



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale o Specialistica in Scienze Economiche e Finanziarie – Analista
Finanziario

**“Portfolio Management: Focus sulla riduzione
degli effetti delle code di sinistra”**

**“Portfolio Management: Focus on left-tails
effects reduction”**

Relatore: Chiar.mo

Prof. ssa:

Maria Cristina Recchioni

Tesi di Laurea di:

Matteo Agostini

Anno Accademico 2021-2022

“A Zio Lino, che mi guarda da Lassù ...”

Indice

INTRODUZIONE	5
CAPITOLO 1: NASCITA ED EVOLUZIONE DELLE TECNICHE DI OTTIMIZZAZIONE DI PORTAFOGLIO	7
1.1 RISCHIO E RENDIMENTO.....	8
1.2 CRITERIO MEDIA-VARIANZA.....	12
1.3 FORMALIZZAZIONE MATEMATICA DELLA SELEZIONE DI PORTAFOGLIO.....	15
1.4 SELEZIONE DEL PORTAFOGLIO OTTIMALE.....	20
1.5 LIMITI DEL MODELLO DI MARKOWITZ.....	24
1.6 CAPM.....	26
1.6.1 Indice di Sharpe.....	29
1.6.2 Indice di Sortino.....	30
1.7 VAR E CVAR.....	32
1.7.1 Proprietà del VaR e del CVaR.....	35
1.7.2 Conclusioni.....	39
CAPITOLO 2: NUOVE TECNICHE DI OTTIMIZZAZIONE DI PORTAFOGLIO PER RIDURRE L'EFFETTO DELLE CODE	41
2.1 EDC e EDH.....	41
2.1.1 Extreme Downside Correlation.....	42
2.1.2 Extreme Downside Hedge.....	43
2.2 PROPRIETA', ROBUSTEZZA E PREDITTIVITA' DI EDC E EDH.....	45
2.2.1 Proprietà.....	45

2.2.2 Robustezza	45
2.2.3 Predittività.....	46
2.3 UNA NUOVA METRICA PER LA MISURAZIONE DEL RISCHIO	48
2.3.1 Clustering con l'Expectation Maximization	49
2.3.2 Nuovo data set informativo e definizione delle misure di similarità e volatilità delle code	52
2.3.3 Definizione della misura di variabilità.....	55
2.4 EXPECTATION MAXIMIZATION	57
2.4.1 Breve definizione matematica	58
2.4.2 Expectation Maximization con una mistura gaussiana monodimensionale	59
CAPITOLO 3: ANALISI EMPIRICA DI QUATTRO DIFFERENTI PORTAFOGLI.....	65
3.1 DESCRIZIONE DEI DATI	65
3.2 COSTRUZIONE DEI PORTAFOGLI	70
3.3 CONFRONTO DELLE PERFORMANCE.....	75
CONCLUSIONI.....	84
APPENDICE	86
BIBLIOGRAFIA e SITOGRAFIA	103

INTRODUZIONE

Il famoso approccio media-varianza proposto da Markowitz nel 1952, su cui si basa la Teoria di Portafoglio Moderna, ha delle limitazioni, nel senso che non prende in considerazione l'asimmetria del rischio. Questo perché la covarianza, cioè la misura considerata dal premio Nobel come rappresentativa del rischio, non fa distinzioni tra “downside risk” e “upside risk”, rispettivamente le possibilità di avere un rendimento di portafoglio al di sotto o al di sopra della media. Le misure basate sui quantili della distribuzione sviluppate successivamente, come il VaR, il CVaR o l'Extreme Downside Correlation, superano le limitazioni della covarianza e sono quindi in grado di catturare ed isolare il “downside risk”, cioè quella porzione di variabilità che effettivamente l'individuo percepisce come rischiosa. Tuttavia, queste misure sono piuttosto insensibili alla forma delle code delle distribuzioni, dato che dipendono dalla scelta che si fa a priori dell' α -quantile di riferimento. Viene così presentata una nuova misura di variabilità, proposta da Mariani F, Polinesi G, Recchioni MC (2022), definita endogenamente, che cattura il rischio sia dalla volatilità degli asset sia dalle code delle distribuzioni dei rendimenti.

Nel primo capitolo viene presentata in maniera sintetica la teoria di ottimizzazione di portafoglio di Markowitz, si accenna al Capital Asset Pricing Model per poi introdurre gli indici di Sharpe e Sortino. Infine, vengono messi a confronto VaR e

CVaR, spiegando perché quest'ultima risulta una preferibile misura di rischio rispetto la prima.

Nel secondo capitolo si introducono l'Extreme Downside Correlation e l'Extreme Downside Hedge come misure di rischio che tentano di tenere in considerazione i rischi di coda, successivamente si spiega la metodologia di costruzione della nuova misura di variabilità che cattura il rischio delle code sinistre, per poi soffermarsi in maniera più specifica sull'algoritmo, l'Expectation Maximization, con cui viene computato il dataset rappresentativo delle code dei rendimenti.

Nel terzo e ultimo capitolo, si profila un'analisi empirica, utilizzando più di 3000 osservazioni di 44 ETF, sulle performance di quattro diversi portafogli: un portafoglio calcolato minimizzando la nuova misura di rischio, un portafoglio calcolato tenendo conto dell'Extreme Downside Risk e due portafogli calcolati rispettivamente in funzione del massimo indice di Sharpe e di Sortino.

CAPITOLO 1: NASCITA ED EVOLUZIONE DELLE TECNICHE DI OTTIMIZZAZIONE DI PORTAFOGLIO

Nel 1952 Harry Markowitz pose le basi della cosiddetta “*Modern Portfolio Theory*” con il suo articolo “*Portfolio Selection*”, pubblicato nel “*Journal of Finance*” che gli valse il premio Nobel 38 anni dopo, nel 1990. Egli fu il primo a dare una formalizzazione matematica alla nozione di diversificazione degli investimenti e riuscì a dimostrare analiticamente come, a parità di rendimento, investire in molteplici asset diminuiva il rischio dell’operazione. L’obiettivo della teoria di portafoglio è quindi quello di spiegare per quale motivo un investitore dovrebbe preferire un portafoglio ben diversificato ad uno composto da pochi o addirittura un solo titolo.

Le ipotesi sotto le quali opera la teoria di Markowitz sono:

- *Durata uniperiodale degli investimenti* → il tempo che trascorre tra il momento dell’investimento e quello del disinvestimento viene considerato come un periodo.
- *Assenza di imposte* → non si considerano le detrazioni spettanti allo Stato
- *Divisibilità illimitata dei titoli* → l’investitore può detenere una qualsiasi frazione di un asset-
- *Assenza di costi di transazione* → non vengono considerati i costi di organizzazione, studio e ricerca dell’attività.

- *Perfetta competitività dei mercati* → non ci sono asimmetrie informative, numero elevato di imprese che operano nel mercato, tutti i prodotti offerti sono simili, ecc.

Queste ipotesi non limitano la descrizione della teoria, ma piuttosto la semplificano.

1.1 RISCHIO E RENDIMENTO

Un titolo viene definito rischioso quando esiste la possibilità che esso realizzi dei rendimenti diversi dal rendimento atteso. Tuttavia, ciò che l'individuo percepisce effettivamente come rischio è quell'insieme di eventualità in cui il rendimento effettivo risulta al di sotto del rendimento atteso. Inoltre, da un punto di vista più rigoroso, si parla di rischio solo nel momento in cui si conoscono con esattezza le probabilità con cui possano accadere gli eventi. Negli altri casi, in cui non sono né in grado di prevedere il futuro, né di conoscere le probabilità di realizzazione degli avvenimenti, si parla di *incertezza*. La situazione normale è quella per cui ci troviamo in una situazione di incertezza, visto che risultano ignote le probabilità per cui un titolo abbia un certo rendimento, ma la descriviamo mediante una situazione di rischio.

Il *rendimento semplice*, cioè il rendimento del titolo alla fine del periodo si definisce come:

$$R = \frac{P' - P + D}{P},$$

dove:

P = prezzo che il titolo possiede oggi.

P' = prezzo che il titolo avrà alla fine del periodo.

D = eventuali dividendi che stacca il titolo.

Ciò che fa sì che il titolo venga descritto da una variabile aleatoria sono proprio P' e D .

R sarà quindi una variabile aleatoria con delle determinazioni associate a delle probabilità:

$$\begin{cases} R_1 R_2 R_3 \dots R_n \\ p_1 p_2 p_3 \dots p_n \end{cases}$$

a cui associo il *rendimento atteso*:

$$\mu = E(R) = \sum_{i=1}^n p_i \cdot R_i,$$

che viene utilizzato come indice di rendimento. Come indice di rischio, invece, viene preso in considerazione lo scostamento dal rendimento atteso, la *varianza*:

$$\sigma^2 = V(R) = \sum_{i=1}^n (R_i - \mu)^2 \cdot p_i.$$

La varianza può essere considerata un discreto, seppur non perfetto, indice di rischiosità dell'investimento perché riesce a dare un'informazione sulla variabile aleatoria in questione senza dover conoscere l'intera distribuzione. È chiaro che sarebbe più coerente al concetto di rischio considerare solamente gli scostamenti negativi, cioè al di sotto, dalla media, ma l'analisi risulta comunque valida e utile anche considerando l'intera varianza.

È richiesto però che la variabile aleatoria R , a cui vengono associate media e varianza, si distribuisca normalmente, perché è noto che quei due parametri descrivono completamente una variabile casuale quando essa è normale. Sarà necessario quindi il passaggio da rendimenti semplici R a rendimenti logaritmici r :

$$r = \log\left(\frac{P' + D}{P}\right) = \log\left(\frac{P' - P + D + P}{P}\right) = \log(1 + R),$$

per cui:

$$r = \log(1 + R) \cong R^1$$

se $|R| < 15\%$.

¹ È dimostrabile, usando la formula di Taylor, che il logaritmo di $1+R$ è circa R quando ho rendimenti semplici compresi tra -15% e 15% .

Il rendimento semplice in un istante t , si può indicare, ad esempio, con $R_t(1)$, cioè come il rendimento semplice di un titolo per 1 giorno dal giorno t al giorno $t+1$, a cui è associabile il corrispondente rendimento logaritmico $r_t(1)$.

Adesso, supponendo di prendere in considerazione un periodo di n giorni, e che alla fine di ogni giorno ogni pagamento viene reinvestito, il rendimento alla fine degli n giorni sarà:

$$1 + R_t(n) = [1 + R_t(1)] \cdot [1 + R_{t+1}(1)] \cdot \dots \cdot [1 + R_{t+n-1}(1)],$$

e considerando il rendimento logaritmico si otterrà:

$$r_t(n) = \log[1 + R_t(n)] = \log \left[\prod_{i=0}^{n-1} [1 + R_{t+i}(1)] \right] = \sum_{i=0}^{n-1} r_{t+i}(1).$$

Essendo il logaritmo di un prodotto uguale alla somma dei logaritmi, si avrà che il rendimento logaritmico lungo tutto il periodo è uguale alla somma dei rendimenti logaritmici dei singoli periodi, cioè alla somma di n variabili casuali. Supponendo queste variabili casuali, indipendenti e identicamente distribuite, a prescindere dal tipo di distribuzione che esse hanno, con stessa media e stessa varianza, allora al divergere di n , la somma delle variabili (quindi il rendimento logaritmico dell'intero

periodo) convergerà, per il teorema centrale del limite², ad una variabile che si distribuisce normalmente.

$$r_t(n) \rightarrow N(n\mu; n\sigma^2).$$

Quindi, anche se è vero che è possibile osservare giornalmente rendimenti che si distribuiscono diversamente da una normale, per intervalli di tempo più ampi la distribuzione dei rendimenti logaritmici converge a distribuzioni normali. Tuttavia, è necessario ribadire che i rendimenti logaritmici “coincidono” con quelli semplici se si considerano rendimenti in valore assoluto minori del 15%.

1.2 CRITERIO MEDIA-VARIANZA

Il criterio decisionale utilizzato da Markowitz nella selezione del portafoglio ottimale è quello *media-varianza*, che confronta due o più titoli usando proprio il rendimento atteso μ e la varianza σ^2 come indici di rendimento e rischio.

Considerando due variabili aleatorie X e Y (rappresentative di due diversi asset), si potrà associare ad ognuna di loro valore atteso μ e varianza σ^2 , cioè una coppia di

² Afferma, nella sua più nota formulazione, quella di Lindeberg-Lévy, che la somma di n variabili indipendenti aventi identica distribuzione è una variabile che si distribuisce normalmente a prescindere dal tipo di distribuzione delle variabili di partenza.

numeri ben precisi. Tale coppia può essere rappresentata in un piano, avente come asse delle ordinate μ e come asse delle ascisse σ^2 , il cosiddetto piano media-varianza. Confrontando X e Y e supponendo abbiano la medesima varianza, si preferirà la variabile con rendimento maggiore. Nel caso in cui esse abbiano lo stesso rendimento, la scelta ricadrà sulla variabile con varianza minore. Nel caso in cui un titolo sia migliore in entrambi gli indici la decisione sarà piuttosto ovvia. Rimangono dei casi in cui però X e Y potrebbero non essere confrontabili e ciò perché i punti in un piano non sono sempre perfettamente ordinabili. È in quel caso che si va a formare la *frontiera efficiente*, cioè l'insieme dei punti da cui, se ci allontanassimo, andremmo a peggiorare uno dei nostri due indici (obiettivi, in ottimizzazione matematica) senza migliorare l'altro.

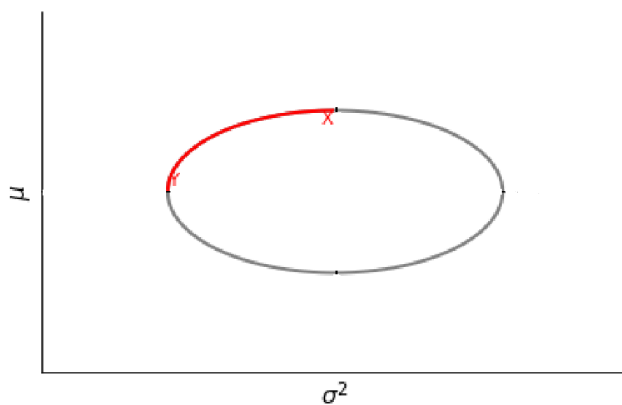


Figura 1.1. Esempio grafico delle combinazioni di media e varianza ottenibili da un portafoglio. Prodotta dall'autore.

Si osserva che nella Figura 1.1, in cui è rappresentato un ipotetico insieme di tutte le combinazioni di rendimento e varianza ottenibili investendo in un portafoglio di due titoli, il segmento che unisce Y ad X costituisce la frontiera efficiente, cioè l'insieme di punti costituiti da combinazioni lineari convesse³ tra X e Y, da cui se ci si allontanasse si peggiorerebbe uno dei due obiettivi. Come si sceglie però, tra questo insieme ottimale di punti, qual è il migliore su cui investire? In questo caso la scelta diventa soggettiva e dipende dall'avversione al rischio dell'investitore, cioè dall'inclinazione delle curve di utilità dell'investitore. Per sapere se è preferibile un punto con alto rendimento ma alta varianza o uno con un rendimento più basso ma anche una varianza più bassa, è necessario conoscere le preferenze soggettive dell'investitore, che si manifestano geometricamente come delle curve, aventi delle specifiche proprietà. Infatti, nel disegnarle sul piano media-varianza, (dove si ricorda l'asse delle ascisse corrisponde alla varianza e l'asse delle ordinate alla media), si suppone che:

- *Sono crescenti ed hanno concavità rivolta verso l'alto* → normalmente, non cresce linearmente la disposizione ad accettare più rischio a fronte di un maggior rendimento. A mano a mano che aumenta il rischio da sopportare, sarà richiesto un rendimento più che proporzionale per compensare la

³ È una combinazione lineare di elementi fatta con coefficienti non negativi a somma 1 $\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n$, dove la somma dei λ_i è uguale a 1, con $i = 1, 2, \dots, n$.

propria esposizione. Un individuo razionale non sarà mai disposto ad accettare quantità illimitate di rischio. la differenza tra la propensione al rischio di due individui sta su quanto risultano essere più o meno concave le loro curve di indifferenza. Chiaramente, secondo questa logica, il decisore sarà disposto ad accettare una varianza maggiore a patto di aumentare anche, seppur in maniera infinitesima, il rendimento medio; per questo motivo non si presuppone la concavità verso il basso.

- *Non si possono mai intersecare* → ogni curva è associata a livelli di utilità differenti, se si intersecassero si avrebbe un punto in cui per due curve si ottiene lo stesso grado di utilità.
- *Spostandosi verso l'alto e verso sinistra, aumenta l'utilità* → ciò perché i livelli massimi di utilità si ottengono in corrispondenza della frontiera efficiente, che come già spiegato, si trova in alto a sinistra.

1.3 FORMALIZZAZIONE MATEMATICA DELLA SELEZIONE DI PORTAFOGLIO

Quando si investe in due o più titoli e si cerca l'insieme delle combinazioni più efficienti, non si fa altro che risolvere un problema di ottimizzazione vincolata, in cui l'obiettivo è quello di massimizzare il rendimento o, in alternativa, minimizzare la varianza, fissato un certo livello di, rispettivamente, rischio o rendimento. Sarà

necessario impostare il problema dal punto di vista formale scrivendone le condizioni.

Impostando il problema di minimo si avrà:

- $\mu_p = \sum_{i=1}^n x_i \cdot \mu_i \rightarrow$ il rendimento atteso del portafoglio, che sarà uguale alla somma delle singole quote investite in ogni titolo per il loro rendimento atteso.
- $\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n x_i \cdot x_j \cdot \sigma_{i,j} \rightarrow$ la varianza dell'intero portafoglio, dove $\sigma_{i,j} = cov(Y_i, Y_j)$ rappresenta la covarianza tra due titoli.
- $\sum_{i=1}^n x_i = 1 \rightarrow$ si suppone che venga investito tutto il capitale a disposizione.

Ed il problema da risolvere sarà:

$$\left\{ \begin{array}{l} \text{Min } \sigma_p^2 \\ \sum_{i=1}^n x_i \cdot \mu_i = \bar{\mu}_p \\ \sum_{i=1}^n x_i = 1 \end{array} \right.$$

Si minimizza la varianza del portafoglio, imponendo il rendimento uguale ad un certo valore e che tutto il capitale venga investito.

La funzione Lagrangiana relativa a questo tipo di problema sarà:

$$L(x, \lambda) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} + \lambda_1 \cdot \left[\sum_{i=1}^n x_i \mu_i - \bar{\mu}_p \right] + \lambda_2 \cdot \left[\sum_{i=1}^n x_i - 1 \right].$$

$\lambda = [\lambda_1, \lambda_2]$ è la variabile ausiliaria della Lagrangiana e possiede due componenti perché i vincoli sono due. Con $\bar{\mu}_p$ si indica il rendimento fissato di portafoglio che vogliamo raggiungere.

Ora le condizioni che ci permettono di risolvere il problema saranno:

$$\left\{ \begin{array}{l} \frac{\delta L}{\delta x_1} = 2x_1\sigma_{11} + 2x_2\sigma_{12} + \dots + 2x_n\sigma_{1n} + \lambda_1\mu_1 + \lambda_2 = 0 \\ \vdots \\ \frac{\delta L}{\delta x_n} = 2x_1\sigma_{n1} + 2x_2\sigma_{n2} + \dots + 2x_n\sigma_{nn} + \lambda_1\mu_n + \lambda_2 = 0 \\ \frac{\delta L}{\delta \lambda_1} = \sum_{i=1}^n x_i \mu_i - \bar{\mu}_p = 0 \\ \frac{\delta L}{\delta \lambda_2} = \sum_{i=1}^n x_i - 1 = 0 \end{array} \right. \quad 4$$

⁴ Condizioni di 1° ordine di Kuhn-Tucker per la determinazione dei punti di minimo.

Si impone che il gradiente della funzione obiettivo, composto da n derivate prime, sia uguale a zero, così come le derivate rispetto ai due moltiplicatori. Sarà necessario quindi risolvere questo sistema ad $n + 2$ incognite, risolvibile con le proprietà del prodotto matrici-vettori. Questo grazie al fatto che il sistema può essere interpretato in forma vettoriale, in particolar modo nella forma $A \cdot \underline{x} = \underline{b}$, dove A rappresenta la matrice delle varianze e covarianze orlata⁵ con i vettori dei vincoli, \underline{x} il vettore delle incognite e \underline{b} il vettore dei termini noti.

Per ricavare il vettore incognito inverti la matrice A (che risulta, appunto, invertibile, per costruzione della matrice orlata):

$$\underline{x} = A^{-1} \cdot \underline{b}.$$

Una volta risolto il sistema è possibile trovare i punti candidati ad essere di minimo, perché soddisfano le condizioni di 1° ordine su L . Per poi poter concludere si prendono in considerazione le condizioni di 2° ordine, utilizzando l'Hessiana orlata. La condizione di 2° ordine richiede che gli ultimi $n - m$ minori principali⁶ di nord-ovest dell'Hessiana orlata siano a segni alterni con il primo di segno $(-1)^{m+1}$, dove

⁵ L'Hessiana orlata è una matrice costruita usando: l'Hessiana di L rispetto ad \underline{x} (che è una matrice quadrata), la Jacobiana dei vincoli verificati con l'uguaglianza, la sua trasposta ed una matrice di zeri.

⁶ Il primo minore di nord-ovest è quello che considera la prima riga e la prima colonna, il secondo le prime due righe e le prime due colonne, il terzo le prime tre righe e le prime tre colonne e così via. L'ultimo minore corrisponderà a tutta la matrice.

n è il numero delle variabili ed m è il numero dei vincoli verificati con l'uguaglianza. In tal modo, è possibile affermare che la matrice Hessiana è definita negativa e la seconda condizione è rispettata.

Facendo poi variare il livello di rendimento, si trova l'insieme dei portafogli a varianza minima, da cui si determina la frontiera efficiente. Quest'ultima sarà rappresentata dalla parte superiore di una parabola con asse orizzontale e concavità rivolta verso destra nel piano media-varianza (σ^2, μ) , mentre nel piano media-deviazione standard (σ, μ) , la curva risulterà essere un'iperbole.

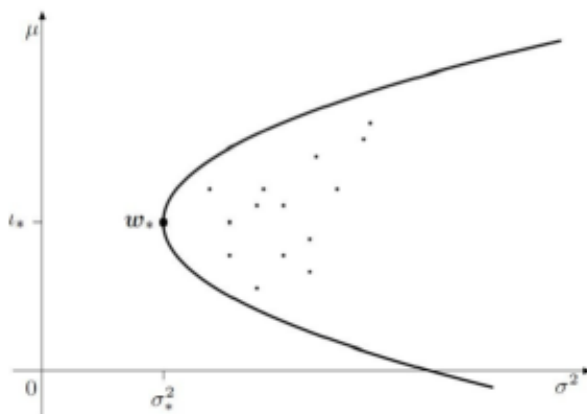


Figura 1.2. Parabola nel piano media-varianza (σ^2, μ) , tratta da di M.Micocci, G.B.Masala (2012) "Manuale di matematica finanziaria. Metodi e strumenti quantitativi per il risk management".

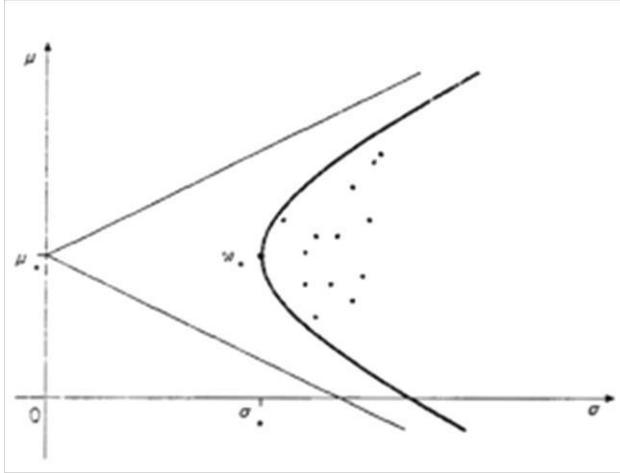


Figura 1.3. Iperbole nel piano media-deviazione standard (σ, μ) , tratta da M.Micocci, G.B.Masala (2012) "Manuale di matematica finanziaria. Metodi e strumenti quantitativi per il risk management".

Va ricordato inoltre, che è stata considerata una situazione in cui sono ammesse vendite allo scoperto, cioè in cui i singoli pesi x_i facenti somma a 1, possono assumere valori negativi. In una situazione in cui non siano ammesse operazioni del genere il problema sarebbe leggermente più complicato, ma la forma non cambierebbe in maniera sostanziale.

1.4 SELEZIONE DEL PORTAFOGLIO OTTIMALE

Terminata la parte oggettiva dell'analisi, è arrivato il momento di far intervenire il decisore, che, portato a massimizzare la propria utilità, sceglierà il punto sulla frontiera efficiente che realizza il suo obiettivo, che è quello di massimizzare la sua

ricchezza finale W_t , partendo da una ricchezza iniziale W_0 . La sua utilità sarà quindi una qualche funzione della ricchezza finale, della ricchezza iniziale o del rendimento R , che in questo caso è esprimibile come $R = \frac{W_t - W_0}{W_0}$.

Avendo rappresentato i portafogli su un piano media-varianza (σ^2, μ) , per tracciare le curve di indifferenza rappresentanti l'utilità dell'individuo è necessario esprimere la loro funzione in termini di σ e μ . Per fare ciò, si considera una funzione d'utilità quadratica perché è uno dei casi in cui le scelte che scaturiscono dal criterio media-varianza sono coerenti con quelle fatte utilizzando il criterio dell'utilità attesa. Una funzione d'utilità di questo tipo garantisce che le curve di isoutilità tracciate sul piano media-varianza non si intersechino mai. Una funzione d'utilità quadratica è del tipo:

$$u(x) = x + ax^2,$$

utilizzando quindi una funzione quadratica che è funzione della ricchezza finale W_t ⁷ si considera:

$$u(W_t) = W_t + b(W_t)^2, \quad \text{con } b < 0$$

⁷ Il decisore investirà una certa ricchezza iniziale W_0 per ottenere una certa ricchezza finale W_t . Per capire in quali titoli investire cercherà di utilizzare una qualche funzione della ricchezza finale, che normalmente chiamiamo funzione d'utilità.

con la costante b negativa perché nel piano $(u(W_t), W_t)$ la concavità della funzione deve essere rivolta verso il basso. Considerando ora l'utilità attesa si ottiene:

$$E[u(W_t)] = E[W_0(1 + R)] + b \cdot E[W_0^2(1 + R)^2],$$

con la ricchezza finale espressa in funzione della ricchezza iniziale e del rendimento, perché $W_t = W_0(1 + R)$.

Calcolando il valore atteso e sostituendo $E[R^2] = V[R] + E^2[R] = \sigma^2 + \mu^2$ si ottiene:

$$E[u(W_t)] = W_0(1 + \mu) + b \cdot W_0^2(1 + \sigma^2 + \mu^2 + 2\mu),$$

si genera il prodotto notevole:

$$E[u(W_t)] = W_0(1 + \mu) + b \cdot W_0^2[\sigma^2 + (1 + \mu)^2],$$

raccogliendo W_0 si ottiene:

$$E[u(W_t)] = W_0 + W_0 \cdot [\mu + b \cdot W_0(\sigma^2 + (1 + \mu)^2)].$$

In tal modo, l'utilità attesa è espressa in termini della ricchezza iniziale, del rendimento μ e della varianza σ^2 , così da poter essere espressa nella forma di curve di indifferenza sul piano media-varianza, che, nel caso di funzioni di utilità di tipo quadratico, rivolgono la concavità verso l'alto.

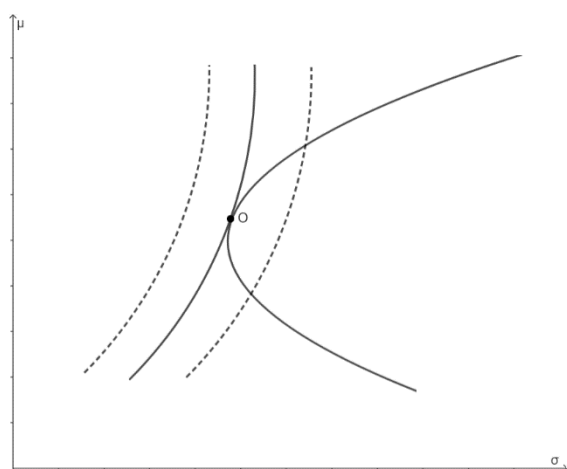


Figura 1.4. intersezione tra la curva su cui giace la frontiera efficiente e il set di curve di indifferenza. Prodotta dall'autore.

Di solito i decisori mostrano tutti una certa avversione al rischio, che si sostanzia nell'inclinazione positiva, cioè concavità rivolta verso l'alto. Nella Figura 1.4, si osserva come nel punto O avviene l'intersezione tra la parte superiore della curva (la frontiera efficiente) e la curva di indifferenza situata il più a "nord-ovest" possibile. Esso costituisce soluzione del problema per il particolare individuo che possiede quel set di curve di indifferenza. È importante ricordare che l'inclinazione delle curve cambia a seconda della propensione al rischio del soggetto preso in

esame. Esse risultano orizzontali nel caso di individui neutrali al rischio, cioè che massimizzano solo il rendimento senza tener conto della varianza e addirittura concave nel caso (teorico) di individui propensi al rischio, cioè che preferiscono rischiare di più al costo di un rendimento inferiore.

1.5 LIMITI DEL MODELLO DI MARKOWITZ

Nonostante la teoria di Markowitz sia stata unica e rivoluzionaria nel suo genere e costituisca tutt'oggi la base della Modern Portfolio Theory, è possibile ravvederne una serie di problematiche:

- Innanzitutto, come già accennato, la varianza risulta essere una misura piuttosto approssimativa del rischio: il fatto che sia una misura simmetrica rispetto alla media fa sì che i rendimenti al di sopra di quest'ultima e quindi positivi non vengano in qualche modo distinti da quelli negativi. Un individuo, al contrario, tende a percepire come rischio solo le oscillazioni al di sotto della media (o di un qualche livello minimo accettabile di rendimento). Markowitz propose la semi-varianza come misura più robusta del rischio, cioè la somma dei quadrati degli scostamenti dalla media al di sotto del valore atteso ma che, nonostante i suoi pregi, risulta più complessa da calcolare. Un altro limite della varianza è che sembra non tenere conto

dell'esistenza delle "fat tails", cioè di quegli accadimenti rischiosi che si concentrano ai margini della distribuzione di probabilità.

- Il modello assume i rendimenti normalmente distribuiti, quando in realtà questa assunzione non è sempre verificata. La non-normalità può tranquillamente sorgere da eventi inattesi, come le crisi economiche (la crisi dei mutui sub-prime, ad esempio), portando il modello a sottostimare il rischio complessivo di portafoglio.
- il criterio media-varianza ignora gli errori di stima legati all'incertezza dei rendimenti attesi, i cui reali valori possono essere osservati solo ex-post, portando a volte alla costruzione di portafogli inefficienti, visto che il modello sembra essere molto sensibile agli errori di stima (piccoli errori nella stima dei rendimenti attesi possono portare in enormi cambiamenti nei pesi di portafoglio). Inoltre, il modello non lascia spazio ad opinioni personali⁸ dell'investitore circa il prezzo futuro dei titoli, perché incorpora tutta l'informazione nell'andamento storico delle quotazioni.

Questi limiti hanno spinto la comunità nella ricerca di misure di rischio e modelli che siano più completi ed il più aderenti possibile alle esigenze di un investitore.

⁸ Quelle che vengono definite "views" nel più recente lavoro di Black e Littermann, sviluppato presso la Goldman Sachs nei primi anni '90.

1.6 CAPM

Il CAPM è un modello di equilibrio dei mercati, sviluppato, nella sua versione base da Sharpe, Lintner e Mossin, tra il 1964 e il 1966, a partire proprio dalla teoria di portafogli di Markowitz. Sebbene esistano formulazioni alternative che ottengono il CAPM a prescindere dalla frontiera efficiente di Markowitz, l'approccio base è fortemente legato all'esistenza di quest'ultima. Esso consente di individuare una relazione tra rendimento e rischio attesi per tutte le attività rischiose presenti sul mercato. Si fonda sulle seguenti ipotesi:

- *Gli individui sono avversi al rischio e tendono a massimizzare la propria utilità attesa $E(U)$.*
- *Gli individui selezionano i portafogli attraverso il criterio media-varianza*
- *L'orizzonte di investimento è uniperiodale.*
- *Ogni individuo può investire o prendere fondi in prestito allo stesso tasso che è privo di rischio.*
- *Le aspettative sui rendimenti attesi, le varianze e le covarianze delle attività rischiose sono omogenee.*
- *Non sono presenti né costi di transazione né tasse.*

L'assunzione di queste ipotesi permette di concludere che la frontiera efficiente stimata è la stessa per tutti gli investitori presenti sul mercato.

Una delle ipotesi fondanti del modello è quella per cui non è previsto un limite all'indebitamento, ad un dato tasso privo di rischio r_f . Considerando un portafoglio di due soli titoli, che include tra essi un titolo privo di rischio, avremo le seguenti caratteristiche in termini di rendimento atteso e varianza:

- $E(r_p) = x \cdot r_f + (1 - x) \cdot E(r_i)$.
- $\sigma_p = (1 - x) \cdot \sigma_i$.

Esplicitando la x dalla seconda equazione e sostituendola alla prima, è possibile ottenere l'equazione di una retta, sul piano media-varianza, che rappresenta la linea dei portafogli ottenibili combinando attività rischiose con attività non rischiose, avente equazione:

$$E(r_p) = r_f + \frac{E(r_i) - r_f}{\sigma_i} \cdot \sigma_i$$

il cui coefficiente angolare è interpretabile come una misura della remunerazione per unità di rischio sostenuto.

Combinando l'investimento in attività prive di rischio con tutti i possibili portafogli di attività rischiose che si trovano sulla frontiera efficiente, accade che una specifica combinazione domina le altre. Essa viene individuata a partire dal punto di tangenza tra la retta che ha intercetta pari a r_f e la frontiera efficiente stessa. Tale

combinazione individua il portafoglio di mercato, un portafoglio perfettamente diversificato, cioè esposto esclusivamente al rischio sistematico. La retta su cui giace il portafoglio di mercato viene detta *Capital Market Line* e costituisce l'insieme dei portafogli che offrono il più alto rendimento atteso per unità di rischio. l'equazione della CML è data da:

$$E(r_p) = r_f + \frac{E(r_m) - r_f}{\sigma_m} \cdot \sigma_p,$$

dove $E(r_m)$ e σ_m rappresentano rispettivamente il rendimento atteso e la volatilità del portafoglio di mercato.

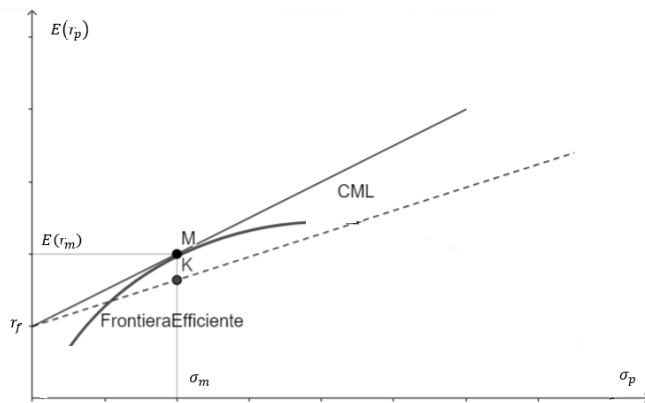


Figura 1.5. la Capital Market Line nel piano media-varianza (σ, μ). Prodotta dall'autore.

1.6.1 Indice di Sharpe

La CML è identica per ogni investitore, dato che una delle assunzioni del modello è quella che ogni individuo abbia le stesse aspettative in termini di rendimenti attesi e di rischio delle attività. Per ognuno di loro il portafoglio ottimale si troverà lungo la CML, sarà quindi composto da una qualche combinazione tra il portafoglio di mercato e titoli risk free e sarà individuato dal punto di tangenza tra la CML e la più alta curva di indifferenza del decisore. Tale retta rappresenta, in una situazione di equilibrio, la miglior combinazione ottenibile tra rendimento atteso e rischio, consentendo di individuare l'ideale livello di rendimento da associare ad ogni grado di rischiosità. Infatti, il coefficiente angolare della CML, misura di quanto deve aumentare il rendimento atteso affinché l'investitore sia disposto ad assumersi un'unità di rischio aggiuntiva, cioè quanto esso venga premiato in termini di rendimento per assumersi del rischio in più. La quantità:

$$\frac{E(r_m) - r_f}{\sigma_m},$$

in letteratura, prende il nome di *indice di Sharpe* e costituisce una misura della performance di portafoglio. È una metrica che misura l'appetibilità di un investimento rischioso dividendo l'eccesso del rendimento atteso rispetto al rendimento di un'attività considerata priva di rischio (comunemente dei titoli di

stato) per la deviazione standard dell'investimento. Pur avendo discreta utilità nella valutazione della bontà di un investimento possiede degli evidenti limiti. L'indice di Sharpe, infatti, non distingue tra volatilità "buona" e "cattiva", intese come oscillazioni positive e negative dei rendimenti. Ad esempio, l'avvento di un rendimento estremamente positivo, avrebbe l'effetto di aumentare la deviazione standard e quindi il denominatore più di quanto possa aumentare il numeratore, facendo scendere il valore dell'indice, nonostante l'evento sia favorevole all'investitore. Inoltre, è un indice che fornisce valutazioni piuttosto scarse quando ci troviamo di fronte a distribuzioni di rendimenti non normali, in maniera particolare per quelle asimmetriche a destra (positivamente).

1.6.2 Indice di Sortino

L'*indice di Sortino* rappresenta una alternativa per molti versi più valida nella valutazione di performance di portafoglio all'indice di Sharpe, specialmente quando si tratta di rendimenti che presentano asimmetrie nella loro distribuzione. L'indice di Sortino si differenzia da quello di Sharpe perché viene presa in considerazione, piuttosto che la deviazione standard, il "*downside risk*", per cui solo quei rendimenti al di sotto di una determinata soglia vengono considerati rischiosi. Lo stesso Markowitz si rese conto, più tardi, nel 1959, che l'utilizzo di una tale misura di variabilità sarebbe stata più consona, ma, nonostante ciò, utilizzò la

varianza per motivi di praticità computazionale. L'indice si presenta in questo modo:

$$S = \frac{(E(r) - T)}{TDD},$$

dove:

$E(r)$ è il rendimento medio nel periodo considerato.

T è il rendimento obiettivo o richiesto dall'investitore. Nei primi lavori di Sortino, T era noto come il *MAR (Minimum Acceptable Return)*, cioè quel livello rendimento al di sotto del quale il soggetto avrebbe considerato un fallimento il proprio investimento.

TDD è la "downside deviation" obiettivo, definita come la media quadratica delle deviazioni delle underperformance dei rendimenti realizzati dal rendimento obiettivo T , dove tutte le deviazioni maggiori di 0 ($X_i > T$) vengono trattate come nulle.

$$TDD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Min}(0, X_i - T))^2},$$

dove:

X_i è l' i -esimo rendimento ed N è il numero totale dei rendimenti.

Rispetto alla classica deviazione standard, nella TDD :

- Vengono misurati gli scostamenti tra i rendimenti target e gli X_i , mentre nella deviazione standard, gli scostamenti sono tra gli X_i e la sua media.

- Tutte le deviazioni maggiori di zero ($X_i - T > 0$) sono poste uguali a 0, ma questi zero vengono comunque inclusi nella somma.

La deviazione standard è una misura di dispersione dei dati attorno alla loro media, sia sopra che sotto, mentre la *TDD* è una misura di scostamento al di sotto di un livello minimo accettabile deciso a priori dal soggetto interessato.

1.7 VAR E CVAR

Il *Value at Risk* rappresenta una misura del rischio della perdita per un investimento. Riesce a stimare, con una data probabilità, l'ammontare della perdita di un investimento, in condizioni normali di mercato e durante un prestabilito intervallo temporale. È una tecnica tipicamente utilizzata dalle aziende per valutare l'importo necessario a coprire eventuali perdite.

Ad esempio, se un portafoglio di azioni ha un VaR ad un giorno al 95% di 1000, significa che esiste una possibilità del 5% che il portafoglio perda almeno 1000 in un giorno. Praticamente, è attesa una perdita superiore di 1000 nel portafoglio una volta ogni 20 giorni. Pur essendo uno strumento utile, semplice, intuitivo e facile

da sottoporre a procedure di backtesting⁹, non è esente da limitazioni, anzi: esso, oltre a non essere una misura coerente di rischio, non fornisce alcuna informazione su cosa possa accadere oltre i limiti del VaR stesso. In sostanza, se con un VaR a un giorno del 95% di 1000 si è sicuri che 95 volte su 100 la perdita giornaliera sarà inferiore a 1000, non vengono fornite informazioni precise su come si distribuisca lungo la coda la perdita (che si sa essere sì superiore a 1000, ma non di quanto) nei restanti 5 casi. Non è in grado di catturare il cosiddetto “*tail effect*”, quel fenomeno per cui possono concentrarsi oltre la soglia del VaR delle perdite significative. Il VaR non fornisce informazioni circa l’aspettativa media delle perdite che superano la soglia ed in certi casi potrebbe valutare scenari potenzialmente catastrofici al pari di scenari più “normali”.

La più recente *Conditional VaR (CVaR o Expected Shortfall o Expected Tail Loss)* supera queste limitazioni. Costituisce una misura di rischio alternativa al VaR che è più sensibile alla forma delle code della distribuzione delle perdite. Il CVaR al 5% rappresenta il rendimento atteso del portafoglio nel peggior 5% dei casi. Stima quindi il rischio di un investimento concentrandosi sulla media delle peggiori performance, considerandone l’entità. Viene considerata una misura di rischio migliore del VaR anche perché coerente con la definizione di rischio.

⁹ Il metodo con il quale viene testata la capacità esplicativa di un modello, attraverso l’applicazione dello stesso su periodi di tempo passati, per i quali sono disponibili le serie storiche ed è possibile giungere a conclusioni certe.

Si introducono formalmente i concetti di VaR e CVaR e ne si osservano nel dettaglio alcune proprietà e differenze.

Sia Y una variabile casuale di costo e sia F_Y la sua funzione di distribuzione $F_Y(u) = P\{Y \leq u\}$. Sia $F_Y^{-1}(v)$ la sua inversa continua a destra $F_Y^{-1}(v) = \inf\{u: F_Y(u) > v\}$. Per un dato livello di α , definiamo il VaR come l' α -quantile:

$$VaR_\alpha(Y) = F_Y^{-1}(\alpha).$$

Il Conditional Value at Risk CVaR si definisce come la soluzione di un problema di ottimo:

$$CVaR_\alpha(Y) := \inf\left\{a + \frac{1}{1-\alpha} E[Y - a]^+ : a \in \mathbb{R}\right\},$$

dove $[Z]^+ = \max(z, 0)$.

Uryasev e Rockafellar (1999) hanno mostrato che il CVaR corrisponde al valore atteso condizionale di Y considerato $Y \geq VaR_\alpha$:

$$CVaR_\alpha(Y) = E(Y | Y \geq VaR_\alpha(Y)).$$

Formulazioni alternative del CVaR sono:

$$CVaR_\alpha(Y) = E(Y|Y \geq F_\alpha^{-1}(\alpha)),$$

$$CVaR_\alpha(Y) = \frac{1}{1-\alpha} \int_\alpha^1 F_\alpha^{-1}(v) dv,$$

$$CVaR_\alpha(Y) = \frac{1}{1-\alpha} \int_{F_\alpha^{-1}(\alpha)}^\infty u dF(u).$$

1.7.1 Proprietà del VaR e del CVaR

Le proprietà delle misure di rischio in generale possono essere formulate, secondo Fishburn (1980) in termini di strutture di preferenze che scaturiscono dalle relazioni di dominanza. Lo vedremo applicato al nostro caso.

Siano Y_1 e Y_2 due variabili casuali:

- *Dominanza stocastica di primo ordine*: affermiamo che la relazione

$$Y_1 \prec_{SD(1)} Y_2$$

vale se e solo se:

$$E[\varphi(Y_1)] \leq E[\varphi(Y_2)],$$

per tutte le funzioni φ monotone e integrabili.

- *Dominanza stocastica di secondo ordine*: affermiamo che la relazione

$$Y_1 \prec_{SD(2)} Y_2$$

vale se e solo se:

$$E[\varphi(Y_1)] \leq E[\varphi(Y_2)],$$

per tutte le funzioni φ concave, monotone e integrabili.

- *Dominanza monotona di secondo ordine*: affermiamo che la relazione

$$Y_1 \prec_{MD(2)} Y_2$$

vale se e solo se:

$$E[\varphi(Y_1)] \leq E[\varphi(Y_2)],$$

per tutte le funzioni φ concave e integrabili.

Si avranno le seguenti conseguenze:

$$Y_1 \prec_{SD(1)} Y_2 \text{ implica } Y_1 \prec_{SD(2)} Y_2$$

$$Y_1 \prec_{SD(2)} Y_2 \text{ che vale se e solo se } Y_1 \prec_{SD(1)} Y_2 \text{ e } Y_1 \prec_{MD(2)} Y_2.$$

$$Y_1 \prec_{SD(2)} Y_2 \text{ è equivalente a } \int_{-\infty}^x F_{Y_1}(u) du \leq \int_{-\infty}^x F_{Y_2}(u) du \text{ per ogni } x.$$

Fatte tali premesse, il CVaR possiede le seguenti proprietà:

- 1) *Il CVaR soddisfa l'invarianza rispetto alla traslazione:*

$$CVaR_\alpha(Y + c) = CVaR_\alpha(Y) + c.$$

- 2) *Il CVaR soddisfa l'omogeneità positiva:*

$$CVaR_\alpha(cY) = cCVaR_\alpha(Y),$$

se $c > 0$

- 3) *Se Y ha densità, risulta che:*

$$E(Y) = (1 - \alpha)CVaR_\alpha(Y) - \alpha CVaR_{1-\alpha}(-Y).$$

- 4) $CVaR_\alpha$ è convesso nel senso che, considerata una coppia di variabili casuali Y_1, Y_2 e un parametro λ tale che $0 < \lambda < 1$, risulta:

$$CVaR_\alpha(\lambda Y_1 + (1 - \lambda)Y_2) \leq \lambda CVaR_\alpha(Y_1) + (1 - \lambda)CVaR_\alpha(Y_2).$$

- 5) $CVaR_\alpha$ è monotono secondo la $SD(2)$ (e di conseguenza anche secondo la $SD(1)$), quindi se $Y_1 \prec_{SD(2)} Y_2$, allora:

$$CVaR_\alpha(Y_1) \leq CVaR_\alpha(Y_2).$$

- 6) $CVaR_\alpha$ è monotono secondo $MD(2)$, quindi se $Y_1 \prec_{MD(2)} Y_2$, allora:

$$CVaR_\alpha(Y_1) \leq CVaR_\alpha(Y_2).$$

Il CVaR viene definita una misura coerente di rischio¹⁰, secondo la definizione data da Artzner, Delbaen, Eber e Heath (1999), perché convessa, monotona secondo la

¹⁰ Secondo Artzner (1999) una metrica per essere considerata misura coerente di rischio deve rispettare gli assiomi di:

- *Monotonicità*: se X ha rendimenti migliori di Y, la misura di rischio deve far sì che sia migliore investire in X piuttosto che in Y
- *Sub-additività* (convessità): il rischio di un portafoglio non è mai maggiore della somma dei rischi dei singoli titoli.
- *Omogeneità positiva*: se aumenta l'investimento in un titolo, anche il rischio dovrà aumentare della stessa entità.

SD(1) e rispettosa delle proprietà di invarianza rispetto la traslazione e di omogeneità positiva. Al contrario, il VaR non è coerente con questa definizione, dato che non è una misura convessa (o sub-additiva). Ciò implica che il VaR di un portafoglio potrebbe essere maggiore della somma del VaR dei suoi singoli componenti, il che va contro il principio base della diversificazione di portafoglio su cui si fonda la Modern Portfolio Theory. D'altra parte, il VaR possiede la proprietà della *comonotonia additiva*:

Def: due variabili casuali Y_1 e Y_2 definite sullo stesso spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ si dicono *comonotone*, se per ogni $\omega, \omega' \in \Omega$,

$$[Y_1(\omega) - Y_2(\omega)][Y_1(\omega') - Y_2(\omega')] \geq 0$$

o, in modo equivalente, Y_1 e Y_2 sono comonotone, se esiste una rappresentazione $Y_1 = f(U)$ e $Y_2 = g(U)$ con f e g che aumentano in maniera monotona e U uniforme in $[0,1]$.

Il VaR possiede le seguenti proprietà:

1) Il VaR_α soddisfa l'invarianza rispetto la traslazione:

$$VaR_\alpha(Y + c) = VaR_\alpha(Y) + c.$$

- *Invarianza rispetto la traslazione:* aggiunta una componente certa α ad una componente aleatoria X , la misura di rischio si riduce di α .

2) Il VaR_α soddisfa l'omogeneità positiva

$$VaR_\alpha(cY) = cVaR_\alpha(Y), \quad \forall c > 0.$$

3) $VaR_\alpha(Y) = -VaR_{1-\alpha}(-Y)$.

4) VaR_α è monotono secondo SD(1) se

$$Y_1 \prec_{SD(1)} Y_2,$$

allora

$$VaR_\alpha(Y_1) \leq VaR_\alpha(Y_2).$$

5) VaR_α rispetta la comonotonia additiva, quindi se Y_1 e Y_2 sono comonotone, allora

$$VaR_\alpha(Y_1 + Y_2) = VaR_\alpha(Y_1) + VaR_\alpha(Y_2).$$

1.7.2 Conclusioni

Analizzando le funzionalità e le principali proprietà delle due misure di rischio è possibile trarre delle conclusioni. Il VaR, pur essendo uno strumento largamente utilizzato nel mondo della finanza per la sua capacità di sintetizzare il rischio in un singolo valore numerico, non è esente da difetti.

Infatti, esso non risulta essere consistentemente una misura coerente di rischio, dato che dei quattro assiomi fondamentali, non rispetta la proprietà della sub-additività, a meno che i rendimenti non seguano una distribuzione ellittica (come la distribuzione normale), opzione che non frequentemente viene verificata, visto che spesso la distribuzione di perdita tende a mostrare “code grasse”. Non rispettando questo assioma, nell’utilizzo del VaR si potrebbe incappare in situazioni in cui un portafoglio poco diversificato potrebbe essere preferito ad uno ben diversificato. Al contrario, il CVaR, rispettando tutti gli assiomi risulta una misura di rischio coerente e maggiormente aderente ad una generica situazione. Inoltre, il VaR, non misura con precisione la grandezza delle perdite potenziali nel momento in cui si verificano eventi di coda, a differenza del CVaR, che ci dà una misura reale, seppur media, dell’entità delle perdite derivanti dai fenomeni di coda.

CAPITOLO 2: NUOVE TECNICHE DI OTTIMIZZAZIONE DI PORTAFOGLIO PER RIDURRE L'EFFETTO DELLE CODE

2.1 EDC e EDH

Un più recente approccio verso la direzione della misurazione del rischio di coda è avvenuto con le misure di *Extreme Downside Correlation* (EDC) e *Extreme Downside Hedge*, proposte in Harris et al. (2019), proprio nel tentativo, come avvenuto con il VaR e il CVaR, di superare i limiti dell'utilizzo della covarianza come misura di rischio in quanto incapace di catturare il downside risk. Vengono presentate queste due nuove misure di rischio sistematico¹¹ di coda, costruite utilizzando due differenti approcci. La prima, l'EDC, è una misura non parametrica che quantifica la tendenza di un titolo a crollare in termini di prezzo simultaneamente ad un crollo di mercato. La seconda, l'EDH, misura la necessità degli investitori di coprirsi contro gli eventi (estremi) di coda. Entrambe le misure sono associate ad un premio per il rischio¹² positivo.

¹¹ È il rischio associato all'andamento del mercato nel suo complesso. Si contrappone al rischio specifico o idiosincratico, legato al singolo titolo.

¹² Differenziale tra il rendimento atteso di un'attività ed un'attività priva di rischio. Esso è pari, inoltre, alla differenza tra il rendimento atteso del mercato e l'attività risk free, moltiplicata per il beta del titolo rischioso.

2.1.1 Extreme Downside Correlation

Un asset che performa quando il mercato crolla risulta più appetibile, a differenza di uno che subisce forti variazioni negative di prezzo e per cui un investitore tenderà a chiedere un premio per il rischio più alto per detenere questi ultimi.

Su questo concetto si basa la *LTD*, misura di Chabi-Yo et al. (2018), che quantifica la propensione di un asset ad affrontare eventi di coda dato il presentarsi di un evento di coda del mercato (un crollo). È una misura computazionalmente complessa e per questo non possiede un particolare utilizzo pratico. Viene proposta quindi una misura non-parametrica più semplice per misurare il rischio sistematico di coda, basata sui comomenti¹³ tra i rendimenti dell'asset e i rendimenti di mercato, adattati in modo che ci si focalizzi sulla coda della distribuzione congiunta dei rendimenti. Il *k-esimo* momento misto dell'asset *i* è dato da:

$$k - \text{esimo comomento dell'asset } i = \frac{E[(R_i - \mu_i)(R_m - \mu_m)^{k-1}]}{\text{Var}(R_i)^{\frac{1}{2}} \text{Var}(R_m)^{\frac{k-1}{2}}}$$

L'approccio modifica il dominio dei comomenti, passando dall'intera distribuzione all' α -quantile della distribuzione (considero cioè solo i rendimenti fino a quel quantile sinistro). Per $k = 2$ avremmo così l'EDC, definito come:

¹³ I comomenti sono una famiglia di statistiche di ordine superiore utilizzate per descrivere la struttura e la distribuzione dei dati. Sono una generalizzazione dei momenti, che sono le statistiche di ordine inferiore utilizzate per descrivere posizione e varianza di un insieme di dati.

$$EDC = \frac{E \left[(R_i - \mu_i)_\alpha (R_m - \mu_m)_\alpha \right]}{E[(R_i - \mu_i)_\alpha^2]^{1/2} E[(R_m - \mu_m)_\alpha^2]^{1/2}}$$

dove R_i e R_m rappresentano l'extra-rendimento dell'asset i e del mercato, μ_i e μ_m le rispettive medie, $(R_i - \mu_i)_\alpha$ è uguale a $(R_i - \mu_i)$ se R_i è più piccolo del suo α -quantile e 0 altrimenti, così come $(R_m - \mu_m)_\alpha$.

Un asset che tende a crollare in prezzo quando il mercato attraversa un momento negativo avrà un alto EDC e viceversa.

2.1.2 Extreme Downside Hedge

La seconda misura sviluppata è l'EDH. Essa si basa sul principio che gli investitori sono capaci di coprirsi dagli eventi estremamente sfavorevoli del mercato e che ogni asset che funge da copertura per questo tipo di rischio dovrebbe essere tanto richiesto quanto sono le probabilità che il mercato crolli. L'EDH viene stimato regredendo i rendimenti dell'asset su una misura di rischio di coda. Per questa misura viene usato il CVaR, piuttosto che il VaR, perché misura coerente di rischio. Il rischio sistematico di coda viene stimato seguendo la seguente regressione:

$$R_{i,t} = c_i + EDH_i \times \Delta CVaR_{m,t} + \varepsilon_{i,t},$$

dove $R_{i,t}$ rappresenta l'extra-rendimento dell'asset i nel giorno t , e c_i e $\varepsilon_{i,t}$ sono rispettivamente, l'intercetta e l'errore. L'EDH costituisce il coefficiente che quantifica la risposta del rendimento dell'asset a cambiamenti nel rischio di coda del mercato. Il CVaR, in questo caso, viene calcolato come il quantile (non negativo, come avviene normalmente) della distribuzione standardizzata dei rendimenti¹⁴ e quindi un basso CVaR implica un alto rischio. Un asset con un alto EDH offre quindi bassi rendimenti quando aumenta il rischio nelle code, e di conseguenza gli investitori sarebbero portati a chiedere un maggior rendimento medio per detenere uno strumento finanziario di quel tipo. Dall'altra parte, un asset con un basso EDH funge da copertura contro gli aumenti dei rischi di coda nel mercato, quindi gli investitori sarebbero disposti a "pagare" in termini di rendimento per questa caratteristica. Uno dei vantaggi dell'approccio EDH è che agendo sui CVaR giornalieri, può richiedere solamente un mese di osservazioni per ottenere una stima robusta.

¹⁴ Si assume che i residui dei rendimenti di mercato si distribuiscano come una "Skewed t-Student" di Hansen (1994).

2.2 PROPRIETA', ROBUSTEZZA E PREDITTIVITA' DI EDC E EDH

2.2.1 Proprietà

Nel paper di riferimento¹⁵, viene dimostrata, utilizzando portafogli composti da stock del NYSE e del NASDAQ dai vari rendimenti, ordinati secondo un ordine crescente in termini di rischio espresso da EDC e EDH, una significativa relazione positiva tra le misure di rischio sistematico di coda e rendimenti medi. All'aumentare quindi della rischiosità di coda, aumenta in maniera quasi monotona il rendimento offerto dal portafoglio. Inoltre, si evidenzia che gli asset a cui sono associate alte misure di rischio, hanno una capitalizzazione media maggiore e che, cambiando di una deviazione standard le misure di EDC e EDH, si osservano scostamenti maggiori nei rendimenti medi rispetto ai casi in cui si variasse di una deviazione standard altri parametri più canonici di misurazione del rischio come beta, book-to-market, rendimenti passati, volatilità o co-simmetria¹⁶.

2.2.2 Robustezza

Eseguendo poi dei check sulla robustezza delle misure, si verifica che:

¹⁵ L'analisi empirica da cui si estrapolano valutazioni circa le proprietà, la robustezza e la predittività delle misure di EDC e EDV viene effettuata in HARRIS RD, NGUYEN LH, STOJA E (2019) *Systematic extreme downside risk*, "Journal of International Financial Markets, Institutions and Money".

¹⁶ Quest'ultima valutazione viene fatta sulla base della regressione Fama-MacBeth (1973), per riportare l'influenza del rischio di coda sui rendimenti, rispetto ai canonici fattori di rischio utilizzati nella regressione (beta, capitalizzazione, book-to-market, ecc.).

- Alterando gli alfa quantili, passando da un 5% ad un 1% o 10%, rimane la tendenza degli stock a crollare assieme al mercato. Il limite rimane quello per cui, considerando valori soglia troppo bassi, l'EDC viene misurato su un insieme troppo ristretto di informazioni. L'EDH, al contrario, non soffre di questo problema, dato che possiamo ottenere CVaR per qualsiasi quantile.
- Stimando la CVaR del mercato con differenti metodologie, quali GARCH, GJR-GARCH e EGARCH¹⁷, oppure utilizzando il VaR al posto del CVaR stesso, la misura di EDH continua ad avere una relazione positiva e statisticamente significativa con i rendimenti. Ciò significa che i risultati, per quanto concerne l'EDH, sono robusti alle modificazioni dei modelli per misurare il rischio di coda di mercato.

2.2.3 Predittività

Viene infine esaminata la possibilità di usare EDC e EDH come misure predittive per i rendimenti futuri degli stock. Per fare ciò, viene stimata la misura di rischio all'inizio di ogni mese usando i dati dei precedenti 12 mesi e viene previsto, tramite un'apposita regressione, l'extra-rendimento (rispetto al rendimento dei buoni del Tesoro Americano ad 1 anno) dello stock nel mese successivo. Viene ripetuta

¹⁷ Sono vari tipi di modelli autoregressivi generalizzati ad eteroschedasticità condizionata, molto usati nella finanza per modellare serie storiche che esibiscono varianza non costante nel tempo.

questa regressione per ogni mese ed esaminata la serie storica dei coefficienti. L'EDC prevede positivamente i rendimenti del mese successivo, dato che si osserva che il coefficiente stimato è altamente significativo. Quello relativo all'EDH invece, non lo è. Lo diventa dal momento in cui si eliminano dal dataset le migliori 10% e le peggiori 10% stock in termini di EDH. Questo perché l'EDH subisce variazioni significative per stock con moltissima o pochissima esposizione al rischio di coda. Con questo metodo anche il coefficiente dell'EDH risulta significativo. Esso quindi, possiede potere predittivo non per tutto il cross-sectional dataset, ma per buona parte dei dati, a differenza dell'EDC, che rimane una misura più flessibile in termini di predittività. Inoltre, attraverso uno studio dell'EDH su diversi orizzonti temporali, visto che utilizzando il CVaR giornaliero può essere stimato anche per serie mensili, si nota come stock con alti rischi di coda in un solo mese tendono ad avere bassi rendimenti, mentre lungo un orizzonte temporale più ampio, offrono rendimenti migliori; quindi, serve del tempo prima che l'investitore che si sia accollato tale rischio venga ripagato. In termini più pratici, il trade-off tra un rendimento positivo e la presenza di alto rischio di coda necessita di almeno 6 mesi per materializzarsi, mentre su periodi brevi, si osserva una relazione negativa tra rischio di coda e rendimento.

2.3 UNA NUOVA METRICA PER LA MISURAZIONE DEL RISCHIO

Nonostante queste misure siano capaci di catturare la componente di variabilità che effettivamente viene considerata dall'individuo come rischio, cioè quella appartenente alle code sinistre della distribuzione, soffrono di alcuni difetti. Quello principale, è che sono piuttosto insensibili alla forma delle code, visto che dipendono fortemente dalla scelta a priori che è necessaria compiere sui quantili. Si propone quindi, una misura di variabilità per il rischio di coda che supera questo limite, basandosi sulla procedura di stratificazione di Mariani (2020). Questa metrica ha il vantaggio di essere definita in maniera endogena, senza alcuna scelta a priori, e cattura il rischio dalla volatilità degli asset e dalle distribuzioni delle code sinistre dei rendimenti degli asset. La nuova misura di variabilità consiste in una combinazione lineare convessa tra la classica matrice varianza-covarianza ed una nuova matrice varianza-covarianza, definita “left-tail-covariance-like matrix” che tiene conto solamente della variabilità relativa alle code delle distribuzioni dei rendimenti.

In seguito, verrà spiegata più nel dettaglio la metodologia utilizzata per il calcolo di questa nuova misura di variabilità.

2.3.1 Clustering con l'Expectation Maximization

Per il calcolo di questa nuova misura di variabilità e per applicare la procedura di stratificazione sopracitata, si parte dai rendimenti lordi, cioè l'esponenziale dei log-rendimenti degli asset della relativa serie temporale dei prezzi. Sia $N > 0$ il numero degli asset, per $i = 1, 2, \dots, N$, il rendimento lordo associato con l' i -esimo asset al tempo t sarà:

$$y_{t,i} = \exp \left\{ \log \left(\frac{p_{t,i}}{p_{t-\Delta t,i}} \right) \right\} = \frac{p_{t,i}}{p_{t-\Delta t,i}}, \quad \text{per } i = 1, 2, \dots, N,$$

dove $p_{t,i}$ rappresenta il prezzo giornaliero dell' i -esimo asset al tempo t e Δt è l'intervallo temporale in cui i prezzi vengono osservati. L'utilizzo dei rendimenti lordi è conveniente perché:

- Sono strettamente positivi ma comunque legati ai rendimenti "classici".
- Oscillano intorno al valore 1.
- I rendimenti lordi di asset che appartengono a diverse asset class mostrano un comportamento costante nel tempo, rendendoli degli ottimi candidati per processi stazionari autoregressivi (AR)¹⁸.

¹⁸ È un tipo di processo stocastico, rappresenta un modello lineare in cui le variabili dipendenti sono i ritardi della variabile indipendente. Si presenta nella forma $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_n z_{t-n} + \varepsilon_t$.

Iniziando quindi dai rendimenti lordi osservati, la procedura di stratificazione procede in maniera iterativa, dividendo un selezionato sottoinsieme dei rendimenti in due gruppi distinti ad ogni iterazione. La funzione di densità dell'insieme scelto è estratta da una mistura di log-normali¹⁹ e la divisione avviene quando viene identificato un rendimento per cui si osserva un “significativo” cambiamento nella mistura. In maniera più specifica, nella prima iterazione, la procedura parte dall'insieme di tutti i rendimenti $S_n = \{y_1, y_2, y_3, \dots, y_n\}$ ordinati in maniera ascendente. Viene poi identificato un valore soglia a^1 , cioè il primo “change point” per dividere l'insieme S_n in due gruppi. Il gruppo di sinistra $K_1 = \{\min(y \in S_n; y \in (0, a^1])\}$, composto da rendimenti più piccoli o uguali al valore soglia ed il gruppo di destra R_1 composto da valori maggiori del valore soglia. Nella seconda iterazione, la procedura considera l'insieme R_1 ottenuto nella prima iterazione, identifica un nuovo valore soglia $a^2 > a^1$ e divide a sua volta R_1 in due gruppi separati, $K_2 = \{\min(y \in S_n; y \in (a^1, a^2])\}$ e $R_2 = \{\min(y \in S_n; y \in (a^2, +\infty))\}$. Tutto ciò si ripete fino alla k -esima iterazione.

Si definisce poi la funzione $g_k(x)$ come la funzione di densità associata al campione dei rendimenti traslato $\tilde{R}_{k-1} = \{\min(x = y - a^{k-1}; y \in R_{k-1})\}$, assumendo che la funzione sia data da una mistura di due funzioni di densità log-normali:

¹⁹ È una distribuzione mista, che combina due diverse distribuzioni log-normali, ognuna avente propria media e deviazione standard, con pesi che tipicamente sommano a uno.

$$g_k(x) = \pi_k f_{1,k}(x) + (1 - \pi_k) f_{2,k}(x) \quad \text{con } x \in \mathbb{R},$$

dove $f_{1,k}(x)$ e $f_{2,k}(x)$ rappresentano le densità log-normali, aventi parametri $\mu_{1,k}$, $\mu_{2,k}$, $\sigma_{1,k}$, $\sigma_{2,k}$, associate alle due componenti della mistura, rispettivamente quella di sinistra e quella di destra (quindi $\mu_{1,k} < \mu_{2,k}$) e π_k rappresenti la probabilità a priori, compresa tra 0 e 1, che il punto del campione sia estratto dalla prima componente $f_{1,k}(x)$. Il vettore dei parametri da stimare è rappresentato quindi da $\Theta_k = (\pi_k, \mu_{1,k}, \mu_{2,k}, \sigma_{1,k}, \sigma_{2,k})$ e viene calcolato, ad ogni k -esimo step, a partire dai rendimenti dell'insieme R_{k-1} attraverso l'algoritmo "Expectation Maximization".

Le assunzioni per cui l'algoritmo converga, sono quelle per cui:

- Non ci sono intervalli temporali in cui i prezzi sono costanti.
- Non ci sono outlier estremi.
- Il campione è sufficientemente ampio.

Il valore soglia della k -esima iterazione viene determinato usando la seguente regola

$$a^k = \min\{y \in R_{k-1}; \pi_k f_{1,k}(y - a^{k-1})\}.$$

a^k divide il campione in due sottogruppi con distribuzioni diverse. La procedura dell'algoritmo termina quando esso non riesce più a determinare un nuovo "change

point”. Un'altra informazione calcolata dall'algoritmo ad ogni iterazione è il “misidentification error”, cioè la probabilità di classificare, in maniera erranea, i rendimenti sotto il valore soglia come membri del gruppo di destra. Esso è calcolato come:

$$F_2^k(a^k - a^{k-1}),$$

dove $F_2^k(x)$ è la distribuzione cumulata associata alla seconda componente $f_{2,k}(x)$ della mistura. Viene anche calcolata la probabilità di classificare incorrettamente i rendimenti sopra il valore soglia, cioè

$$1 - F_1^k(a^k - a^{k-1}),$$

dove F_1^k è la distribuzione cumulata associata alla prima componente $f_{1,k}(x)$ della mistura.

2.3.2 Nuovo data set informativo e definizione delle misure di similarità e volatilità delle code

Al termine della procedura, viene creato un nuovo data set informativo che include il vettore dei parametri $\Theta_k, k = 1, 2, \dots$, i “misidentification errors” e i “change points” per ogni asset (viene reputato sufficiente considerare i risultati provenienti dalle prime due iterazioni $k = 1, 2, \dots$, visto che è di interesse determinare i parametri

che caratterizzano la distribuzione di sinistra di ogni asset) e che costituirà l'elemento chiave per costruire la nuova misura di rischio.²⁰

In questo modo, al tempo t ogni asset sarà identificato da 12 parametri:

- 1) I “change points” delle distribuzioni dei rendimenti determinati nelle prime due iterazioni: $a_{t,i}^1, a_{t,i}^2$.
- 2) I parametri (media e volatilità) della componente di sinistra della mistura determinati nelle prime due iterazioni: $\sigma_{1,t,i}^1, \mu_{1,t,i}^1, \sigma_{1,t,i}^2, \mu_{1,t,i}^2$.
- 3) L'errore di prima e di seconda specie, cioè il complemento della distribuzione cumulata associata alla componente di sinistra della mistura e la distribuzione cumulata associata alla componente di destra della mistura, nella prima e nella seconda iterazione: $1 - F_{1,t,i}^1, 1 - F_{1,t,i}^2, F_{2,t,i}^1, F_{2,t,i}^2$.
- 4) Le probabilità a priori associate alla mistura nella prima e nella seconda iterazione della procedura: $\pi_{t,i}^1, \pi_{t,i}^2$.

Per ogni tempo t , i 12 parametri calcolati per ogni asset i , vengono raccolti nella matrice X_t , definita come il *Set Informativo delle code sinistre al tempo t* .

²⁰ Questo data set informativo può essere calcolato solamente se:

- 1) La distribuzione dei rendimenti lordi può essere approssimata da una mistura di due log-normali.
- 2) L'algoritmo EM converge.

Viene quindi definita la *Matrice di correlazione delle code sinistre*, dati gli asset i e j al tempo t identificati rispettivamente con il loro data set informativo $\underline{x}_{t,i}$ e $\underline{x}_{t,j}$ (la i -esima e j -esima riga della matrice X_t):

$$\bar{I}_{t,i,j} = \bar{\rho}_{t,i,j} = \frac{\underline{x}'_{t,i} \underline{x}'_{t,j}}{\|\underline{x}_{t,i}\| \|\underline{x}_{t,j}\|^{21}} \quad \text{con } i, j = 1, 2, \dots, N.$$

La matrice di correlazione in questione è quindi una matrice di similarità ed i coefficienti al suo interno misurano la similarità nelle distribuzioni delle code sinistre dell' i -esimo e j -esimo asset al tempo t . Più è alto il valore del coefficiente e maggiore sarà il rischio di comportamenti simili nelle code sinistre degli asset presi in considerazione. Ciò significa che allocare asset con rischi di coda troppo simili, esporrebbe l'investitore ad un maggiore rischio nel caso di importanti shock e crisi esterne, aumentando quindi l'esposizione complessiva del portafoglio.

La volatilità della coda sinistra viene definita come:

$$\bar{\sigma}_{t,i} = \sqrt{\frac{\pi_{t,i}^1}{1 - \pi_{t,i}^1} \sigma_{1,t,i}^1} \quad \text{con } i = 1, 2, \dots, N,$$

²¹ $\|z\|$ identifica la norma euclidea del vettore z .

dove $\sigma_{1,t,i}^1$ è la volatilità della componente di sinistra della mistura di log-normali.

La quantità sotto radice $\frac{\pi_{t,i}^1}{1-\pi_{t,i}^1}$, rappresenta un odds-ratio che ci fornisce informazioni circa il peso di questa componente della mistura rispetto all'altra. Più è alto il valore di $\pi_{t,i}^1$ più sarà importante nel calcolo il ruolo della componente di sinistra.

Da queste due misure si costruisce la *Matrice di covarianza delle code sinistre*, definita come:

$$\bar{C}_{t,i,j} = \bar{\sigma}_{t,i} \bar{\rho}_{t,i,j} \bar{\sigma}_{t,j} \quad \text{con } i = 1, 2, \dots, N^{22}.$$

2.3.3 Definizione della misura di variabilità

La misura di variabilità finale viene definita come una combinazione lineare convessa della classica matrice varianza-covarianza e questa nuova misura di covarianza che tiene conto del rischio dovuto alla similarità e alla volatilità delle code sinistre. Impostando il classico problema di minimizzazione del rischio di Markowitz, si avrà che il rendimento del portafoglio sarà una variabile casuale data da:

$$\Pi_{\underline{w}} = \sum_{i=1}^N w_i \log Y_i,$$

²² La matrice risulta essere semi-definita positiva.

dove Y_i è la variabile casuale che rappresenta i rendimenti lordi degli asset, \underline{w} è il vettore dei pesi, che sono non negativi (non sono ammesse quindi vendite allo scoperto) e la cui somma è 1. Il problema di asset allocation sarà formulato come un problema di ottimizzazione quadratica lineare, il cui obiettivo è quello di minimizzare la misura di rischio $R()$ dato un determinato livello di rendimento ed a patto che i pesi sommino ad 1.

$$\begin{cases} \min_{\underline{w}} R(\Pi_{\underline{w}}) \\ \sum_{i=1}^N w_i \log Y_i = \bar{y} \\ \sum_{i=1}^N w_i = 1 \quad \text{con } w_i \geq 0 \end{cases} .$$

Nel caso del portafoglio di Markowitz, la funzione obiettivo del problema sarà la varianza di portafoglio. In questo caso, ciò che si andrà a minimizzare, mantenendo la struttura del problema di ottimo, sarà l'equazione:

$$\underline{w}^T C_{\alpha} \underline{w},$$

dove $C_{\alpha} = \alpha C + (1 - \alpha)\bar{C}$ con $0 \leq \alpha \leq 1$,

C_α è la combinazione lineare convessa tra la matrice varianza-covarianza C e la matrice di covarianza delle code sinistre \bar{C} (vengono eliminati i pedici per semplicità di notazione). In questo modo, a seconda dell' α scelto, si andrà a minimizzare una misura di rischio che tiene conto sia della volatilità sia della covarianza tra gli asset in senso assoluto, ma anche della propensione degli asset ad avere simili comportamenti nelle code sinistre ed alla volatilità di queste. Questa misura, infatti, può essere vista come una misura della vulnerabilità del portafoglio a shock di mercato. La presenza di numerosi asset con alta volatilità ed al contempo alta similarità e volatilità nelle code potrebbe implicare un'alta esposizione del portafoglio visto che un crollo di mercato potrebbe causare un effetto a cascata. Minimizzando questa misura per un determinato livello di α , ovviamente compreso e diverso da 0 e 1, verrà diversificato il portafoglio tentando di ridurre il più possibile gli effetti dei comportamenti di coda degli asset.

2.4 EXPECTATION MAXIMIZATION

L'Expectation Maximization è un metodo iterativo utilizzato per trovare stime di parametri in modelli statistici che dipendono da variabili "latenti", cioè non osservate. L'iterazione dell'algoritmo si alterna tra due fasi:

- *Expectation step*: viene calcolato il valore atteso della log-verosimiglianza dato l'insieme dei valori osservati ed una prima stima dei parametri.
- *Maximization step*: vengono calcolati dei nuovi parametri massimizzando la funzione di log-verosimiglianza attesa trovata al punto precedente.

I parametri trovati vengono poi utilizzati per determinare una nuova funzione di log-verosimiglianza nel prossimo Expectation step. Ad esempio, una mistura può essere descritta dall'algoritmo assumendo che ogni punto osservato ha una corrispondente variabile latente che specifica la componente della mistura a cui il punto appartiene.

2.4.1 Breve definizione matematica

Dato un modello statistico che genera un insieme X di dati osservabili, un insieme di dati non osservabili o latenti Z e un vettore di parametri sconosciuti ϑ , assieme ad una funzione di verosimiglianza $L(\vartheta; X, Z) = p(X, Z|\vartheta)$, la stima di massima verosimiglianza dei parametri è determinata massimizzando la verosimiglianza marginale dei dati osservati,

$$L(\vartheta; X) = p(X|\vartheta) = \int p(X, Z|\vartheta)dZ = \int p(X|\vartheta, Z)p(Z|\vartheta)dZ.$$

Tuttavia, questa quantità risulta intrattabile dato che Z non è osservabile e la sua distribuzione è sconosciuta prima di calcolare ϑ . L'algoritmo EM cerca la stima di massima verosimiglianza applicando iterativamente questi due passi:

- *Expectation step*: definisce $Q(\vartheta|\vartheta^{(t)})$ come il valore atteso della log-verosimiglianza di ϑ , rispetto alla corrente distribuzione condizionale di Z dato X e le correnti stime dei parametri ϑ .

$$Q(\vartheta|\vartheta^{(t)}) = E_{Z|X, \vartheta^{(t)}}[\log L(\vartheta; X, Z)].$$

- *Maximization step*: trova i parametri che massimizzano la quantità:

$$\vartheta^{(t+1)} = \arg \max_{\vartheta} Q(\vartheta|\vartheta^{(t)}).$$

Nel comune caso in cui sia Z che ϑ siano sconosciuti, prima si inizializzano i parametri di ϑ su dei valori casuali, viene calcolata la probabilità di ogni possibile valore di Z dato ϑ , e poi questi valori vengono usati per calcolare una migliore stima dei parametri di ϑ . Tutto ciò, inizializzazione a parte, viene ripetuto fino a convergenza dell'algoritmo.

2.4.2 Expectation Maximization con una mistura gaussiana monodimensionale

Un modo più semplice di vedere l'EM è quello di considerarlo come un metodo di clustering. Esso è applicabile facilmente ad una mistura (normale, log-normale, multinomiale, ecc.) perché ogni cluster può essere associato ad una distribuzione di probabilità della mistura e l'algoritmo in questione può aiutare a trovare i parametri di ognuna delle distribuzioni della mistura.

Supponendo di avere x_n dati, di cui non si conoscono né media μ né deviazione standard σ , ma di cui si sa siano provenienti da due distribuzioni normali differenti.



Figura 2.1²³. Prodotta dall'autore.

Questa ultima informazione permette di ricavare con certezza media e varianza e stimare le due gaussiane



Figura 2.2. Prodotta dall'autore.

Ma come si procede se non si conosce a priori la provenienza dei dati?

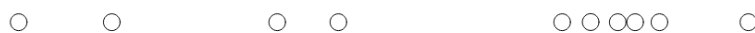


Figura 2.3. Prodotta dall'autore.

²³ Le figure dalla 2.1 alla 2.7 sono state prodotte con GeoGebra Classic.

In questo caso si possiede un insieme di dati, che si presume provengano da due distribuzioni differenti, ma non si conosce con esattezza quale punto appartiene a quale distribuzione. È possibile però stimare la probabilità che un punto appartenga ad una gaussiana piuttosto che all'altra, costruendole proprio attraverso l'algoritmo EM. Esso inizia calcolando in maniera del tutto casuale due distinte distribuzioni con due diverse coppie di parametri (μ, σ^2) e per ogni punto del data set, stabilisce se è più probabile che esso provenga da una distribuzione piuttosto che dall'altra, assegnandolo²⁴ all'una o all'altra in base alla probabilità calcolata.

Una volta assegnati i punti, vengono ricalcolati i parametri delle distribuzioni, sulla base dei nuovi dati a disposizione. A questo punto si ricomincia con il processo di assegnazione fino a convergenza dell'algoritmo.

Si inizia con dei dati distribuiti casualmente, di cui non si conosce l'esatta provenienza ma che si sa con certezza appartenere a due distribuzioni normali. Si posizionano casualmente due distribuzioni normali, la nera e la rossa.

²⁴ In questo caso non abbiamo un "hard assignment" (come avviene ad esempio nel k-Means), ma un "soft assignment". Ciò significa che una volta stabilito che quel punto provenga dalla distribuzione 1, piuttosto che dalla 2, non è detto che nelle successive iterazioni questa decisione possa essere cambiata. La probabilità assegnata che il punto provenga da una specifica distribuzione non è mai 0 o 1, ma sempre compresa tra 0 e 1.

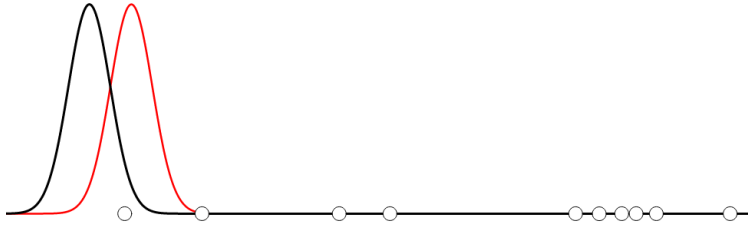


Figura 2.4. Prodotta dall'autore.

Nel primo step dell'algoritmo, si vorrà capire quanto è probabile che ognuno dei punti venga dalla distribuzione nera o rossa. La probabilità che un punto x_i venga dalla distribuzione rossa verrà calcolata come:

$$P(x_i|b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(x_i - \mu_b)^2}{2\sigma_b^2}\right).$$

In questo caso, la probabilità che, ad esempio, il punto più a destra provenga dalla distribuzione rossa è molto bassa, ma è ancora più bassa quella che provenga dalla distribuzione nera, trovandosi ancora più a sinistra.

Quindi, calcolando probabilità a posteriori²⁵, assumendo che le probabilità a priori²⁶ siano uguali, la probabilità che quel punto venga dalla rossa sarà quasi 1.

²⁵ La probabilità che un evento A accada condizionato al fatto che un altro evento B è già accaduto (o si ha la certezza che accada).

$$b_i = P(b|x_i) = \frac{P(x_i|b)P(b)}{P(x_i|b)P(b) + P(x_i|a)P(a)}$$

La b_i verrà calcolata per ogni punto, così come il suo complemento ad 1, cioè la probabilità a posteriori che il punto provenga dalla nera.

$$a_i = P(a|x_i) = 1 - b_i.$$

Questi numeri rappresentano graficamente una colorazione dei punti. Più un punto è colorato, più la probabilità sarà polarizzata verso una distribuzione piuttosto che un'altra.



Figura 2.5. Prodotta dall'autore.

Dopo questa fase, è possibile ricalcolare le distribuzioni (calcolando media e varianza) sulla base dei nuovi punti. La media verrà calcolata sommando i prodotti tra la massa dei singoli punti e la loro probabilità a posteriori.

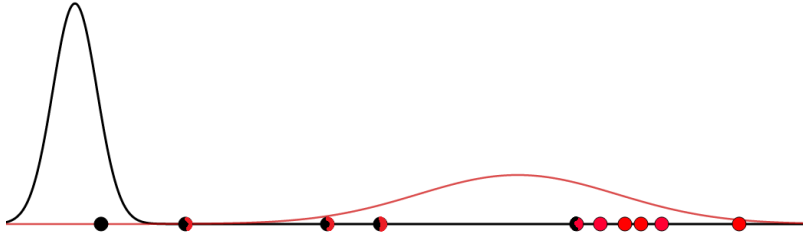


Figura 2.6. Prodotta dall'autore.

La gaussiana rossa si muoverà sensibilmente verso destra, visto che al primo step la maggior parte dei punti della destra erano stati assegnati alla gaussiana rossa.

A questo punto l'algoritmo reitererà la procedura, riassegnando le probabilità ai punti sulla base delle nuove distribuzioni e ricalcolando nuove distribuzioni.

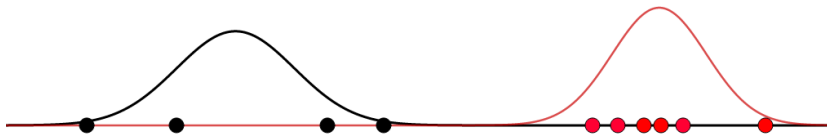


Figura 2.7. Prodotta dall'autore.

CAPITOLO 3: ANALISI EMPIRICA DI QUATTRO DIFFERENTI PORTAFOGLI

In questo capitolo, verranno messe a confronto le performance di quattro diversi portafogli calcolati prendendo in considerazione: la misura di rischio bi-dimensionale presentata nel capitolo precedente, l'Extreme Downside Correlation e gli indici di Sharpe e di Sortino. A tal scopo, viene utilizzato un dataset di 3174 osservazioni giornaliere di 44 ETF.

3.1 DESCRIZIONE DEI DATI

Si considera un dataset composto dalle serie storiche dei prezzi di 44 ETF relative al periodo gennaio 2006 - febbraio 2018, per un totale di 3174 osservazioni. La Tabella 3.1 mostra gli ETF classificati in 5 categorie, secondo la classificazione fornita da Borsa Italiana, mettendone in evidenza valore minimo, massimo, media, deviazione standard, curtosi e simmetria. Da notare come media e deviazione standard siano diverse per ogni asset class, ma i valori relativi agli asset appartenenti alle categorie "Emerging" risultano sensibilmente più grandi, sottolineandone le maggiori potenzialità in termini di rendimento e di volatilità. La Figura 3.1 mostra l'andamento dei prezzi degli ETF, suddivisi in base alla loro categoria. Da notare la presenza di *FXC IM* e *BRA IM*, rispettivamente nelle categorie *Emerging Equity*

Asia ed *Emerging Equity America* aventi un valore decisamente più alto rispetto la media. Nella Tabella 3.2 vengono mostrate invece le statistiche riassuntive dei log-rendimenti. I valori medi si approssimano non sorprendentemente a 0, considerato il lungo periodo di osservazione, così come le deviazioni standard, mentre osserviamo dei valori di curtosi piuttosto elevati per le *categorie Emerging Equity Asia* ed *Emerging Equity America*, proprio a causa della loro natura, per cui osservare rendimenti che si discostano sensibilmente dalla media è molto più frequente rispetto alle altre categorie d'investimento.

Tabella 3.1, Statistiche riassuntive per i prezzi degli ETF. Prodotta dall'autore.

CLASSE ETF	MIN	MAX	MEDIA	DEV.STD.	CURTOSI	SIMMETRIA
AGGREGATE BONDS	139.07	492.86	249.24	104.63	-0.19	1.14
COMMODITIES	64.81	1722.22	387.47	309.05	2.94	1.89
CORPORATE BONDS	133.95	248.70	187.53	32.42	-1.22	0.14
EMERGING EQUITY ASIA	81.81	38909.60	2503.96	6314.18	9.45	3.23
EMERGING EQUITY AMERICA	81.81	87429.27	7973.74	18481.67	4.18	2.41

Nella Figura 3.2 vengono mostrati i *boxplot* per ogni prezzo di ETF appartenente alla rispettiva categoria d'investimento. Si nota anche qui la differenza, specialmente per le classi emergenti, tra i prezzi all'interno della stessa classe,

spiegando così gli elevati valori di deviazione standard nella Tabella 3.1. La Figura 3.3 mostra invece i boxplot relativi ai log-rendimenti. Vediamo qui come il passaggio da prezzo a rendimento ci permette di eliminare il trend stagionale tipico delle serie storiche di lungo periodo e di rendere le serie stazionarie intorno allo zero.

Si dimostra che il set informativo delle code ottenuto da queste serie storiche, utilizzato per calcolare la nuova matrice di similarità delle code, è attendibile, nel senso che ogni variabile utilizzata per descrivere le code degli asset (change points, media e volatilità delle code, probabilità a priori, ecc.) risulta debolmente correlata l'una con l'altra²⁷. Inoltre, applicando un metodo di clustering (utilizzando una distanza calcolata a partire dal set informativo delle code), il dataset degli ETF viene classificato in maniera coerente alla categorizzazione fatta da Borsa Italiana.²⁸

²⁷ La correlazione tra le variabili del set informativo delle code viene verificata attraverso una Principal Component Analysis.

²⁸ Per maggiori informazioni riguardo i metodi di clustering si rimanda alla lettura del paper di riferimento: “F.Mariani, G.Polinesi, M.C.Recchioni (2020). A tail-revisited Markowitz mean-variance approach and a portfolio network centrality”.

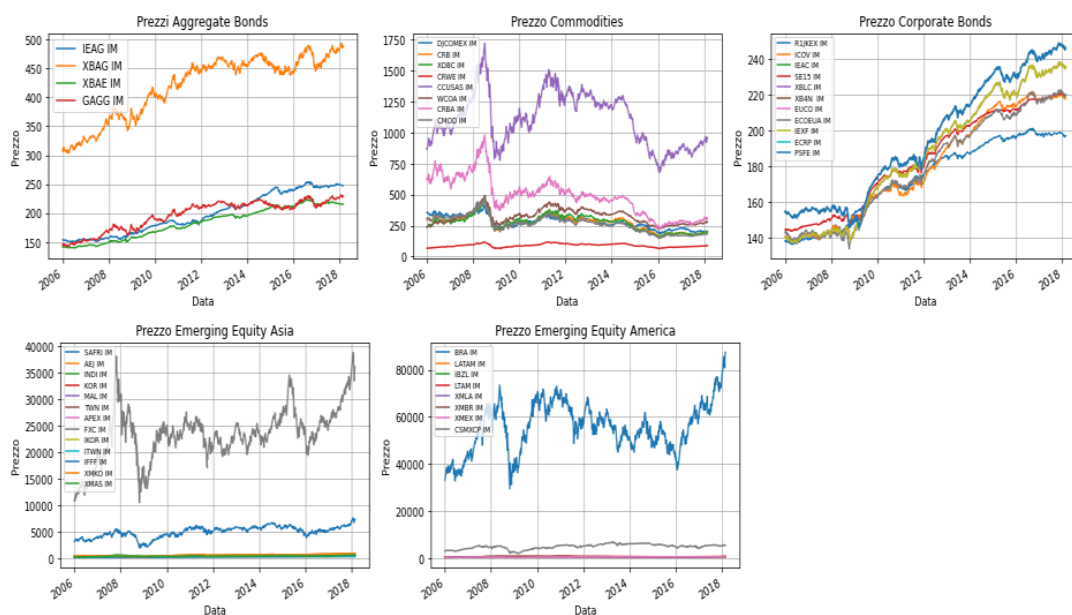


Figura 3.1. Andamento dei prezzi degli ETF tra il 2006 e il 2018, suddivisi per categoria d'investimento (Aggregate Bonds, Commodities, Corporate Bonds, Emerging Equity Asia, Emerging Equity America). Prodotta dall'autore.

Tabella 3.2. Statistiche riassuntive dei log-rendimenti degli ETF. Prodotta dall'autore.

CLASSE ETF	MEDIA	DEV.STD.	CURTOSI	SIMMETRIA
AGGREGATE BONDS	0	0.0026	6.67	0.11
COMMODITIES	0	0.0105	3.64	-0.31
CORPORATE BONDS	0	0.0015	3.35	-0.49
EMERGING EQUITY ASIA	0	0.0155	14.43	-0.19
EMERGING EQUITY AMERICA	0	0.0189	8.95	-0.29

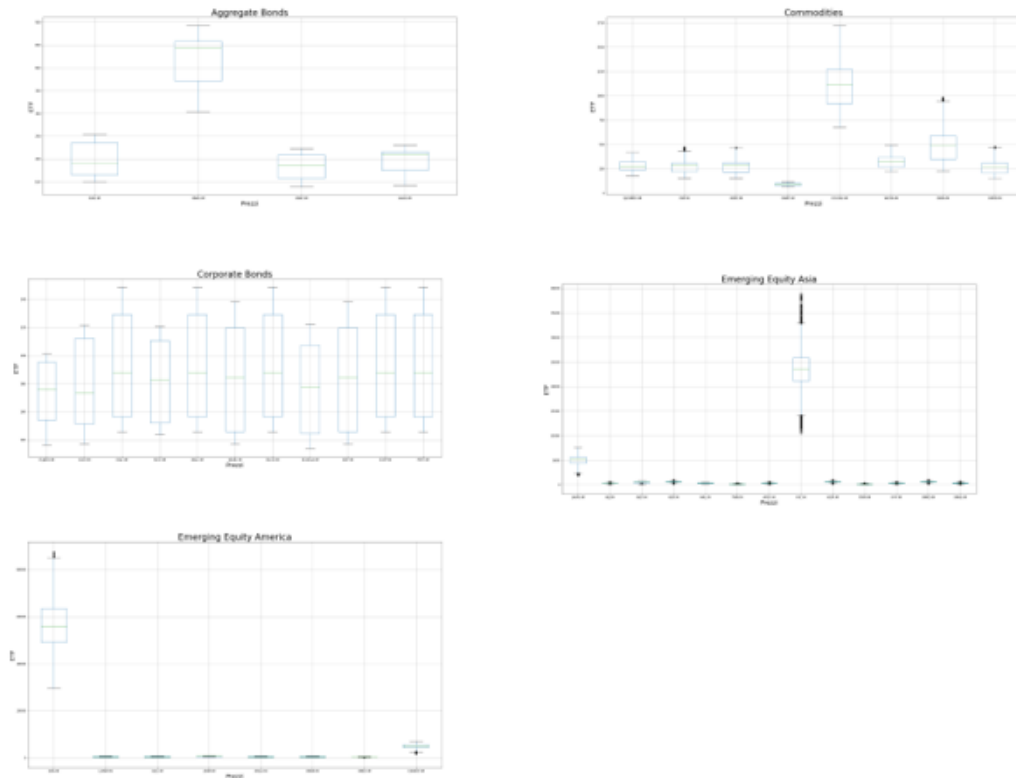


Figura 3.2. Boxplot dei prezzi per ogni categoria d'investimento (Aggregate Bonds, Commodities, Corporate Bonds, Emerging Equity Asia, Emerging Equity America). Prodotta dall'autore.

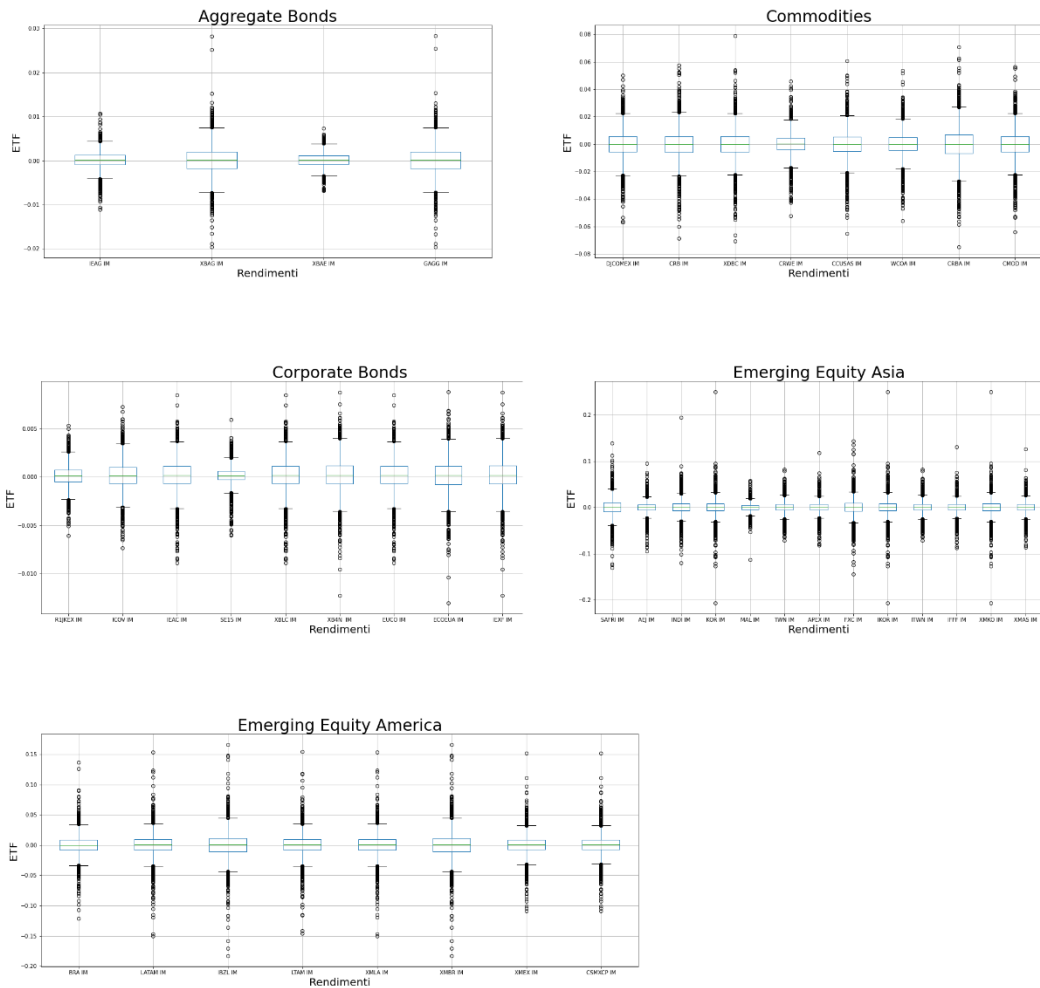


Figura 3.3. Boxplot dei rendimenti per ogni categoria d'investimento (Aggregate Bonds, Commodities, Corporate Bonds, Emerging Equity Asia, Emerging Equity America). Prodotta dall'autore.

3.2 COSTRUZIONE DEI PORTAFOGLI

Ora verranno messi a confronto i risultati derivanti da 4 differenti tipi di portafoglio. Nello specifico, i pesi dei portafogli in questione vengono calcolati nelle seguenti maniere:

- *Combined Portfolio*: I pesi vengono calcolati utilizzando nel problema di allocation, come matrice delle varianze-covarianze $C_\alpha = \alpha C + (1 - \alpha)\bar{C}$, cioè la matrice bi-dimensionale che tiene conto della volatilità e della covarianza tra gli asset in senso assoluto (C) e della propensione degli asset ad avere simili comportamenti nelle code sinistre (\bar{C}). A seconda del valore di α scelto, teniamo più in considerazione un aspetto piuttosto che l'altro. Con un $\alpha = 0$ avremo un portafoglio che tiene in considerazione solamente gli effetti delle code sinistre, con un $\alpha = 1$ avremo il classico portafoglio di Markowitz, mentre con un valore compreso tra 0 e 1 una combinazione di queste due misure di variabilità. Nel nostro caso, viene utilizzato un valore di $\alpha = 0.7$, perché dall'analisi di differenti valori di α , specificatamente $\alpha = \alpha_k = 0.1k$, $k = 0,1, \dots, 10$, si osserva che per tale valore si ottiene il profitto maggiore.
- *EDC Portfolio*: i pesi vengono calcolati utilizzando $\alpha = 0$, ma viene considerato l'Extreme Downside Correlation dei rendimenti: nel dettaglio, la matrice varianza-covarianza viene calcolata utilizzando la differenza dei rendimenti dalla loro media filtrata, in cui, se il valore del rendimento è maggiore del 5° percentile, la differenza viene impostata a 0, altrimenti viene lasciata invariata. In questo modo si andrà a minimizzare una misura di variabilità che tiene conto dei rischi più estremi, quelli al di sotto del 5° percentile.

- *Sharpe Portfolio*: i pesi vengono calcolati minimizzando il negativo dello Sharpe Ratio, inteso come il rapporto tra la somma della media ponderata per i pesi dei rendimenti $\sum_{i=1}^N E(r) \cdot w_i$ e la forma quadratica $w' C w$, dove C rappresenta la matrice varianza-covarianza dei rendimenti, sotto i vincoli che la somma dei pesi sia 1 e che essi siano maggiori di 0 $\sum_{i=1}^N w_i = 1$ con $w_i \geq 0$. L'ottimizzazione avviene attraverso il metodo numerico "*Sequential Least Squares Programming (SLSQP)*", partendo da un set di pesi randomico w_0 e giungendo alla soluzione dopo un certo numero di iterazioni.
- *Sortino Ratio*: i pesi vengono calcolati minimizzando il negativo del Sortino Ratio, inteso come il rapporto tra la somma della media ponderata per i pesi dei rendimenti $\sum_{i=1}^N E(r) \cdot w_i$ e la forma quadratica $w' C_1 w$, sotto i vincoli che la somma dei pesi sia 1 e che essi siano maggiori di 0 $\sum_{i=1}^N w_i = 1$ con $w_i \geq 0$. In questo caso, C_1 rappresenta la matrice varianza-covarianza costruita a partire dalla *downside deviation*, cioè la deviazione standard calcolata considerando solo i rendimenti al di sotto del *Minimum Acceptable Return*, in questo caso 0. Anche qui l'ottimizzazione avviene attraverso il *SLSQP*.

Per i primi due casi si risolve il seguente problema di asset allocation di Markowitz:

$$\left\{ \begin{array}{l} \min \underline{w}' C_{\alpha} \underline{w}, \quad 0 < \alpha < 1 \\ \sum_{i=1}^N w_i \log Y_i = \bar{y} \\ \sum_{i=1}^N w_i = 1 \quad \text{con } w_i \geq 0 \end{array} \right. ,$$

dove Y_i con $i = 1, 2, \dots, N$ fa riferimento ai rendimenti lordi (l'esponenziale dei logaritmici) dei 44 ETF e \bar{y} viene scelto come la media dei rendimenti del periodo considerato. L'analisi di portafoglio viene eseguita su 16 finestre temporali consecutive e per dei tratti sovrapposte di N_g giorni consecutivi di negoziazione. N_g viene scelto pari a 1290, cioè circa 5 anni e due finestre consecutive differiscono di circa 6 mesi (l'equivalente di circa 125 giorni di negoziazione). Viene utilizzato $\tau_{j-1} = 1 + 125(j - 1)$ e $\tau_j = 1290 + 125(j - 1)$ per indicare il tempo della prima e ultima osservazione della j -esima finestra, con $j = 1, 2, \dots, 16$. Si risolvono sedici problemi di ottimo, utilizzando però solo gli ultimi 252 rendimenti osservati in ognuna delle sopracitate finestre temporali, ottenendo 16 set di pesi, per $\alpha = 0.7$, il Combined Portfolio e per l'EDC Portfolio²⁹.

²⁹ Questi due risultati sono stati tratti direttamente dal codice di programmazione relativo al paper "A tail-revisited Markowitz mean-variance approach and a portfolio network centrality" di F.Mariani, G.Polinesi, M.C.Recchioni.

Per gli ultimi due casi, la metodologia è la medesima. Ciò che cambia, è ovviamente il metodo di ottimizzazione, che, come già spiegato è il SLSQP, ma anche i rendimenti utilizzati, che passano dall'essere lordi a logaritmici. Nelle Figure 3.4 e 3.5 vengono riportati i rendimenti percentuali annualizzati dei portafogli con Max Sharpe e Max Sortino per ognuna delle sedici finestre temporali.

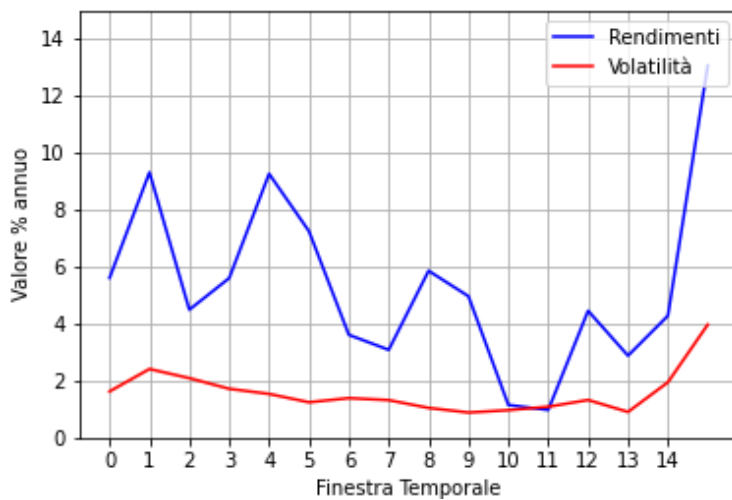


Figura 3.4. Rendimenti percentuali annui per finestra temporale del Max Sharpe Portfolio. Prodotta dall'autore.

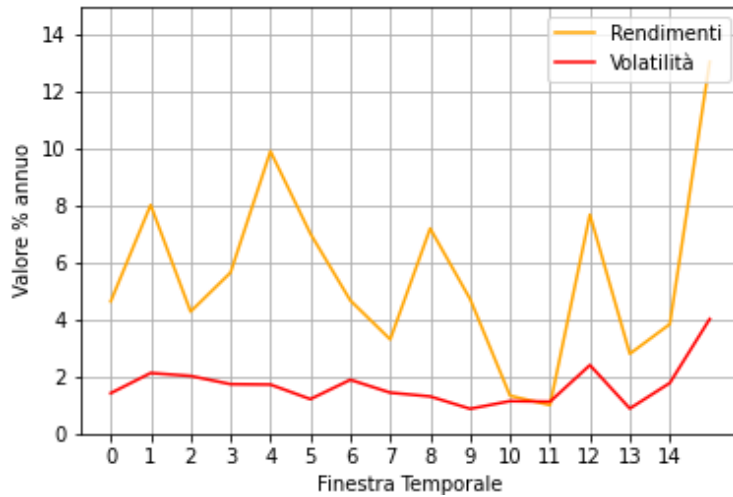


Figura 3.5. Rendimenti percentuali annui per finestra temporale del Max Sortino Portfolio. Prodotta dall'autore.

3.3 CONFRONTO DELLE PERFORMANCE

Quindi, vengono comparate le performance dei portafogli ottenuti guardando ai rendimenti dei portafogli dopo 6 mesi. Nel dettaglio, per $j = 1, 2, \dots, 16$, si definiscono i rendimenti di portafoglio come

$$r_{j,t} = \frac{(\sum_{i=1}^N w_{i,t} p_{i,t}) - P_j}{P_j}, \quad t = \tau_{j+1}, \tau_{j+2}, \dots, \tau_{j+125},$$

dove $p_{i,t}$ è il prezzo dell' i -esimo asset con t che varia da τ_{j+1} a τ_{j+125} mentre P_j è il valore del j -esimo portafoglio nell'ultima data della j -esima finestra.

P_j corrisponde a

$$P_j = \sum_{i=1}^N w_{i,t} p_{i,\tau_j}$$

I rendimenti $r_{j,t}$ sono quindi rendimenti *out-of-sample*, nel senso che sono calcolati usando come riferimento il valore del portafoglio da 1 a 125 giorni fuori dalla relativa finestra temporale utilizzata per calcolare i pesi. Le statistiche descrittive vengono riportate nella Tabella 3.3

Tabella 3.3, Statistiche descrittive per i quattro portafogli. Prodotta dall'autore.

	MEDIA	DEV.STD.	MIN	1° QUANT	2° QUANT	3° QUANT	MAX
COMBINED	0.0072	0.0281	-0.1166	-0.0088	0.0041	0.0245	0.1093
EDC	0.0009	0.0479	-0.1769	-0.0294	0.0001	0.0247	0.1388
MAX	0.0091	0.0312	-0.1287	-0.0039	0.0064	0.0194	0.1512
SHARPE							
MAX	0.0057	0.0305	-0.1247	-0.0074	0.0048	0.0185	0.1013
SORTINO							

La tabella mostra come il portafoglio Combined offra un buon compromesso tra volatilità, che risulta essere la più bassa tra tutti e quattro i portafogli, e rendimento medio, oltre a presentare le minori perdite. Il portafoglio Max Sharpe garantisce il maggior rendimento medio, al costo di una maggiore volatilità, ed appare come il

portafoglio con il massimo picco di profitto, nonostante il valore del 3° quantile sia inferiore a quelli dei portafogli Combined e EDC. Il Max Sortino sembra non reggere il confronto con i due precedenti portafogli, data la consistente volatilità a fronte di un rendimento inferiore di quasi il 50% rispetto al portafoglio Max Sharpe. Il portafoglio EDC invece mostra un rendimento medio pressoché nullo, se messo a paragone con gli altri e, coerentemente con la sua natura, garantisce le perdite maggiori ma anche dei cospicui profitti, come si può notare dai valori Min e Max.

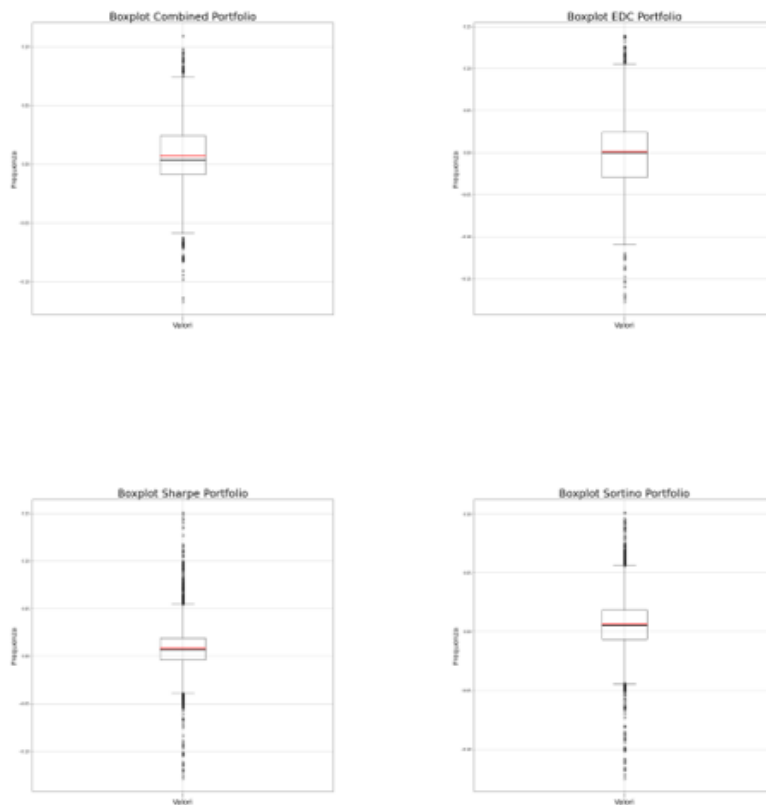


Figura 3.6. Boxplot dei rendimenti aggregati per Combined Portfolio, EDC Portfolio, Sharpe Portfolio e Sortino Portfolio. Prodotta dall'autore.

La Figura 3.6 mostra i boxplot dei rendimenti aggregati per i 4 portafogli, dove la linea nera e rossa rappresentano rispettivamente mediana e media dei valori. Qui è possibile avere una visualizzazione grafica dei rendimenti medi di portafoglio e delle loro distribuzioni, evidenziando il miglior rendimento medio del portafoglio Max Sharpe e l'alto range di variazione dei portafogli EDC e Sharpe.

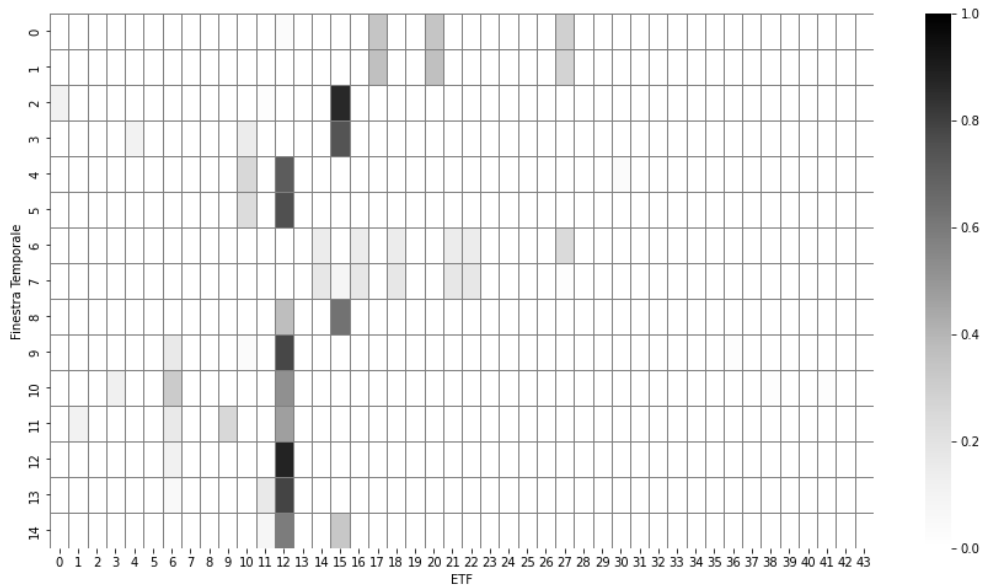


Figura 3.7. Pesi ottimali del portafoglio Combined. Prodotta dall'autore.

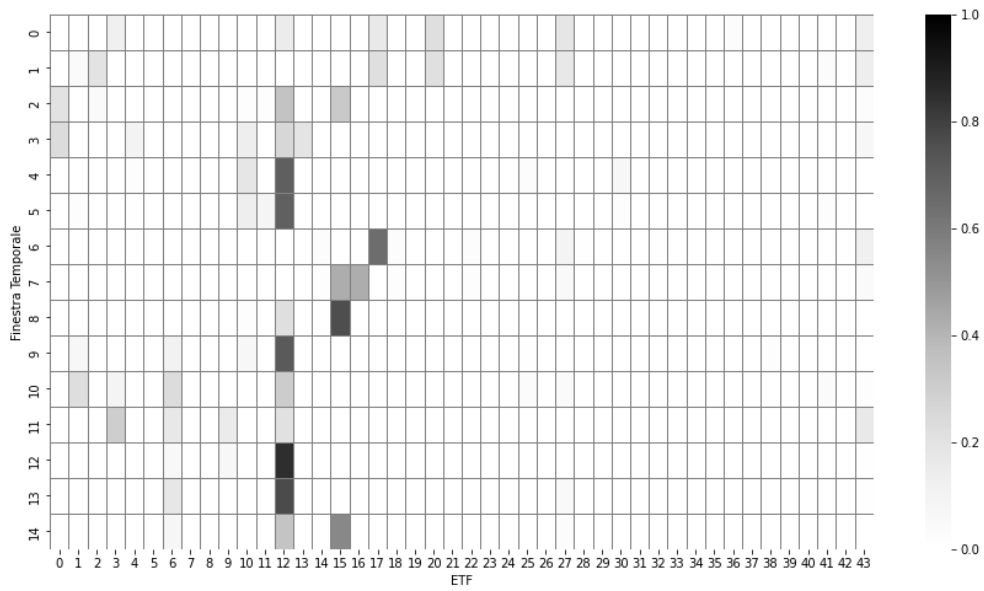


Figura 3.8. Pesi ottimali del portafoglio EDC. Prodotta dall'autore.

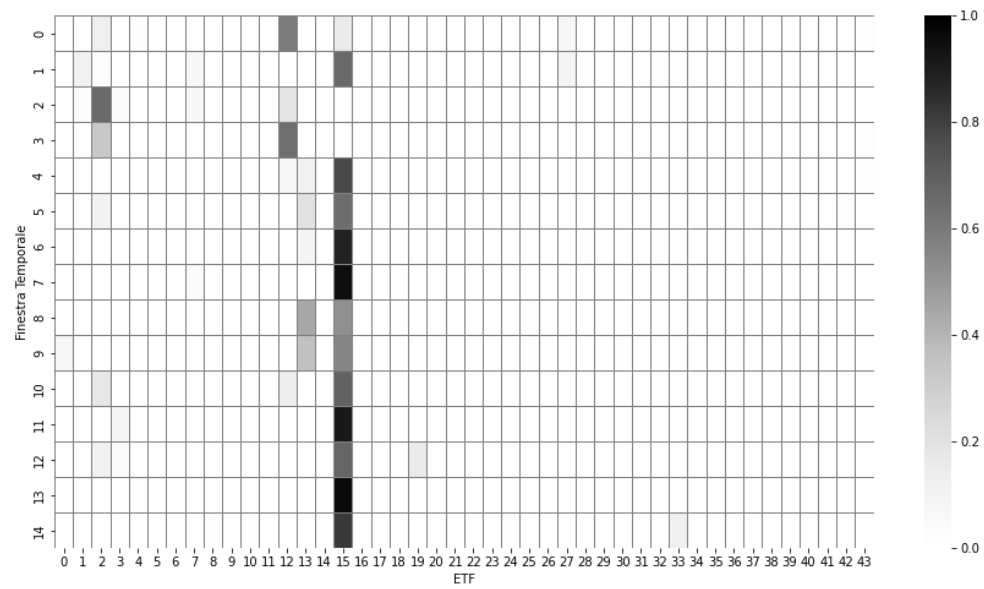


Figura 3.9. Pesi ottimali del portafoglio Max Sharpe. Prodotta dall'autore.

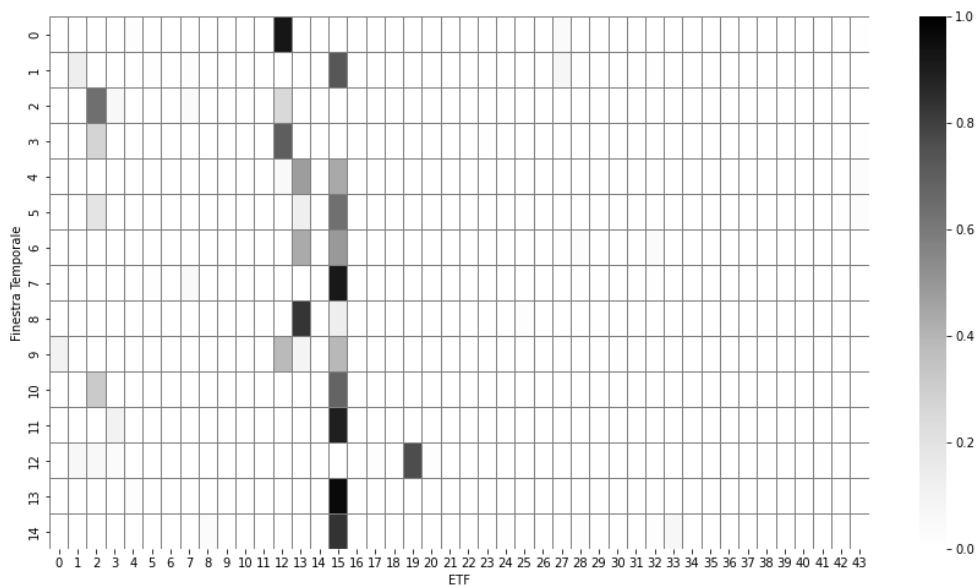


Figura 3.10. Pesi ottimali del portafoglio Max Sortino. Prodotta dall'autore.

Nelle Figure 3.7, 3.8, 3.9, 3.10 vengono mostrate in forma di *heatmap* le distribuzioni dei pesi per ognuno dei 4 portafogli. Sull'asse delle ordinate vengono poste le finestre temporali, mentre sull'asse delle ascisse gli asset, denominati con un indice da 0 a 43. Mentre i pesi dei portafogli Combined ed EDC sembrano essere più distribuiti lungo i sedici periodi considerati, quelli dei portafogli Sharpe e Sortino sono più polarizzati intorno gli asset di indice 12, 13 e 15 (rispettivamente *RIJKEX IM*, *ICOV IM* e *SE15 IM*, tutti e 3 Corporate Bonds e con un andamento piuttosto lineare).

Per quanto riguarda i costi di transazione, facciamo riferimento al *Turnover di portafoglio* definito come

$$Turnover = \sum_{t=t_0+1}^{t_1} \sum_{i=1}^N |w_{i,t} - w_{i,t-1}|,$$

dove t_0 e t_1 sono il primo e l'ultimo istante all'interno della finestra temporale.

Tabella 3.4, Turnover dei portafogli Combined, EDC, Sharpe e Sortino. Prodotta dall'autore.

COMBINED TURNOVER	EDC TURNOVER	SHARPE TURNOVER	SORTINO TURNOVER
13.10	15.98	11.35	18.55

Un elevato valore di turnover sta ad indicare una maggiore necessità di riadattare i pesi del portafoglio nel giro di sei mesi e quindi un aumento dei costi di transazione. Si osserva, nella Tabella 3.4, come il Turnover del portafoglio Max Sharpe sia di gran lunga inferiore a quello del portafoglio Max Sortino, nonostante graficamente, ad un primo colpo d'occhio, le distribuzioni dei pesi possano sembrare simili. Nel portafoglio Max Sortino però, tra la 5° e la 9° finestra temporale subentra con un'incidenza maggiore l'investimento nell'asset *ICOV IM*, giustificando così il netto aumento del Turnover. Il portafoglio Max Sharpe conserva anche il Turnover minore in assoluto tra i quattro portafogli, persino del Combined, che nonostante risulti più frazionato mantiene un valore non molto distante a quello del Max Sharpe. Tuttavia, valutazioni più approfondite andrebbero fatte verificando l'effettiva incidenza dei costi di transazione sull'ammontare finale dei profitti

generati dai portafogli, tenendo sempre a mente che i portafogli che costerebbero meno mantenere negli anni sono il Max Sharpe e il Combined, oltre ad essere quelli che offrono i migliori rapporti rendimento-volatilità tra i quattro analizzati.



Figura 3.11. Rendimenti out-of-sample dei portafogli Combined, EDC, Sharpe e Sortino, dal 2010 al 2018. Prodotta dall'autore.

La Figura 3.11 contiene una rappresentazione grafica dei rendimenti out-of-sample calcolati precedentemente su tutto l'intervallo temporale analizzato. La linea blu rappresenta l'andamento del portafoglio con Max Sharpe, quella arancione il portafoglio Max Sortino, quella verde il portafoglio Combined e quella rossa il portafoglio EDC. Osserviamo che il portafoglio Combined risulta essere un ottimo compromesso in termini di picchi di profitto e di perdita, risultando nel corso degli anni piuttosto bilanciato, eccezion fatta per il periodo della seconda metà del 2015, dovuto probabilmente alla crisi petrolifera. Il portafoglio EDC risulta essere il più

volatile, con delle grandi perdite durante il periodo della crisi del debito sovrano (tra il 2011 e il 2012), ma con anche delle ottime performance negli anni appena successivi. I Portafogli Max Sharpe e Sortino hanno invece un andamento piuttosto simile, con il Sortino che è sembrato essere preferibile allo Sharpe sino al crollo avvenuto nella metà del 2013. Da lì in poi il portafoglio Max Sharpe ha ottenuto quasi sistematicamente performance migliori. Da notare come negli ultimi anni considerati, a partire dalla seconda metà del 2017 fino all'ultima osservazione, la loro performance abbia superato quella del portafoglio Combined, con dei sorprendenti picchi positivi di rendimento, ad inizio 2018, del portafoglio Max Sharpe.

La Tabella 3.5 mostra la somma delle perdite (come somma dei rendimenti inferiori allo 0) e dei profitti (come somma dei rendimenti maggiori allo 0). Qui si nota meglio la superiorità a livello di performance del portafoglio Sharpe, che in termini di differenza profitto-perdita rende meglio rispetto agli altri portafogli, riuscendo in maniera particolarmente brillante a contenere le perdite. Il portafoglio EDC mostra elevatissime perdite, ma anche degli elevati profitti, garantendo a fine periodo un differenziale molto basso e vicino allo 0, mentre gli ultimi due si assestano su un livello intermedio, mantenendo comunque una performance più che positiva alla fine del periodo considerato.

Tabella 3.5, Tabella delle perdite e dei profitti dei portafogli Combined, EDC, Max Sharpe e Max Sortino

		COMBINED	EDC PORTFOLIO	SHARPE	SORTINO
		PORTFOLIO		PORTFOLIO	PORTFOLIO
SOMMA DELLE		13.1363	33.0842	11.905	15.400
PERDITE					
SOMMA DEI		26.6527	34.7956	29.016	26.167
PROFITTI					
PROFITTI-PERDITE		13.5164	1.71146	17.111	10.766

CONCLUSIONI

L'obiettivo di questo lavoro è stato quello di presentare, partendo dalla teoria classica del portafoglio di Markowitz, l'evoluzione delle tecniche di portafoglio e delle misure di rischio e variabilità utilizzate, concentrandosi in modo particolare su una recente tecnica utilizzata per tenere in considerazione gli effetti delle code sinistre, tema che nella letteratura finanziaria sta avendo, specialmente negli ultimi anni, sempre più rilevanza. Questa tecnica, si basa sulla costruzione, attraverso un algoritmo iterativo, di un dataset di parametri legato alla distribuzione delle code sinistre degli asset e sul conseguente calcolo della relativa matrice varianza-covarianza. In seguito, combinando linearmente questa nuova misura di variabilità alla classica matrice varianza-covarianza dei rendimenti, si mostra che è possibile

ottenere una nuova misura di variabilità che tiene conto anche della distribuzione delle code degli asset e della loro volatilità.

Nell'analisi finale, si sono messi a confronto alcuni dei risultati ottenuti nel paper di riferimento, con le performance di portafogli che massimizzano lo Sharpe ed il Sortino Ratio, calcolati su un dataset di 12 anni circa di osservazioni giornaliere su 44 ETF, mostrando la superiorità dell'utilizzo del portafoglio Max Sharpe per la massimizzazione dei profitti di portafoglio. Per migliorare e completare lo studio, si potrebbe utilizzare un differente dataset e osservare se, sullo stesso intervallo temporale, i risultati rimangono invariati. Si potrebbero inoltre, calcolare gli indici di Sharpe e Sortino utilizzando differenti valori di tasso Risk-Free e Minimum Acceptable Return, entrambi supposti pari a 0 ed osservare come variano i risultati. Infine, vista la tendenza dei portafogli Max Sharpe e Sortino ad avere delle migliori performance nel periodo finale (tra il 2017-2018), sarebbe interessante osservare il loro andamento oltre quel periodo di riferimento.

APPENDICE

Nel seguente appendice si mostrerà il codice Python utilizzato per svolgere la gran parte dei calcoli e per comporre alcuni dei grafici presenti nella tesi. La restante parte di codice, utilizzata per produrre la gran parte delle figure, non verrà mostrata per non appesantire l'elaborato.

'Codice Python, Agostini Matteo, Calcolo e Creazione figure'

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
plt.style.use('fast')
import pandas as pd

df = pd.read_excel("36etf_4_classipyth.xls", sheet_name=0, )
stock_data1 = df.set_index('date')

stock_returns = np.log(stock_data1.pct_change()+1)

shifted_returns={}
meanReturns={}
cov_matrix = {}
std_neg_vect = {}
corr_mat= {}
cov_matrix_neg= {}
```

```

for i in range(16):
    shifted_returns[(i+1)]= stock_returns[(1+i*125+1037):(1290+i*125)]
    meanReturns[(i+1)]= shifted_returns[i+1].mean()
    cov_matrix[(i+1)]= shifted_returns[i+1].cov()
    std_neg_vect[(i+1)] =shifted_returns[i+1][shifted_returns[i+1]<0].std()
    corr_mat[(i+1)]= shifted_returns[i+1].corr()
    cov_matrix_neg[(i+1)]=
np.dot(np.dot(np.diag(std_neg_vect[i+1]),corr_mat[i+1]),np.diag(std_neg_vect[i+
1]))

from scipy.optimize import minimize
#definiamo le variabili della funzione di minimizzazione
#definiamo il negative Sharpe Ratio per ognuna delle 16 finestre temporali
def negativeShR(w, i):
    w = np.array(w)
    R = np.sum(meanReturns[i+1]*w)
    V = np.sqrt(np.dot(w.T,np.dot(cov_matrix[i+1],w)))
    ShR = R/V
    return -1*ShR

#funzione che prende i pesi e ci fornisce 0 se sommano a 1
def checkSumToOne(w):
    return np.sum(w)-1
#definiamo il vettore dei pesi da cui partirà l'algoritmo
w0 = np.array(np.random.random(44))
w0 = w0/np.sum(w0)
#il vincolo che i pesi siano compresi tra 0 e 1

```

```

bounds = ((0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),
          (0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),
          (0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),
          (0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1))
#definiamo i vincoli
constraints = ({'type':'eq','fun':checkSumToOne})
#mando le funzioni che minimizzano gli SR per ogni finestra temporale
shR_funcs=[]
w_opts = []
OptimalSharpeWeights = []

for i in range(16):
    shR_funcs.append(lambda x, i=i: negativeShR(x, i))
    w_opt = minimize(shR_funcs[i], w0, method='SLSQP', bounds=bounds,
constraints=constraints)
    w_opts.append(w_opt)
    OptimalSharpeWeights.append(w_opts[i].x)

#metto i pesi in un'unica matrice
SharpeWeights = pd.DataFrame(OptimalSharpeWeights).T
SharpeWeights.columns = list(range(16))

#calcolo i rendimenti dei portafogli con max Sharpe
SharpeReturns1=pd.DataFrame(0, index=range(16), columns=['Rendimenti'])

for xx in range(16):

```



```

    SharpeReturns1.iloc[xx,0]=np.dot(SharpeWeights.iloc[:,xx].T,
meanReturns[xx+1])*252

#calcolo le volatilità dei portafogli con Max Sharpe
SharpeVolatilities=pd.DataFrame(0, index=range(16), columns=['Volatilità'])

for indexx in range(16):
    SharpeVolatilities.iloc[indexx,0]=np.sqrt(np.dot(SharpeWeights.iloc[:,indexx].T,
np.dot(cov_matrix[indexx+1],
        SharpeWeights.iloc[:,indexx]))) * np.sqrt(252)

fig, ax = plt.subplots()
# Aggiungi la linea dei rendimenti al grafico
ax.plot(SharpeReturns1.index,
SharpeReturns1['Rendimenti']*100,label='Rendimenti',color='blue')
# Aggiungi la linea delle volatilità al grafico
ax.plot(SharpeVolatilities.index,
SharpeVolatilities['Volatilità']*100,label='Volatilità', color='red')
# Imposta i titoli degli assi del grafico
ax.set_xlabel('Finestra Temporale')
ax.set_ylabel('Valore % annuo')
ax.legend(loc='upper right')
# Imposta il titolo del grafico
plt.ylim(0,15)

```

```

plt.grid()
plt.xticks(range(0,15))
plt.show()
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Rendimento annuo
Sharpe.png')

#Rifaccio lo stesso procedimento con il Sortino Ratio

def negativeSoR(w, i):
    w = np.array(w)
    R = np.sum(meanReturns[i+1]*w)
    DV = np.sqrt(np.dot(w.T,np.dot(cov_matrix_neg[i+1],w)))
    SoR = R/DV
    return -1*SoR

soR_funcs=[]
w_opts1 = []
OptimalSortinoWeights = []

for i in range(16):
    soR_funcs.append(lambda x, i=i: negativeSoR(x, i))
    w_opt1 = minimize(soR_funcs[i], w0, method='SLSQP', bounds=bounds,
constraints=constraints)
    w_opts1.append(w_opt1)
    OptimalSortinoWeights.append(w_opts1[i].x)

#metto i pesi in un'unica matrice

```

```

SortinoWeights = pd.DataFrame(OptimalSortinoWeights).T
SortinoWeights.columns = list(range(16))

#plotto le heatmaps dei pesi del portafoglio combined, del max EDC e del Sharpe
e Sortino
#plotto i pesi delle 16 finestre del portafoglio ottimo con alfa=0,7 e del portafoglio
con max EDC
SharpeWeightsx = SharpeWeights.drop(columns=15)
SortinoWeightsx = SortinoWeights.drop(columns=15)
combinedweights = pd.read_excel("combined_weights.xlsx", sheet_name=0, )
combinedweights.columns=('0','1','2','3','4','5','6','7','8','9','10','11','12','13','14')
edcweights = pd.read_excel("my_w_edc.xlsx", sheet_name=0, )
edcweights.columns=('0','1','2','3','4','5','6','7','8','9','10','11','12','13','14')
#heatmap
plt.figure(figsize=(15, 8))
sns.heatmap(SharpeWeightsx.T, vmin=0, vmax=1,linewidths=1,linecolor='grey',
cmap='Greys')
plt.xlabel('ETF')
plt.ylabel('Finestra Temporale')
plt.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Pesi Sharpe.png')
plt.show()
#####
plt.figure(figsize=(15, 8))
sns.heatmap(combinedweights.T, vmin=0, vmax=1,linewidths=1,linecolor='grey',
cmap='Greys')
plt.xlabel('ETF')

```

```

plt.ylabel('Finestra Temporale')
plt.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Pesi Combined.png')
plt.show()
#####
plt.figure(figsize=(15, 8))
sns.heatmap(edcweights.T, vmin=0, vmax=1,linewidths=1,linecolor='grey',
cmap='Greys')
plt.xlabel('ETF')
plt.ylabel('Finestra Temporale')
plt.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Pesi EDC.png')
plt.show()
#####
plt.figure(figsize=(15, 8))
sns.heatmap(SortinoWeightsx.T, vmin=0, vmax=1,linewidths=1,linecolor='grey',
cmap='Greys')
plt.xlabel('ETF')
plt.ylabel('Finestra Temporale')
plt.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Pesi Sortino.png')
plt.show()

```

```
SortinoReturns1=pd.DataFrame(0, index=range(16), columns=['Rendimenti'])
```

```
for xx in range(16):
```

```
    SortinoReturns1.iloc[xx,0]=np.dot(SortinoWeights.iloc[:,xx].T,
meanReturns[xx+1])*252
```

```
SortinoReturns1*100
```

```

SortinoVolatilities=pd.DataFrame(0, index=range(16), columns=['Volatilità'])

for indexx in range(16):

SortinoVolatilities.iloc[indexx,0]=np.sqrt(np.dot(SortinoWeights.iloc[:,indexx].T,
          np.dot(cov_matrix[indexx+1],
SortinoWeights.iloc[:,indexx])))*np.sqrt(252)

fig, ax = plt.subplots()
# Aggiungi la linea dei rendimenti al grafico
ax.plot(SortinoReturns1.index,
SortinoReturns1['Rendimenti']*100,label='Rendimenti',color='orange')
# Aggiungi la linea delle volatilità al grafico
ax.plot(SortinoVolatilities.index,
SortinoVolatilities['Volatilità']*100,label='Volatilità', color='red')
# Imposta i titoli degli assi del grafico
ax.set_xlabel('Finestra Temporale')
ax.set_ylabel('Valore % annuo')
ax.legend(loc='upper right')
# Imposta il titolo del grafico
plt.ylim(0,15)
plt.grid()
plt.xticks(range(0,15))
plt.show()

```

```
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/Rendimento annuo
Sortino.png')
```

```
#Calcolo i rendimenti out of sample per le 16 finestre relative allo sharpe ratio
```

```
RetOutSharpe =
```

```
pd.DataFrame(index=range(1,126),columns=list(range(0,16)),dtype=float)
```

```
for column in range(0,15):
```

```
    for row in range(1,126):
```

```
RetOutSharpe.loc[row,column]=float((np.dot(stock_data1.iloc[1289+row+125*co
lumn:1290+row+125*column,:],
```

```
SharpeWeights.iloc[:,column])-
```

```
np.dot(stock_data1.iloc[1289+125*column:1290+125*column,:],
```

```
SharpeWeights.iloc[:,column]))/
```

```
np.dot(stock_data1.iloc[1289+125*column:1290+125*column,:],
```

```
SharpeWeights.iloc[:,column]))
```

```
#Calcolo i rendimenti out of sample per il Sortino
```

```
RetOutSortino =
```

```
pd.DataFrame(index=range(1,126),columns=list(range(0,16)),dtype=float)
```

```
for column in range(0,15):
```

```
    for row in range(1,126):
```

```

RetOutSortino.loc[row,column]=float((np.dot(stock_data1.iloc[1289+row+125*column:1290+row+125*column,:],
SortinoWeights.iloc[:,column]) -

np.dot(stock_data1.iloc[1289+125*column:1290+125*column,:],
SortinoWeights.iloc[:,column]))/

np.dot(stock_data1.iloc[1289+125*column:1290+125*column,:],
SortinoWeights.iloc[:,column]))

dates = pd.read_excel("my_date.xlsx", sheet_name=0, )

datacombined = pd.read_excel("my_sampleB1.xlsx", sheet_name=0, ).head(1874)
dataEDC = pd.read_excel("my_sampleedc1.xlsx", sheet_name=0, ).head(1874)
datacombined1=pd.concat([dates,datacombined],axis=1,)
dataEDC1=pd.concat([dates,dataEDC],axis=1,)
datacombined1.columns=['Date','Values']
dataEDC1.columns=['Date','Values']
#metto tutti i valori di Sharpe in una colonna
all_values = []
for cols in RetOutSharpe:
    this_column_values = RetOutSharpe[cols].tolist()
    all_values += this_column_values

one_column_RetOutSharpe = (pd.DataFrame(all_values)).head(1874)
print(one_column_RetOutSharpe)

```

```

#metto tutti i valori di Sortino in una colonna
all_values1 = []
for colss in RetOutSharpe:
    this_column_values1 = RetOutSortino[colss].tolist()
    all_values1 += this_column_values1
#aggiusto i rendimenti affiancando le date
one_column_RetOutSortino = (pd.DataFrame(all_values1)).head(1874)
aOCRSOR=pd.concat([dates,one_column_RetOutSortino],axis=1,)
aOCRSOR.columns =['Date', 'Values']
axax=aOCRSOR.set_index('Date')
aOCRSHR=pd.concat([dates,one_column_RetOutSharpe],axis=1,)
aOCRSHR.columns =['Date', 'Values']
axax1=aOCRSHR.set_index('Date')
axax2=datacombined1.set_index('Date')
axax3=dataEDC1.set_index('Date')

plt.figure(figsize=(25,12))
plt.plot(axax1.index,axax1.iloc[:,0])
plt.plot(axax.index,axax.iloc[:,0])
plt.plot(axax2.index,axax2.iloc[:,0])
plt.plot(axax3.index,axax3.iloc[:,0])
plt.xlabel('Tempo',fontsize=20)
plt.ylabel('Curva Profit-Losses',fontsize=20)
plt.legend(['Sharpe Ratio','Sortino Ratio','Combined','EDC'],loc='lower right')
plt.grid()

```



```
plt.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/CurvaLossProfit.png')
plt.show()
```

```
#calcolo i turnover Ratio
```

```
Turnover_Sharpe =
pd.DataFrame(index=range(1,15),columns=list(range(0,1)),dtype=float)
Turnover_Sortino =
pd.DataFrame(index=range(1,15),columns=list(range(0,1)),dtype=float)
Turnover_Combined =
pd.DataFrame(index=range(1,15),columns=list(range(0,1)),dtype=float)
Turnover_EDC =
pd.DataFrame(index=range(1,15),columns=list(range(0,1)),dtype=float)
```

```
for turn in range(1,15,1):
```

```
    Turnover_Sharpe.loc[turn]=sum(abs(SharpeWeights.iloc[:,turn]-
SharpeWeights.iloc[:,turn-1]))
```

```
for turn in range(1,15,1):
```

```
    Turnover_Sortino.loc[turn]=sum(abs(SortinoWeights.iloc[:,turn]-
SortinoWeights.iloc[:,turn-1]))
```

```
for turn in range(1,15,1):
```

```
    Turnover_Combined.loc[turn]=sum(abs(combinedweights.iloc[:,turn]-
combinedweights.iloc[:,turn-1]))
```

```
for turn in range(1,15,1):
```

```
Turnover_EDC.loc[turn]=sum(abs(edcweights.iloc[:,turn]-edcweights.iloc[:,turn-1]))
```

```
TShR = pd.DataFrame.sum(Turnover_Sharpe)
```

```
TSoR = pd.DataFrame.sum(Turnover_Sortino)
```

```
TCom = pd.DataFrame.sum(Turnover_Combined)
```

```
TEdc = pd.DataFrame.sum(Turnover_EDC)
```

```
print(TShR,TSoR,TCom,TEdc)
```

```
#somma valori positivi e somma valori negativi dei return
```

```
Profit_losses = pd.read_excel('result_Profit_loss.xlsx',sheet_name=0)
```

```
ShrProfit = float(pd.DataFrame.sum(abs(axax1[axax1['Values'] > 0])))
```

```
SorProfit = float(pd.DataFrame.sum(abs(axax[axax['Values'] > 0])))
```

```
ShrLosses = float(pd.DataFrame.sum(abs(axax1[axax1['Values'] < 0])))
```

```
SorLosses = float(pd.DataFrame.sum(abs(axax[axax['Values'] < 0])))
```

```
ShrPL = ShrProfit - ShrLosses
```

```
SorPL = SorProfit - SorLosses
```

```
Profit_lossesNEW=pd.DataFrame(np.array([[ShrLosses,SorLosses],  
[ShrProfit,SorProfit], [ShrPL,SorPL]]))
```

```
#Tabella che mostra la somma dei profitti e delle perdite
```

```

Profit_lossesFIN=pd.concat([Profit_losses,Profit_lossesNEW],axis=1)
Profit_lossesFIN.columns=('Combined','EDC Portfolio','Sharpe Portfolio','Sortino
Portfolio')
Profit_lossesFIN.index = ('Somma delle perdite','Somma dei profitti','Profitti-
Perdite')

#Combined alfa=0.7
axax2.describe()
#Edc
axax3.describe()
#Sharpe
axax1.describe()
#Sortino
axax.describe()

#####Combined#####
fig, ax = plt.subplots(figsize=(15, 15))
bp = ax.boxplot(axax2['Values'], meanline=True, showmeans=True,
                meanprops=dict(linestyle='-', linewidth=2.5, color='red'),
                medianprops=dict(linestyle='-', linewidth=2.5, color='black'))

# imposta il titolo del grafico e le etichette degli assi
ax.set_title('Boxplot Combined Portfolio',fontsize=30)
ax.set_xlabel('Valori',fontsize=20)
ax.set_ylabel('Frequenza',fontsize=20)

```

```

plt.grid()
plt.show()
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/BoxCombined.png')
#####EDC#####
fig, ax = plt.subplots(figsize=(15, 15))
bp = ax.boxplot(axax3['Values'], meanline=True, showmeans=True,
                meanprops=dict(linestyle='-', linewidth=2.5, color='red'),
                medianprops=dict(linestyle='-', linewidth=2.5, color='black'))

# imposta il titolo del grafico e le etichette degli assi
ax.set_title('Boxplot EDC Portfolio',fontsize=30)
ax.set_xlabel('Valori',fontsize=20)
ax.set_ylabel('Frequenza',fontsize=20)
plt.grid()
plt.show()
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/BoxEDC.png')
#####Sharpe#####
fig, ax = plt.subplots(figsize=(15, 15))
bp = ax.boxplot(axax1['Values'], meanline=True, showmeans=True,
                meanprops=dict(linestyle='-', linewidth=2.5, color='red'),
                medianprops=dict(linestyle='-', linewidth=2.5, color='black'))

# imposta il titolo del grafico e le etichette degli assi
ax.set_title('Boxplot Sharpe Portfolio',fontsize=30)
ax.set_xlabel('Valori',fontsize=20)
ax.set_ylabel('Frequenza',fontsize=20)
plt.grid()

```

```

plt.show()
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/BoxSharpe.png')
#####Sortino#####à
fig, ax = plt.subplots(figsize=(15, 15))
bp = ax.boxplot(axax['Values'], meanline=True, showmeans=True,
                meanprops=dict(linestyle='-', linewidth=2.5, color='red'),
                medianprops=dict(linestyle='-', linewidth=2.5, color='black'))

# imposta il titolo del grafico e le etichette degli assi
ax.set_title('Boxplot Sortino Portfolio',fontsize=30)
ax.set_xlabel('Valori',fontsize=20)
ax.set_ylabel('Frequenza',fontsize=20)
plt.grid()
plt.show()
fig.savefig('C:/Users/agost/Desktop/Cartella immagini tesi/BoxSortino.png')

from PIL import Image
# Apri le tue immagini
img1 = Image.open('C:/Users/agost/Desktop/Cartella immagini
tesi/BoxCombined.png')
img2 = Image.open('C:/Users/agost/Desktop/Cartella immagini tesi/BoxEDC.png')
img3 = Image.open('C:/Users/agost/Desktop/Cartella immagini
tesi/BoxSharpe.png')
img4 = Image.open('C:/Users/agost/Desktop/Cartella immagini
tesi/BoxSortino.png')
result_width = img1.width + img2.width + img3.width + img4.width
result_height = max(img1.height, img2.height, img3.height, img4.height)*3

```

```
result = Image.new('RGB', (result_width, result_height), (255, 255, 255))
# Aggiungi le immagini alla nuova immagine
result.paste(img1, (0, 0))
result.paste(img2, (img1.width+150, 0))
result.paste(img3, (0, img1.height+250))
result.paste(img4, (img1.width+150, img1.height+250))
# Salva l'immagine risultante
result.save('C:/Users/agost/Desktop/Cartella          immagini
tesi/BoxPlotRendimentiOS.png')
```

BIBLIOGRAFIA e SITOGRAFIA

ARTZNER PH, DELBAEN F, EBER JM, HEATH D (1999) *Coherent measures of risk*, “Mathematical Finance” 9, 203-228.

CAPARRELLI F (1998) *Economia dei mercati finanziari*, Mc Graw Hill, 564.

CHABI-YO, RUENZI S, WEIGERT F (2018) *Crash Sensitivity and the Cross Section of Expected Stock Returns*, “Journal of Financial and Quantitative Analysis” 53(3), 1059-1100.

DEMPSTER AP, LAIRD NM, RUBIN DB (1977) *Maximum likelihood from incomplete data via the EM algorithm*, “Journal of the Royal Statistical Society: Series B” 39(1): 1-38.

FISHBURN PC (1980) *Stochastic Dominance and Moments of Distributions*, “Mathematics of Operations Research” 5, 94-100.

GARBADE K (1994) *Teoria dei mercati finanziari*, Il Mulino, Bologna, 584.

HARRIS RD, NGUYEN LH, STOJA E (2019) *Systematic extreme downside risk*, “Journal of International Financial Markets, Institutions and Money” 61:128-142.

MARIANI F, CIOMMI M, CHELLI FM, RECCHIONI MC (2020). *An iterative approach to stratification: Poverty at closed-form formula supporting the shadow price approach*, “European Journal of Operational Research” 275(3): 1178-1189.

MARIANI F, POLINESI G, RECCHIONI MC (2022) *A tail-revisited Markowitz mean-variance approach and a portfolio network centrality*, “Computational Management Science” 19(3): 425-455.

MARKOWITZ H (1952) *Portfolio selection*, “Journal of Finance” 7(1): 77-91.

MASON F (1992) *Metodi quantitativi per le decisioni*, Giappichelli, Torino, 272.

MICOCCI M, MASALA GB (2012) *Manuale di matematica finanziaria. Metodi e strumenti quantitativi per il risk management*, Carocci, Roma, 655.

PFLUG GC (2000) *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk*, “Journal of Risk Finance” 1(4).

ROLLINGER TN, HOFFMAN ST (1994) *Sortino: A ‘Sharper’ Ratio*, “The Journal of Investing”.

SHARPE WF (1964) *Capital Asset Prices: a theory of market equilibrium under conditions of risk*, “Journal of Finance 19(3), 425-442.

SHARPE WF (1994) *The Sharpe Ratio*, “Journal of Portfolio Management Fall” 21(1), 49-58.

SORTINO FA, SACHELL S (2001), *Managing downside risk in financial markets: Theory, practice and implementation*, Butterworth-Heinemann, 237.

TASCHE D (2002) *Expected shortfall and beyond*, “Journal of Banking & Finance” 26, 1519-1533.

URYASEV S (2000) *Conditional Value-at-Risk: Optimization Algorithms and Applications*, “Financial Engineering News” 14.

URYASEV S, ROCKAFELLAR RT (1999) *Optimization of Conditional Value-at-Risk*, “Research Report” 99(4).

https://it.wikipedia.org/wiki/Indice_di_Sharpe

https://it.wikipedia.org/wiki/Indice_di_Sortino

https://it.kamiltaylan.blog/conditional_value_at_risk/

https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm

<https://towardsdatascience.com/expectation-maximization-explained-c82f5ed438e5>