



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

Ottimizzazione di portafogli: Confronto tra modelli quadratici e lineari

Relatore: Prof. Domenico Salvagnin

Laureando: Filippo-Massimo Niero

ANNO ACCADEMICO 2022-2023

Data di laurea - 29/09/2023

Abstract

La portfolio optimization è un problema centrale nel campo della finanza e ha lo scopo di allocare gli asset in maniera ottimale per massimizzare il rendimento atteso, tenendo conto anche del livello di rischio associato. Uno dei primi modelli, creato da Harry Markowitz nel 1952, è conosciuto come “mean-variance model”. Successivamente sono state introdotte altre misure di rischio, che hanno portato alla creazione di modelli lineari. Nella tesi vengono presentate le formulazioni matematiche dei modelli di Markowitz e Conditional Drawdown, evidenziando le loro differenze. Vengono poi condotti degli esperimenti utilizzando dati storici del mercato azionario per confrontare i rendimenti e le caratteristiche di rischio-rendimento dei portafogli generati dai due modelli. I risultati degli esperimenti, come ad esempio le performance dei portafogli e i tempi di esecuzione dei modelli, vengono analizzati e confrontati. Questo confronto fornisce una prospettiva sulle prestazioni dei modelli lineari rispetto al modello di Markowitz, valutando la loro validità come strumento per l’ottimizzazione di portafoglio in un contesto finanziario. In sintesi, la tesi si concentra sull’ottimizzazione di portafoglio attraverso modelli lineari e quadratici, fornendo una panoramica storica, una presentazione dei modelli e una valutazione comparativa delle loro prestazioni.

Indice

Abstract	1
1 Introduzione	3
1.1 Portfolio Optimization	3
2 Modern Portfolio Theory	5
2.1 Concetti generali	5
2.2 Il modello di Markowitz	8
3 Modelli lineari	10
3.1 Scenario e discretizzazione	10
3.2 Misure lineari	11
3.3 Modello Conditional Drawdown	12
4 Dati e campionamento	14
5 Analisi sperimentale	17
5.1 Prestazioni computazionali	17
5.2 Simulazioni out of sample	22
5.2.1 Analisi portfolio generati il 01/01/2020	23
5.2.2 Analisi portfolio generati il 15/06/2022	26
6 Conclusioni	30
6.1 Limitazioni e possibili miglioramenti	30
6.2 Conclusioni	31
Bibliografia	32

Capitolo 1

Introduzione

In questo capitolo vengono introdotti importanti concetti e termini riguardanti la Portfolio Optimization.

1.1 Portfolio Optimization

Investire il capitale in eccesso in modo ottimale è un problema comune per aziende e individui che richiede una soluzione specifica e personalizzata. La portfolio optimization fornisce uno strumento utile ai *portfolio manager* per prendere decisioni basandosi sulle informazioni correnti e sulle aspettative future.

Un portfolio è un insieme di asset, investimenti finanziari che comprendono azioni, obbligazioni, liquidità, materie prime e altri strumenti. Per valutare la performance dell'asset si guarda al suo prezzo di mercato (*market price*), il prezzo a cui viene comprato o venduto, determinato dall'incontro tra domanda e offerta.

Si considera il problema di costruire un portfolio dato un certo capitale e un insieme di asset disponibili (*available assets*), cioè di scegliere la quantità di capitale da allocare in ogni asset (*share*).

La *Portfolio optimization* è il processo con il quale si seleziona il portfolio migliore sulla base di obiettivi predefiniti, generalmente si punta a massimizzare il rendimento minimizzando i rischi. Ogni tecnica di ottimizzazione usa diverse misure come funzioni obiettivo e presenta vantaggi e svantaggi.

Bisogna poi tenere conto di alcuni parametri che modificano la scelta del portfolio ottimale, come la tolleranza al rischio, l'orizzonte temporale e lo scopo dell'investimento che cambiano da investitore a investitore. È possibile dividere i portfolio in aggressivi, difensivi, da rendita,

speculativi e ibridi [4] .

Il rischio specifico è l'esposizione di un particolare asset ad alcuni eventi che lo influenzano, mentre il rischio sistematico è un rischio che influenza tutti gli asset.

Due asset si dicono correlati se i loro prezzi tendono a muoversi nella stessa direzione nello stesso periodo. Tendenzialmente gli asset nello stesso settore sono molto correlati tra loro e condividono alcuni rischi specifici.

La diversificazione sta alla base del processo di ottimizzazione: i rischi specifici degli asset possono compensarsi a vicenda, in questo modo è possibile abbassare il livello di rischio complessivo del portfolio, riducendo appunto il rischio specifico. Il rischio sistematico non può essere mitigato con la diversificazione, quindi non è nemmeno possibile eliminare totalmente il rischio. L'ottimizzazione mirerà quindi a trovare la combinazione di asset che, per un determinato livello di rischio ritenuto accettabile, minimizzi la componente specifica del rischio.

Questa tesi tratterà di modelli che seguono la strategia *buy and hold*, nella quale gli asset vengono comprati in un momento preciso, detto *investment time* con l'obiettivo di massimizzare il valore del portfolio a un certo *target time*. Non è consentito effettuare alcuna operazione dopo l'*investment time*.

Capitolo 2

Modern Portfolio Theory

Nel 1952, il premio Nobel Harry Markowitz pubblica un articolo di rilievo nel *Journal Of Finance* [7] che getta le basi della *Modern Portfolio Theory* (MPT).

La MPT fornisce il fondamento concettuale per la costruzione di portafogli ottimali in modo efficiente, tenendo in considerazione le restrizioni e gli obiettivi specifici degli investitori.

Nel suo lavoro, Markowitz per la prima volta formalizza i concetti di rischio e di ritorno atteso usando un linguaggio matematico e propone un primo modello di ottimizzazione, usando la varianza come *risk measure*. Negli anni successivi sono poi stati introdotti diversi modelli che usano *risk measure* diverse, come verrà approfondito nella sezione "Misure lineari".

2.1 Concetti generali

Parte delle formule presenti in questo elaborato provengono dalle analisi di Mansini, Ogryczak e Speranza presenti in "Linear and Mixed Integer Programming for Portfolio Optimization" [6].

Prima di analizzare il modello proposto da Markowitz, si espone un modello generico per la portfolio optimization, in questo modo vengono introdotte diverse variabili che saranno utili in seguito.

Si indica l'insieme dei potenziali asset su cui investire con $N = \{1, 2, \dots, n\}$ e con x_j il peso in percentuale dell'asset $j = 1, \dots, n$. Ogni possibile vettore di pesi $\mathbf{x} = (x_j)_{j=1, \dots, n}$ rappresenta un diverso portfolio e deve soddisfare alcuni vincoli: il primo banalmente indica che non è possibile investire più del capitale a disposizione:

$$\sum_{j=1}^n x_j = 1 \tag{2.1}$$

Inoltre si impone che non sia possibile effettuare la vendita allo scoperto:

$$x_j \geq 0 \quad \forall j = 1, \dots, n \quad (2.2)$$

È possibile aggiungere altri vincoli a x_j per rispettare dei requisiti specifici, come per esempio un peso minimo o massimo per ogni asset, oppure garantire di non essere troppo esposti in un determinato settore o area geografica. In ogni caso tutti questi vincoli sono lineari. Un portfolio \mathbf{x} è detto *feasible* se rispetta tutti questi vincoli. Si definisce \mathcal{Q} come l'insieme dei *feasible portfolios*.

Per valutare la performance storica di un asset si usa il rendimento

$$r_t = (q_{t2} - q_{t1})/q_{t1}$$

dove q_t è la quotazione al tempo t , mentre $t1$ e $t2$ sono rispettivamente il tempo iniziale e finale del periodo che ci interessa. In generale dovremmo includere anche eventuali dividendi pagati nel periodo, ma per semplicità in questa tesi saranno ignorati.

Al fine di stimare la performance futura, data l'incertezza dei rendimenti, il rendimento dell'asset j viene modellato con la variabile aleatoria R_j , di media $\mu_j = \mathbb{E}\{R_j\}$.

A ogni portfolio si associa quindi una variable

$$R_{\mathbf{x}} = \sum_{j=1}^n R_j x_j$$

e si chiama ritorno atteso del portfolio

$$\mu(\mathbf{x}) = \mathbb{E}\{R_{\mathbf{x}}\}$$

Uno degli obiettivi dell'investitore è massimizzare il ritorno atteso, l'altro è proteggere il suo investimento dall'incertezza.

L'incertezza del rendimento viene misurata con una *risk measure* associata ad ogni portfolio $\varrho(\mathbf{x})$. Modelli diversi usano funzioni diverse per $\varrho(\mathbf{x})$. Generalmente $\varrho(\mathbf{x})$ vale 0 nel caso del *risk-free portfolio*, e un numero positivo per ogni altro portfolio. Si definiscono queste misure *shift independent dispersion measures*.

Un altro modo per esprimere l'incertezza del rendimento è dato dalle *safety measure*, che esprimono la probabilità di raggiungere un determinato rendimento. Le *safety measure* sono indicate con $\mu(\mathbf{x}) - \varrho(\mathbf{x})$

Come sarà esaminato in seguito, ogni *risk measure* ha una *safety measure* corrispondente e viceversa. Quindi è sufficiente modellare il problema della *portfolio optimization* usando le *risk measure*, il problema in questo caso viene chiamato *mean-risk*:

$$\max\{[\mu(\mathbf{x}), -\varrho(\mathbf{x})] : \mathbf{x} \in Q\} \quad (2.3)$$

Come si può notare, nel *mean risk problem* sono presenti due obiettivi che sono in conflitto tra loro, non è possibile trovare un portfolio che abbia un alto rendimento atteso senza avere anche un alto livello di rischio. Per questo ci interessa trovare l'insieme dei portfolio *efficienti*, starà poi all'investitore, in base alla sua propensione al rischio, sceglierne uno. Un portfolio \mathbf{x}_0 è detto μ/ϱ -*efficient* se non esiste alcun portfolio $\mathbf{x} \in Q$ tale che $\mu(\mathbf{x}) \geq \mu(\mathbf{x}_0)$ e $\varrho(\mathbf{x}) \leq \varrho(\mathbf{x}_0)$. Nessun investitore razionale dovrebbe accettare un portfolio che non sia μ/ϱ -*efficient*.

È possibile creare un grafico, chiamato *risk/return frontier* che rappresenta il ritorno atteso più alto ottenibile per ogni livello di rischio. La figura 2.1 rappresenta un esempio di questo grafico. L'insieme di portfolio efficienti compone la *efficient frontier*, evidenziata in figura. La regione al di sotto della *efficient frontier* rappresenta l'insieme dei *feasible portfolios*. Ogni punto rappresenta un portfolio realizzabile. Nello stesso grafico si possono evidenziare anche i singoli asset. La regione al di fuori della parabola rappresenta lo spazio dei portfolio che non è possibile costruire.

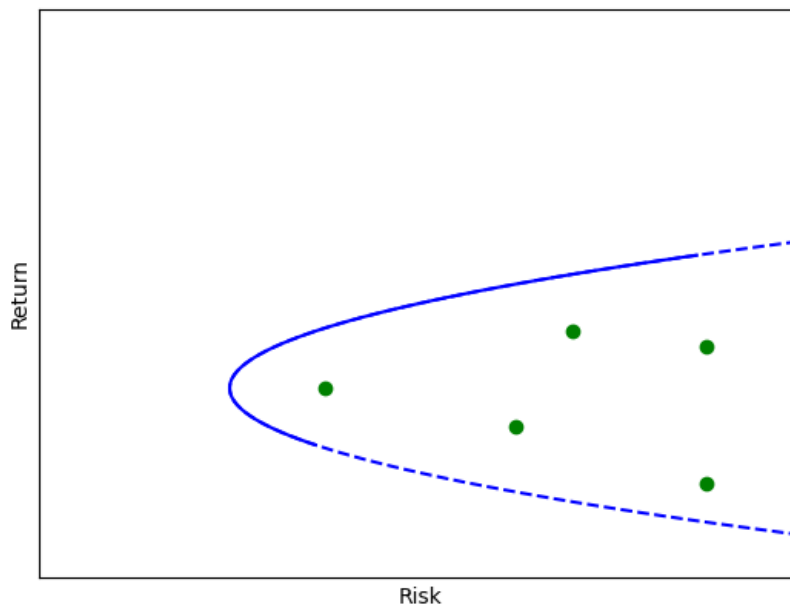


Figura 2.1: In blu l'efficient frontier, in verde i singoli asset

Dall'immagine è evidente che la efficient frontier sia sempre al di sopra e a sinistra di ogni

asset preso singolarmente. Questo è dovuto ai benefici della diversificazione, in quanto combinando asset scorrelati tra loro, il rendimento è la media dei loro rendimenti attesi, mentre il rischio è minore della media dei rischi[13]. Inoltre è possibile notare la legge dei rendimenti decrescenti: dopo un certa soglia, il rendimento aumenta sempre più lentamente a mano a mano che il rischio aumenta.

Per risolvere un problema a due criteri, il metodo più semplice è il *bounded approach*, dove si impone un lower bound μ_0 sul ritorno atteso. Facendo variare μ_0 si ottengono diversi portfolio efficienti. Questo è anche un modo di parametrizzare la efficient frontier.

$$\min\{\varrho(\mathbf{x}) : \mu(\mathbf{x}) \geq \mu_0 \mathbf{x} \in Q\} \quad (2.4)$$

2.2 Il modello di Markowitz

Come anticipato in precedenza, Markowitz sceglie di usare la varianza come *risk measure*.

La varianza di un asset j è definita come

$$\sigma_j^2 = \mathbb{E}\{R_j - \mathbb{E}\{R\}^2\} \quad (2.5)$$

Sfruttare la correlazione tra asset ci permette di diminuire il rischio del portfolio. C'è bisogno di calcolare la covarianza tra i vari asset.

La covarianza σ_{ij} tra due asset i e j è data da:

$$\sigma_{ij} = \mathbb{E}\{(R_i - \mu_i)(R_j - \mu_j)\} \quad (2.6)$$

Normalizzando la covarianza si ottengono i coefficienti di correlazione $\bar{\rho}_{ij}$

$$\bar{\rho}_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (2.7)$$

Questi coefficienti possono essere interpretati così:

- $\bar{\rho}_{ij} = +1$ se i due asset si sono sempre mossi nella stessa direzione
- $\bar{\rho}_{ij} = 0$ se i due asset si sono mossi in maniera indipendente tra loro
- $\bar{\rho}_{ij} = -1$ se i due asset si sono sempre mossi nella direzione opposta

La varianza dell'intero portfolio \mathbf{x} è:

$$\sigma^2(\mathbf{x}) = \sum_{i=1}^n \sigma_i^2 x_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \sigma_{ij} x_i x_j \quad (2.8)$$

Per semplificare la scrittura, la varianza dell'asset i viene indicata con σ_{ii} . La varianza quindi diventa

$$\sigma^2(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \quad (2.9)$$

Usando quindi la varianza come funzione obiettivo, e ponendo un lower bound μ_0 sul ritorno atteso, si ottiene la forma classica del modello di Markowitz:

$$\begin{aligned} \min \quad & \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \\ & \sum_{j=1}^n \mu_j x_j \geq \mu_0 \\ & \sum_{j=1}^n x_j = 1 \\ & x_j \geq 0 \quad j = 1, \dots, n \end{aligned}$$

Come si può notare, questo è un problema quadratico, dato che la risk measure, e quindi la funzione obiettivo, è quadratica. Dopo Markowitz, sono state introdotte diverse misure di rischio lineari, in quanto la LP solvability è richiesta per alcune applicazioni pratiche nel mondo reale, come ad esempio i costi di transazione. Provare ad aggiungere questi vincoli direttamente a questo modello produce un problema *Mixed-Integer Nonlinear Programming*, che è NP-Hard.

Nel capitolo seguente si analizzano alcune di queste misure lineari e verrà introdotta la *Conditional Drawdown* per il confronto con il modello di Markowitz.

Capitolo 3

Modelli lineari

I modelli di ottimizzazione lineare offrono diversi vantaggi: innanzitutto, gli algoritmi e i software di risoluzione per i modelli lineari sono più evoluti e affidabili. In secondo luogo, i modelli lineari si caratterizzano per una notevole flessibilità: è infatti molto semplice combinare tra loro vincoli e funzioni obiettivo lineari per ottenere modelli ibridi che si adattino al problema specifico; inoltre, come anticipato nel capitolo precedente, è possibile aggiungere dei vincoli molto utili nel mondo reale. Ignorare, ad esempio, i costi di transazione potrebbe portare a non avere un portfolio veramente efficiente.

In questo capitolo si introducono alcune accortezze necessarie per i modelli lineari di portfolio optimization, successivamente si forniscono alcuni esempi di misure lineari, da cui sarà ricavata la *Conditional Drawdown*, *risk measure* che verrà usata nelle simulazioni.

3.1 Scenario e discretizzazione

Per avere delle *risk/safety measure* lineari è necessario discretizzare i rendimenti, usando il concetto degli *scenario*. Uno *scenario* è una realizzazione dell'insieme delle variabili casuali R_j , cioè un insieme dei possibile valori al *target time*. Ogni scenario $t = 1, \dots, T$ può capitare con una certa probabilità p_t , con $\sum_{t=1}^T p_t = 1$.

Si definisce con r_{jt} la realizzazione dell'asset j nello scenario t . Lo scenario t è quindi definito completamente dall'insieme dei r_{jt} per ogni $j = 1, \dots, n$ e dalla sua probabilità p_t .

Il ritorno atteso dell'asset j può essere calcolato come $\mu_j = \sum_{t=1}^T p_t r_{ij}$.

Può essere interessante definire e visualizzare la funzione di distribuzione dei ritorni discretizzati di ogni asset, in quanto dovrebbe essere simile a quella della variabile aleatoria di partenza R_j . Possono essere richiesti numerosi scenari per avere una buona approssimazione.

Generalmente le distribuzioni dei rendimenti degli asset tendono a seguire una gaussiana.

Con y_t viene indicato il rendimento del portfolio \mathbf{x} nello scenario t

$$y_t = \sum_{j=1}^n r_{jt} x_j \quad (3.1)$$

e il rendimento atteso del portfolio $\mu(\mathbf{x})$

$$\mu(\mathbf{x}) = \mathbb{E}\{R_{\mathbf{x}}\} = \sum_{t=1}^T p_t y_t = \sum_{t=1}^T p_t \left(\sum_{j=1}^n r_{jt} x_j \right) = \sum_{j=1}^n x_j \sum_{t=1}^T p_t r_{jt} = \sum_{j=1}^n \mu_j x_j \quad (3.2)$$

Si dice che i ritorni sono *discretized* se sono definiti dalle loro realizzazioni nei vari scenari. Nel caso di ritorni *discretized* e se il modello è in forma lineare, si dice che la *risk/safety measure* utilizzata nel modello è *LP computable*.

3.2 Misure lineari

I primi tentativi di avere un modello lineare per la portfolio optimization consistevano nel cercare di approssimare la varianza, attraverso funzioni lineari a tratti.

Successivamente vennero introdotte delle *risk measure* lineari, prime tra tutte la *Gini's mean difference* (GMD), definita come la media del valore assoluto della differenza dei rendimenti dei portfolio y_t nei vari scenario:

$$\Gamma(\mathbf{x}) = \frac{1}{2} \sum_{t'=1}^T \sum_{t''=1}^T |y_{t'} - y_{t''}| p_{t'} p_{t''} \quad (3.3)$$

Γ è *LP computable*, e ha un numero di vincoli proporzionale a T , cosa che rende la risoluzione molto veloce.

Un'altra *risk measure* che è stata introdotta è la *Mean Absolute Deviation* (MAD), che considera la media del valore assoluto della differenza tra la variabile e il suo valore atteso. Può quindi essere interpretata come una versione della varianza che utilizza la distanza assoluta anziché quella quadratica.

$$\delta(\mathbf{x}) = \mathbb{E}\{|R_{\mathbf{x}} - \mathbb{E}\{R_{\mathbf{x}}\}|\} \quad (3.4)$$

Dalla definizione di MAD è possibile derivare una sua formulazione lineare.

Sia usando la MAD che la GMD, nel caso in cui i ritorni siano normalmente distribuiti, condizione che rappresenta una buona approssimazione della realtà per quanto riguarda alcune asset class, la misura di rischio diventa proporzionale alla varianza. Ciò implica che il nuovo modello produrrebbe un risultato identico al modello di Markowitz.

Sono state introdotte anche alcune *safety measure*, la più semplice tra tutte la *worst realization*, in cui si vuole massimizzare il rendimento al caso peggiore:

$$M(\mathbf{x}) = \min_{t=1, \dots, T} y_t \quad (3.5)$$

Per generalizzare questa misura si introduce il concetto usato in economia del quantile. Dato $\beta \in [0, 1]$, il β -quantile di una variabile aleatoria R è quel valore q tale che:

$$\mathbb{P}\{R < q\} \leq \beta \leq \mathbb{P}\{R \leq q\}$$

Il quantile è una misura statistica ampiamente utilizzata in ambito finanziario per rappresentare il valore della perdita massima prevista, a un determinato livello di probabilità o confidenza (*Value at Risk, VaR*).

Per generalizzare $M(\mathbf{x})$, si usa la *Conditional VaR* $M_\beta(\mathbf{x})$, definita come il rendimento atteso del portfolio calcolato a un certo livello di confidenza $\beta \in [0, 1]$ dei peggiori scenari.

Se gli scenari sono equiprobabili e si sceglie $\beta = k/T$, $M_\beta(\mathbf{x})$ è la media dei rendimenti nei k casi peggiori:

$$M_{\frac{k}{T}}(\mathbf{x}) = \frac{1}{k} \sum_{s=1}^k y_{t_s} \quad (3.6)$$

Con t_1, \dots, t_k i k scenari peggiori per il rendimento del portfolio.

È possibile generalizzare $M_\beta(\mathbf{x})$ con $\beta \neq k/T$,

$$M_\beta(\mathbf{x}) = \min_{u_t} \left\{ \frac{1}{\beta} \sum_{t=1}^T y_t u_t : \sum_{t=1}^T u_t = \beta, 0 \leq u_t \leq p_t, t = 1, \dots, T \right\} \quad (3.7)$$

All'ottimo $u_t = 0$ per ogni scenario t non inclusi nell'insieme degli scenari peggiori, $u_t = p_t$ per gli scenari inclusi nell'insieme degli scenari peggiori, e in un solo caso $0 < u_t < p_t$.

La definizione di M_β è lineare, ma il corrispondente problema di portfolio optimization non lo è. Per ovviare a questo problema si usa il duale della formulazione del CVar, aggiungendo la variabile η senza vincoli, che prende il posto dell'equazione $\sum_{t=1}^T u_t = \beta$ e le variabili d_t^- che servono come upper bounds di u_t . Si ottiene quindi:

$$M_\beta(\mathbf{x}) = \max_{\eta, d_t^-} \left\{ \eta - \frac{1}{\beta} \sum_{t=1}^T p_t d_t^- : d_t^- \geq \eta - y_t, d_t^- \geq 0, t = 1, \dots, T \right\} \quad (3.8)$$

3.3 Modello Conditional Drawdown

La CVar è una *safety measure*, come anticipato nel capitolo "Concetti generali", è possibile passare da *risk measure* a *safety measure* e viceversa. La *Conditional Drawdown* $\Delta_\beta(\mathbf{x}) =$

$\mu(\mathbf{x}) - M_\beta(\mathbf{x})$ è la *risk measure* corrispondente alla CVar:

$$\Delta_\beta(\mathbf{x}) = \min \left\{ \sum_{j=1}^n \mu_j x_j - \eta + \frac{1}{\beta} \sum_{t=1}^T d_t^- p_t : d_t^- \geq \eta - y_t, d_t^- \geq 0, t = 1, \dots, T \right\} \quad (3.9)$$

Si introduce la variabile $d_t^+ = d_t^- - \eta + y_t$, il modello completo diventa:

$$\begin{aligned} & \min \sum_{t=1}^T (d_t^+ + \frac{1-\beta}{\beta} d_t^-) p_t \\ & d_t^- - d_t^+ = \eta - y_t && t = 1, \dots, T \\ & y_t = \sum_{j=1}^n r_{jt} x_j && t = 1, \dots, T \\ & \sum_{j=1}^n \mu_j x_j \geq \mu_0 \\ & d_t^-, d_t^+ \geq 0 \\ & \mathbf{x} \in Q \end{aligned}$$

Nella simulazione verrà ripetuto l'esperimento usando valori di β diversi. Questo ci consentirà di esaminare attentamente l'impatto di β sulla generazione dei portafogli ottimali. Inoltre, sarà di interesse valutare se vi siano differenze significative nei tempi di esecuzione al variare di questa soglia, in modo da comprendere meglio la relazione tra il parametro β e le prestazioni dell'algoritmo di ottimizzazione del portafoglio.

Capitolo 4

Dati e campionamento

Nei capitoli precedenti sono stati introdotti i due modelli con cui verranno effettuate le simulazioni: Markowitz, o Mean-Variance (MV) e Conditional Drawdown (CDD). L'efficacia di questi modelli è legata alla qualità e alla quantità dei dati disponibili. In particolare è necessario creare degli scenari che rappresentino i rendimenti degli asset.

Il processo di creazione degli scenari è chiamato *scenario generation*, e i metodi specifici per ottenere gli scenari sono detti *scenario generation techniques*. La maggior parte di queste tecniche, originariamente concepite per altre applicazioni, sono comunque valide o facilmente adattabili anche per i problemi di portfolio optimization. [Vedi 3, p. 21]

Tra le varie *scenario generation techniques* si prendono in considerazione le seguenti:

- *Historical data*. Assume che i rendimenti e le correlazioni del passato si ripetano anche nel futuro; solitamente ogni scenario è considerato equiprobabile. È uno dei metodi più semplici, la difficoltà sta nel reperire abbastanza dati storici per avere un numero di scenari sufficiente. In ogni caso, la performance passata di un asset non dà alcuna garanzia sul futuro di quell'asset.
- *Bootstrapping technique*. Questa tecnica effettua diversi campionamenti dei dati storici, fino a raggiungere un numero di scenari prefissato, così da creare istanze di qualsiasi dimensione. Questa tecnica viene spesso usata se il numero di dati storici è minore del numero di scenari richiesto e prevede di dover mantenere la correlazione tra i vari asset.
- *Block Bootstrapping technique*. È una variante della precedente, in cui si cerca di preservare oltre alla correlazione per singolo scenario, anche una correlazione del periodo usando blocchi di scenari. Un blocco è un insieme temporalmente continuo di scenari. È necessario scegliere una lunghezza del blocco, e a ogni campionamento viene preso un

blocco contiguo della lunghezza scelta, fino a ottenere il numero di scenari richiesto.

Per condurre le simulazioni, è stato scelto l'indice americano *Standard & Poor's 500*, un indice ponderato per capitalizzazione di mercato che rappresenta una selezione delle 500 più grandi società quotate nelle borse statunitensi. È ampiamente considerato uno dei migliori indicatori delle prestazioni del mercato azionario statunitense e, di conseguenza, ha un'enorme rilevanza nell'ambito finanziario e della ricerca.

Lo script per scaricare i dati storici può essere trovato nella repository di questa tesi [9] nel file `scripts/download_data.py`. I *ticker* vengono letti da Wikipedia usando BeautifulSoup [12] e Request [11] e viene usato il pacchetto `yfinance`[10] per scaricare i dati giornalieri di ogni ticker. I grafici sono stati creati usando `Matplotlib`[8]

Sono stati scaricati tutti i dati degli ultimi 10 anni ¹, se disponibili, in quanto non tutte le aziende presenti attualmente nell'indice sono state quotate per tutta la durata del periodo richiesto. In particolare, molte aziende prese in esame hanno una data di inizio diversa. La soluzione più comune, e quella che verrà adottata in questo elaborato, è di troncare i dati fino ad avere la stessa data iniziale per tutti gli asset presi in considerazione.

Questa non è la scelta ottimale in quanto, come espresso da Stambaugh [14], questo riduce la possibilità di catturare importanti relazioni tra gli asset che potrebbero non essere apparenti in periodi più brevi. Nell'articolo viene proposto un metodo che sfrutta le reti Bayesiane per stimare i possibili valori mancanti. Questo esula dagli obiettivi della presente tesi, e si è deciso comunque di applicare un troncamento ai dati.

Per non ridurre ulteriormente il periodo, si è deciso invece di escludere tutti gli asset che non hanno uno storico sufficiente, abbassando il numero di asset a 470.

Nel contesto della generazione di scenari, è importante fare una distinzione tra *in-sample data*, cioè i dati che vengono usati nel modello, e *out-of-sample data*, cioè i dati che vengono usati solo per valutare le performance del modello.

Per valutare l'efficacia di un modello di portfolio optimization si effettuano diversi *backtest*, in cui si sceglie un *investment time t*, si crea un portfolio con un modello usando solo dati disponibili prima di *t*, e successivamente si valuta come si sarebbe comportato il portfolio generato usando dati reali nel periodo successivo a *t*. In questo caso i dati *in-sample* saranno tutti i dati precedenti a *t*, e i dati *out-of-sample* i rendimenti effettivi degli asset nel periodo successivo a *t*.

Per poter procedere a confrontare i modelli, inoltre, è necessario valutare come calcolare la

¹Sono stati scaricati i dati storici dal 01/09/2013 al 01/09/2023.

matrice della covarianza per il modello di Markowitz: visto che i ritorni sono stati resi discreti, e gli scenari equiprobabili è possibile semplificare la formula:

$$\sigma_{ij} = \frac{1}{T} \sum_{t=1}^T (r_{it} - \mu_i)(r_{jt} - \mu_j) \quad (4.1)$$

L'analisi sarà composta da due parti: nella prima saranno valutati le performance computazionali dei due modelli, esaminando tempo di esecuzione e memoria utilizzata da ciascun modello, nella seconda sarà valutata l'efficacia dei modelli eseguendo dei *backtest*.

Per l'analisi delle prestazioni non c'è bisogno di *out-of-sample data*, quindi viene applicata la tecnica *bootstrap* su tutti i rendimenti storici e si generano istanze via via crescenti. Nell'analisi dei rendimenti invece verrà usata sia la tecnica *bootstrap* che *block bootstrap* per valutare se quest'ultima porta dei vantaggi nei *backtest*. Mantenendo un periodo di un anno come dati *out-of-sample*, saranno effettuati confronti dei rendimenti dei portfolio generati dai modelli con l'indice S&P500 stesso. L'obiettivo sarà quello di verificare se i portafogli ottimizzati tramite i modelli oggetto di studio siano in grado di ottenere performance superiori rispetto al benchmark di mercato.

In questo capitolo, sono state introdotte le tecniche di generazione di scenari. Nel prossimo capitolo si procederà con l'analisi delle prestazioni e dell'efficacia dei modelli.

Capitolo 5

Analisi sperimentale

In questo capitolo verranno presentati i risultati delle simulazioni condotte per valutare e mettere a confronto le performance dei due modelli di ottimizzazione studiati: il modello di Markowitz e il modello Conditional Drawdown.

Sono stati calcolati i rendimenti giornalieri di un sottoinsieme dell'indice S&P500 usando i prezzi di apertura e chiusura nel periodo che va dal 01/09/2013 al 01/09/2023. I dividendi sono stati ignorati.

Tutte le simulazioni sono state effettuate in una macchina dotata di processore AMD Ryzen 7 Pro 3700U, 2.3 GHz con 8 core.

Per realizzare ed eseguire i modelli è stato usato CPLEX [1] versione 22.10, un potente e flessibile risolutore di problemi di ottimizzazione. In particolare i modelli sono stati realizzati usando le librerie di C++.

5.1 Prestazioni computazionali

Nella prima parte della analisi l'obiettivo è valutare i tempi d'esecuzione e la memoria utilizzata dai 2 modelli. Per fare questo sono state create delle istanze chiamate *performance_XXXXXX* il cui numero di scenari aumenta esponenzialmente. Gli scenari sono stati generati usando la tecnica del *bootstrapping* su tutto il periodo preso in considerazione.

La simulazione sarà effettuata sul modello di Markowitz e 6 volte sul modello CDD, con valori di β diversi, per verificare se questo parametro impatta il tempo di esecuzione.

Per diminuire il tempo di I/O, nei file di output creati sono state salvate solo le informazioni riguardanti il tempo di preprocessing, il tempo totale e una stima della memoria utilizzata, non il risultato del modello in sé.

Tempi

Per valutare i tempi si è deciso di effettuare due simulazioni distinte: nella prima si fissa un certo rendimento target μ_0 mentre si aumenta il numero di scenari, mentre nella seconda si mantiene il numero di scenari fisso e vengono simulati 100 μ_0 diversi, andando da 0 al rendimento massimo possibile per quell'insieme di asset. Per fare questo il modello dovrà essere eseguito 100 volte.

Per tutte le simulazioni è stato poi imposto un tempo limite dell'esecuzione del modello di 100 secondi, una volta passata questa soglia il programma termina.

Nella prima simulazione si è fissato $\mu_0 = 5\%$ (annualizzato), $\beta = \{0.05, 0.1, 0.2, 0.5, 0.9, 1.0\}$, e il numero di scenari T parte da 10 e raddoppia a ogni simulazione.

Nella tabella 5.1 sono mostrati i tempi di preprocessing. Si può notare come in Markowitz, aumenta circa linearmente con l'aumentare di T . Questo lungo tempo di precalcolo è dovuto al calcolo della matrice di covarianza.

T	MV	CDD 0	CDD 1	CDD 2	CDD 3	CDD 4	CDD 5
10	0.05	0.0	0.01	0.0	0.0	0.0	0.0
20	0.09	0.0	0.00	0.0	0.0	0.0	0.0
40	0.16	0.01	0.01	0.01	0.01	0.01	0.01
80	0.29	0.01	0.01	0.01	0.01	0.02	0.01
160	0.57	0.03	0.03	0.02	0.02	0.03	0.03
320	1.06	0.04	0.04	0.04	0.05	0.04	0.05
640	2.12	0.09	0.09	0.09	0.09	0.09	0.1
1280	4.30	0.20	0.19	0.21	0.35	0.42	0.44
2560	8.10	0.35	0.33	0.35	0.37	0.4	0.55
5120	16.27	0.73	0.86	0.91	0.93	0.95	0.92
10240	36.17	1.68	1.79	1.75	1.73	1.63	1.67
20480	69.83	3.25	3.17	2.89	3.03	3.04	3.07

Tabella 5.1: Tempi di precalcolo, in secondi

Si nota poi come il modello CDD abbia bisogno di una frazione di tempo rispetto al MV per il precalcolo. Date le dimensioni delle istanze, il tempo per leggere il file di input inizia a non essere più trascurabile. Inoltre, è possibile osservare che il tempo di precalcolo non è impattato significativamente da variazioni di β .

Dalla tabella 5.2 e dal grafico 5.1 è possibile vedere che anche il tempo totale, per entrambi i modelli, cresce linearmente all'aumentare del numero degli scenari. Markowitz sembra avere

una pendenza inferiore nella crescita, anche se per risolvere istanze piccole richiede più tempo.

T	MV	CDD 0	CDD 1	CDD 2	CDD 3	CDD 4	CDD 5
10	0.33	0.04	0.01	0.01	0.01	0.01	0.01
20	0.27	0.01	0.01	0.01	0.01	0.01	0.01
40	0.52	0.06	0.03	0.03	0.03	0.03	0.02
80	0.57	0.14	0.07	0.06	0.05	0.07	0.04
160	0.81	0.20	0.17	0.14	0.15	0.17	0.09
320	1.31	0.38	0.36	0.32	0.31	0.34	0.17
640	2.40	0.95	0.94	0.85	0.94	0.83	0.49
1280	4.60	4.12	4.29	5.94	5.85	4.1	1.96
2560	8.39	7.98	8.72	10.26	9.28	9.04	3.89
5120	16.62	20.82	20.11	18.12	18.45	14.92	7.82
10240	36.57	38.04	35.1	37.38	32.68	26.7	19.48
20480	70.36	59.54	56.5	55.02	54.26	53.78	42.18

Tabella 5.2: Tempi totali, in secondi

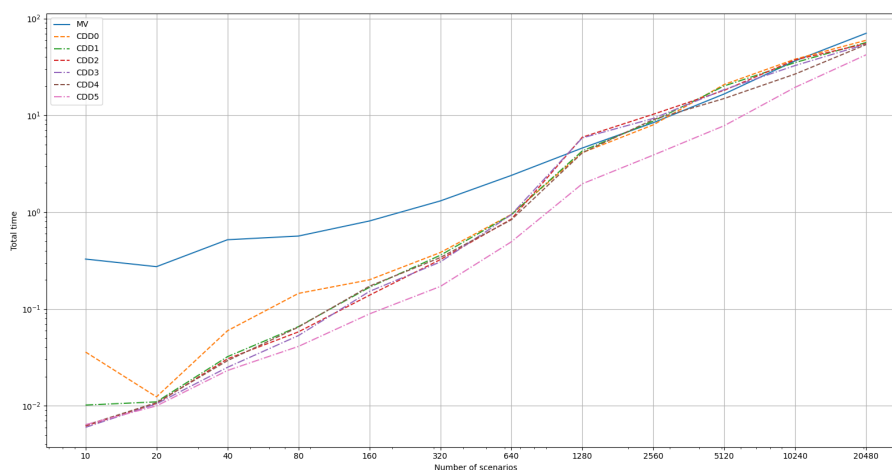


Figura 5.1: Tempi totali, in secondi

Confrontando la tabella 5.2 con la 5.1 per quanto riguarda il modello MV si può notare come la maggior parte del tempo impiegato non sia dovuta al modello ma bensì al precalcolo.

In ogni caso, aumentando il numero di asset ci si aspetterebbe un aumento quadratico della complessità e del costo computazionale per Markowitz, cosa che non succede per il modello Conditional Drawdown.

Si sono viste alcune simulazioni per risolvere un problema di portfolio optimization, dato il rendimento minimo fissato. In realtà però il problema iniziale della portfolio optimization prevederebbe di trovare la soluzione ottima per ogni valore di μ_0 . È stata svolta un'altra simulazione con lo scopo di valutare le performance dei modelli nel caso in cui fosse necessario risolvere completamente il problema, usando 100 valori di μ_0 .

Per questa seconda simulazione è stata usata l'istanza *performance_002560*. Dal grafico 5.2 è possibile notare come il tempo di esecuzione per un singolo valore μ_0 del modello di Markowitz è molto inferiore rispetto agli altri. Questo è dovuto al fatto che, una volta costruita la matrice di covarianza la prima volta, è possibile riutilizzarla anche per le istanze successive. Si può inoltre notare come i vari β in realtà modificano i tempi. Un β più alto porta a considerare solo un sottoinsieme degli scenari, scartando quelli peggiori, e quindi tempi di computazione leggermente inferiori. Infine nel grafico si nota un rapido declino nei tempi di esecuzione intorno a $\mu_0 = 0.22$. Pretendendo un rendimento così alto, il numero di asset tra cui è possibile scegliere per realizzare il portfolio ottimo scende e quindi si trova più velocemente una soluzione.

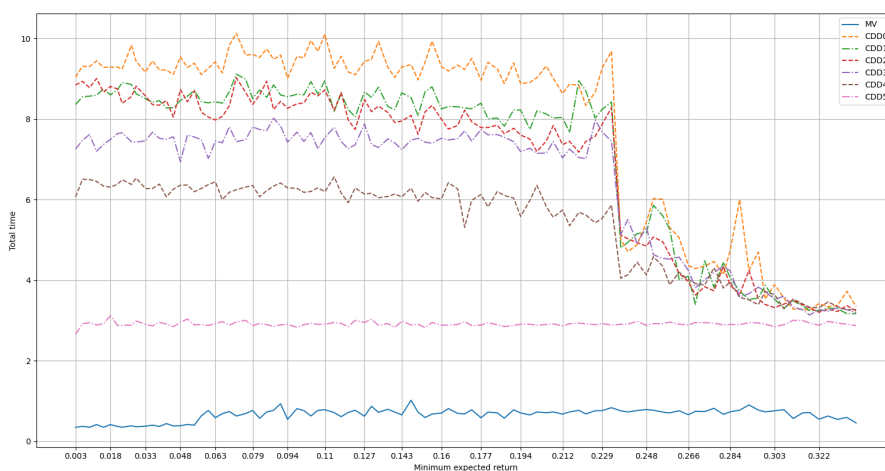


Figura 5.2: Tempi per risolvere il modello al variare di μ_0 , in secondi

In ogni caso, se è richiesto di risolvere il problema per ogni μ_0 è sempre possibile procedere a risolvere le varie istanze in parallelo, riducendo di molto il tempo totale di esecuzione.

Nella tabella 5.3, è ancora più evidente la superiorità del modello MV a livello temporale, anche includendo il calcolo della matrice di covarianza.

MV	CDD 0	CDD 1	CDD 2	CDD 3	CDD 4	CDD 5
74.38	793.82	722.42	701.18	650.92	548.10	291.91

Tabella 5.3: Tempi totali, in secondi

Memoria

Nella simulazione non è stato utilizzato un profiler, però monitorando a mano usando il comando `top` durante le istanze più grandi (`performance_020480`) il modello MV non ha superato 1 GB di memoria allocata, mentre il modello CDD ha usato circa 4 GB. La memoria non appare un fattore limitante. In una pagina della documentazione di CPLEX [2] è riportato un modo per avere una stima molto grossolana della memoria utilizzata dal modello a partire dal numero di righe e colonne ¹, queste stime sono riportate nella tabella 5.4

T	MV	CDD 0	CDD 1	CDD 2	CDD 3	CDD 4	CDD 5
10	220.90	0.01	0.01	0.01	0.01	0.01	0.01
20	220.90	0.02	0.02	0.02	0.02	0.02	0.02
40	220.90	0.04	0.04	0.04	0.04	0.04	0.04
80	220.90	0.08	0.08	0.08	0.08	0.08	0.08
160	220.90	0.16	0.16	0.16	0.16	0.16	0.16
320	220.90	0.32	0.32	0.32	0.32	0.32	0.32
640	220.90	0.64	0.64	0.64	0.64	0.64	0.64
1280	220.90	1.28	1.28	1.28	1.28	1.28	1.28
2560	220.90	2.56	2.56	2.56	2.56	2.56	2.56
5120	220.90	5.12	5.12	5.12	5.12	5.12	5.12
10240	220.90	10.24	10.24	10.24	10.24	10.24	10.24
20480	220.90	20.48	20.48	20.48	20.48	20.48	20.48

Tabella 5.4: Stima della memoria utilizzata dai modelli, in MB

Si nota come la memoria utilizzata nel modello CDD aumenta linearmente con T , mentre il modello di Markowitz ha sempre 2 soli vincoli e il numero di asset rimane costante in tutte le simulazioni, quindi anche il risultato della stima è sempre lo stesso.

¹Per i modelli lineari è sufficiente prendere il numero di vincoli e dividere per 1000, per quelli quadratici prendere il numero di variabili all'interno della funzione obiettivo e sommare il numero di vincoli, e dividere anche in questo caso per 1000.

5.2 Simulazioni out of sample

In questa sezione, dopo aver scelto 2 date significative per il *backtest*, verranno eseguite 4 simulazioni: per ogni data sarà generata un'istanza usando il metodo *Bootstrap* e una usando il *Block Bootstrapping*, con un numero di scenari $T = 2500$. In ogni simulazione verrà confrontato il portfolio generato dal modello di Markowitz con i portfolio generati dal modello Conditional Semideviation con $\beta = \{0.01, 0.1, 0.5\}$. Tutti i modelli avranno un target $\mu_0 = 7\%$ (annualizzato).

Le due date selezionate sono il 01/01/2020 e il 15/06/2022. Queste date sono state scelte perché immediatamente precedenti a momenti di forte volatilità.

Una volta generati i portfolio, viene valutata la loro performance usando i dati *out-of-sample* per tutto l'anno successivo. I rendimenti dei portfolio saranno anche messi a confronto con le performance dell'indice S&P500 stesso nello stesso periodo, e sono calcolate alcune metriche che verranno introdotte a breve.

Nel block bootstrapping oltre a T , c'è bisogno di scegliere una lunghezza per i blocchi. Per un'analisi della scelta della lunghezza del blocco ottima, si rimanda a Hall [5]. Per focalizzarsi sulla distribuzione dei rendimenti giornalieri, Hall ha dimostrato che la scelta migliore sia avere i blocchi di lunghezza $n^{1/5}$ nel caso di funzioni di distribuzione a due lati, dove n è la lunghezza della serie di dati. Dovendo però eseguire un'analisi realistica, è necessario scartare tutti i dati delle date successive all'*investment time*: per la simulazione del 01/01/2020 avendo a disposizione 1593 campioni la *block length* da scegliere è 4, mentre per quanto riguarda la simulazione del 15/06/2022 avendo 2453 dati la lunghezza del blocco è pari a 5.

Nelle prossime sezioni vengono introdotte due metriche che verranno usate per valutare i portfolio generati.

Sortino Ratio

Per tutti i portfolio della simulazione viene calcolato il *Sortino Ratio*, che è una metrica utilizzata per valutare il rapporto rischio/rendimento di un asset o portfolio.

Si calcola con la seguente formula:

$$S = \frac{R_p - r_f}{\sigma_d} \quad (5.1)$$

R_p indica il rendimento dell'asset/portfolio preso in considerazione

r_f è il *risk free return*, generalmente si usa il rendimento annualizzato dei titoli di debito

americani. In questo elaborato, è fissato a 1.5%.

σ_d è la *Downside Deviation*, una misura di rischio che si concentra sui rendimenti minori di una soglia minima, che per il calcolo del *Sortino Ratio* è sempre r_t . Per calcolare la *Downside Deviation* di un singolo asset si procede nel seguente modo: dall'insieme delle realizzazioni possibili per il rendimento degli asset si prende il sottoinsieme di rendimenti minori di r_t . σ_d è la deviazione standard di questo sottoinsieme.

Per un portfolio composto da diversi asset, la sua *Downside Deviation* è la media ponderata delle σ_d dei singoli asset.

Confrontare il *Sortino Ratio* tra i portfolio è importante in quanto ci dà una misura oggettiva del risultato ottenuto.

Maximum Drawdown

Il *Maximum Drawdown* (MDD) è un'altra metrica utilizzata in ambito finanziario per valutare la qualità di un portfolio. L'MDD quantifica la massima perdita percentuale subita da un portafoglio dal suo picco al minimo successivo, prima di stabilire un nuovo massimo. Questa misura è di grande interesse per investitori e gestori di portfolio poiché fornisce un'indicazione sul potenziale calo massimo che un portafoglio potrebbe sperimentare. Un MDD elevato indica che quel particolare asset o portfolio è particolarmente rischioso. Il MDD viene calcolato per tutti i portfolio generati durante la simulazione.

5.2.1 Analisi portfolio generati il 01/01/2020

Nel periodo considerato, poco dopo l' *investment time*, si è verificato un crollo di più del 30% dell'indice. I risultati del *backtest* usando gli scenari generati usando le tecniche di *Bootstrap* e di *Block bootstrap* sono riassunti rispettivamente nelle tabelle 5.5 e 5.6.

	MV	CDD 0	CDD 1	CDD 2	S&P500
<i>R</i>	8.10%	9.24%	8.43%	6.81%	9.91%
<i>S</i>	0.18	0.21	0.19	0.14	0.27
<i>MDD</i>	38.27%	39.56%	39.36%	37.52%	33.92%

Tabella 5.5: Riassunto metriche del backtest usando la tecnica *Bootstrap*

	MV	CDD 0	CDD 1	CDD 2	S&P500
R	6.29%	6.94%	7.33%	5.57%	9.91%
S	0.13	0.14	0.15	0.11	0.27
MDD	39.11%	39.05%	39.59%	39.29%	33.92%

Tabella 5.6: Riassunto metriche del backtest usando la tecnica *Block Bootstrap*

Come si può notare in entrambi gli scenari tutti i portfolio generati hanno sottoperformato l'indice in tutte le metriche.

Nei grafici 5.3 e 5.4 è possibile confrontare l'andamento dei portfolio rispetto all'indice di riferimento.



Figura 5.3: Confronto con i rendimenti del S&P500 (dati di partenza: instance/boot0101)



Figura 5.4: Confronto con i rendimenti del S&P500 (dati di partenza: instance/block0101)

Le tabelle 5.7 e 5.8 riportano i ticker degli asset con peso più elevato per ogni portfolio. Questa vista è utile a evidenziare un grande problema dei modelli di portfolio optimization, cioè la tendenza a concentrare molto le allocazioni in pochi asset. Questo è particolarmente evidente, in questo caso, nei portfolio generati con il modello *Conditional Drawdown*, dove più di un quarto dell'intero capitale è stato allocato a una singola azione (AMCR).

Per ovviare a questo problema è necessario introdurre vincoli aggiuntivi nel modello per imporre un limite a quanto è possibile allocare a un singolo asset, aumentando così la diversificazione.

Un'altra osservazione che è possibile fare è la seguente: nonostante i modelli MV e CDD siano diversi, tendono ad allocare comunque la maggior parte del capitale allo stesso gruppo ristretto di asset. Questo continua ad accadere anche se i dati di partenza sono campionati in maniera differente, come si può vedere confrontando le tabelle.

MV		CDD 0		CDD 1		CDD 2	
AMCR	16.10%	AMCR	25.97%	AMCR	25.54%	AMCR	25.57%
CBOE	6.46%	RSG	5.00%	EG	4.52%	EVRG	7.50%
SO	5.17%	NEE	4.53%	RSG	4.42%	CBOE	5.05%
KO	4.84%	PG	4.52%	MCD	4.42%	T	4.55%
DUK	4.60%	MCD	4.39%	PG	4.39%	KO	3.68%

Tabella 5.7: Asset principali e relativo peso nel portfolio (dati di partenza: instance/boot0101)

MV		CDD 0		CDD 1		CDD 2	
AMCR	17.10%	AMCR	21.38%	AMCR	21.61%	AMCR	26.27%
CBOE	5.94%	MCD	5.86%	MCD	5.70%	PSA	4.73%
MCD	5.72%	NEE	5.18%	KDP	4.40%	KO	4.68%
SO	4.82%	KDP	4.25%	CBOE	4.22%	CBOE	4.46%
DUK	4.49%	TFC	3.46%	SO	3.85%	PG	4.21%

Tabella 5.8: Asset principali e relativo peso nel portfolio (dati di partenza: instance/block0101)

A prima vista, i risultati di questa simulazione potrebbero sembrare deludenti, con rendimenti inferiori al *benchmark*. Tuttavia, è importante contestualizzare adeguatamente questi risultati, considerando attentamente le condizioni e i limiti metodologici. In particolare, i dati *in-sample* usati coprono un periodo di circa 7 anni particolarmente positivo per il mercato azionario americano. Di conseguenza, i modelli assumono incorrettamente che queste dinamiche rimangano invariate nel tempo. Espandere i dati storici di partenza a un arco temporale più ampio, cercando di includere sia periodi positivi che negativi dei mercati, potrebbe portare a simulazioni meno soggette a *overfitting*.

5.2.2 Analisi portfolio generati il 15/06/2022

In questa simulazione l'investment time è avvenuto dopo una significativa discesa del mercato, approssimativamente del 20%. I risultati dei portfolio generati e dell'indice sono riassunti nelle tabelle 5.9 e 5.10.

	MV	CDD 0	CDD 1	CDD 2	S&P500
<i>R</i>	23.53%	16.29%	14.33%	13.33%	9.91%
<i>S</i>	0.71	0.44	0.38	0.35	0.27
<i>MDD</i>	14.14%	29.47%	29.34%	28.92%	33.92%

Tabella 5.9: Riassunto metriche del backtest usando la tecnica *Bootstrap*

	MV	CDD 0	CDD 1	CDD 2	S&P500
<i>R</i>	33.96%	19.78%	18.57%	18.24%	9.91%
<i>S</i>	0.98	0.53	0.50	0.50	0.27
<i>MDD</i>	23.58%	27.87%	28.57%	24.82%	33.92%

Tabella 5.10: Riassunto metriche del backtest usando la tecnica *Block Bootstrap*

In questa simulazione tutti i portfolio generati hanno battuto l'indice considerato. Anche le metriche sono positive: un *Sortino Ratio* più elevato ci indica che il portfolio è riuscito a ottenere rendimenti maggiori del tasso di riferimento (1.5%) minimizzando l'esposizione al ribasso.

In tutti i casi, inoltre, la *MDD* è ridotta rispetto all'andamento dell'indice. Questo è evidente anche guardando l'andamento dei grafici 5.5 e 5.6.

È possibile notare come il portfolio generato con Markowitz nel caso degli scenari ottenuti con *Bootstrap* abbia protetto di più dal ribasso, mentre nel caso degli scenari ottenuti con *Block Bootstrap* si ha avuto un ritorno e un *Sortino Ratio* maggiore.

In questa simulazione, si osserva una diminuzione del rendimento finale all'aumentare del parametro β . Tale risultato appare in contrasto con le attese teoriche, secondo cui β più elevati dovrebbero tradursi in obiettivi di rendimento più ambiziosi e quindi maggiori ritorni, pur con un rischio maggiore.

Come detto in precedenza nella sezione 5.1, questo probabilmente è dovuto al fatto che all'aumento di β le scelte vengono ridotte.

Nei grafici 5.5 e 5.6 è possibile confrontare l'andamento dei portfolio rispetto all'indice di riferimento.



Figura 5.5: Confronto con i rendimenti del S&P500 (dati di partenza: instance/boot1506)



Figura 5.6: Confronto con i rendimenti del S&P500 (dati di partenza: instance/block1506)

Le tabelle 5.11 e 5.12 riportano i ticker degli asset con peso più elevato per ogni portfolio. Anche in questa simulazione si può notare la tendenza dei portfolio a concentrare le scelte in un numero ristretto di asset, nonostante la generazione di scenari diversa e i diversi modelli.

MV		CDD 0		CDD 1		CDD 2	
CLX	11.87%	AMCR	16.12%	AMCR	16.43%	AMCR	18.29%
AMCR	9.19%	RSG	6.86%	CBOE	6.99%	CBOE	6.27%
VZ	8.97%	CBOE	6.85%	RSG	6.23%	ED	5.99%
ED	6.21%	CLX	6.53%	ED	5.53%	CLX	5.46%
NEM	5.73%	WMT	4.72%	CLX	5.25%	WMT	5.05%

Tabella 5.11: Asset principali e relativo peso nel portfolio (dati di partenza: instance/boot1506)

MV		CDD 0		CDD 1		CDD 2	
AMCR	11.12%	AMCR	17.67%	AMCR	16.79%	AMCR	18.74%
CLX	10.26%	CLX	9.40%	CLX	9.50%	ED	7.48%
VZ	7.91%	CBOE	7.90%	CBOE	7.47%	PG	5.82%
KO	6.36%	PG	6.23%	ED	5.52%	CBOE	4.86%
CHRW	5.90%	ED	4.72%	PG	5.46%	WMT	4.65%

Tabella 5.12: Asset principali e relativo peso nel portfolio (dati di partenza: instance/block1506)

Capitolo 6

Conclusioni

In questo ultimo capitolo, prima di trarre le conclusioni, si evidenziano alcune limitazioni e bias dell'analisi effettuata e si offrono spunti per eventuali sviluppi futuri.

6.1 Limitazioni e possibili miglioramenti

Alcune importanti limitazioni di questa analisi riguardano i bias e le semplificazioni che sono stati introdotti:

- *survivorship bias*: avendo usato la composizione dell'indice S&P500 attuale è stata introdotta una selezione importante: sono stati esclusi a priori asset che all'investment time facevano parte dell'indice, ma che oggi non lo sono più, generalmente perché la loro valutazione è diminuita nel tempo, portandoli a uscire dalla lista delle 500 aziende più grandi. Questo può essere dovuto al fatto che le loro valutazioni sono diminuite nel tempo, portandoli ad uscire dall'elenco delle 500 aziende più grandi. Di conseguenza, le aziende che rimangono nell'indice tendono ad apparire in modo più positivo dal punto di vista delle prestazioni.
- *selection bias*: sono state prese in considerazione solo aziende con più di 10 anni di dati storici. Questo fa in modo che le aziende giovani non vengano prese in considerazione: queste potrebbero essere più propense a fallire, ma dall'altro lato hanno anche avere il potenziale di avere rendimenti elevati.
- nella sezione 3.1 è stato ipotizzato che i rendimenti dei vari asset seguano una distribuzione gaussiana. In realtà non solo la distribuzione generalmente è asimmetrica, ma è presente il fenomeno detto *long tail*: gli eventi estremi sono più comuni rispetto a quanto dovrebbe accadere con la distribuzione normale.

Per quanto riguarda i modelli, questi potrebbero essere ampliati aggiungendo dei vincoli per mitigare il problema della concentrazione degli asset selezionati dal portfolio e modellare i costi di transazione degli intermediari per valutare come questi ultimi impattino i risultati.

L'analisi delle performance computazionali può essere ampliata in diversi modi: sarebbe interessante sfruttare la parallelizzazione per risolvere il problema per tutti i μ_0 in maniera efficiente, o l'implementazione di una strategia di "warm start" al fine di sfruttare soluzioni preesistenti, visualizzando il cambiamento nei pesi finali del portfolio in seguito a leggere modifiche dei parametri.

Inoltre sarebbe interessante vedere come le performance dei modelli di Markowitz e Conditional Drawdown possano scalare aumentando il numero di asset, in quanto nel presente elaborato si è valutato unicamente il cambiamento del numero di scenari.

6.2 Conclusioni

In questa tesi, dopo un'introduzione al problema della portfolio optimization, sono stati definiti e confrontati i modelli Mean Variance e Conditional Drawdown. Sebbene i risultati ottenuti dagli esperimenti non siano sufficienti a determinare in modo definitivo l'efficacia dei modelli lineari nell'ottimizzazione dei portafogli, essi costituiscono un solido punto di partenza.

Sviluppi futuri potrebbero includere l'analisi di un maggior numero di modelli, sia lineari che non, per ampliare il confronto. L'integrazione di vincoli e costi operativi realistici consentirebbe di avvicinare i risultati all'implementazione pratica.

Bibliografia

- [1] *CPLEX documentation*. URL: <https://www.ibm.com/docs/en/icos/22.1.1>.
- [2] *Estimating cplex memory requirements based on problem size*. URL: <https://www.ibm.com/support/pages/guidelines-estimating-cplex-memory-requirements-based-problem-size>.
- [3] Gianfranco Guastaroba. “Portfolio Optimization: Scenario Generation, Models and Algorithms”. PhD thesis. UNIVERSITÀ DEGLI STUDI DI BERGAMO, 2010.
- [4] Abhishek Gunjan and Siddhartha Bhattacharyya. “A brief review of portfolio optimization techniques”. In: *Artificial Intelligence Review* (2022).
- [5] PETER HALL, JOEL L. HOROWITZ and BING-YI JING. “On blocking rules for the bootstrap with dependent data”. In: *Biometrika* 82.3 (Sept. 1995), pp. 561–574.
- [6] Renata Mansini, Włodzimierz Ogryczak and M. Grazia Speranza. *Linear and Mixed Integer Programming for Portfolio Optimization*. Springer International Publishing, 2015. DOI: 10.1007/978-3-319-18482-1. URL: <https://doi.org/10.1007/978-3-319-18482-1>.
- [7] Harry Markowitz. “Portfolio Selection”. In: *The Journal of Finance* (1952).
- [8] *Matplotlib*. URL: <https://matplotlib.org/>.
- [9] Filippo-Massimo Niero. *Portfolio Optimization*. URL: <https://github.com/FilippoNiero/Portfolio-optimization>.
- [10] ranaroussi. *yFinance*. URL: <https://github.com/ranaroussi/yfinance>.
- [11] Kenneth Reitz. *Request*. URL: <https://pypi.org/project/requests/>.
- [12] Leonard Richardson. *BeautifulSoup Documentation*. URL: <https://beautiful-soup-4.readthedocs.io/en/latest/>.

-
- [13] Lawrence S. Speidell, Deborah H. Miller and James R. Ullman. "Portfolio Optimization: A Primer". In: *Financial Analysts Journal* (1989).
- [14] Robert F. Stambaugh. "Analyzing investments whose histories differ in length". In: *Journal of Financial Economics* 45.3 (Sept. 1997).