo caso la curva presenta frequenze distribuite lungo tutte le classi, quindi si noterà un maggiore spessore lungo tutta la distribuzione, le code saranno più evidenti rispetto alla curva leptocurtica.
- $C_{Kurtosis} = 0$ , la curva si definisce normocurtica.  
Le sembianze della distribuzione tendono alla normale standard.



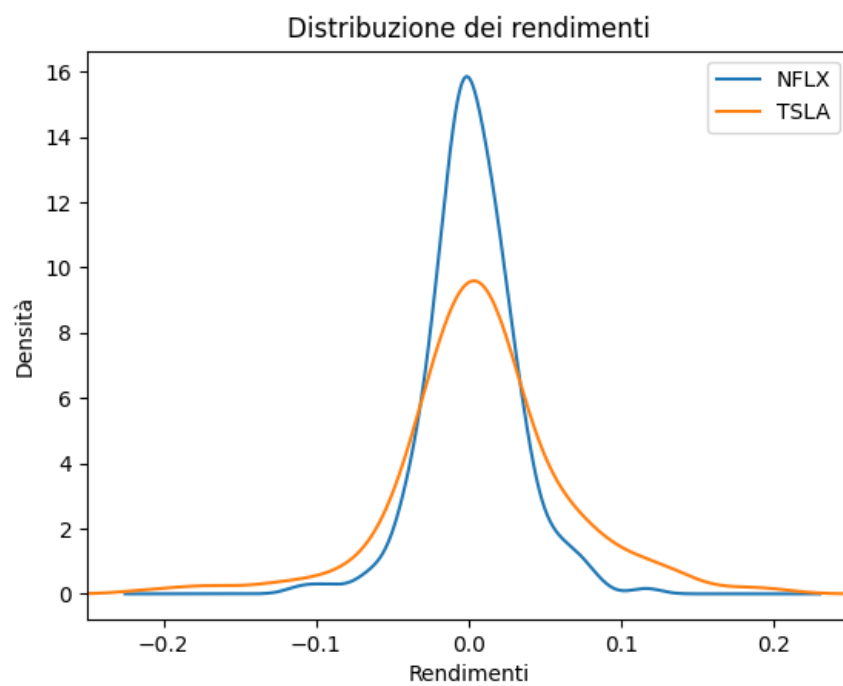


Figura 1.2: Viene proposto un esempio di comparison tra la distribuzione del titolo NFLX kurtosis = 2.0935 e il titolo TSLA kurtosis = 2.5978

Una precisazione per quanto riguarda la formula di calcolo della curtosi utilizzata, il coefficiente  $-3$  è presente solo quando si ha intenzione di utilizzare la definizione di curtosi proposta da Fisher considerando dunque il coefficiente di curtosi come sopra proposto. Nel caso invece si voglia computare la curtosi secondo Pearson tale costante non sarà presente, il coefficiente assumerà valore 3 come punto di distribuzione normocurtica.

Attraverso la considerazione di questi indici assieme alle misure basilari è possibile avere una panoramica estesa di come si è comportato il titolo nell'orizzonte considerato.

Basandosi sullo studio dei momenti centrali è opportuno ricercare in un titolo qualità come una tendenza nel tempo a overperformare rispetto al proprio rischio, coda destra più densa rispetto alla coda sinistra, distribuzione leptocurtica e asimmetria destra.

## 1.2 Costruzione del portafoglio

Con il termine portafoglio si intende un paniere di beni, in questo caso titoli azionari che variano reciprocamente con l'obiettivo di trarre un profitto dal *dividend yield* e/o al momento del *capital gain*. L'investitore accetta un trade-off tra rischio e rendimento in base alla propria funzione di utilità, l'obiettivo è quello di minimizzare la componente di rischio per massimizzare il rendimento estraibile dal portafoglio [3].

Il calcolo della performance di portafoglio avviene tramite il rendimento atteso, questo si compone del vettore dei pesi  $w_i$  e del vettore dei rendimenti attesi dei titoli  $r_i$ , viene calcolato attraverso la formula:

$$E(R_P) = \sum w_i r_i$$

Ogni asset partecipa nella creazione del rendimento di portafoglio, infatti questo viene ponderato per il peso  $w_i$ , rappresentate la quota di capitale assegnatagli rispetto al totale investito nel portafoglio che, una volta moltiplicato per il rendimento dell'asset danno vita ad una sorta di media pesata.

È chiaro come sotto questo punto di vista si potrebbe pensare di inserire solo titoli con alti rendimenti attesi in modo da alzare il rendimento di portafoglio, si ricorda però che tali rendimenti sono associati a rischi proporzionali, perciò non si potrà avere un alto rendimento e un basso rischio di portafoglio comparato alla frontiera efficiente.

Per l'analisi del rischio di portafoglio attraverso la deviazione standard si introduce la modalità di costruzione più comune nel ruolo didattico per illustrare il funzionamento di come la componente rischiosa sia distribuita tra gli asset, ovvero l'*Equally Weighted Portfolio* [4]. Un portafoglio in questo senso costruito avrà una pari componente di peso per ciascun asset selezionato. La standard deviation di portafoglio nel caso di 2 asset sarà così ricavata:

$$\sigma_P = \sqrt{w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2\rho_{AB} \sigma_A w_A \sigma_B w_B} =$$

$$\sqrt{w_A^2 \sigma_A^2 + (1 - w_A^2) \sigma_B^2 + 2\rho_{AB} \sigma_A w_A (1 - w_A) \sigma_B}$$

A questo punto è necessario fare una considerazione su  $\rho_{AB}$ , questo si riferisce al coefficiente di correlazione di Pearson e spiega la tendenza dei titoli, in questo caso  $A$  e  $B$ , a variare reciprocamente. Una precisazione sulla tendenza riguarda il fatto che non vi è una relazione di causa-effetto ma solo la rappresentazione di una relazione sistematica tra i due asset.

Questo coefficiente verrà approfondito nella sezione sulle matrici di correlazione, per ora è opportuno sapere che può assumere valori all'interno di un range  $[-1; 1]$ . Tale coefficiente modifica il rischio di portafoglio in modo importante, verranno analizzati i casi limite del valore del coefficiente, infatti:

- $\rho_{AB} = 0$ , i titoli non sono relazionati tra loro;

Nel caso di nessuna relazione non vi è alcuna variazione reciproca dei titoli per il periodo di tempo analizzato. Nel caso in cui uno di questi sia in fase di *draw down* l'altro titolo tenderà a non esserne influenzato, il portafoglio sarà condizionato dalla performance negativa del titolo

ponderata per il proprio peso. É evidente come un valore del coefficiente pari a 0, oltre ad essere improbabile, non presenti ne difetti ne pregi per il portafoglio.

- $\rho_{AB} = 1$ , i titoli sono perfettamente correlati positivamente tra loro. In questo caso si ha la combinazione lineare del rischio degli investimenti inseriti in portafoglio. Nel caso di crisi di uno dei titoli, l'altro tenderà ad essere in una situazione simile, perciò in portafoglio non vi sarà alcuna attenuazione di tale rischio.
- $\rho_{AB} = -1$ , i titoli sono perfettamente correlati negativamente tra loro. La riduzione del rischio è nella componente massima poiché vi sarà una divergenza tra i titoli. In una situazione del genere è sempre possibile trovare un portafoglio efficiente con rischio pari a zero.

Il ruolo del coefficiente di correlazione presenta quindi un'importante componente nel rischio di portafoglio, ovviamente il concetto appena introdotto ha un'incidenza maggiore poiché il portafoglio a titolo esemplificativo è formato da solo 2 asset con pari peso.

$$\sigma_P^2 = \sum_{i=1}^N \left(\frac{1}{N}\right)^2 \sigma_i^2 + \sum_{i=1}^N \sum_{j \neq i}^N \left(\frac{1}{N}\right) \left(\frac{1}{N}\right) \sigma_{ij}$$

Variando il numero di titoli con  $N \rightarrow \infty$  si vede la prima parte della formula, rappresentativa della componente diversificabile di rischio, tendere a 0 mentre nella seconda parte, rappresentativa della componente sistematica del portafoglio si osserverà la tendenza alla covarianza media.

L'inserimento di più titoli induce ad una riduzione del rischio fino ad un limite corrispondente alla covarianza media del portafoglio, sotto il quale, diversificando ulteriormente, il rischio non potrà essere ridotto.

In generale quindi, i benefici ottenuti dalla diversificazione dei titoli è tanto maggiore quanto è minore la correlazione tra i titoli scelti.

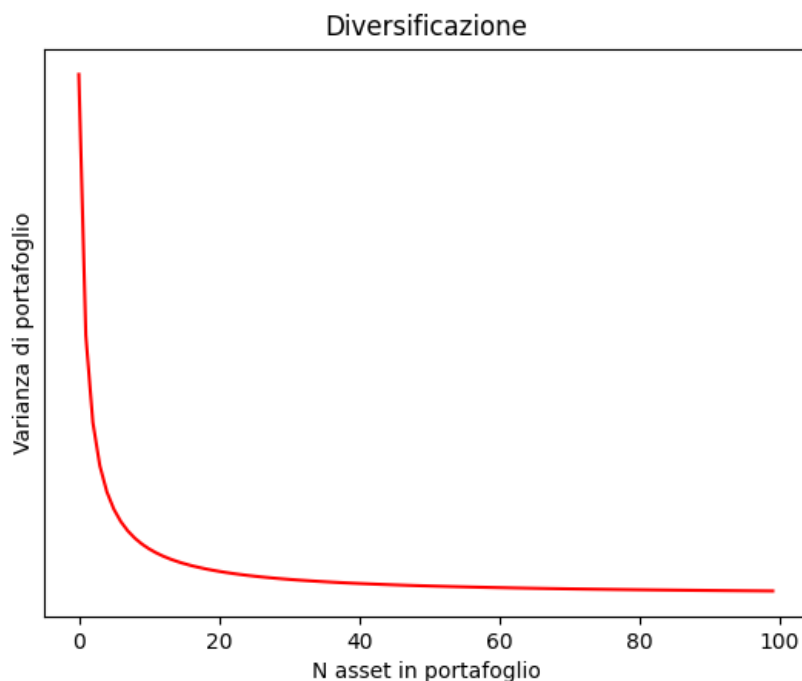


Figura 1.3: Viene rappresentata graficamente la varianza di portafoglio al variare del numero di asset inseriti.

Attraverso la notazione vettoriale la deviazione standard può essere ricavata come segue:

$$\sigma_P = \sqrt{w' COV w} = \begin{bmatrix} w_1, \dots, w_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

Per fini di calcolo è molto più comodo utilizzare la notazione matriciale soprattutto quando vi sono molti titoli in portafoglio.

Giunti a questo punto è opportuno introdurre un indice di performance che rapporterà l'efficienza tra rendimento e rischio. Si tratta dell'indice di Sharpe e viene calcolato utilizzando gli strumenti fino ad ora introdotti, ovvero, rendimento atteso di portafoglio, deviazione standard di questo è il rendimento risk free  $r_f$ . Quest'ultimo viene spesso associato al rendimento di un bond

governativo, solitamente lo stato più solido del mercato o ad una struttura di tassi accettati dal mercato come parametro risk free. Lo *Sharpe Ratio* viene così calcolato:

$$\text{SharpeRatio} = SR = \frac{(E(R_p) - rf)}{\sigma_P}$$

Per avere una panoramica più estesa è possibile calcolare anche i percentili della distribuzione dei rendimenti in modo da dividere la distribuzione in gruppi e poter trarre informazioni per un'ipotetica analisi più approfondita di performance e di loss. In tema di portafogli viene spesso utilizzato un indice chiamato *Value at Risk* che permette di identificare il valore assunto dal portafoglio nel caso di perdite ad un certo percentile, ovvero nell'estrema coda sinistra della distribuzione, in poche parole, oltre la soglia delle perdite più significative.

$$VAR_{1-\alpha} = -(\bar{R}_P - q_{1-\alpha}\sigma(R_P))V$$

Attraverso questo indice di performance viene stimato il valore massimo della loss di portafoglio in un determinato periodo di tempo per una determinata soglia. Il coefficiente  $q_{1-\alpha}$  rappresenta il livello di confidenza attraverso il quale l'analisi verrà fatta, in caso di una stima del VAR al 95% la perdita stimata sarà la minore perdita di portafoglio superando il percentile del 5% in direzione della coda sinistra.

Una misura integrativa rispetto al VAR è il C-VAR o *Expected shortfall*, questo indice rappresenta la perdita media tra le perdite più importanti caratterizzanti l'*expected loss* del portafoglio in un certo intervallo di confidenza per un certo orizzonte temporale. Il C-VAR viene calcolato come segue:

$$ES(x) = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_u(x) du$$

Vi sono innumerevoli strumenti e indici di performance da utilizzare e analizzare per una gestione del portafoglio ottimale, l'obiettivo è utilizzare indici

che permettano una chiara rappresentazione dei propri investimenti sotto il profilo del rendimento e del rischio. È chiaro come un investitore amatoriale non disporrà né del tempo né dei mezzi di un investitore istituzionale, perciò, è bene focalizzare le risorse sugli indici principali che permettano di esprimere una gestione degli investimenti consapevole e trasparente.

## Capitolo 2

# HRP e algoritmi per la creazione di portafogli

Tra i tanti metodi di costruzione di portafoglio ne verranno citati in questa sede solo alcuni, quali il Critical Line Algorithm (CLA), Inverse Variance Portfolio (IVP) e il soggetto dell'analisi Hierarchical Risk Parity (HRP). É importante l'introduzione di questi metodi poiché, malgrado non ci si soffermerà sulla comparazione di essi nei vari aspetti caratterizzanti un'analisi di portafoglio, è bene portare al lettore la conoscenza di quali siano i vantaggi della scelta dell'HRP.

Il CLA teorizzato da Markowitz è il modello attualmente più comune e insegnato nelle scuole di economia, si basa sul presupposto che esista una frontiera efficiente composta dall'insieme di portafogli efficienti tali per cui il rapporto tra rischio e rendimento è in equilibrio e la scelta ricadrà sul soggetto in base discrezionale tenendo conto della propria funzione di utilità. Gli obiettivi del soggetto investitore saranno quindi di minimizzazione in termini di rischio a parità di rendimento, oppure di massimizzazione del rendimento a parità di rischio.

L'IVP viene assemblato in modo più efficiente rispetto al modello teorizzato da Markowitz, il concetto di fondo riguarda l'assegnazione del peso in modo inversamente proporzionale alla propria varianza. La media pesata corri-



spondente all'*inverse variance* è la minore tra tutte le medie pesate. I due aspetti caratterizzanti l'IVP sono, la distribuzione omogenea dei pesi tra tutti gli asset per ridurre il rischio e l'ignorare la matrice di correlazione tra gli asset. Il risultato di queste due caratteristiche sarà una più elevata varianza in casi di shock idiosincratici.

L'HRP cerca di intermediare tra i due metodi sopra accennati, infatti non concentrerà la maggior parte del peso su pochi assets come potrebbe capitare per il CLA, di contro, non distribuirà i pesi tanto omogeneamente sugli asset poco performanti in cerca di minor rischio quanto l'IVP ma troverà una via di mezzo che permetterà di fatto una migliore performance in termini di rischio rispetto agli altri due metodi.

I dati utilizzati per l'analisi dei metodi di portafoglio annunciati precedentemente devono rispettare determinate condizioni per far sì che l'analisi in questa sede sia quantomeno corretta dal punto di vista formale. Per validare la tesi secondo il quale la varianza del metodo HRP sia minore degli altri metodi (IVP-CLA) basterebbe utilizzare una serie di dati che utilizzata come input nell'algoritmo restituisca una composizione di portafoglio che validi la tesi.

Un procedimento come quello appena descritto viene chiamato *In Sample Analysis*, purtroppo questo metodo non consente di comparare o confermare i risultati dell'algoritmo in un contesto più globale. Dato questo pretesto, l'analisi formale dell'algoritmo verrà svolta con serie di dati *Out of Sample*, ovvero generando questi dati secondo un input pseudo randomico rappresentante una normale standard, successivamente, sommati in funzione di un drift arbitrario  $\sigma = 0.25$  e poi testati svolgendo simulazioni di Monte Carlo. I dati generati random distribuiti come una normale standard<sup>1</sup> simuleranno 10000 osservazioni rappresentanti rendimenti su base giornaliera per cinque asset fittizi. I restanti cinque assets saranno invece il risultato di un'equazione creata appositamente per indurre una correlazione con i primi cinque. Il

---

<sup>1</sup>Con distribuzione normale si intende una serie di dati distribuiti in modo standardizzato con media = 0 e deviazione standard = 1.

risultato del procedimento di generazione saranno quindi dieci assets correlati tra loro in modo da simulare il mercato di titoli reali. Per testare in modo più completo gli algoritmi di IVP e HRP vengono indotti shock all'interno della serie per simulare un contesto di alta volatilità e mettere alla prova la reazione degli algoritmi di portafoglio.

## 2.1 Costruzione delle matrici di covarianza e correlazione

Il processo di costruzione delle matrici di covarianza e correlazione non è semplice come possa apparire, molti fattori intervengono quando si analizzano serie storiche e soprattutto quando si utilizzano grandi mole di dati. Se si parla di costruzione standard avremo una matrice di covarianza di  $N$  colonne e  $N$  righe con  $N$  il numero di asset selezionati. La formula per trovare la covarianza tra un elemento  $x$  e un elemento  $y$  è la seguente:

$$COV(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Questa indica la capacità di un elemento  $x$  di variare assieme all'elemento  $y$ . In forma matriciale si avrà un output corrispondente alle covarianze di ogni asset per ogni altro asset facente parte del pool di titoli. Sulla diagonale si avranno invece le varianze dei singoli asset, infatti riprendendo la formula della covarianza applicandola allo stesso elemento si otterrà la formula della varianza:

$$COV(x, x) = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{n} = VAR(x)$$

È cruciale la caratteristica della diagonale poiché fondamentale per il calcolo della varianza di portafoglio in forma matriciale soprattutto se gli asset selezionati costituiscono un numero importante.

È importante distinguere la covarianza dalla correlazione poiché potrebbero

essere confuse data la struttura molto simile. Entrambe le misure determinano la dipendenza tra due variabili, nel caso della correlazione si ha un indice standardizzato ovvero che varia tra  $-1$  (massima correlazione negativa) e  $1$  (massima correlazione positiva) mentre la covarianza non ha un range attraverso il quale esprime la massima e minima variazione reciproca. Il coefficiente che esprime il grado di correlazione tra due variabili, in questo caso  $x, y$  è così calcolato:

$$\rho_{x,y} = \frac{COV(x,y)}{\sigma_x \sigma_y}$$

La matrice di correlazione che verrà modificata durante la computazione dell'algoritmo pre e post clusterizzazione viene rappresentata tramite heatmap per portare al lettore una chiara visione della struttura stessa:

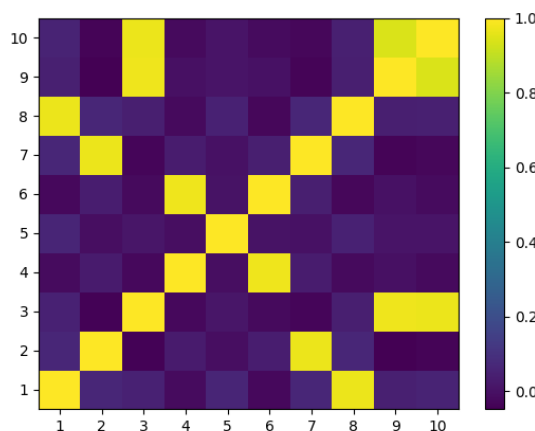


Figura 2.1: Heatmap della matrice di correlazione tra 10 asset generati secondo distribuzione normale. Cinque di questi asset sono arbitrariamente correlati agli altri cinque tramite una funzione di addizione alla distribuzione normalizzata con drift  $= \sigma = 0.25$ .

Come si può notare dalla figura 2.1 la diagonale è composta da valori unitari poiché c'è perfetta correlazione positiva al variare dello stesso asset, si

può notare d'altra parte che vi sono valori tendenti all'uno riguardanti asset diversi, dimostrazione di come il processo di correlazione indotta abbia funzionato. Ovviamente il processo di dati standardizzato si discosta dalla realtà poiché come si può notare ci sono molti asset dove la correlazione è prossima allo zero, nei mercati è molto difficile avere correlazioni così basse.

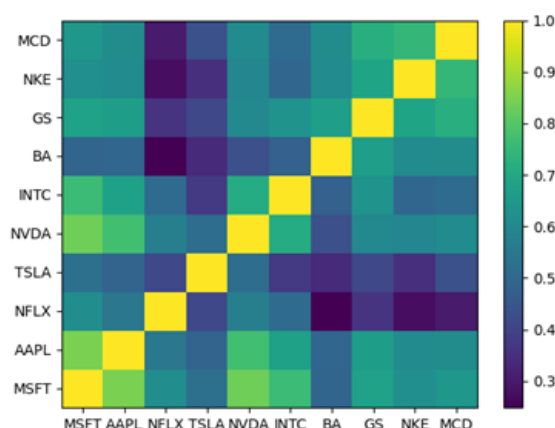


Figura 2.2: Un esempio di heatmap della matrice di correlazione tra 10 assets selezionati dal mercato US.

Viene portato un esempio di titoli scelti dal mercato americano, si tratta di titoli per la maggior parte inerenti al settore tecnologico con più alta capitalizzazione. Si nota che rispetto alla matrice di correlazione proposta precedentemente valori sotto a 0.4 sono particolarmente rari. Sempre per la validità formale della tesi si specifica che gli asset sono stati selezionati arbitrariamente per portare al lettore un'idea chiara della matrice di correlazione e successivamente verrà svolta con gli stessi un'analisi affiancata da quella formale *Out of Sample*.

Ottenuta la chiave di lettura per l'interpretazione della correlazione è necessario introdurre il concetto di cross-correlation [5] tra time series, ovvero una misura che traccia i movimenti di serie storiche e che determini quanto queste si incrocino e si sovrappongano nel tempo. Per chiarire l'idea della cross-correlation si può esprimere il concetto sotto forma di variabili dipen-

denti  $(Y, Z)$  e indipendente  $(X)$ . Ipotizzando che il valore di  $X$  vari, di conseguenza anche il valore di  $Y$  varierà per funzione della relazione di dipendenza. Se si ipotizza che vi sia una relazione di dipendenza simile anche tra  $X$  e  $Z$  allora al variare di  $X$  varieranno anche  $Y$  e  $Z$ . Di fatto si può dire che le variabili  $Y$  e  $Z$  sono cross-correlated perché sono correlate come risultato individuale di una relazione di dipendenza con  $X$ .

Per la validità della tesi si è scelto di utilizzare al posto della correlazione standard un metodo innovativo che prenda in considerazione la cross-correlation ovvero il metodo della *Detrended Fluctuation Analysis (DFA)*. All'interno di questo metodo vi sono diversi processi di stima della correlazione come, il coefficiente di Spearman, *Detrending Moving-Average* e il coefficiente di Pearson. Nel nostro caso si utilizzerà la *Detrended Cross Correlation Analysis (DCCA)* [6] [7] [8] [9] per catturare la cross-correlation nel caso di serie storiche non stazionarie.

Il processo di stima della matrice di correlazione DCCA avviene attraverso cinque passaggi:

1. Vengono prelevati due asset con le relative serie storiche  $x_i, y_i$  con  $i = 1, 2, \dots, N$ , dove  $N$  sono le osservazioni totali della time series. Viene poi computato lo scarto dalla media per ogni asset selezionato:

$$X_k = \sum_{i=1}^k (x_i - \bar{x})$$

$$Y_k = \sum_{i=1}^k (y_i - \bar{y})$$

2. Per questo step è necessario introdurre il concetto di metodo dei minimi quadrati (OLS), ovvero la tecnica che permette di trovare una funzione, rappresentata da una curva, che minimizzi la somma dei quadrati delle distanze delle osservazioni da essa. Questa funzione è in via analitica riportata tramite equazione composta da una variabile dipendente e da una o più variabili indipendenti. Tramite l'assegnazione di valori osservati alla variabile indipendente e risolta l'equazione si ottiene un

valore per la variabile dipendente, questo valore è chiamato *fitted value*. Queste time series  $x_i, y_i$  vengono divise in  $(N - n)$  sottogruppi di uguale lunghezza  $n$ , dove  $4 \leq n \leq \frac{N}{4}$ .

Vengono calcolati i fitted values tramite OLS in ogni sottogruppo ottenendo  $\{\tilde{X}_{k,i}\}, \{\tilde{Y}_{k,i}\}$ , ovvero i local trends per ogni serie.

3. Viene calcolata la covarianza dei residui in ogni sottogruppo:

$$f_{(n,i)}^2 = \frac{1}{(n+1)} \sum_{k=1}^{i+n} (X_k - \tilde{X}_{k,i}) (Y_k - \tilde{Y}_{k,i})$$

4. Viene calcolata la media di tutti gli  $(N - n)$  sottogruppi ottenendo la funzione *detrended covariance*:

$$F_{xy}^2(n) = \frac{1}{(N - n)} \sum_{i=1}^{N-n} f_{xy}^2(n, i)$$

5. Come ultimo passo si calcola il coefficiente di *detrended correlation*:

$$\rho_{DCCA}(n) = \frac{F_{xy}^2(n)}{F_{xx}(n) F_{yy}(n)}$$

I coefficienti al denominatore dell'ultima equazione rappresentano la radice quadratica media delle fluttuazioni degli asset prelevati nel primo passaggio. Il coefficiente della DCCA è standardizzato e assumerà valori:

- $\rho_{DCCA}(n) = 1$ , se gli asset sono perfettamente correlati in modo positivo;
- $\rho_{DCCA}(n) = 0$ , se gli asset non sono correlati;
- $\rho_{DCCA}(n) = -1$ , se gli asset sono perfettamente correlati negativamente.

Attraverso l'utilizzo di un metodo come il DCCA si prende in considerazione la non stazionarietà, infatti si otterrà una matrice di correlazione che non è dipendente dall'assunzione che le time series siano costanti nel tempo.

All'interno del processo per la computazione del metodo DCCA viene proposto un approccio integrativo attraverso l'utilizzo di una sliding window [10] [11] [12]. Essenzialmente il processo prevede la creazione di un buffer all'interno della serie storica di grandezza  $w$ . Selezionata arbitrariamente la grandezza appena proposta le osservazioni saranno inserite all'interno del buffer in relazione al periodo, l'algoritmo incrementerà la posizione del buffer over time ottenendo in fine un array di coefficienti DCCA.

$$\rho_{DCCA}(n, w, T) = \frac{F_{xy}^2(n, w, T)}{F_{xx}(n, w, T) F_{yy}(n, w, T)}$$

## 2.2 Metodi di clustering

Ottenute le matrici di covarianza e correlazione è necessario un ultimo passo prima della clusterizzazione ovvero la creazione di una matrice funzione della correlazione per il calcolo delle distanze [13]. Con calcolo delle distanze si intende il processo Euclideo per il calcolo della distanza tra due punti. La motivazione che risiede dietro questo procedimento viene spiegata dal processo computazionale dell'algoritmo di clustering, infatti il modus operandi generale di questo prevede l'allocazione in un determinato gruppo tramite la distanza dell'oggetto identificato da tutti gli altri oggetti esistenti nell'universo specificato. L'intuizione alla base dell'algoritmo di clustering è simile a quanto introdotto riguardo all'OLS.

Il primo passo inerente il tree clustering considerando  $N$  asset per  $T$  osservazioni è trasformare la precedentemente descritta matrice di correlazione in una matrice appropriata per il calcolo delle distanze. In input verrà quindi utilizzata la matrice DCCA ( $N \times N$ ) computando per ogni coppia di elementi  $i, j$ :

$$d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$$

Ottenuta la matrice distanze  $D$  è possibile ora calcolare le distanze Euclidee

| Dist. | MSFT | AAPL | NFLX | TSLA | NVDA | INTC | BA   | GS   | NKE  | MCD  |
|-------|------|------|------|------|------|------|------|------|------|------|
| MSFT  | 0    | 0.31 | 0.45 | 0.47 | 0.36 | 0.45 | 0.51 | 0.48 | 0.47 | 0.51 |
| AAPL  | 0.31 | 0    | 0.51 | 0.52 | 0.40 | 0.48 | 0.53 | 0.47 | 0.50 | 0.53 |
| NFLX  | 0.45 | 0.51 | 0    | 0.50 | 0.44 | 0.57 | 0.61 | 0.62 | 0.60 | 0.63 |
| TSLA  | 0.47 | 0.52 | 0.50 | 0    | 0.45 | 0.56 | 0.51 | 0.54 | 0.54 | 0.48 |
| NVDA  | 0.36 | 0.40 | 0.44 | 0.45 | 0    | 0.46 | 0.48 | 0.51 | 0.42 | 0.47 |
| INTC  | 0.45 | 0.48 | 0.57 | 0.56 | 0.46 | 0    | 0.50 | 0.47 | 0.51 | 0.52 |
| BA    | 0.51 | 0.53 | 0.61 | 0.51 | 0.48 | 0.50 | 0    | 0.40 | 0.36 | 0.32 |
| GS    | 0.48 | 0.47 | 0.62 | 0.54 | 0.51 | 0.47 | 0.40 | 0    | 0.42 | 0.40 |
| NKE   | 0.47 | 0.50 | 0.60 | 0.54 | 0.42 | 0.51 | 0.36 | 0.42 | 0    | 0.34 |
| MCD   | 0.51 | 0.53 | 0.63 | 0.48 | 0.47 | 0.52 | 0.32 | 0.40 | 0.34 | 0    |

Tabella 2.1: Esempio di matrice di distanze tra asset selezionati dal mercato.

tra i diversi asset:

$$\tilde{d}[D_i, D_j] = \sqrt{\sum_{n=1}^N (d_{n,i} - d_{n,j})^2}$$

Questo ci permette di ottenere una matrice di distanze  $(N \times N)$  input essenziale per la funzione di linkage [14]. Per clustering si intende una procedura attraverso il quale diversi oggetti vengono raggruppati in modo gerarchico in base a particolari funzioni.

Nel caso generale si pensi a due oggetti  $i$  e  $j$ , in input  $d(i, j)$ , l'algoritmo inizia con l'analizzare i cluster a  $t = 0$  ancora singoli elementi. Quando due cluster da questo universo vengono combinati assieme in funzione delle distanze formandone uno solo  $u$ , si otterrà questa situazione, gli oggetti singoli in questo caso  $i$  e  $j$  verranno rimossi dall'universo mentre verrà aggiunto il cluster  $u$ . L'output che si ottiene da questo primo step è l'universo rimasto invariato tranne per la sottrazione degli elementi  $i$  e  $j$  e per l'aggiunta di un nuovo elemento  $u$ . È chiaro come il nuovo elemento non avrà la stessa distanza dei due elementi utilizzati come input ma, in base al metodo di linkage scelto, avrà una distanza che in via generale sarà un compromesso tra i due oggetti iniziali. La matrice delle distanze riceverà un update per riflettere la nuova distanza degli elementi nativi con il nuovo cluster  $u$ . Il processo viene



iterato fino a quando non vi saranno più cluster nell'universo. Per portare al lettore un'idea visiva di come i cluster siano organizzati viene rappresentato di seguito un dendrogramma [15] [16], ovvero un particolare grafico in grado di mostrare le connessioni tra cluster gerarchici.

Come si può vedere dalla figura 2.2 gli asset scelti arbitrariamente sono raggruppati dall'algoritmo di linkage. Nel primo step dell'algoritmo si vede il clustering tra i titoli (foglie) *MSFT* e *AAPL*, per poi in istanti successivi convogliare in un altro cluster  $[NVDA \cup u(MSFT \cup AAPL)]$ . L'altezza delle candele rappresenta la dissimilarità tra ogni foglia. Negli ultimi step si possono invece notare titoli come *NFLX*, *INTC*, *TSLA* che non hanno trovato similarità con gli altri cluster.

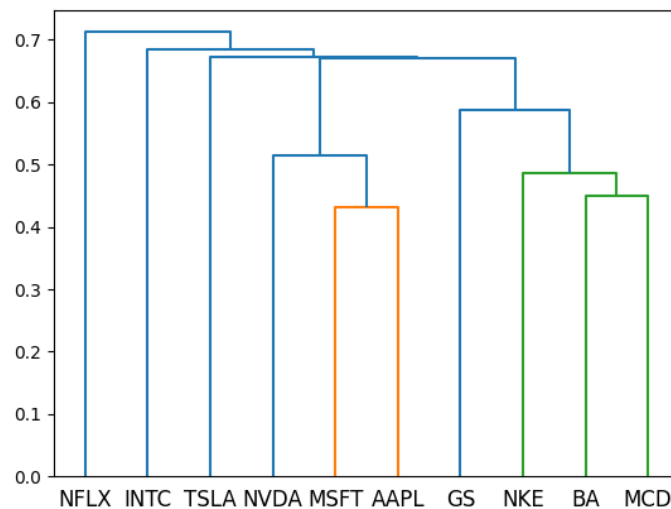


Figura 2.3: Esempio di dendrogramma rappresentante i cluster formatisi attraverso l'algoritmo di linkage. I 10 asset sono gli stessi scelti arbitrariamente anche nella figura 2.2.

Avendo portato al lettore una chiara rappresentazione di come i cluster vengono creati si procederà con la descrizione di ogni metodo di linkage:

- *Single linkage method:*

$$d(u, v) = \min (dis(u[i], v[j]))$$

$$\forall i \in u \wedge j \in v$$

Attraverso questo metodo la distanza tra i due cluster è definita come la distanza tra i suoi due membri più vicini. Spesso porta a cluster dove gli elementi vengono aggiunti singolarmente in modo sequenziale. Conosciuto anche come *Nearest Point Algorithm*, è il più semplice ma può portare a outliers notevoli [17].

- *Complete linkage method:*

$$d(u, v) = \max (dis(u[i], v[j]))$$

$$\forall i \in u \wedge j \in v$$

Conosciuto anche come *Farthest Point Algorithm*.

- *Average linkage method:*

$$d(u, v) = \sum_{ij} \frac{d(u[i], v[j])}{(|u| * |v|)}$$

La distanza tra i due gruppi è definita come la media delle distanze tra ogni elemento.

- *Centroid linkage method:*

$$d(s, t) = ||c_s - c_t||_2$$

Dove  $c_s$  e  $c_t$  sono i centroidi o baricentri dei rispettivi cluster,  $s$  e  $t$ . Quando questi vengono combinati in un nuovo cluster  $u$ , il nuovo centroide viene calcolato in relazione agli elementi originali ( $s$  e  $t$ ). La distanza computata è quella Euclidea tra il centroide di  $u$  e il centroide del rimanente cluster nell'universo  $v$ . Essenzialmente viene definita una distanza tra i due gruppi come la distanza tra i propri baricentri.

- *Median linkage method*: La computazione del metodo è simile al *Centroid Linkage Method*, in questo caso viene definita la distanza tra i due gruppi come la distanza pesata tra i due baricentri, il peso assegnato è proporzionale al numero di elementi dentro ciascun gruppo.
- *Ward linkage method*:

$$d(u, v) = \sqrt{\frac{|v| + |s|}{T} d(v, s)^2 + \frac{|v| + |t|}{T} d(v, t)^2 - \frac{|v|}{T} d(s, t)^2}$$

Dove  $u$  è il nuovo cluster formato da  $s$  e  $t$ ,  $v$  è un cluster non utilizzato nell'universo,  $T = |v| + |s| + |t|$ . È noto anche come *incremental algorithm* poiché aumenta man mano che i cluster si uniscono tra di loro. Il metodo di ward [18] è il più performante in termini di *Expected Shortfall* comparato all'utilizzo del metodo *Single* e *Average* [19] con una stima della matrice di covarianza inferiore.

## 2.3 Quasi diagonalizzazione

La fase di quasi diagonalizzazione della matrice di covarianza si concretizza attraverso la modifica delle righe e delle colonne in modo tale da ottenere come risultato una nuova matrice, riorganizzata, con una proprietà particolare al fine del portafoglio, infatti ci sarà la collocazione di tutti gli investimenti simili vicini, mentre verranno allontanati quelli diversi. Si ricorda che la discriminazione in tale senso viene svolta dall'algoritmo utilizzando la matrice di covarianza come metodo di comparazione tra i titoli, si ottiene quindi una sorta di discriminazione in senso puramente analitico.

Sarebbe opportuno dal punto di vista manageriale creare una sorta di algoritmo ibrido che utilizzi le impartizioni da parte del top management opportunamente giustificate dal punto di vista di *forecasting* e di sconto al tempo attuale insieme ad una revisione da parte dell'algoritmo sugli investimenti proposti. Una accezione del genere potrebbe essere replicata anche nel contesto di diversificazione. Si ricorda che nel contesto della tesi è stato preso in considerazione solamente il mercato azionario nonostante le opportunità

di sviluppo sono estremamente allargabili a strumenti di varia tipologia con caratteristiche altrettanto ampie.

Introdotta il concetto attraverso il quale la quasi diagonalizzazione avviene si propone un esempio concreto step by step utilizzando i titoli estratti dal mercato come nelle sezioni precedenti.

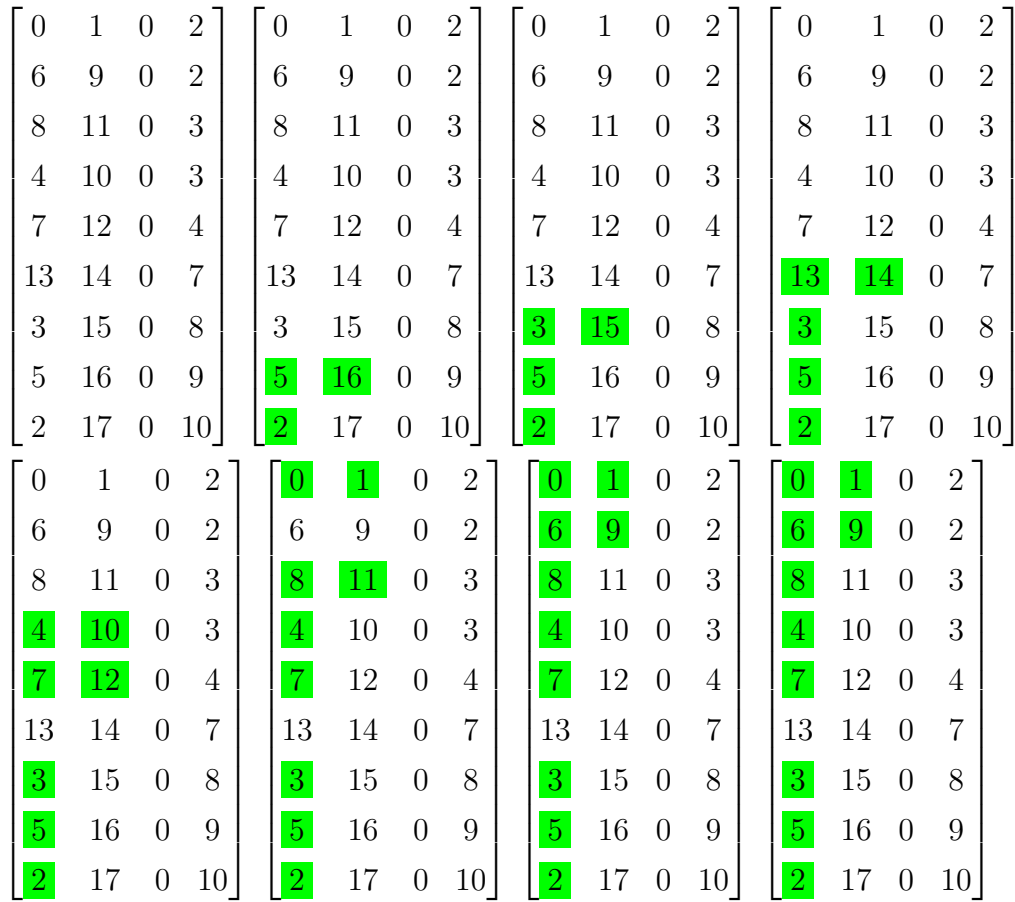


Figura 2.4: Viene rappresentato il processo di quasi diagonalizzazione a livello numerico. Ogni matrice corrisponde ad uno step che viene svolto iterativamente. In verde sono evidenziati gli asset selezionati in base al clustering e poi ordinati e restituiti come lista. I numeri sono la sostituzione analitica dei simboli dei titoli azionari. Una proprietà riguardante questo processo è il fatto che tutti gli asset evidenziati risiedono sopra l'antidiagonale della matrice.

Come risultato di questo processo si otterrà una lista, in questo caso di titoli, ordinata in base al processo di clustering precedente per la quale le proprietà rimangono immutate nonostante il processo di riarrangiamento e di rimpiazzo. L'output di questa fase verrà successivamente dato in pasto al processo di bipartizione ricorsiva.

## 2.4 Bipartizione ricorsiva

La fase di bipartizione ricorsiva consiste nell'assegnazione dei pesi di portafoglio attraverso le composizioni create ad hoc attraverso la bipartizione degli asset ricevuti in input nella fase precedente.

Come annunciato precedentemente è di fondamentale importanza la preservazione dell'ordine dei cluster che rimane inalterato nella fase di quasi diagonalizzazione poiché in questo modo si potrà sfruttare la matrice di covarianza e indirettamente quella di correlazione per quanto riguarda i procedimenti che costruiscono l'algoritmo di HRP descritti in precedenza e per il processo che verrà illustrato in questa sezione. Gli step sono:

1. Vengono inizializzati gli asset dentro un solo cluster  $L = \{L_0\}$  all'interno del quale si hanno tutti gli asset.
2. Viene successivamente inizializzato il vettore dei pesi  $w_i = 1$ . Si ricorda che il vettore dei pesi assumerà in ultima istanza tutti i singoli pesi assegnati ad ogni asset.

In questa sede non verranno analizzati altri vincoli per quanto riguarda i pesi. Si porta il lettore però a ragionare su una ipotetica applicazione di vincolo di pesi negativi, assimilabili in gergo come posizione *short* e quindi con segno negativo. Altri vincoli potrebbero essere intromessi per quanto riguarda doveri normativi istituzionali o per necessità di liquidità, di immagine e di performance. Ovviamente non è questo il contesto per un'analisi sui relativi vincoli possibili nell'ambiente di gestione del portafoglio, ma viene portata al lettore un'idea sulla vastità dell'argomento.

3. Condizione necessaria per lo svolgimento dell'algoritmo è la presenza di più cluster. Di primo impatto potrebbe sembrare fuorviante come condizione, ma è necessario specificare che essendo un algoritmo iterativo il check di questa condizione viene fatta ripetutamente e considerando che nei prossimi passi il cluster  $L_0$  verrà diviso in molteplici cluster, la condizione viene ora ad assumere senso logico.

4. Per ogni  $L_i \in L$  che rispetti  $|L_i| > 1$ :

- $L_i$  viene bipartito in due *sub-cluster* tale che  $L_i = L_i^1 \cup L_i^2$ , dove si ha che  $|L_i^1| = \text{int} \left[ \frac{1}{2} |L_i| \right]$ .

In questo passaggio è importante l'arrotondamento ad un numero intero per evitare problemi di tipo analitico e a livello di codice. L'intuizione è che in caso di cluster con asset dispari non si creino problemi di codice dove non avvenga l'assegnazione di un asset ad un determinato cluster a causa di un ipotetico *float*.

- Viene calcolata la varianza dei componenti del cluster. In questo caso è necessaria una precisazione a livello di costruzione del portafoglio. Rimane chiaro come avendo il vincolo elementare di una distribuzione totalitaria del peso di portafoglio attraverso i titoli per la quale  $w_P = \sum w_i = 1$  questo sia, nel caso di un portafoglio formato da due soli asset, invertibile per ottenerne i pesi dei singoli per complementarità,  $w = w_1 + w_2$  quindi  $w_2 = 1 - w_1$ . Tenendo a mente questo breve concetto l'algoritmo struttura in identico modo la distribuzione del peso verso i sub-cluster. Nonostante in un contesto generalizzato di gestione del portafoglio questo concetto possa essere coadiuvato da diversi input quali capitale disponibile, vincoli di diversificazione e così via, l'unico metodo di assegnazione del peso tramite algoritmo è l'analisi della varianza attraverso la matrice di covarianza tra cluster [20] [21]. Qui si riporta il lettore a quanto annunciato precedentemente in riguardo alla proprietà del cluster e all'ordinamento degli asset facenti parte di questo.

Il calcolo della varianza avviene come riportato di seguito:

$$L_i^j \text{ con } j = 1, 2$$

$$\tilde{V}_i^j = w_i^{j'} V_i^j w_i^j$$

$$w_i^j = \frac{\text{tr} [V_i^j]^{-1}}{\sum_i \text{tr} [V_i^j]^{-1}}$$

La varianza degli elementi facenti parte del sub cluster viene calcolata implementando l'algoritmo di *inverse variance*, tramite questo si ha la costruzione del vettore pesi opportunamente modificato che sarà l'input per il calcolo della varianza tramite operazioni di prodotto matriciale, per semplificare, il vettore dei pesi trasposto viene moltiplicato per la matrice di covarianza degli elementi presi in esame e moltiplicata per il vettore dei pesi.

Si ricorda al lettore che per svolgere una moltiplicazione matriciale è necessaria una *shape* particolare, infatti, in questo caso avremo un vettore pesi idealmente  $w = 1 \times N$  dove  $N$  sono gli asset e una matrice delle covarianze  $v = N \times N$ . Riferendosi al caso generale e non a quello specifico del sub-cluster si ottiene post moltiplicazione  $V = 1 \times 1$  come risultato di  $(\mathbf{1} \times N \times N \times N \times N \times \mathbf{1})$ .

- Viene calcolato il fattore  $\alpha_i$  responsabile dell'assegnazione del peso per quanto riguarda il primo sub-cluster.

$$\alpha_i = 1 - \frac{\tilde{V}_i^1}{\tilde{V}_i^1 + \tilde{V}_i^2}$$

- Le allocazioni del peso vengono ponderate per il fattore  $\alpha_i$  per quanto riguarda il primo cluster  $L_i^1$ .
- Le allocazioni del peso vengono ponderate per il fattore  $(1 - \alpha_i)$  per quanto riguarda il secondo cluster  $L_i^2$ .

5. L'iterazione riprende dallo step 2.

| Bipartizione ricorsiva |            |                                |       |       |       |       |       |       |       |       |       |
|------------------------|------------|--------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Step                   | N° Cluster | Pesi step by step bipartizione |       |       |       |       |       |       |       |       |       |
|                        |            | NFLX                           | INTC  | TSLA  | NVDA  | MSFT  | AAPL  | GS    | NKE   | BA    | MCD   |
| 1                      | 1          | 46.7%                          | 46.7% | 46.7% | 46.7% | 46.7% |       |       |       |       |       |
|                        | 2          |                                |       |       |       |       | 53.3% | 53.3% | 53.3% | 53.3% | 53.3% |
| 2                      | 1          | 26.5%                          | 26.5% |       |       |       |       |       |       |       |       |
|                        | 2          |                                |       | 20.3% | 20.3% | 20.3% |       |       |       |       |       |
|                        | 3          |                                |       |       |       |       | 24.4% | 24.4% |       |       |       |
|                        | 4          |                                |       |       |       |       |       |       | 28.8% | 28.8% | 28.8% |
| 3                      | 1          | 15.1%                          |       |       |       |       |       |       |       |       |       |
|                        | 2          |                                | 11.3% |       |       |       |       |       |       |       |       |
|                        | 3          |                                |       | 4.5%  |       |       |       |       |       |       |       |
|                        | 4          |                                |       |       | 15.8% | 15.8% |       |       |       |       |       |
|                        | 5          |                                |       |       |       |       | 13.6% |       |       |       |       |
|                        | 6          |                                |       |       |       |       |       | 10.9% |       |       |       |
|                        | 7          |                                |       |       |       |       |       |       | 15.4% |       |       |
|                        | 8          |                                |       |       |       |       |       |       |       | 13.5% | 13.5% |
| 4                      | 1          |                                |       |       | 5.8%  |       |       |       |       |       |       |
|                        | 2          |                                |       |       |       | 10.0% |       |       |       |       |       |
|                        | 3          |                                |       |       |       |       |       |       |       | 2.4%  |       |
|                        | 4          |                                |       |       |       |       |       |       |       |       | 11.1% |
| Pesi finali            |            | 15.1%                          | 11.3% | 4.5%  | 5.8%  | 10.0% | 13.6% | 10.9% | 15.4% | 2.4%  | 11.1% |
|                        |            | NFLX                           | INTC  | TSLA  | NVDA  | MSFT  | AAPL  | GS    | NKE   | BA    | MCD   |

Tabella 2.2: Viene mostrato l'algoritmo di bipartizione ricorsiva per quanto riguarda i titoli azionari selezionati in modo discrezionale. Da notare come ad ogni step dell'algoritmo i cluster si separano ottenendo un valore di peso specifico per ogni titolo.





## Capitolo 3

# Hrp in approccio stock market

L'algoritmo di HRP fin'ora descritto ha un'applicazione puramente sperimentale e teorica basandosi sulla generazione random di osservazioni, invece per un approccio più concreto che si avvicini al mercato azionario si è deciso di modificare il codice in modo da permettere all'utente in fase ex-ante di creare una lista di titoli che, successivamente, verranno dati in pasto all'algoritmo di HRP restituendo come output la composizione di portafoglio con annessi i pesi. Questa modalità di computazione HRP ha lo scopo di immedesimare l'utilizzatore in una vera e propria gestione di portafoglio attraverso il picking di asset reali con dati di mercato aggiornati.

Si ricorda al lettore che l'applicativo inerente il mercato azionario non si sostituisce a una qualsiasi piattaforma per la gestione del portafoglio ma è un ibrido verso un approccio accademico e concreto.

Il codice per il selezionamento arbitrario degli asset di input deriva da un progetto passato riguardante una piattaforma di gestione portafoglio semi condizionata da algoritmi base di *robo-advisoring* sempre inerenti la gestione degli asset in portafoglio.

Per introdurre l'applicativo bisogna rendere noto il concetto di *application program interface (API)* [22], questa è un insieme di procedure che permette di interfacciarsi con un ambiente esterno come internet e fare un download o un upload di informazioni. Le API sono spesso rese disponibili dalla casa

produttrice che si interfaccia con una propria applicazione, oppure fatte in modo amatoriale da soggetti su permesso della società.

In questa sede sono state utilizzate API che si interfacciano con il sito di Investing [23], nota piattaforma che rende disponibili dati finanziari ed economici riguardanti diversi mercati, valute, strumenti ecc. Tramite questa modalità è possibile ricevere grandi mole di dati in input e di immagazzinarle temporaneamente, a meno di uno *storing* su un database o server, e di svolgere determinate funzioni per gli scopi presentati in questo elaborato. In prima battuta è opportuno specificare che queste informazioni vengono richieste tramite API ad ogni compilazione dell'applicativo in modo da avere informazioni sempre aggiornate alla data corrente, infatti il *buffer* di osservazioni sarà sempre concernente il  $t_0 - \text{buffer}(\text{discrezionale})$ .

I dati raccolti non vengono immagazzinati in nessun luogo in ottica permanente, questo approccio porta con sé dei vantaggi e degli svantaggi:

- Non si hanno costi computazionali inerenti altre piattaforme quali server, domini ecc.
- Si ha un costo computazionale in termini di latency di compilazione nel momento dell'esecuzione del programma. Se si prende in considerazione lo *scaling* della piattaforma ovviamente i vantaggi sarebbero dalla parte dello *storing* di dati su strutture più complesse, ma considerando il fine accademico e di tesi non è il caso di complicare l'esperimento con lo sviluppo di strutture più complesse quali database, server e annessi.
- I dati aggiornati sono in via temporanea concernenti l'integrità della serie storica al contrario di una piattaforma fissa quale dovrebbe eseguire il download di solo i dati mancanti dall'ultima esecuzione al tempo della successiva. È chiaro come nel contesto attuale, avendo l'infinita possibilità di richiesta di dati a costo zero sia un vantaggio importante per lo sviluppo dell'applicativo, sarebbe diverso se per analogia si fosse presentata l'occasione di un costo fisso o variabile per mole di transazioni effettuate, rapportandosi al contesto del trading.

Avendo a mente la modalità attraverso il quale questi dati vengono ricevuti in input è stata creata una struttura che permette di interpretare i dati in modo semplice, specialmente dati strutturati in serie storiche, utilizzando una libreria di Python chiamata Pandas [24].

La richiesta di dati verso Investing può essere effettuata secondo diverse modalità, è possibile richiedere dati riguardanti la giornata di borsa, come le informazioni principali, settore ecc, oppure vi sono anche modalità secondo il quale è possibile ottenere i dati su periodizzazioni diverse, nel caso dell'elaborato è stata utilizzata una periodicità giornaliera, utilizzando come output la chiusura del giorno lavorativo precedente la richiesta. Vi è una precisazione da fare in questo contesto poiché, per fine puramente accademico è stato preso in esame solo il prezzo di chiusura del titolo sulla quale verrà calcolato il rendimento. Vi sono diverse modalità attraverso il quale un prezzo su *daily basis* possa essere opportunamente modificato per l'utilizzo su piattaforme di gestione del portafoglio, si parla di modifiche intra-giornaliere che tengano conto di misure di media per quanto riguarda il calcolo dei rendimenti [25] e di misura di autocorrelazione e covarianza per quanto riguarda la volatilità [26] [27].

Come annunciato precedentemente il processo di acquisizione dati terrà conto solo della chiusura relativa all'ultimo giorno di apertura del mercato precedente la richiesta, il codice si presenta come segue:

```
1 def getDailyFromStock(isin):
2
3     years_obs = timedelta(days=365.24) * 1           #~250 obs
4     endus = datetime.now()
5     endeu = endus.strftime("%d/%m/%Y")               #EU to US
6     date_conversion
7     startus = endus - years_obs
8     starteu = startus.strftime("%d/%m/%Y")
9
10    country_iniziali = isin[:2]                       #ISIN conversion
11    country = country_isin.get(country_iniziali)
```

```

12     periodicita = "Daily"                #time basis from API,
daily-weekly-monthly...

13

14     stock = inv.stocks.get_stocks(country = country)
15     info_gen = stock.loc[stock["isin"] == isin]
16     #info_tech = inv.stocks.get_stock_information(info_gen["
symbol"].values[0], country, as_json=False)
17     df = inv.get_stock_historical_data(stock = info_gen["
symbol"].values[0], country = country,
18         from_date = starteu, to_date = endeu, as_json=False,
order='ascending', interval = periodicita)
19
20     return df, info_gen, #info_tech

```

Listing 3.1: Chiamata ad Investing tramite API che restituisce i dati riguardante il titolo azionario ricercato tramite ISIN.

La funzione restituisce la struttura dati  $df$  e le informazioni generali riguardanti il titolo e viene ripetuto per ogni asset che si vuole inserire in portafoglio. In questo primo passo la discrezionalità del soggetto utilizzatore ha la scelta dei titoli inseribili in portafoglio utilizzando il codice ISIN<sup>1</sup> del titolo voluto. Successivamente nell'elaborato verrà introdotto un modo parallelo per la selezione dei titoli nella quale l'utente finale non avrà arbitrio nel *picking* ma tutto verrà lasciato al *machine learning* di base.

Il passo successivo riguarda la trasformazione dei dati nativi in una struttura più vicina all'utilizzo dell'algoritmo, questi verranno quindi “puliti” e calcolati i rendimenti. Il calcolo del rendimento viene effettuato tramite la formula:

$$r_{A_i} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Successivamente questi vengono opportunamente trasformati e riordinati per ottenere come output una matrice di  $r_{Obs} \times N$  con  $N$  il numero di titoli. Si ricorda che i rendimenti saranno sempre  $Obs - 1$  dato che è impossibile per

<sup>1</sup>È l'acronimo di International Securities Identification Number e rappresenta il codice identificativo dei valori mobiliari a livello internazionale.

via analitica calcolare il rendimento della prima osservazione della serie storica analizzata.

Ottenuti i rendimenti si ricava da questi la deviazione standard come misura di rischio implementandola successivamente a livello di portafoglio. Da questo punto, l'algoritmo di HRP è pressoché identico a come è stato presentato nella parte di elaborato precedentemente esposta, in fase finale vengono aggiunte funzioni di performance, ovvero viene mostrato a schermo quali sono le misure di rendimento di portafoglio, *standard deviation* e *Sharpe ratio* attese su base giornaliera. Si ricorda che per quanto riguarda l'applicativo il soggetto ha arbitrio sui titoli da inserire in portafoglio, sulla *rolling window* riguardante il computo della DCCA nel caso sia il metodo scelto per la stima della matrice di correlazione e sul tasso *risk free* adottato nel calcolo dell'indice di Sharpe.

Per quanto riguarda i titoli da inserire in portafoglio è necessario fare una precisazione, l'applicativo funge da intermediario tra una attenta e coscienziosa decisione di investimento verso precisi titoli in ex-ante il processo di investimento, quindi, in questa sezione non si propone come chiave di lettura per interpretare il mercato restituendo un output sulla migliore composizione di portafoglio tra le possibili dal mercato ma più semplicisticamente svolge una sorta di analisi fondamentale di base per una comprensione del portafoglio più integrale.

### 3.1 Algoritmo iterativo per massimizzazione SR

Per ovviare al problema del pick iniziale degli asset in portafoglio si è pensato ad una soluzione analitica di combinazioni e di massimizzazione. È necessario fare una precisazione in questo senso poiché chiaramente l'applicativo svolge solo una funzione sperimentale di ciò che sul mercato potrebbe essere controproducente. In parole più semplici, tenendo a mente una degli assiomi di efficienza per la quale non è possibile battere il mercato, non è

detto che sia una buona scelta affidarsi solamente e completamente ai dati quale strumento per la composizione migliore di portafoglio. Si ricorda in aggiunta che si tratta di una serie storica e che quindi la condizione di replica del passato nel futuro rimane condizionale, perciò i dati potrebbero rappresentare una situazione non reale.

Avendo a mente queste assunzioni l'applicativo è un'integrazione del HRP in approccio stock market, il ragionamento di fondo si basa sulla necessità di massimizzare lo *Sharpe Ratio* come misura standardizzata di performance del portafoglio. Per costruzione matematica le variabili su cui si può agire per aumentare la performance attesa di portafoglio sono, la volatilità e il rendimento di questo e il tasso risk free. È chiaro come non è possibile modificare il tasso risk free poiché dipendente dall'economia in generale, fermo restando la discrezionalità sulla struttura a termine utilizzata, ma in questo caso si tratta di una scelta tecnica quale solo l'operatore prenderà in causa. Le due variabili rimanenti possono essere modificate indirettamente poiché funzione degli asset in portafoglio. Utopicamente, l'investitore punta ad una minimizzazione del rischio o a una massimizzazione del rendimento, non è possibile ottenere entrambe le situazioni al di fuori della frontiera efficiente dei portafogli e alla propria funzione di utilità.

Come già annunciato l'algoritmo iterativo riprende le sembianze dell'HRP in approccio stock market nonostante in questo caso non sia possibile utilizzare come input la lista degli asset scelti, si ha invece una query verso Investing che permette di ricevere le informazioni riguardante tutto il mercato azionario americano. L'idea sottostante è quella di ripulire i dati di tali titoli ottenendo così osservazioni utilizzabili per il calcolo dei rendimenti. Il secondo passo riguarda la selezione di un numero di asset da detenere in portafoglio e di asset selezionabili dal mercato. L'algoritmo prevede la realizzazione di diversi array costituiti dalle possibili combinazioni non ripetute costituite da  $k$  elementi estratti in  $n$  oggetti distinti.

Una specificazione in tale senso riguarda l'impossibilità computazionale dell'esperimento su scala imponente dal momento che il mercato americano otte-

nibile tramite API e considerando le società con ISIN appartenente al mercato US, quindi con sede legale negli Stati Uniti, si hanno circa 4140 osservazioni. L'impossibilità tecnica non deriva dall'estrazione di dati da queste società quanto dal calcolo combinatorio per la quale, considerando un portafoglio ben diversificato con circa 30 titoli estratti da un pool di soli 100 titoli, le combinazioni semplici senza ripetizione sono  $2.9372339821610947e + 25$ . Per arginare in parte questo problema sarebbe possibile trovare un trade-off tra gruppi  $k$  ed oggetti  $n$ , ma in ogni caso, improbabile l'utilizzo su larga scala. Si pensi che anche in un contesto più piccolo come il mercato Italiano, i titoli estraibili dal mercato sarebbero circa 370, infattibile anche in questo caso. Prendendo coscienza di questa impossibilità tecnica si procede utilizzando un numero di oggetti inferiore, pari a 15 e un numero di elementi facenti parte il portafoglio pari a 10. Dato il contesto sperimentale vengono selezionati solo i primi 15 elementi in ordine alfabetico tra i titoli estratti dalla richiesta ad Investing. Si precisa che la selezione arbitraria dei primi titoli in ordine alfabetico viene identificato come un processo di selezione da parte del management per replicare fittiziamente una scelta quantomeno ponderata su titoli ipoteticamente più performanti o sottostanti vincoli quali faranno parte i 15 ( $n$ ) scelti dal mercato, per una selezione dei soli 10 ( $k$ ) migliori in relazione alla massimizzazione dello  $SR$ .

```
1 def getStocksIsin(n_asset_mkt):
2     df = pd.DataFrame(inv.get_stocks(country = country_isin.
3     get("US"))) #get infos about US market
4     df = np.array(df["isin"])
5     array_pulito = []
6
7     for i in df:
8         if i[:2] == "US":
9             array_pulito.append(i)
10
11     df = pd.DataFrame(array_pulito)
12     df.columns = ["Isin"]
```



```

12     head = df[:n_asset_mkt].values.tolist()           #get first
n_asset_mkt
13     head_lista = []
14     for i in head:
15         head_lista.append(i[0])
16     return head_lista #First 15obs from mkt

```

Listing 3.2: Selezione N asset dal mercato da richiamare tramite API.

Il passo seguente è il calcolo delle combinazioni ottenendo come output una lista di array contenenti i simboli degli asset estratti dalla query ed ovviamente i dati degli stessi. Ovviamente non è necessario fare la richiesta tramite API per ogni combinazione ma utilizzare i rendimenti già calcolati computando solo l'HRP e di conseguenza gli indici di portafoglio. Questi ultimi vengono inseriti all'interno di un array che una volta terminato il processo iterativo avrà la funzione di veicolo per la comparazione e quindi per la massimizzazione dello *Sharpe Ratio*.

```

1 def getWomboCombo(dataframe_returns, n_asset_port, riskfree,
    rolling_window, corr_est_method, linkage_method):
2
3     array_wombocombos = [] #array with best comb and infos(
hrp, std, rend, SR)
4     array_posizionale_sr = []
5     comb = list(combinations(dataframe_returns.columns,
n_asset_port))
6     comb = np.asarray(comb) #get combinations from
assets retrieved
7     for i in comb:
8
9         df_combo = dataframe_returns[i]
10        cov, corr = getCovCorr(df_combo, rolling_window,
corr_est_method)
11        hrp = getHRP(cov, corr, linkage_method)
12        std_p, rend_p, sr_p = getPortfolioInfos(hrp, cov,
df_combo, riskfree) #get infos about simulated portfolio
13        array_wombocombos.append(np.array([hrp, std_p, rend_p
, sr_p]))

```

```
14         array_posizionale_sr.append(np.array(sr_p))
15
16     return array_wombocombos, array_posizionale_sr
```

Listing 3.3: Calcolo delle combinazioni; HRP; e ricerca dello SR maggiore.

Lo step di massimizzazione avviene tramite una funzione di Python *argmax()* [28], questa permette di trovare il valore massimo di un array e di restituirne la posizione. In questo caso, restituendo la posizione si otterranno sia i dati riguardanti la composizione di portafoglio ovvero i pesi distribuiti per i vari asset, sia gli indici di performance e gestione.



## Capitolo 4

### Dati e metodologia

L'algoritmo di HRP termina nel momento in cui i pesi vengono assegnati ad ogni asset avendo quindi come output un vettore di pesi che il management si preoccuperà di inserire in portafoglio allocando il proprio capitale. Ovviamente vi sono alcune condizioni che dal punto di vista sperimentale non sono concordi alla realtà del mondo dei mercati. Si porta il lettore a pensare a condizioni come i costi di transizione, quali i costi fissi per la negoziazione, diretti al broker o piattaforma di trading e le tasse che porteranno in fase di cash-in una riduzione del reddito lordo accumulato dalla sessione di investimento. Per quanto riguarda condizioni propriamente più tecniche relative ai titoli, non viene presa in considerazione l'opportunità di shorting del titolo, vincolo già introdotto precedentemente, per la quale i pesi assegnati ad un determinato asset non possono essere negativi pur rispettando il peso totale di portafoglio  $w_p = 1$ .

Non viene considerato il pagamento dei dividendi, questo potrebbe inficiare sul rendimento del titolo poiché componente accessoria nella formula classica per il calcolo dello stesso. L'inclusione dei dividendi all'interno del calcolo dei rendimenti deve tenere conto anche delle diverse frequenze di pagamento, nel caso americano 2 volte all'anno, quindi opportunamente scontati e inclusi nella formula.

Un'altra condizione riguarda la divisibilità dei titoli, risolvibile in parte ap-

prossimando la quota di peso assegnato per il prezzo del titolo ad un valore realmente investibile. Per portare un esempio concreto si pensi ad un peso assegnato all'asset  $A$  per un valore di 500 quando il titolo quota a 750 per azione. In questo caso non sarebbe possibile assegnare il valore risultante dall'algoritmo all'asset richiesto.

Una condizione riguardante il calcolo dello *Sharpe Ratio*, prendendo in considerazione il tasso risk free, questo potrebbe essere dedotto a discrezione dell'utilizzatore o implementando una funzione che prenda in input il tasso di un certo bond statale, o i tassi OIS. In ogni caso è necessario ponderare bene il tasso utilizzato come risk free poiché gli asset in portafoglio potrebbero essere molto diversificati, quindi non è semplice avere un termine di paragone come reward del mercato verso il titolo privo di rischio.

Per quanto riguarda l'approccio *Out of Sample* verrà utilizzato il metodo Monte Carlo per effettuare delle simulazioni in riguardo all'algoritmo di HRP e IVP comparando il rischio dei rispettivi portafogli così creati. In questo caso vi saranno diverse variabili da osservare:

- La sliding window varierà sempre nell'ordine di 250, 125 e 50 osservazioni, malgrado le osservazioni totali generate secondo una normale siano 520, una ulteriore finestra di *back testing* verrà generata nella simulazione di Monte Carlo includendo 260 osservazioni.  
Il numero di osservazioni facenti parte della sliding window rappresenterà in entrambe le analisi rispettivamente l'integrità delle osservazioni  $\sim 250$ ,  $1/2$  e  $1/5$  di queste.
- Il portafoglio verrà ribilanciato dopo determinati periodi per simulare un vero portafoglio e quindi un costante impegno dati i cambiamenti del mercato, soprattutto in funzione degli shock indotti nel mercato. Il rebalancing sarà effettuato su base giornaliera, settimanale e mensile, ovvero ogni 1, 5 e 22 giorni.

Le variabili sotto la quale viene effettuata l'analisi *Out of Sample* sono state descritte ma vi è una precisazione riguardante il fatto che questo processo

verrà iterato un numero di volte pari a 1000, costituendo quindi i presupposti per una simulazione di Monte Carlo. Sotto questo presupposto vi sarà il confronto tra IVP e HRP per quanto riguarda il rischio di portafoglio attraverso l'utilizzo dei metodi proposti.

Per quanto riguarda la fase di testing, si è deciso per l'approccio *In Sample* l'utilizzo di diverse sliding windows descritte nella sezione 2.1 con valori rispettivamente di 250, 125, e 50 osservazioni attraverso il quale viene modificata la finestra di tempo su cui la DCCA è calcolata. Le osservazioni ricevute in input saranno corrispondenti ad un anno di apertura di borsa, ovvero circa 252 osservazioni assimilabili a circa 251 rendimenti giornalieri di ogni asset inserito in portafoglio. Verrà comparato successivamente il metodo di stima della matrice di correlazione normale e DCCA sotto il profilo della composizione, del rendimento atteso, rischio e sharpe ratio. Verranno illustrate anche le diverse composizioni di portafoglio tra le varie sliding window all'interno del metodo DCCA.

Si introducono anche delle costanti per l'analisi sopra descritta, il metodo di clustering utilizzato è quello di *ward* e nel caso dell'analisi *In Sample* viene utilizzato un tasso risk free pari a  $-0.00586$  ricavato dalla curva dei tassi LIBOR a 1 giorno. Viene scelta questa frequenza per il tasso risk free poiché si ricorda che tutta l'analisi è condotta sotto il profilo giornaliero per la quale, dal punto di vista simulativo, si immagina un costante impegno da parte del management per la migliore performance del portafoglio mentre per un aspetto di coerenza dei rendimenti su base giornaliera si esprimeranno le variazioni e le performance come risultato della giornata di borsa. Per fare chiarezza, un titolo che esprime su base giornaliera una media dei rendimenti pari a  $0.00245$  ovvero  $0.245\%$  sarà interpretato come una variazione attesa del prezzo dell'azione pari a tale importo. La trasformazione analitica dei rendimenti su base giornaliera ad una frequenza maggiore può essere espressa tramite formula per l'interesse composto [29].

$$r_n = (1 + r_{daily})^n - 1, \text{ con } n \text{ frequenza obbiettivo}$$

## 4.1 Out of Sample Analysis

Le simulazioni di Monte Carlo hanno confermato alcuni degli obiettivi prefissati. Sotto il profilo del rebalancing di portafoglio vengono analizzati i due algoritmi IVP e HRP in termini percentuali tramite la formula:

$$\bar{\sigma}_{IVP}^2 \% = \frac{\sigma_{IVP}^2}{\sigma_{HRP}^2 - 1}$$

I risultati sono,  $\bar{\sigma}_{IVP_1}^2 \% = \frac{\sigma_{IVP_1}^2}{\sigma_{HRP_1}^2 - 1} = 0.34933 = 34.9\%$ ,  $\bar{\sigma}_{IVP_5}^2 \% = \frac{\sigma_{IVP_5}^2}{\sigma_{HRP_5}^2 - 1} = 0.39076 = 39\%$  e  $\bar{\sigma}_{IVP_{22}}^2 \% = \frac{\sigma_{IVP_{22}}^2}{\sigma_{HRP_{22}}^2 - 1} = 0.30831 = 30.8\%$ .

Da questo primo test si evince che la massima differenza tra varianze di portafoglio è ottenuta con un rebalancing settimanale. L'interpretazione in riguardo a questo aspetto può essere assimilabile al fatto che IVP soffrendo maggiormente dagli shock di volatilità non riesca a ribilanciare i pesi degli asset in portafoglio in modo opportuno, ne discende che la varianza aumenta registrando un valore più elevato rispetto all'algoritmo di HRP di quasi il 40%. Dal lato opposto si prende in considerazione il ribilanciamento mensile dove l'IVP sottoperforma del 30.8% rispetto all'HRP. In un orizzonte temporale di ribilanciamento più esteso l'IVP tende a sottoperformare meno rispetto a periodi di ribilanciamento più corti. Questo aspetto può essere interpretato nella capacità dell'IVP di assimilare lo shock del periodo intercorrente tra ogni ribilanciamento e di distribuire la volatilità recente in modo più omogeneo.

| Rapporto varianze IVP-HRP       |       |       |        |
|---------------------------------|-------|-------|--------|
| Rebalancing                     | 1     | 5     | 22     |
| Varianza media rapporto         | 34.9% | 39.1% | 30.8%  |
| Delta riferiti al rebalancing 1 |       | 11.9% | -11.7% |

Tabella 4.1: Viene mostrata la tabella riguardante i rapporti per la comparazione dell'algoritmo IVP e HRP.

Un altro test è stato effettuato tra i diversi periodi di rebalancing in relazione alle finestre utilizzate nell'HRP. Comparando le varianze ottenute

dall'utilizzo delle diverse finestre  $SW$  si nota una differenza importante per quanto riguarda la  $SW250$ . Questa registra una varianza per il rebalancing giornaliero pari a  $\sigma_{HRP_1}^2 = 0.08126 = 8\%$ , in relazione a questa si ha per i due periodi successivi una riduzione del  $\sigma_{HRP_5}^2 \% = \frac{\sigma_{HRP_5}^2 - \sigma_{HRP_1}^2}{\sigma_{HRP_1}^2} = -13\%$  per il rebalancing settimanale e  $\sigma_{HRP_{22}}^2 \% = \frac{\sigma_{HRP_{22}}^2 - \sigma_{HRP_1}^2}{\sigma_{HRP_1}^2} = -15\%$  per il rebalancing mensile. Le variazioni per la finestra  $SW250$  sono circa 4 volte più imponenti rispetto alle variazioni delle altre finestre sempre in relazione ai periodi di ribilanciamento. In questo senso si ha quindi una riduzione importante di varianza tra periodi per la finestra  $SW250$  con l'aumentare del periodo di ribilanciamento malgrado non registri la varianza di portafoglio più bassa in termini assoluti.

Si può interpretare la riduzione della varianza dell'HRP su periodi di rebalancing più estesi rispetto al giornaliero nello stesso modo dell'algoritmo IVP. Lo shock di volatilità registrato invece che ponderarlo e distribuirlo in una visione di minimo locale viene distribuito in modo più omogeneo quando ponderato per un periodo di stress più lungo sfruttando il vantaggio delle correlazioni con gli asset. In questo modo il peso di portafoglio verrà tendenzialmente indirizzato verso gli asset meno correlati a quelli sotto shock.

| Tipo corr | Window (SW) | Varianza tra Ribilanciamenti |        |  |        |        |              |        |        |              |
|-----------|-------------|------------------------------|--------|--|--------|--------|--------------|--------|--------|--------------|
|           |             | 1                            |        |  | 5      |        |              | 22     |        |              |
|           |             | IVP                          | HRP    |  | IVP    | HRP    | $\Delta HRP$ | IVP    | HRP    | $\Delta HRP$ |
| Normale   | Nessuna     | 0.0948                       | 0.0690 |  | 0.0914 | 0.0653 | -5%          | 0.0964 | 0.0732 | 6%           |
|           | 250         | 0.1058                       | 0.0813 |  | 0.0955 | 0.0705 | -13%         | 0.0887 | 0.0687 | -15%         |
|           | 125         | 0.0956                       | 0.0708 |  | 0.0967 | 0.0680 | -4%          | 0.0907 | 0.0691 | -2%          |
|           | 50          | 0.0976                       | 0.0712 |  | 0.0946 | 0.0682 | -4%          | 0.0904 | 0.0690 | -3%          |
| Media     |             | 0.0985                       | 0.0731 |  | 0.0945 | 0.0680 | -7%          | 0.0916 | 0.0700 | -4%          |

Tabella 4.2: Tabella riportante la varianza dei risultati in termini di delta e media. Le variazioni sono espresse rispetto al rebalancing giornaliero nella colonna azzurra.

In ultimo viene comparato il metodo DCCA e il metodo normale di costruzione della matrici di correlazione. Sotto questo punto di vista l'unica



piccola differenza si nota sotto il rebalancing mensile. Il metodo normale in questo caso registra un aumento della varianza al contrario delle finestre DCCA che registrano una riduzione.

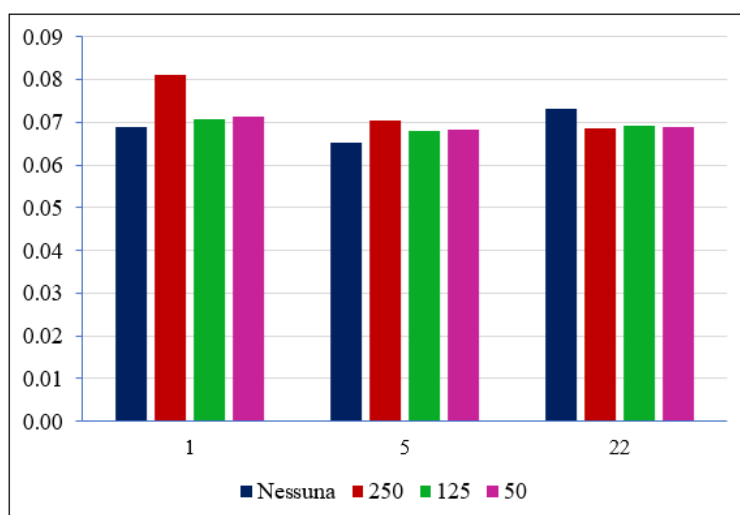


Figura 4.1: In questo grafico vengono mostrate le varianze del portafoglio costruito secondo HRP in base al metodo DCCA e normale.

Come si può notare dalla figura 4.1 le differenze sono veramente minime. Si osserva un aumento della varianza per il metodo normale per quanto riguarda il ribilanciamento mensile ma negli altri casi questa è inferiore a qualsiasi altra finestra DCCA.

L'analisi di questi dati può essere interpretata sotto due aspetti. Il primo riguarda il fatto che il metodo classico potrebbe sottostimare il rischio dato dalla correlazione degli asset. In questo senso viene giustificata la minore varianza del portafoglio. Si ricorda però che per costruzione della correlazione, la media delle osservazioni nell'orizzonte considerato potrebbe considerare degli outliers, in questo modo verrebbe sottostimata l'effettiva varianza di portafoglio computata secondo metodo classico.

L'aspetto opposto invece considera il metodo DCCA più realistico secondo il quale lo stesso orizzonte di osservazioni viene scomposto in sottogruppi e analizzato. Attraverso questo processo si potrebbe annullare in parte l'ef-

fetto outlier sulla media riportando l'effettiva situazione tramite matrice di correlazione.

## 4.2 In Sample Analysis

L'analisi con dati del mercato ottenuti dalle API è stata svolta con modalità diverse rispetto a quella con simulazioni di Monte Carlo, infatti in questo caso IVP e HRP verranno confrontati solo sotto l'aspetto della composizione di portafoglio mostrando i pesi dei vari asset. Successivamente verrà analizzato l'HRP sotto il punto di vista del metodo di correlazione DCCA con le finestre  $SW$  e tramite il metodo normale.

È necessario fare una precisazione in riguardo alla composizione del portafoglio nei diversi test sulle correlazioni, in input all'algoritmo sono stati passati solo gli ISIN dei titoli e considerando che le composizioni finali non sono modificabili dall'investitore, le variazioni nei pesi dei titoli sono variazioni significative nell'interpretare il comportamento dell'algoritmo dietro le letture delle correlazioni e delle variabili di mercato.

Per dare un'idea chiara al lettore si presenta il grafico delle composizioni di portafoglio tra l'algoritmo classico HRP e IVP.

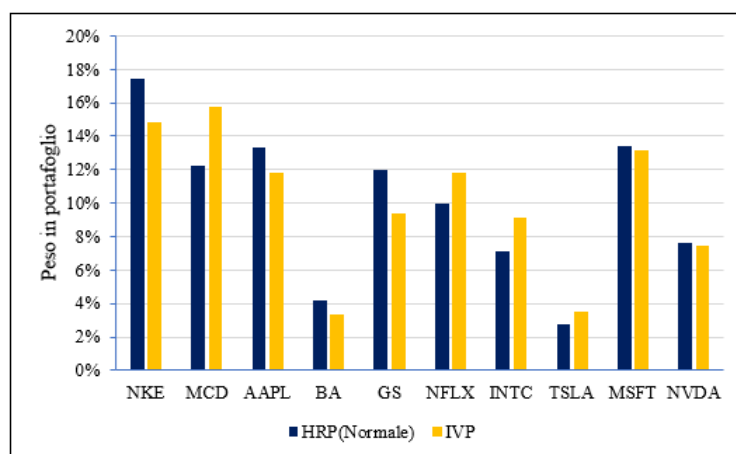


Figura 4.2: Grafico rappresentante le diverse composizioni di portafoglio tra HRP metodo normale e IVP.

Si nota in questo caso la differenza tra i due algoritmi nell'allocare il peso dei diversi asset nonostante la serie storica sia la stessa, e quindi gli stessi rendimenti. Il punto focale è che nell'algoritmo di HRP viene presa in considerazione anche la matrice di correlazione come espressione di possibile rischio tra asset.

In secondo luogo viene introdotto nell'analisi anche il metodo HRP-DCCA con le diverse finestre.

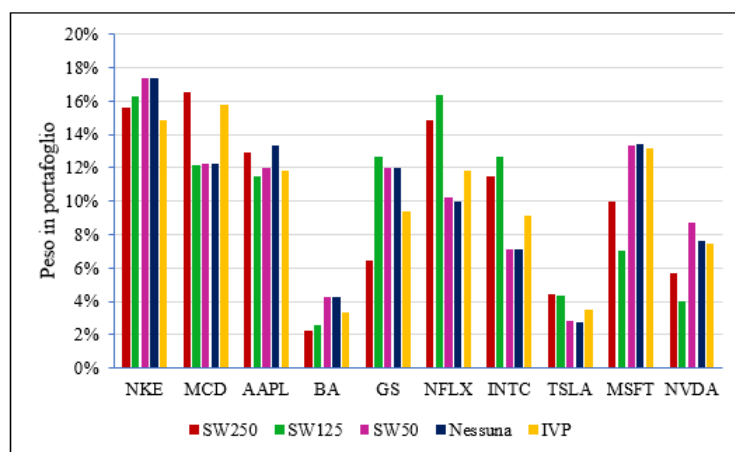


Figura 4.3: Grafico rappresentante le diverse composizioni di portafoglio sia tra diversi algoritmi che intra HRP utilizzando il DCCA.

Nella figura 4.3 si notano diverse composizioni assegnate dall'algoritmo HRP al variare delle finestre utilizzate nella computazione del DCCA. Attraverso questa figura si può dedurre che il metodo DCCA vada a catturare valori di correlazione tra asset diversi in base alla finestra utilizzata scorrendo con il buffer di diverse grandezze. Questi pesi diversamente assegnati sono originati dal processo di clustering ordinato e successivamente dalla quasi-diagonalizzazione e bipartizione.

Viene riportata la matrice con i cluster degli asset ordinati in modalità diverse dall'algoritmo a sostegno del fatto che l'algoritmo di linkage abbia, secondo il metodo *ward*, organizzato i cluster in modo diverso per ogni finestra.

| Cluster ordinati HRP |       |       |      |         |
|----------------------|-------|-------|------|---------|
| N° asset             | SW250 | SW125 | SW50 | Nessuna |
| 1                    | NKE   | NFLX  | GS   | BA      |
| 2                    | MCD   | INTC  | BA   | GS      |
| 3                    | AAPL  | NKE   | NKE  | NKE     |
| 4                    | BA    | MCD   | MCD  | MCD     |
| 5                    | GS    | BA    | INTC | INTC    |
| 6                    | NFLX  | GS    | AAPL | NVDA    |
| 7                    | INTC  | TSLA  | MSFT | MSFT    |
| 8                    | TSLA  | AAPL  | NVDA | AAPL    |
| 9                    | MSFT  | MSFT  | NFLX | NFLX    |
| 10                   | NVDA  | NVDA  | TSLA | TSLA    |

Tabella 4.3: Viene mostrato l'ordine degli asset in base alle diverse modalità di stima matrice di correlazione.

La varianza usata come metodo per l'assegnazione del peso e l'ordine degli asset post clusterizzazione portano l'algoritmo di HRP a diverse composizioni di peso in portafoglio e a diverse performance.

| HRP Investing |         | E(Rp)  | STD     | VAR     | SR     |
|---------------|---------|--------|---------|---------|--------|
| Normale       | Nessuna | 0.0024 | 0.02724 | 0.00074 | 0.3021 |
|               | 250     | 0.0016 | 0.02586 | 0.00067 | 0.2878 |
| DCCA          | 125     | 0.0024 | 0.02590 | 0.00067 | 0.3180 |
|               | 50      | 0.0022 | 0.02711 | 0.00073 | 0.2992 |

Tabella 4.4: Viene mostrata la tabella comprendente le misure di performance dell'HRP *In Sample*.

Come si può notare dalla tabella 4.4 il rendimento di portafoglio più alto è ottenuto attraverso l' $\bar{R}_{Normale} = 0.0024 = 0.24\%$  e  $\bar{R}_{125} = 0.0024 = 0.24\%$ , osservando la deviazione standard di portafoglio invece si nota che  $\sigma_{Normale} = 0.02724$  è la più alta tra le deviazioni standard confermando l'ottima performance della finestra *SW125* con  $\sigma_{125} = 0.02590$  che si aggiudica anche il migliore sharpe ratio  $SR_{125} = 0.3180$ .

Una considerazione sui risultati ottenuti riguarda i valori molto simili tra

loro e considerando la costruzione dello  $SR$ , la deviazione standard di portafoglio si trova al denominatore e quindi anche una piccola differenza tra i valori delle  $SW$  può portare questo indice a preferire una finestra come  $SW125$ . Fatta questa precisazione è opportuno specificare che, stando alla lettura degli strumenti utilizzati fin ora, l'approccio al metodo DCCA porta un beneficio rispetto alla stima della correlazione normale nell'analisi *In Sample*.

Si ricorda al lettore che i risultati *In Sample* potrebbero non avere valenza generale in quanto non facenti parte di un metodo simulativo come potrebbe esse Monte Carlo ma prelevati dal mercato in modo discrezionale.

## Capitolo 5

### Spunti per lavori futuri

Un possibile sviluppo potrebbe riguardare la rigorosità tecnica di un approccio direzionato alla praticità di business, con questo si intende che la commercializzazione dell'idea sottostante l'elaborato potrebbe essere implementata solo aggiungendo diverse qualità tecniche e prese in considerazione diverse variabili. L'inserimento di API riguardanti per esempio il book di negoziazione per la quale la piattaforma non si sostituisce ad un applicativo di trading ma potrebbe collaborare con esso per ricevere informazioni riguardanti per esempio, il prezzo effettivo di incrocio tra domanda e offerta. In questo modo non si avrebbe come punto di acquisto il dato prezzo giornaliero ricevuto da Investing ma la sostituzione di tale con il prezzo ricevuto dalla contrattazione API del book di negoziazione.

Uno spunto riguardante l'algoritmo di  $MAX(SR)$  potrebbe essere come annunciato in tale sezione una preselezione di determinati titoli scelti dal management per poi essere inseriti nell'algoritmo di massimizzazione dello sharpe ratio. In questo modo si potrebbe ovviare ad indecisioni sui titoli risolvendo in modo analitico per le performance attese degli stessi.

L'imposizione di vincoli diversi dai pesi positivi potrebbe essere imposta all'algoritmo in modo da trovare altri portafogli efficienti che prevedano lo shorting del titolo. Tenendo conto del fatto che su molte piattaforme di trading questa modalità di investimento è permesso potrebbe dunque rivelarsi

una grande opportunità per l'investitore singolo o per la società.

Nell'analisi del portafoglio potrebbe essere opportuno utilizzare altri indici di performance, per esempio potrebbe essere inserita una simulazione di loss del portafoglio con una minimizzazione del VAR ad un determinato percentile sensibile agli obiettivi del management.

Si potrebbe tenere conto della distribuzione dei rendimenti dei titoli in modo da formare una simulazione di Monte Carlo con annessi indici di valutazione per ottimizzare le funzioni di  $MAX(SR)$  o  $MIN(VAR)$ . Si pensi che la distribuzione simulata tramite MC in questo elaborato è considerata normale, ma dal momento che i titoli nel mercato spesso non hanno distribuzioni normali sarebbe opportuno tenerne conto per il fit delle simulazioni sulla distribuzione osservata nella serie storica. Aprendo una parentesi su tale questione, sarebbe possibile trovare una sorta di opportunità di arbitraggio nella differenza tra le valutazioni svolte dal mercato basate sulla distribuzione normale e da una valutazione sull'effettiva distribuzione dei rendimenti. L'assunzione di una normale standard è molto forte e in alcuni casi potrebbe portare importanti vantaggi, si pensi per esempio alle Opzioni scritte sul valore del sottostante, il *payout* di queste viene settato analizzando la probabilità che il sottostante superi una certa soglia della distribuzione normale trovando le fasce di prezzo atteso su cui le opzioni sono scritte. Uno scostamento significativo da tali soglie per approssimazione teorica mediante l'utilizzo di una distribuzione diversa potrebbe portare un beneficio.

L'implementazione di altri strumenti nella piattaforma sarebbe un notevole sviluppo soprattutto da un punto di vista di diversificazione degli strumenti rientranti nel portafoglio. L'adattamento a tale modifica dovrebbe riguardare anche il sistema di input per la quale l'ISIN non sarebbe più chiave univoca nella selezione di tutti gli strumenti. Da un punto di vista amatoriale sarebbe molto soddisfacente combinare tramite algoritmo HRP diversi strumenti per ottenere una massimizzazione di un determinato indice, anche costruito appositamente quale potrebbe essere un'equazione sottostante a determinati parametri di performance e di composizione.

# Conclusioni

Lo scopo della tesi è di trovare una giustificazione nell'utilizzo del metodo migliore tra IVP e HRP per la gestione del portafoglio, in secondo luogo è stata proposta una modifica nella metodologia di stima della matrice di correlazione utilizzando la *Detrended Cross Correlation Analysis* per stimare i coefficienti e successivamente costruire la matrice che verrà analizzata attraverso diverse finestre per trovare variazioni significative. In ultimo è stato proposto un metodo di analisi di portafoglio più concreto con titoli estratti dal mercato azionario americano e come sviluppo di tale metodo un'implementazione che mira alla massimizzazione di un coefficiente di performance per ottenere la migliore composizione di portafoglio HRP.

L'analisi delle differenze tra i due algoritmi è risultata positiva per l'algoritmo di HRP che guadagna una riduzione nella varianza di portafoglio confrontata all'IVP pari a  $\bar{\sigma}_{HRP(1)}^2 = -34.9\%$ ,  $\bar{\sigma}_{HRP(5)}^2 = -39.1\%$  e  $\bar{\sigma}_{HRP(22)}^2 = -30.8\%$ . Per quanto riguarda il metodo DCCA si nota una differenza tra le varianze delle finestra *SW250* in confronto al metodo normale e alle restanti finestre. Questa registra variazioni più imponenti di circa 4 volte registrando una riduzione della varianza con l'incrementarsi del periodo di rebalancing. In generale il metodo DCCA non esprime un vantaggio in termini di varianza di portafoglio in confronto al metodo normale, questo risultato è interpretabile sotto due punti di vista, una sottostima delle correlazioni da parte del metodo classico oppure una presa di coscienza da parte del metodo DCCA restituendo effettivamente le correlazioni tra asset durante l'orizzonte temporale considerato. Per quanto riguarda l'analisi *In Sample* si nota una



evidente differenza in termini di composizioni di portafoglio sia tra i due algoritmi IVP e HRP, sia intra algoritmo HRP con le diverse finestre. L'idea sottostante risiede nell'effettiva descrizione delle correlazioni tra asset. A sostegno di questo si vede performare la finestra  $SW_{125}$  presentando il migliore  $SR_{125} = 0.3180$ .

# Elenco delle figure

|     |  |    |
|-----|--|----|
| 1.1 | Esempio di distribuzione dei rendimenti del titolo NFLX, presenta una skewness pari a $-0.005563$ . . . . .  | 15 |
| 1.2 | Viene proposto un esempio di comparison tra la distribuzione del titolo NFLX kurtosis = 2.0935 e il titolo TSLA kurtosis = 2.5978 . . . . .  | 16 |
| 1.3 | Viene rappresentata graficamente la varianza di portafoglio al variare del numero di asset inseriti. . . . .   | 20 |
| 2.1 | Heatmap della matrice di correlazione tra 10 asset generati secondo distribuzione normale. Cinque di questi asset sono arbitrariamente correlati agli altri cinque tramite una funzione di addizione alla distribuzione normalizzata con drift = $\sigma = 0.25$ . . . . . | 26 |
| 2.2 | Un esempio di heatmap della matrice di correlazione tra 10 assets selezionati dal mercato US. . . . .  | 27 |
| 2.3 | Esempio di dendrogramma rappresentante i cluster formati attraverso l'algoritmo di linkage. I 10 asset sono gli stessi scelti arbitrariamente anche nella figura 2.2. . . . .  | 32 |

- 2.4 Viene rappresentato il processo di quasi diagonalizzazione a livello numerico. Ogni matrice corrisponde ad uno step che viene svolto iterativamente. In verde sono evidenziati gli asset selezionati in base al clustering e poi ordinati e restituiti come lista. I numeri sono la sostituzione analitica dei simboli dei titoli azionari. Una proprietà riguardante questo processo è il fatto che tutti gli asset evidenziati risiedono sopra l'antidiagonale della matrice. . . . . 35
- 4.1 In questo grafico vengono mostrate le varianze del portafoglio costruito secondo HRP in base al metodo DCCA e normale. . 56
- 4.2 Grafico rappresentante le diverse composizioni di portafoglio tra HRP metodo normale e IVP. . . . . 57
- 4.3 Grafico rappresentante le diverse composizioni di portafoglio sia tra diversi algoritmi che intra HRP utilizzando il DCCA. . 58

# Elenco delle tabelle

|     |   |    |
|-----|---|----|
| 2.1 | Esempio di matrice di distanze tra asset selezionati dal mercato.   | 31 |
| 2.2 | Viene mostrato l'algoritmo di bipartizione ricorsiva per quanto riguarda i titoli azionari selezionati in modo discrezionale. Da notare come ad ogni step dell'algoritmo i cluster si separano ottenendo un valore di peso specifico per ogni titolo. . . . . | 39 |
| 4.1 | Viene mostrata la tabella riguardante i rapporti per la comparazione dell'algoritmo IVP e HRP. . . . .  | 54 |
| 4.2 | Tabella riportante la varianza dei risultati in termini di delta e media. Le variazioni sono espresse rispetto al rebalancing giornaliero nella colonna azzurra. . . . .  | 55 |
| 4.3 | Viene mostrato l'ordine degli asset in base alle diverse modalità di stima matrice di correlazione. . . . .   | 59 |
| 4.4 | Viene mostrata la tabella comprendente le misure di performance dell'HRP <i>In Sample</i> . . . . .   | 59 |



# Bibliografia

- [1] Lopez De Prado. "building diversified portfolios that outperform out of sample". 2016.
- [2] Eugene Fama. "efficient capital market: a review of theory and empirical work". 1970.
- [3] H.M. Markowitz. "portfolio selection". 1952.
- [4] Raman Uppal Victor DeMiguel, Lorenzo Garlappi. "how inefficient is the  $1/n$  asset-allocation strategy?". 2006.
- [5] Adam Hayes. Cross-correlation. "<https://www.investopedia.com/terms/c/crosscorrelation.asp>", 2020.
- [6] G.F. Zebende. "dcca cross correlation coefficient: Quantifying level of cross-correlation". 2010.
- [7] P.J. Ferreira G.F. Zebende E. Guedes, A. Dionisio. "dcca cross correlation in blue-chips companies: A view of the 2008 financial crisis in the eurozone". 2017.
- [8] G.F. Zebende E. Guedes. "dcca cross coefficient with sliding windows approach". 2019.
- [9] H.E. Stanley D. Horvatic, B.Podobnik. "detrended cross-correlation analysis for non-stationary time series with periodic trends". 2011.

- [10] H. Stanley A. Carbone, G.Castelli. "time-dependent hurst exponent in financial time series". 2004.
- [11] P.J. Ferreira G.F. Zebende E. Guedes, A. Dionisio. "a sliding windows approach to analyse the evolution of bank shares in the euopean union". 2018.
- [12] H.J. Ruskin A.A. Sharkasi M. Crane J.A. Matos, S.M. Gama. "time and scale husrt exponent for financial markets". 2008.
- [13] Clustering distance measures. <https://www.datanovia.com/en/lessons/clustering-distance-measures/>.
- [14] scipy.cluster.hierarchy.linkage. "https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html", 2020.
- [15] Hierarchical clustering/dendrogram. "https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Hierarchical\_clustering – Dendrograms.pdf".
- [16] Stephanie. Hierarchical clustering / dendrogram: Simple definition, examples. "https://www.statisticshowto.com/hierarchical-clustering/", 2016.
- [17] Thomas Raffinot. "hierarchical clustering-based asset allocation". 2017.
- [18] Pierre Legendre Fionn Murtagh. "ward's hierarchical clustering method: Clustering criterion and agglomerative algorithm". 2011.
- [19] Shashi Jain Prayut Jain. "can machine learning-based portfolios outperform traditional risk-based portfolios? the need to account for covariance misspecification". 2019.
- [20] Enrico Schumann. "take-the-best in portfolio selection". 2013.
- [21] Vincent Lapointe Bertrand, Philippe. "risk-based strategies: The social responsibility of investment universes does matter". 2018.

- [22] What is an api? (application programming interface).  
”<https://www.mulesoft.com/resources/api/what-is-an-api>”.
- [23] Investing.com. ”<https://www.investing.com/>”.
- [24] Pandas. ”<https://pandas.pydata.org/>”.
- [25] Asger Lunde Hansen, Peter R. ”a forecast comparison of volatility models: Does anything beat a garch(1, 1)?”. 2005.
- [26] Martin Martens. ”measuring and forecasting sp 500 index-futures volatility using high-frequency data”. 2002.
- [27] Eugenie Hol Uspensky Koopman, Siem Jan. ”the stochastic volatility in mean model: Empirical evidence from international stock markets”. 2002.
- [28] Argmax function from numpy. ”<https://numpy.org/doc/stable/reference/generated/numpy.argmax.html>”.
- [29] Motley Fool Staff. How to convert daily returns to annual returns. ”<https://www.fool.com/knowledge-center/how-to-convert-daily-returns-to-annual-returns.aspx>”, 2016.