



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale
in Sviluppo economico e dell'impresa
(ordinamento ex D.M. 270/2004)

Tesi di Laurea

Analisi sugli stimatori corretti della
frontiera efficiente

Relatore

Prof. Claudio Pizzi

Laureando

Francesco Cattai
Matricola 842570

Anno Accademico
2016-2017

INDICE

Indice	i
Elenco delle tabelle	iii
Elenco delle figure	iv
Introduzione	1
1 Modello selezione del portafoglio: l'approccio di Markowitz	3
1.1 Approccio media-varianza	3
1.2 Benefici della diversificazione	8
1.3 Derivazione analitica della frontiera efficiente	11
1.4 Critiche al modello di Markowitz	15
2 Il modello proposto da Bodnar e Bodnar	18
2.1 Derivazione degli stimatori campionari corretti	18
2.2 F-Test sulla Frontiera Efficiente	25
2.3 Limiti dell'approccio di Bodnar e Bodnar	26
3 Analisi delle componenti principali	31
3.1 Derivazione delle componenti principali	31
3.2 Interpretazione grafica	36
3.3 Applicazioni economiche	38
4 Analisi empirica	40
4.1 Analisi dei dati	40
4.1.1 Analisi della normalità	41
4.1.2 Analisi per rendimenti mensili	42

4.1.3	Analisi per rendimenti giornalieri	45
4.2	Confronto fra frontiera efficiente Campionaria e frontiera efficiente à la Bodnar	48
4.2.1	Analisi sulle ipotesi di base di Bodnar e Bodnar	48
4.2.2	Confronto grafico dei modelli	49
4.2.3	Analisi empirica sul parametro $\hat{\sigma}$	51
4.2.4	Comportamento asintotico della frontiera efficiente corretta	59
4.3	Utilizzo della PCA per stimare la frontiera à la Bodnar	60
4.3.1	Definizione delle componenti principali	60
4.3.2	Applicazione della PCA per stimare modello à la Bodnar . .	61
	Conclusioni	66
	Bibliografia	68

ELENCO DELLE TABELLE

4.1	Jarque Bera test e D'Agostino test per rendimenti mensili	43
4.2	Jarque Bera test e D'Agostino test per rendimenti giornalieri	46
4.3	Confronto dei rendimenti mensili per prima e seconda metà	55
4.4	Effetto delle variazioni della matrice di covarianza su \hat{s}	57
4.5	Effetto delle variazioni della numerosità campionaria sulla frontiera à la Bodnar e Bodnar	60
4.6	Confronto fra i valori di R_{GMV} , V_{GMV} e \hat{s} per dati giornalieri	62
4.7	Confronto fra i valori di R_{GMV} , V_{GMV} e \hat{s} per dati mensili	63

ELENCO DELLE FIGURE

1.1	Rappresentazione della non dominanza fra il titolo A e il titolo B	4
1.2	Rappresentazione grafica della CAL per il titolo A e il titolo B	5
1.3	Rappresentazione grafica del cambio del tasso a cui è possibile indebitarsi, che corrisponde ad una diminuzione della pendenza della CAL	6
1.4	Rappresentazione della curva di Utilità tangente alla CAL	8
1.5	Rappresentazione dei benefici della diversificazione al cambio dell'indice di correlazione tra il titolo A e il titolo B	9
1.6	Effetto della numerosità dei titoli sul rischio del portafoglio negli USA (1975)	14
2.1	Frontiera efficiente con e senza aggiustamento, per portafogli creati con numeri diversi di titoli	19
2.2	Limite inferiore per il numero di asset (k) nel portafoglio in funzione di s per il quale la frontiera efficiente campionaria sovrastima la frontiera efficiente della popolazione per tutti i V ($n \in \{60, 120, 240\}$)	23
2.3	Stimatore campionario e corretto della frontiera efficiente	24
2.4	$n = 95, s_{lim2} = 0.1192, s_{lim1} = 0.1060$, Frontiera accettabile $\hat{s} > s_{lim2}$	30
2.5	$n = 60, s_{lim2} = 0.2049, s_{lim1} = 0.1806$, Frontiera non accettabile $\hat{s} < s_{lim1}$.	30
2.6	$n = 85, s_{lim2} = 0.1354, s_{lim1} = 0.1202$, Frontiera non accettabile $s_{lim1} < \hat{s} < s_{lim2}$	30
3.1	Esempio scree plot	36
3.2	Proiezione sul piano R^2 (componente 1; componente 2)	37
3.3	Proiezione sul piano R^3 (componente 1; componente 2; componente 3) .	37

ELENCO DELLE FIGURE

4.1	Istogramma e normal probability plot per i rendimenti mensili di GE e GS	44
4.2	Istogramma e normal probability plot per i rendimenti mensili di JPM e KO	44
4.3	Istogramma e normal probability plot per i rendimenti mensili di MCD e NKE	44
4.4	Istogramma e normal probability plot per i rendimenti mensili di PFE e T	45
4.5	Matrice di correlazione per rendimenti mensili dal 1/1/2006 al 31/12/2017	45
4.6	Istogramma e normal probability plot per i rendimenti giornalieri di GE e GS	46
4.7	Istogramma e normal probability plot per i rendimenti giornalieri di JPM e KO	47
4.8	Istogramma e normal probability plot per i rendimenti giornalieri di MCD e NKE	47
4.9	Istogramma e normal probability plot per i rendimenti giornalieri di PFE e T	47
4.10	Matrice di correlazione per rendimenti giornalieri dal 1/1/2006 al 31/12/2017	48
4.11	Confronto frontiere efficienti campionaria e corretta per rendimenti giornalieri e mensili	50
4.12	Confronto rendimenti giornalieri e mensili del titolo Ford con la normale	52
4.13	Frontiera efficiente campionaria e corretta per rendimenti giornalieri e mensili con il titolo Ford	53
4.14	Confronto frontiere per rendimenti giornalieri	54
4.15	Confronto frontiere per rendimenti mensili	55
4.16	Matrice di covarianza per rendimenti mensili dal 01/03/2006 al 27/12/2011	56
4.17	Matrice di covarianza per rendimenti mensili dal 28/12/2011 al 28/12/2017	56
4.18	Rappresentazione di \hat{s} alle variazioni percentuali della matrice di covarianza	58
4.19	Confronto grafico fra il modello à la Bodnar e la sua stima con l'applicazione della PCA, per dati giornalieri	63

ELENCO DELLE FIGURE

4.20	Confronto grafico fra il modello à la Bodnar e la sua stima con l'applicazione della pca, per dati mensili	64
4.21	Confronto frontiera corretta à la Bodnar con quella ottenuta con l'applicazione della PCA per rendimenti mensili	65

Alla mia famiglia

INTRODUZIONE

La “Modern Portfolio Theory” (MPT) proposta da Harry Markowitz (1952) è diventata una pietra miliare della finanza contemporanea, poiché grazie alle sue ipotesi sulla distribuzione dei rendimenti è riuscita a semplificare il processo di selezione del portafoglio. Infatti, l’assunto statistico di base della MPT circa la distribuzione gaussiana dei rendimenti permette di analizzare ogni attività in base ai suoi indici di rendimento atteso e varianza.

Nonostante la teoria di Markowitz sia ancora oggi il punto di partenza per gli studi sulla selezione del portafoglio, sono state numerose le critiche al modello sia per gli assunti statistici di base sia per le condizioni, ritenute poco realistiche, poste sul mercato. Ad esempio, l’ipotesi di distribuzione normale posta sui rendimenti non trova riscontro nelle analisi empiriche, le quali evidenziano la tendenza, che si accentua per rendimenti giornalieri, ad assumere una distribuzione leptocurtica.

Inoltre, come è stato verificato da Siegel e Woodgate (2007) e Kan e Smith (2008) l’applicazione sul mercato azionario del MPT porta gli investitori a sottostimare il rischio e sovrastimare il rendimento e per questo motivo si sono sviluppati negli ultimi anni nuovi modelli con lo scopo di correggerne l’ottimismo. Tra questi troviamo lo studio proposto da T.Bodnar e O.Bodnar (2010) che si propone di derivare uno stimatore corretto per la frontiera efficiente campionaria; ma anche riguardo quest’ultimo rimangono delle perplessità in quanto si mantengono inalterate sia le ipotesi statistiche che quelle sul mercato del MPT.

La nostra analisi si concentrerà sul confronto fra le due frontiere efficienti, quella campionaria classica e la sua stima corretta, allo scopo di verificare se l’applicazione del modello à la Bodnar e Bodnar porti a dei benefici per l’investitore e riesca a superare le criticità del MPT.

Dunque la trattazione sarà eseguita trattando inizialmente dal punto di vista teorico gli studi di Markowitz e di T.Bodnar e O.Bodnar e successivamente, grazie a delle simulazioni ottenute con il programma Matlab, ne sarà fornita un'analisi empirica. Inoltre, nel terzo capitolo dell'elaborato sarà spiegata la tecnica delle componenti principali (PCA) che grazie ad una trasformazione lineare delle p variabili iniziali permette di ottenere q variabili dette appunto componenti principali caratterizzate dall'essere ortogonali fra loro. Questa tecnica verrà successivamente applicata nell'analisi empirica per verificare il comportamento del modello à la Bodnar nel caso in cui i titoli che compongono la frontiera efficiente corretta siano indipendenti.

Capitolo 1

MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

1.1 Approccio media-varianza

Per descrivere la teoria di selezione del portafoglio proposta da Markowitz partiremo analizzando le due caratteristiche della distribuzione dei rendimenti maggiormente utilizzate: il valore atteso e la deviazione standard. Questi sono i due indici che secondo l'approccio media-varianza guidano l'investitore nella composizione del proprio portafoglio, poiché cercherà sempre il titolo con il miglior pay off tra rischio e rendimento.

Markowitz assume che i rendimenti di un titolo si distribuiscano secondo una gaussiana, si utilizza di conseguenza la media aritmetica dei rendimenti per la stima del rendimento atteso e la varianza come misura della dispersione dalla media, ovvero il primo momento e il secondo momento centrale della distribuzione normale.

Per il calcolo del rendimento atteso di un titolo azionario si utilizzano le serie storiche dei prezzi:

$$\mu = \frac{1}{n} \sum_{t=1}^n \frac{P_{t+1} - P_t}{P_t}$$

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

Mentre come indice di rischiosità si usa la varianza, che indica di quanto i risultati differiscono dalla media, tanto maggiore sarà lo scarto tanto maggiore sarà il rischio dell'investimento nel titolo selezionato.

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2$$

Analizzando il processo di selezione per soli due titoli, calcolati i valori della media e della varianza, un investitore si troverà nella posizione di dover scegliere su quale attività investire il proprio capitale. Intuitivamente a parità di rischio sceglierà l'investimento che gli garantisca un rendimento atteso maggiore, mentre a parità di rendimento sceglierà quello con rischio minore. Questo processo di selezione si può attuare solo quando ci si trova in una situazione in cui un titolo domina l'altro, ovvero riesce a garantire maggior rendimento e minor o uguale rischio, o viceversa. Di norma però a titoli con maggior rendimento è associato maggior rischio, di conseguenza il criterio di dominanza spesso non è applicabile. Quest'ultima situazione è rappresentata nel piano cartesiano rendimento atteso-deviazione standard (Fig.1.1), da cui si può notare che al titolo B corrisponde sia un rendimento atteso sia una deviazione standard maggiore, dunque non sarà possibile definire quale fra le due attività sia quella dominante. Di conseguenza ogni investitore sceglierà il titolo su cui investire in base alla sua personale valutazione del rischio.

Oltre ai titoli rischiosi il mercato offre la possibilità di investire in titoli che non

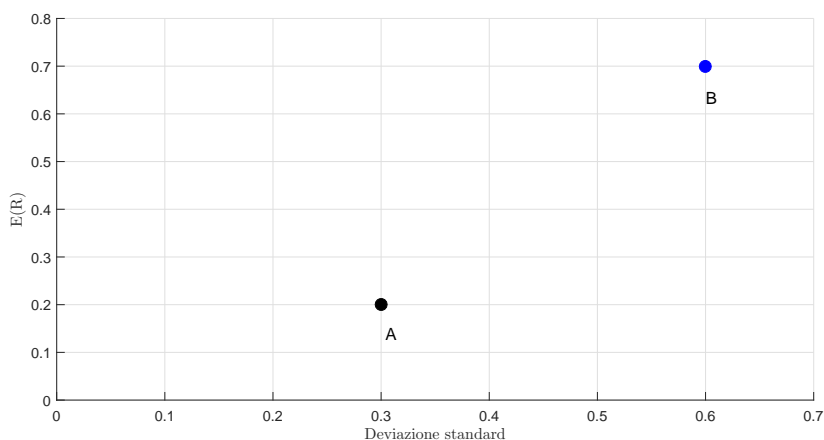


Figura 1.1: Rappresentazione della non dominanza fra il titolo A e il titolo B

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

sono caratterizzati da incertezza e offrono quindi un rendimento sicuro, questi sono detti *risk free* e ne sono un esempio i titoli di stato a breve termine. Proprio per questa loro peculiarità vengono utilizzati dagli investitori come benchmark nelle loro scelte d'investimento. L'investitore sarà allora interessato a rapportare il rendimento e il rischio offerto da un'attività rischiosa con il rendimento offerto dal titolo *risk free*, dovrà dunque ritenere che assumersi il rischio d'investimento sia giustamente remunerato.

Per misurare il maggior rischio assunto dall'investitore scegliendo il titolo rischioso si può utilizzare l'indice di Sharpe:

$$S = \frac{\mu_x - \mu_{rf}}{\sigma_x}$$

Dove μ_x definisce il rendimento medio del titolo x , μ_{rf} quello del titolo *risk free* e σ_x misura la deviazione standard di x , cioè la sua rischiosità.

Analizzando la formula si può vedere come il numeratore sia una misura del *risk premium*, ovvero il compenso che viene dato all'investitore per essersi assunto un rischio extra rispetto a quello del titolo *risk free*.

Graficamente, come rappresentato sul piano cartesiano sottostante (Fig.1.2), l'indice misura la pendenza della retta che unisce il titolo rischioso con il titolo *risk free*.

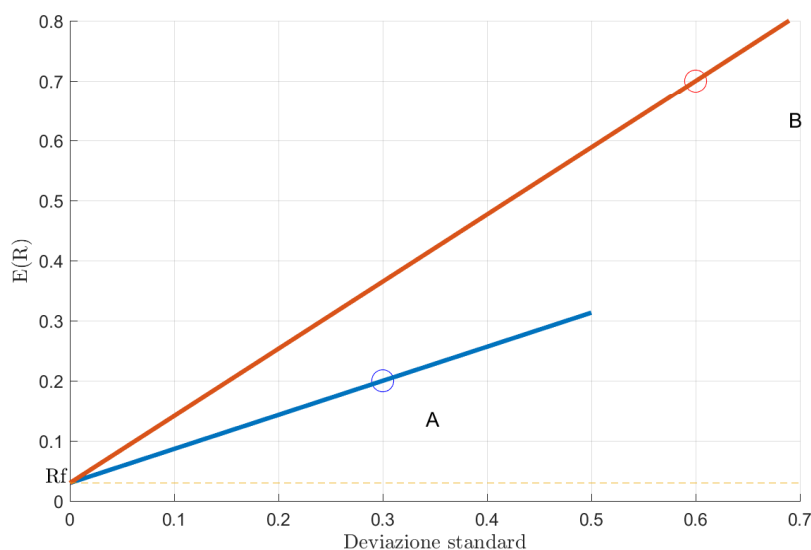


Figura 1.2: Rappresentazione grafica della CAL per il titolo A e il titolo B

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

Il titolo B che offre un rendimento maggiore avrà un indice di Sharpe maggiore e di conseguenza la retta che lo unisce con il titolo privo di rischio avrà pendenza maggiore. Quest'ultima è detta Capital Allocation Line (CAL), e mostra tutte le possibili combinazioni di rischio-rendimento per il portafoglio contenente i due titoli; la sua pendenza denota l'aumento di rendimento atteso per ogni aumento di deviazione standard. I punti della CAL che giacciono oltre il punto B indicano la possibilità per l'investitore di prendere a prestito denaro al tasso risk free e investirlo nel titolo B. Questa situazione però non è rappresentativa della realtà perché gli investitori non possono sottoscrivere prestiti al tasso risk-free, ma ad un tasso superiore. Di conseguenza dal punto B in poi la CAL assumerà una pendenza minore, proprio perché il tasso a cui è possibile indebitarsi è maggiore di quello risk-free (Fig.1.3).

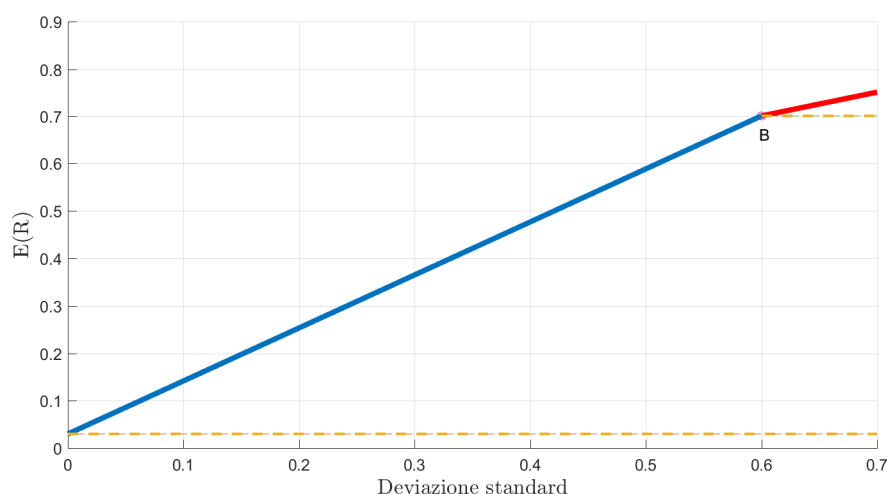


Figura 1.3: Rappresentazione grafica del cambio del tasso a cui è possibile indebitarsi, che corrisponde ad una diminuzione della pendenza della CAL

Ogni investitore dunque concentrerà il proprio capitale su una combinazione di titolo rischioso e risk-free, scegliendo fra tutti i titoli quello che offre un indice di Sharpe migliore. La quota di capitale investito nel titolo rischioso dipenderà dall'avversione al rischio dell'investitore.

Per identificare la combinazione adeguata per ogni investitore si utilizza la funzione di utilità data da:

$$U = E(r) - \frac{1}{2}A\sigma^2$$

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

Dove U è il valore dell'utilità, $E(r)$ e σ^2 sono il rendimento atteso e la varianza del titolo rischioso, e A è un indice di avversione al rischio. Il valore assunto dalla funzione e dunque l'utilità ottenuta dal portafoglio, aumenta all'aumentare del rendimento e diminuisce al crescere della varianza. L'ammontare di quanto un aumento di rischio penalizzi l'utilità dipende dal coefficiente di avversione al rischio (A), di conseguenza investitori con una forte avversione al rischio verranno maggiormente penalizzati.

Dalla funzione di utilità si possono ricavare le curve di indifferenza, lungo le quali il livello di utilità per l'investitore rimarrà costante. Graficamente queste risultano convesse, indicando che al crescere del rischio assunto per mantenere l'utilità costante c'è bisogno di un aumento del rendimento. Date le possibili combinazioni di investimento (CAL) e il coefficiente di avversione al rischio, l'investitore sceglierà il suo portafoglio ottimale. Calcolando il rendimento atteso del portafoglio come $E(r_C) = r_f + yE(r_p - r_f)$ e la sua varianza $\sigma_C^2 = y^2\sigma_p^2$ l'investitore sceglierà di investire nel titolo rischioso la quantità y che massimizzi la sua funzione di utilità. Per risolvere il problema di massimizzazione si può riscrivere la funzione di utilità come segue:

$$\text{Max}_y U = r_f + y[E(r_p) - r_f] - \frac{1}{2}Ay^2\sigma_p^2$$

Per trovare il massimo basterà calcolare la derivata prima della funzione rispetto a y e porla uguale a zero, nota la derivata seconda. Il risultato che si ottiene permette di trovare la quantità da investire nel titolo rischioso che massimizzi l'utilità dell'investitore.

$$y^* = \frac{E(r_C) - r_f}{A\sigma_p^2}$$

Quest'ultima espressione ci mostra come y^* , ovvero la quota di titolo rischioso, sia inversamente proporzionale al livello di avversione al rischio e alla varianza del titolo e direttamente proporzionale al premio al rischio. Dal punto di vista grafico lo stesso problema di ottimizzazione può essere risolto trovando fra tutte le curve di isoutilità quella tangente alla capital allocation line, come si può vedere dalla figura(1.4).

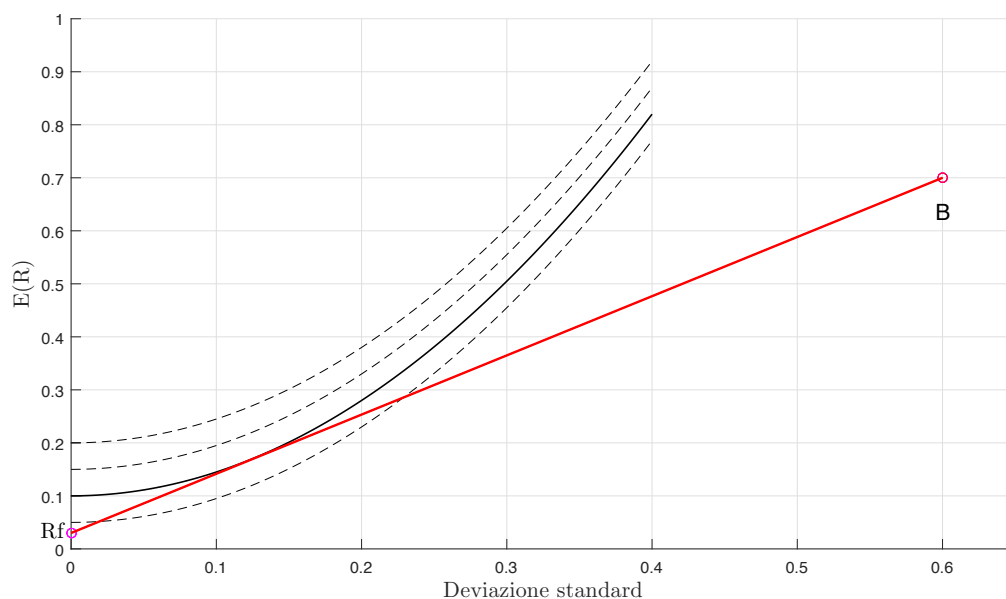


Figura 1.4: Rappresentazione della curva di Utilità tangente alla CAL

1.2 Benefici della diversificazione

Si è visto come combinando un titolo rischioso con uno risk-free si possano ottenere diverse combinazioni di rischio e rendimento, ma lo stesso approccio lo si può applicare a due titoli rischiosi. La diversificazione si basa sul presupposto che il rischio derivante da una combinazione di attività differisca dalla semplice media delle varianze delle singole attività. Infatti, la formula che esprime la varianza di due titoli rischiosi è data da:

$$\sigma_p^2 = W_A^2 \sigma_A^2 + W_B^2 \sigma_B^2 + 2W_A W_B Cov(r_A r_B)$$

Dove $Cov(r_A r_B) = \rho_{AB} \sigma_A \sigma_B$

Di conseguenza la varianza può essere riscritta come:

$$\sigma_p^2 = W_A^2 \sigma_A^2 + W_B^2 \sigma_B^2 + 2W_A W_B \rho_{AB} \sigma_A \sigma_B \quad (1.1)$$

L'espressione ci mostra come la varianza del portafoglio dipenda non solo dalla media pesata delle varianze ma anche dall'indice di correlazione fra i due titoli, il quale dà una misura di quanto si possa ridurre il rischio complessivo combinando le due attività, poiché indica la tendenza con cui queste si muovono,

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

concordemente o meno. L'indice assume sempre un valore compreso tra -1 e +1, e nel caso di correlazione positiva ($\rho > 0$) all'aumentare del rendimento di un titolo aumenta anche l'altro, viceversa in caso di correlazione negativa ($\rho < 0$) i due rendimenti si muovono in verso opposto. Va sottolineato che anche nel caso in cui la correlazione risulti positiva si ha ugualmente un beneficio dalla diversificazione, poiché solo in presenza di perfetta positività ($\rho = 1$) il rischio del portafoglio è uguale alla media delle varianze dei titoli.

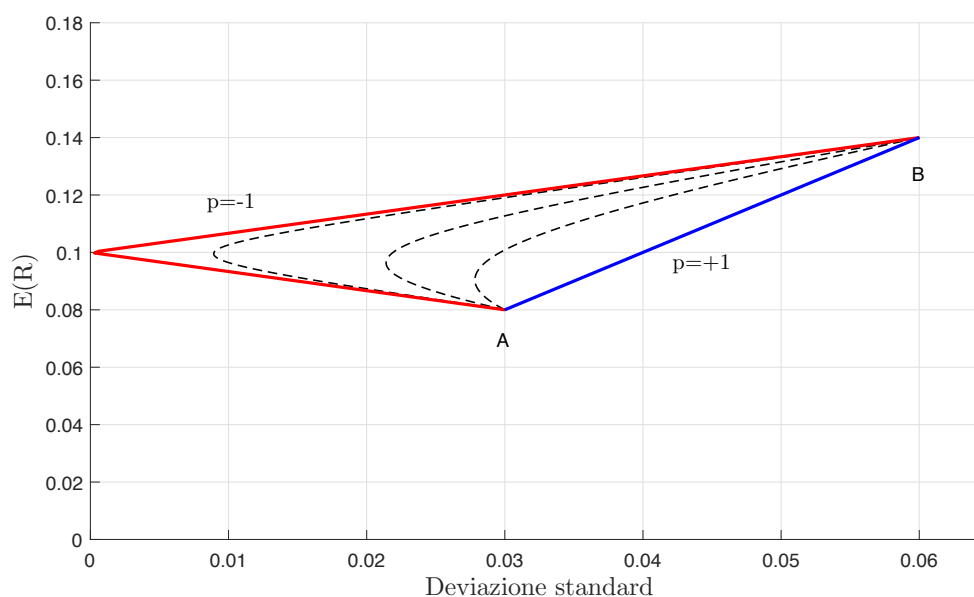


Figura 1.5: Rappresentazione dei benefici della diversificazione al cambio dell'indice di correlazione tra il titolo A e il titolo B

Il grafico (Fig.1.5) mostra come il coefficiente di correlazione vada ad influenzare le possibili combinazioni di rischio e rendimento. Di particolare importanza sono i due casi limite ovvero la perfetta correlazione positiva ($\rho = 1$), rappresentata dal segmento AB che evidenzia l'inefficacia del processo di diversificazione, e la perfetta correlazione negativa ($\rho = -1$) che permette nella teoria un investimento a rischio nullo. Analizzando nel dettaglio il caso di perfetta correlazione positiva ($\rho = 1$) fra il titolo A e il titolo B, si nota come la covarianza fra le due attività diventi il prodotto fra le due deviazioni standard, pertanto la formula della deviazione standard del portafoglio diventa:

$$\sigma_P = [W_A^2 \sigma_A^2 + (1 - W_A)^2 \sigma_B^2 + 2W_A(1 - W_A)\sigma_A\sigma_B]^{1/2}$$

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

Da cui essendo il termine tra parentesi quadre il quadrato di una somma si può riscrivere come:

$$\sigma_P = W_A\sigma_A + (1 - W_A)\sigma_B$$

In questo caso limite sia il rischio sia il rendimento del portafoglio possono essere calcolati come medie ponderate dei rischi e dei rendimenti delle singole attività. Questo significa che non risulta alcun beneficio di diversificazione rispetto all'acquisto di una sola delle due attività. Graficamente le combinazioni delle due attività giacciono su una linea retta.

Il secondo caso limite, ovvero $\rho = -1$, si ha quando i due asset si muovono esattamente in direzione opposta, questo permette di identificare una combinazione delle due attività a rischio nullo. Infatti, sostituendo $\rho = -1$ nella formula (1.1) si ottiene:

$$\sigma_P = [W_A^2\sigma_A^2 + (1 - W_A)^2\sigma_B^2 - 2W_A(1 - W_A)\sigma_A\sigma_B]^{1/2}$$

Dunque σ_P sarà dato da:

$$\sigma_P = W_A\sigma_A - (1 - W_A)\sigma_B$$

Le due curve di perfetta correlazione rappresentano anche i limiti entro i quali giacciono tutti i portafogli con valori intermedi dell'indice di correlazione. Dalla formula della deviazione standard del portafoglio si può notare come al diminuire dell'indice di correlazione la rischiosità dell'investimento diminuisca fino al caso limite di $\rho = -1$.

Riassumendo, sebbene il rendimento atteso di un qualsiasi portafoglio sia dato dalla semplice media ponderata dei rendimenti attesi dei singoli titoli, questo non è valido per la deviazione standard. Infatti, i benefici della diversificazione si hanno ogni volta che la correlazione risulta inferiore a +1 e a valori minori dell'indice corrispondono maggiori vantaggi in termini di riduzione del rischio d'investimento.

1.3 Derivazione analitica della frontiera efficiente

Abbiamo visto fin ora l'analisi per due attività rischiose e come la combinazione di queste possa portare alla diminuzione del rischio complessivo dell'investimento. Ampliando l'analisi ad una molteplicità di titoli si arriva a definire la frontiera efficiente dei portafogli, ovvero l'insieme delle combinazioni degli n titoli che presentano la minima deviazione standard a parità di rendimento atteso.

Graficamente la frontiera si presenta concava al di sopra del portafoglio di minima varianza e convessa al di sotto. Tutti i portafogli che giacciono sulla parte concava della frontiera vengono definiti efficienti, al contrario quelli che sono situati al di sotto del punto di minima varianza sono definiti inefficienti, questo perché a parità di rischio la combinazione che offre il miglior rendimento si trova sempre nella parte superiore della frontiera. Il vertice della frontiera definisce il portafoglio di minima varianza ovvero la combinazione dei titoli selezionati che permette di ottenere il minor rischio possibile.

La derivazione matematica della frontiera efficiente con n titoli rischiosi è più complessa di quella già vista con soli due titoli rischiosi, in quanto prevede la soluzione di un problema di ottimo vincolato. Infatti, l'investitore dopo aver scelto gli n titoli che formeranno il portafoglio potrà o minimizzare il rischio dato un certo livello di rendimento desiderato o viceversa massimizzare il rendimento dato un livello di rischio ideale.

Nella nostra analisi vedremo come identificare la frontiera dei portafogli ottimi dato il livello di rendimento.

Date le seguenti quantità:

- E vettore dei rendimenti attesi degli n titoli selezionati.
- u vettore composto da soli 1.
- Σ matrice di varianza e covarianza fra gli n titoli selezionati.

La funzione obiettivo sarà:

$$\min_{\sigma^2} = \alpha^T \Sigma \alpha$$

Soggetta ai due vincoli:

1) $u^T \alpha = 1$

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

$$2) E^T \alpha = \mu$$

La prima condizione indica che la somma delle quote dei titoli deve sempre sommare ad 1. Il secondo vincolo invece indica che il vettore dei rendimenti attesi (E) moltiplicato per il vettore dei pesi (α) deve dare come risultato il rendimento atteso obiettivo (μ).

La soluzione del problema passa per tre fasi, inizialmente si troveranno le quote di portafoglio attraverso la soluzione del problema di ottimo vincolato, successivamente si calcolerà il moltiplicatore di Lagrange e infine si individuerà la funzione della frontiera efficiente. Si inizia dunque formando il Lagrangiano:

$$\mathcal{L} = \frac{1}{2} \alpha^T \Sigma \alpha + \lambda(1 - u^T \alpha) + \gamma(\mu - E^T \alpha)$$

Derivando la funzione rispetto ad α ed uguagliandola a 0 si ottiene:

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \Sigma \alpha - \lambda u - \gamma E = 0$$

Da cui

$$\Sigma \alpha = \lambda u + \gamma E$$

Si ottengono così le quote di portafoglio ottimale α^* :

$$\alpha^* = \lambda \Sigma^{-1} u + \gamma \Sigma^{-1} E$$

Per ricavare i valori di λ e di γ utilizzeremo i vincoli, dunque basterà risolvere il seguente sistema di equazioni.

$$\begin{cases} 1 = u^T \alpha^* = \lambda u^T \Sigma^{-1} u + \gamma u^T \Sigma^{-1} E \\ \mu = E^T \alpha^* = \lambda E^T \Sigma^{-1} u + \gamma E^T \Sigma^{-1} E \end{cases}$$

Per comodità applichiamo le seguenti sostituzioni:

- $A = u^T \Sigma^{-1} u$,
- $B = u^T \Sigma^{-1} E$,
- $C = E^T \Sigma^{-1} E$,

Possiamo allora riscrivere il problema precedente in forma matriciale come:

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 \\ \mu \end{pmatrix}$$

Da cui possiamo così ricavare il valore dei moltiplicatori:

$$\lambda = \frac{1}{D}(C - \mu B)$$

$$\gamma = \frac{1}{D}(A\mu - B)$$

$$\text{Con } D = AC - B^2$$

Inserendo nella formula della varianza gli α^* ottimali e i valori dei moltiplicatori λ e γ che si sono trovati in precedenza si ottiene la funzione della frontiera efficiente.

$$\sigma^2 = \alpha^{*T} \Sigma \alpha^* = \lambda + \gamma \mu = \frac{1}{D}(A\mu^2 - 2B\mu + C)$$

Quest'ultima è la formula analitica per la derivazione della frontiera efficiente alla Markowitz in presenza di soli titoli rischiosi. Dal punto di vista grafico la funzione verrà rappresentata come una parabola caratterizzata per avere nel vertice il punto di minima varianza (GMV), ovvero il portafoglio ottimale al quale corrisponde il minor livello di rischio. Questo può essere facilmente calcolato, in quanto possiede rendimento atteso pari a $\mu_{GMV} = \frac{B}{A}$ e varianza $\sigma_{GMV}^2 = \frac{1}{A}$.

Come risultato dalla teoria sviluppata da Markowitz detenere un maggior numero di attività nel proprio portafoglio permette di ridurre il rischio complessivo. Infatti, è facile notare come al crescere del numero dei titoli il rischio del portafoglio diminuisca. Nonostante questo, non è possibile arrivare ad un livello di rischio nullo.

Per dimostrarlo esprimiamo la varianza del portafoglio come:

$$\sigma_P^2 = \frac{1}{N} \bar{\sigma}_j^2 + \frac{N-1}{N} \bar{\sigma}_{jk}$$

Da cui si nota come, al crescere di N la parte di varianza del portafoglio che dipende dalle varianze dei singoli titoli va a zero, mentre la parte delle covarianze tende alla covarianza media delle attività ($\bar{\sigma}_{jk}$). Questo dimostra che il contributo delle singole attività al rischio di portafoglio può essere diversificato, ma quello derivante dalle covarianze non potrà essere eliminato del tutto.

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

Il grafico sottostante (Fig.1.6)¹ mostra proprio la situazione appena descritta, si può vedere come il beneficio della diversificazione diminuisca al crescere del numero di attività fino ad arrivare ad un livello oltre il quale non si ottiene alcuna riduzione di rischio nell'aumentare la quantità di titoli.

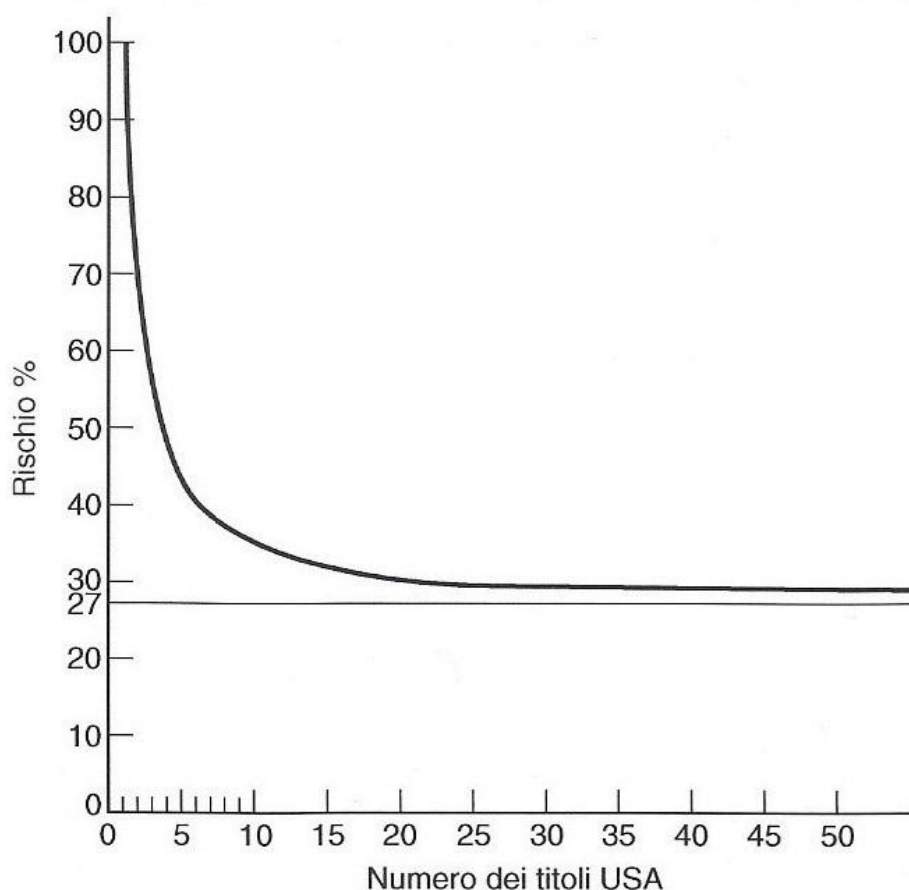


Figura 1.6: Effetto della numerosità dei titoli sul rischio del portafoglio negli USA (1975)

La parte di rischio definito dalla covarianza media delle attività è detto rischio non diversificabile o rischio di mercato, poiché non dipende dai singoli titoli, ma dalle condizioni del mercato di riferimento. Infatti, la quota di rischio non diversificabile risulterà diversa in base al mercato di riferimento e quindi varierà da Paese a Paese.

¹ELTON E., GRUBER M., BROWN S. E GOETZMANN W., Teorie di portafoglio e analisi degli investimenti, *Apogeo Education*, 11-2017

1.4 Critiche al modello di Markowitz

Il modello proposto da Markowitz rimane ad oggi il punto di partenza per lo studio delle dinamiche di selezione del portafoglio, nonostante le critiche al modello siano state numerose. In questo paragrafo andremo a vedere quelle che sono le ipotesi postulate dall'autore sia sugli investitori sia sul mercato e le critiche che sono state mosse al modello.

La teoria proposta da Harry Markowitz si basa sull'ipotesi di normalità dei rendimenti, individuando nei due momenti della distribuzione gaussiana, media e varianza gli unici indici necessari per valutare un investimento.

Le altre ipotesi su cui si basa il modello sono:

- Per acquistare e vendere titoli non sono presenti né costi di transazione né tasse.
- Il mercato è perfettamente liquido, quindi l'investitore potrà acquistare qualsiasi quantità del titolo, anche una frazione.
- Gli investitori sono razionali e avversi al rischio, dunque preferiranno a parità di rendimento atteso sempre l'investimento che offre minor rischio.
- L'investimento è considerato uniperiodale, non è preso in considerazione un eventuale reinvestimento della ricchezza.

Come precedentemente affermato si sono mosse numerose critiche al modello, sia sulle ipotesi statistiche sia su quelle del mercato, perchè definito poco rappresentativo della realtà. Analizzando inizialmente le ipotesi proposte da Markowitz sul mercato si può notare come l'autore definisca l'assenza di costi di transazioni e di tasse per l'acquisto e la vendita di titoli, condizioni che nella realtà non sono verificate. E' stato dimostrato che inserendo queste due componenti nel calcolo della frontiera efficiente si ottiene una diminuzione del rendimento atteso, mentre il rischio dei portafogli rimane invariato. Dunque la frontiera del modello con la presenza di costi di transazione e tassazione risulterà sempre posizionata al di sotto di quella classica, implicando che il modello à la Markowitz sovrastima il

rendimento atteso ².

La perfetta razionalità degli investitori, implica che questi sceglieranno sempre l'investimento che massimizzi la loro utilità e sottointende la presenza di informazione perfetta, poichè se non disponessero di tutte le informazioni non potrebbero compiere la scelta ottima. Una delle maggiori critiche a questo concetto viene dagli studi di Herbert Simon che nel suo articolo " A behavioral model of rational choice" del 1955 sostiene che gli individui non seguano nelle loro scelte i principi economici, ma vengano condizionati dalle loro esperienze passate, dalle informazioni possedute e dalle loro credenze; il risultato è quindi una razionalità limitata dell'investitore. Dallo studio di Simon si sono seguite diverse analisi sui comportamenti e sulle scelte degli agenti economici, che oggi possiamo classificare come studi di finanza comportamentale. Questa scienza attraverso l'applicazione della psicologia cognitiva si pone l'obbiettivo di spiegare e comprendere quelle che sono le scelte economiche e come queste influenzino il mercato. Uno dei maggiori contributi alla finanza comportamentale è stato dato da Daniel Kahneman e Amos Tversky con la loro teoria del prospetto (1979) ³, in cui gli autori dimostrano, grazie ad evidenze empiriche, come gli investitori siano influenzati dal contesto in cui avviene la loro scelta e dalle metodologie con cui questa gli viene proposta (effetto contesto). Inoltre, gli autori hanno riscontrato che gli agenti economici sono molto più sensibili alle perdite che ai guadagni dunque preferiranno evitare una perdita piuttosto che ottenere un guadagno. Di conseguenza non si possono considerare gli investitori come agenti economici perfettamente razionali poichè le loro scelte risulteranno sempre condizionate.

In questa seconda parte dell'analisi sulle ipotesi del modello di Markowitz ci focalizzeremo sulle critiche derivanti dai suoi assunti statistici. Il modello pone le sue basi anche sull'assunzione della distribuzione normale dei rendimenti, ma come evidenziato dagli studi empirici di Mandelbrot (1963) ⁴ e Fama (1965) ⁵

²MARASOVIC B., PIVAC S. E VUKASOVIC V., The impact of transaction costs on rebalancing an investment portfolio in portfolio optimization, *World Academy of Science, Engineering and Technology*, 9(3), 2015

³KAHNEMAN D. E TVERSKY A., Prospect Theory: An Analysis of Decision Under Risk, *The Econometric Society*, 47(2), pp 263-292, 1979

⁴MANDELBROT B., The Variation of Certain Speculative Prices, *The Journal of Business*, 36(4), pp 394-419, 1963

⁵FAMA E., The Behavior of Stock-Market prices, *Journal of Business*, 38(1), pp 34-105, 1965

CAPITOLO 1. MODELLO SELEZIONE DEL PORTAFOGLIO: L'APPROCCIO DI MARKOWITZ

sulla distribuzione dei rendimenti si evidenzia la presenza di leptocurtosi, ovvero code più pesanti e valori centrali più frequenti di quelli della gaussiana. Questo implica che i rendimenti molto piccoli o molto grandi avranno una probabilità di verificarsi maggiore di quello che viene stimato dalla normale.

L'indice di curtosi o momento quarto di una distribuzione, può essere calcolato come:

$$\mathbb{K}[X] = \frac{\mathbb{E}[(X - \mu)^4]}{\sigma^4}$$

Dove μ e σ rappresentano media e deviazione standard di X .

Per la distribuzione normale si ha $\mathbb{K}[X] = 3$ mentre quando $\mathbb{K}[X] > 3$ si è in presenza di un eccesso di curtosi e quindi di una forma leptocurtica. La tendenza dei rendimenti ad assumere questa caratteristica si accentua nel caso di rendimenti giornalieri e diminuisce nel caso di quelli mensili.

Da studi successivi sulla bontà della frontiera efficiente sono emersi dubbi sulla sua effettiva capacità di rappresentare efficacemente un'analisi d'investimento. Tra questi quello di Siegel e Woodgate⁶ e quello di Taras e Olha Bodnar⁷. Nel primo, pubblicato nel 2007, gli autori descrivono come la frontiera efficiente classica, quella proposta da Markowitz, sia distorta in quanto non prenda in considerazione la stima dell'errore derivante dall'uso di un campione al posto della popolazione, che comporta una sovrastima del rendimento atteso e una sottostima del rischio dell'investimento. Il successivo studio pubblicato da Taras e Olha Bodnar nel 2010 riprende il precedente lavoro di Siegel e Woodgate e propone uno stimatore non distorto della frontiera efficiente che corregge le problematiche nella stima dei parametri. Nei capitoli seguenti si analizzerà nel dettaglio il modello proposto dai due autori e si evidenzieranno le problematiche emerse in fase di riproduzione.

⁶SIEGEL A. E WOODGATE A., Performance of Portfolios Optimized with Estimation Error, *Management Science*, 53(6), pp. 1005-1015, 2007

⁷BODNAR T. E BODNAR O., On the Unbiased Estimator of the Sample Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 13(7), pp. 1065-1073, 2010

Capitolo 2

IL MODELLO PROPOSTO DA BODNAR E BODNAR

2.1 Derivazione degli stimatori campionari corretti

Nel corso del nostro elaborato andremo ad analizzare nel dettaglio il modello proposto da Taras e Olha Bodnar(2010), ma per fare questo sarà necessario riprendere l'analisi svolta da Siegel e Woodgate(2007) a cui il modello si ispira.

Gli autori sostengono che le performance che si ottengono utilizzando l'approccio di Markowitz nella selezione del portafoglio risulteranno spesso inferiori alle aspettative, poiché si utilizzano i dati storici dei rendimenti per stimare i parametri ignoti della distribuzione di probabilità dei titoli. Di conseguenza il rendimento atteso e la varianza che verranno calcolati risulteranno una stima dei veri parametri, ma nonostante questo nella definizione della frontiera saranno utilizzati come i veri parametri. Siegel e Woodgate definiscono il procedimento di stima classico come "naive" o ingenuo poiché non prende in considerazione gli errori di stima derivanti dall'utilizzo di un campione al posto della popolazione nel determinare i parametri. L'effetto è quello di definire una frontiera distorta poiché rispetto alla sua forma corretta, proposta dagli autori, risulterà sovrastimare il rendimento atteso e sottostimare la varianza. Dallo studio emerge come l'effetto della correzione sia sulla media sia sulla deviazione standard sia maggiore quando si è in presenza di una numerosità di titoli elevata e ad un breve orizzonte temporale. Infatti, come è presumibile aspettarsi, la correzione è inversamente proporzionale

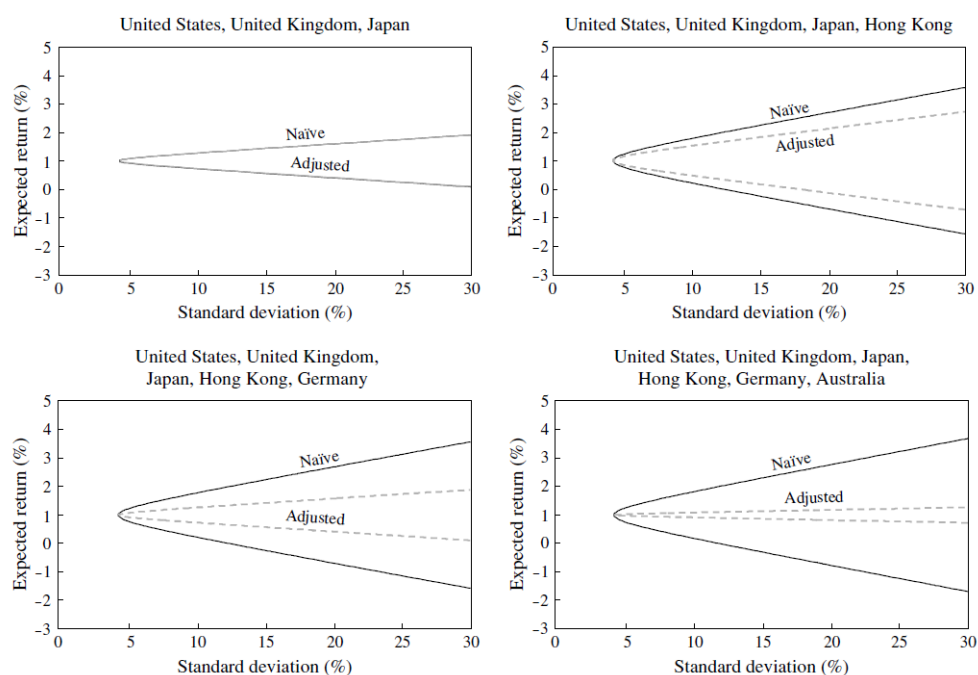


Figura 2.1: Frontiera efficiente con e senza aggiustamento, per portafogli creati con numeri diversi di titoli

al numero di osservazioni, poiché aumentando il numero dei dati la stima diventa più affidabile, e direttamente proporzionale al numero di attività del portafoglio poiché al loro crescere aumenteranno anche le probabilità di commettere errori nel processo di ottimizzazione. Gli autori forniscono la dimostrazione empirica della teoria, come rappresentato in figura (Fig.2.1)¹, dove analizzano le serie storiche dei rendimenti mensili dal 1970 al 2003 su vari mercati globali. L'utilizzo della frontiera efficiente corretta permette all'investitore di ottenere sempre un vantaggio in termini di aumento dell'utilità che sarà direttamente proporzionale al numero di titoli che compongono il portafoglio. Infatti, aumentando il numero di titoli la differenza tra la frontiera "naïve" e quella corretta aumenta e nel caso limite di sole 3 attività non è necessaria alcuna correzione. Gli autori propongono l'analisi con il processo di ricampionamento bootstrap che permette di ottenere risultati robusti all'eventuale distribuzione non normale dei rendimenti.

L'elaborato proposto da Bodnar e Bodnar propone uno stimatore corretto della frontiera efficiente classica, che come descritto da Siegel e Woodgate sovrastima

¹SIEGEL A. E WOODGATE A., Performance of Portfolios Optimized with Estimation Error, *Management Science*, 53(6), pp. 1005-1015, 2007

il rendimento e sottostima il rischio. Gli autori inizialmente forniscono la derivazione della frontiera efficiente alla Markowitz data dalla parte superiore della parabola formata da:

$$g_p(R, V) = (R - R_{GMV})^2 - s(V - V_{GMV}) \quad (2.1)$$

Dove:

$$s = \mu' R \mu \quad \text{con} \quad R = \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1} \mathbf{1}' \Sigma^{-1}}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}$$

s è il parametro che indica la pendenza della frontiera efficiente, mentre:

$$R_{GMV} = \frac{\mathbf{1}' \Sigma^{-1} \mu}{\mathbf{1}' \Sigma^{-1} \mathbf{1}} \quad \text{e} \quad V_{GMV} = \frac{1}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}$$

indicano il rendimento atteso e la varianza del portafoglio a minima varianza (GMV), ovvero il particolare portafoglio con il minimo rischio possibile. Nonostante la frontiera efficiente della popolazione dipenda solamente da 3 parametri, la sua derivazione è complicata. Per questo motivo si utilizza la sua forma campionaria che si ottiene applicando alla (2.1) i rispettivi stimatori campionari $\hat{\mu}$ e $\hat{\Sigma}$:

$$g_s(R, V) = (R - \hat{R}_{GMV})^2 - \hat{s}(V - \hat{V}_{GMV}) \quad (2.2)$$

Questa, come dimostrato da Siegel e Woodgate sovrastima la frontiera della popolazione.

Assumendo che i rendimenti dei titoli siano serialmente indipendenti e normalmente distribuiti si deriva l'esatta distribuzione di \hat{R}_{GMV} , \hat{V}_{GMV} e \hat{s} , come dimostrato da Bodnar e Schmid ².

PROPOSIZIONE 2.1. Assumendo X_1, \dots, X_n un vettore di variabili casuali indipendenti e sia $X_i \sim \mathcal{N}_k(\mu, \Sigma)$ per $i = 1, \dots, n$. Con Σ definita positivamente. Siano $k > 2$ i titoli e $n > k$ le osservazioni.

si ottiene:

²BODNAR T. E SCHMID W., Econometrical Analysis of the Sample Efficient Frontier, *The European Journal of Finance*, 15(3), pp. 317-335, 2009

- (a) \hat{V}_{GMV} è indipendente da (\hat{R}_{GMV}, \hat{s}) .
- (b) $(n-1)\hat{V}_{GMV}/V_{GMV} \sim \chi_{n-k}^2$.
- (c) $\frac{n(n-k+1)}{(n-1)(k-1)}\hat{s} \sim F_{k-1, n-k+1, ns}$.
- (d) $\hat{R}_{GMV}|\hat{s} = y \sim \mathcal{N}(R_{GMV}, \frac{1+\frac{n}{n-1}y}{n}V_{GMV})$.

Ottenuta, grazie alla Proposizione 2.1, l'esatta distribuzione dei parametri $(\hat{R}_{GMV}, \hat{V}_{GMV}$ e $\hat{s})$ della frontiera efficiente campionaria, si dimostra che questa non sia uno stimatore corretto per la stima della frontiera calcolata sulla popolazione.

Si assuma il seguente lemma:

Lemma 2.1 siano X_1, \dots, X_n vettori di variabili casuali indipendenti e sia $X_i \sim \mathcal{N}_k(\mu, \Sigma)$ per $i = 1, \dots, n$. Si assuma Σ definita positivamente. Con $k > 2$ e $n > k$. Si dimostra che:

- (a) $E(\hat{s}) = \frac{n-1}{n-k-1}s + \frac{(n-1)(k-1)}{n(n-k-1)}$.
- (b) $E((R - \hat{R}_{GMV})^2) = (R - R_{GMV})^2 + (\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s)V_{GMV}$.
- (c) $E(\hat{V}_{GMV}) = \frac{n-k}{n-1}V_{GMV}$.

Applicando l'enunciato c della Proposizione 2.1 e la proprietà della distribuzione F non centrata al punto b della Proposizione 2.1 si ottiene:

$$E((R - \hat{R}_{GMV})^2) = (R - R_{GMV})^2 + Var(\hat{R}_{GMV})$$

dove applicando la Proposizione 2.1d si ottiene:

$$\begin{aligned} Var(\hat{R}_{GMV}) &= Var(E(\hat{R}_{GMV}|\hat{s})) + E(Var(\hat{R}_{GMV}|\hat{s})) \\ &= Var(R_{GMV}) + \left(\frac{1}{n} + \frac{1}{n-1}E(\hat{s})\right)V_{GMV} \\ &= \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s\right)V_{GMV} \end{aligned} \quad (2.3)$$

Da quest'ultima uguaglianza e grazie agli altri enunciati del Lemma 2.1 si arriva dimostrare che $E(g_s(R, V)) \neq 0$ per $(R, V) \in \mathbb{R} \times \mathbb{R}_+$ così $g_p(R, V) = 0$. Allora:

$$\begin{aligned} E(g_s(R, V)) &= E((R - \hat{R}_{GMV})^2) - E(\hat{s}(V - \hat{V}_{GMV})) \\ &= E((R - \hat{R}_{GMV})^2) - E(\hat{s})(V - E(\hat{V}_{GMV})) \end{aligned}$$

Questo perchè le varie quantità sono indipendenti, come definito nel Proposizione 2.1a.

Applicando il Lemma 2.1 si dimostra che:

$$E(g_s(R, V)) = (R - R_{GMV})^2 + \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1} s \right) V_{GMV} - \left(\frac{n-1}{n-k-1} s + \frac{(n-1)(k-1)}{n(n-k-1)} \right) \left(V - \frac{n-k}{n-1} V_{GMV} \right)$$

Dunque

$$E(g_s(R, V)) = g_p(R, V) - \left(\frac{k}{n-k-1} s + \frac{(n-1)(k-1)}{n(n-k-1)} \right) V + \left(\frac{2}{n-k-1} s + \frac{(n-k)(k-1) + n-2}{n(n-k-1)} \right) V_{GMV} \quad (2.4)$$

Da questa dimostrazione si può notare come il valore atteso dello stimatore della frontiera efficiente campionaria sia diverso dal suo valore teorico calcolato sulla popolazione. Infatti, $E(g_s(R, V))$ è diverso da $g_p(R, V)$ e quindi lo stimatore non è definibile come corretto. Esiste un solo punto in cui la frontiera efficiente campionaria fornisce una stima corretta, ovvero il portafoglio per cui vale l'uguaglianza $V = \gamma V_{GMV}$ con:

$$\gamma = \frac{\frac{2}{n-k-1} s + \frac{(n-k)(k-1) + n-2}{n(n-k-1)}}{\frac{k}{n-k-1} s + \frac{(n-1)(k-1)}{n(n-k-1)}} = \frac{2ns + (n-k)(k-1) + n-2}{kns + (n-k)(k-1)}$$

Quindi per ogni valore di $V > \gamma V_{GMV}$ la frontiera campionaria sovrastima quella della popolazione, al contrario per i valori di $V < \gamma V_{GMV}$ la sottostima. Inoltre se $1 + \frac{-ns + \sqrt{(ns+2)^2 + 4n-12}}{2} < k < n$ si ottiene che $\gamma < 1$ e dunque in questo caso la frontiera campionaria sovrastima la frontiera efficiente della popolazione per tutti i $V \geq V_{GMV}$. L'ultima equazione ci mostra come i portafogli efficienti siano sovrastimati quando il numero di titoli è ampio rispetto alla dimensione del campione. Questo lo si può vedere dalla figura (Fig.2.2)³, in cui è rappresentato il limite inferiore per il numero di attività nel portafoglio k , in funzione di s per il quale la frontiera efficiente campionaria sovrastima quella sulla popolazione. Dal grafico si vede anche come k sia una funzione decrescente in s .

Si osserva che, anche per portafogli composti da un numero limitato di titoli ($k < 10$) è presente una distorsione per tutti i valori di V indipendentemente dalla

³BODNAR T. E BODNAR O., On the Unbiased Estimator of the Sample Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 13(7), pp. 1065-1073, 2010

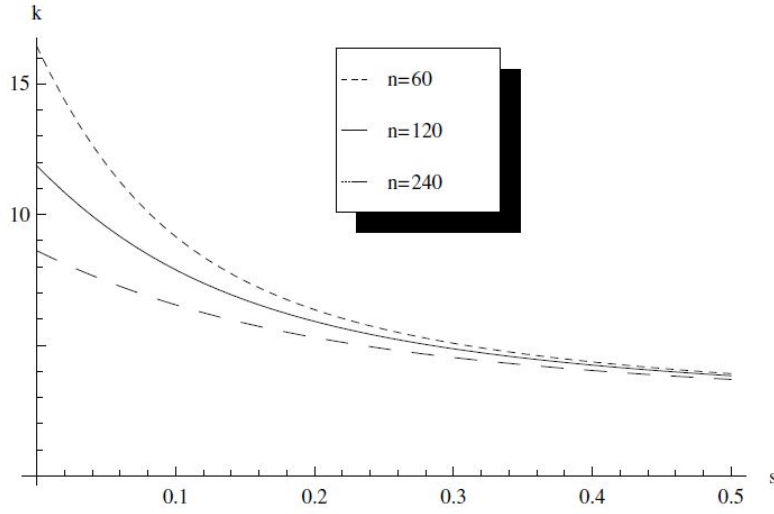


Figura 2.2: Limite inferiore per il numero di asset (k) nel portafoglio in funzione di s per il quale la frontiera efficiente campionaria sovrastima la frontiera efficiente della popolazione per tutti i V ($n \in \{60, 120, 240\}$)

grandezza del campione se $s > 0.2$. Il risultato ottenuto conferma quanto dimostrato da Siegel e Woodgate, ovvero che la frontiera campionaria sia ottimistica e sovrastimi la reale posizione della frontiera efficiente nello spazio media-varianza. Nella fase successiva del loro studio gli autori forniscono uno stimatore corretto per la frontiera efficiente campionaria che vada a correggerne l'ottimismo.

Teorema 2.1: Assumendo X_1, \dots, X_n un vettore di variabili casuali indipendenti e sia $X_i \sim \mathcal{N}_k(\mu, \Sigma)$ per $i = 1, \dots, n$. Con Σ definita positivamente. Siano $k > 2$ e $n > k$. Allora lo stimatore non distorto della frontiera è dato da:

$$g_u(R, V) = (R - \hat{R}_{GMV})^2 - \frac{(n-2)(n-1)}{n(n-k)(n-k-1)} \hat{V}_{GMV} - \left(\frac{n-k-1}{n-1} \hat{s} - \frac{k-1}{n} \right) \left(V - \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)} \hat{V}_{GMV} \right). \quad (2.5)$$

Affinchè il teorema sia valido è necessario che $E(g_u(R, V)) = 0$ per tutti i punti $(R, V) \in \mathbb{R} \times \mathbb{R}_+$ così che $g_p(R, V) = 0$. Si consideri

$$E(g_u(R, V)) = E((R - \hat{R}_{GMV})^2) - E\left(\frac{(n-2)(n-1)}{n(n-k)(n-k-1)} \hat{V}_{GMV} \right) - E\left(\left(\frac{n-k-1}{n-1} \hat{s} - \frac{k-1}{n} \right) \left(V - \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)} \hat{V}_{GMV} \right) \right).$$

Utilizzando quanto esposto nella Proposizione 2.1 e nel lemma 2.1 si può sostenere che:

$$\begin{aligned} E(g_u(R, V)) &= (R - R_{GMV})^2 + \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s \right) V_{GMV} \\ &\quad - \frac{n-2}{n(n-k-1)} V_{GMV} - s \left(V - \frac{n-k-2}{n-k-1} \hat{V}_{GMV} \right) \\ &= (R - R_{GMV})^2 - s(V - V_{GMV}) = g_p(R, V) = 0 \end{aligned}$$

Il risultato del Teorema 2.1 mostra che l'applicazione di \hat{V}_{GMV} e di \hat{s} per stimare la frontiera efficiente porta ad una considerevole distorsione nello stimatore della frontiera.

Un applicazione pratica è mostrata in Fig.2.3⁴ dove lo stimatore corretto della frontiera efficiente $g_u(R, V)$ -linea tratteggiata- è confrontato con lo stimatore campionario $g_s(R, V)$ -linea continua- dato dall'equazione 2.2.

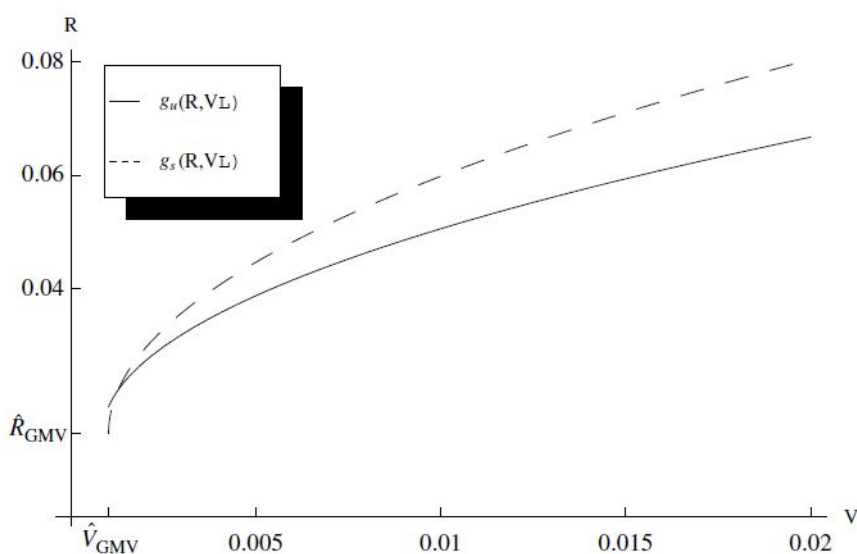


Figura 2.3: Stimatore campionario e corretto della frontiera efficiente

I dati utilizzati per la simulazione sono: $\hat{R}_{GMV} = 0.00145664$, $\hat{V}_{GMV} = 0.0010337$, $\hat{s} = 0.221457$, $k = 5$, e $n = 60$. Per i valori elencati risulta $\gamma = 1.00707$, di conseguenza avremo una sottostima per $V < \gamma V_{GMV} = 0.00104101$, mentre si avrà una sovrastima della frontiera sulla popolazione per $V > 0.00104101$. Dal punto di vista grafico

⁴BODNAR T. E BODNAR O., On the Unbiased Estimator of the Sample Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 13(7), pp. 1065-1073, 2010

si nota come, $g_u(R, V)$ giaccia al di sotto di $g_s(R, V)$, per valori elevati di V , e al di sopra per valori piccoli. Questo indica che lo stimatore non distorto corregge l'effetto ottimismo derivante dall'utilizzo della frontiera campionaria, evitando agli investitori di aspettarsi dei rendimenti maggiori per il rischio assunto.

2.2 F-Test sulla Frontiera Efficiente

Gli autori nella fase finale del loro studio propongono un test statistico sul parametro s che, indicando la pendenza della frontiera efficiente, misura di quanto aumenti il profitto del portafoglio all'aumentare di un unità di varianza. Quando s assume valore uguale a zero, la frontiera efficiente della popolazione degenera in una linea retta, ed in questo caso l'unico portafoglio possibile è quello di minima varianza (GMV).

Riportando il test sviluppato dagli autori si consideri:

$$\begin{cases} H_0 : s = 0 \\ H_1 : s > 0 \end{cases}$$

Dal punto di vista finanziario accettare l'ipotesi nulla significa che il coefficiente s della frontiera efficiente sia uguale a 0, questo implica per l'investitore la mancanza di guadagno nel prendere una pozione rischiosa. L'unica alternativa sarebbe quella di investire nel portafoglio di minima varianza, poichè essendo la frontiera una retta parallela all'asse delle ascisse un investimento in una combinazione diversa da questa porterebbe ad un maggior rischio ma allo stesso rendimento atteso. L'ipotesi nulla nel test viene rifiutata se:

$$T_S = \frac{n(n-k+1)}{(n-1)(k-1)} \hat{s} > F_{k-1, n-k+1; \alpha}$$

Dove $F_{k-1, n-k+1; \alpha}$ indica il quantile di ordine α della distribuzione F non centrale con $k-1$ e $n-k+1$ gradi di libertà. Lo stesso risultato si può ottenere testando $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, poichè in entrambi i casi si ha la presenza di un unico portafoglio ottimo, ed è il portafoglio di minima varianza.

2.3 Limiti dell'approccio di Bodnar e Bodnar

Come abbiamo avuto modo di vedere nel corso del capitolo, il modello presentato da O.Bodnar e T.Bodnar fornisce uno stimatore non distorto per l'analisi della frontiera efficiente, che sebbene riesca a colmare alcune lacune del modello di Markowitz, non riesce a superare i limiti posti sul mercato e sugli investitori. Infatti, gli autori non si esprimono riguardo le ipotesi poco realistiche di perfetta liquidità, investitori perfettamente razionali e assenza di costi di transazione. In aggiunta il modello preserva le stesse condizioni, anche qui poco realistiche, sulla distribuzione dei rendimenti mensili, ovvero la loro normalità e la loro indipendenza.

Da studi empirici si è rilevato che i rendimenti mensili non siano affetti da autocorrelazione e di conseguenza si esclude una loro dipendenza nel tempo. Lo stesso non si può affermare per i rendimenti elevati al quadrato, i quali mostrano valori significativi di autocorrelazione.

Riguardo la condizione di normalità, come si è visto nel capitolo precedente, la sua assunzione può essere accettata per rendimenti di ordine mensile, mentre per dati giornalieri non è sempre verificata.

Inoltre, andando a studiare il comportamento asintotico della frontiera esposta nel modello, si vede come questa all'aumentare del numero di osservazioni e del numero di titoli tenda alla frontiera classica. Infatti possiamo riscrivere la funzione della frontiera come:

$$g_u(R, V) = (R - \hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} - (B \cdot \hat{\sigma} - C)(V - D \cdot \hat{V}_{GMV}) = 0$$

Dove i valori di A , B , C e D sono dati da:

$$\begin{aligned} A &= \frac{(n-2)(n-1)}{n(n-k)(n-k-1)} && \text{con } A \sim \frac{1}{n}, \\ B &= \frac{n-k-1}{n-1} && \text{con } B \sim 1, \\ C &= \frac{k-1}{n} && \text{con } C \sim \frac{1}{n}, \\ D &= \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)} && \text{con } D \sim 1. \end{aligned}$$

Da cui, andando a studiare il comportamento di questi parametri al crescere della numerosità campionaria si ottiene:

$$\lim_{n \rightarrow +\infty} (A) = 0,$$

$$\lim_{n \rightarrow +\infty} (B) = 1,$$

$$\lim_{n \rightarrow +\infty} (C) = 0,$$

$$\lim_{n \rightarrow +\infty} (D) = 1.$$

Allora sostituendo i rispettivi valori limite nella formula dello stimatore proposto da Bodnar avremo:

$$\begin{aligned} g_u(R, V) &= (R - \hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} - (B \cdot \hat{s} - C)(V - D \cdot \hat{V}_{GMV}) = 0 \\ &= (R - \hat{R}_{GMV})^2 - 0 \cdot \hat{V}_{GMV} - (1 \cdot \hat{s} - 0)(V - 1 \cdot \hat{V}_{GMV}) = 0 \\ &= (R - \hat{R}_{GMV})^2 - \hat{s}(V - \hat{V}_{GMV}) = 0 \end{aligned}$$

Quest'ultima è proprio la funzione della frontiera efficiente campionaria, quindi si può affermare che:

$$\lim_{n \rightarrow +\infty} g_u(R, V) = g_s(R, V)$$

Questo implica che, all'aumentare della numerosità campionaria di riferimento, l'effetto di correzione dovuto all'applicazione del modello di O.Bodnar e T.Bodnar tende a ridursi, poichè tanto è più elevata la numerosità campionaria tanto più risulterà efficiente la stima della frontiera campionaria. Al contrario, l'utilizzo del modello per l'analisi di serie storiche composte da un limitato numero di osservazioni porta ad ottenere una correzione non trascurabile.

L'ultimo punto dell'analisi delle criticità del modello prende in considerazione l'andamento della frontiera efficiente à la Bodnar e Bodnar per alcuni valori del parametro \hat{s} . Infatti, dall'analisi empirica è risultato che per alcuni valori assunti dal parametro la frontiera non distorta si comporti in modo anomalo, in quanto la varianza assume valori negativi. Dal punto di vista economico questo comportamento non ha alcuna giustificazione poiché implicherebbe un portafoglio a rischio negativo. Abbiamo visto in precedenza come il parametro \hat{s} dipenda sia dal vettore dei rendimenti attesi dei titoli, sia dalla loro matrice di covarianza.

Di seguito è proposta un'analisi sul parametro, per definirne i limiti per cui la frontiera proposta da Bodnar assume forma inaccettabile.

Si può considerare la frontiera efficiente corretta, ovvero la funzione $g_u(R, V)$, come una parabola con asse di simmetria orizzontale rivolta verso destra, e con coordinate del vertice pari a $(\frac{-\Delta}{4a}; \frac{-b}{2a})$. Di conseguenza, affinché la parte superiore della parabola, che sarà del tipo $x = ay^2 + by + c$, sia contenuta interamente nella parte positiva del piano cartesiano di riferimento dovranno valere contemporaneamente le seguente condizioni:

$$\begin{cases} a > 0 \\ \frac{-\Delta}{4a} > 0 \end{cases}$$

La prima condizione ($a > 0$) indica che la frontiera sia rivolta verso destra, mentre la seconda ($\frac{-\Delta}{4a} > 0$) implica che l'ascissa del vertice sia positiva.

Riprendendo la formula di $g_u(R, V)$

$$g_u(R, V) = (R - \hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} - (B \cdot \hat{s} - C)(V - D \cdot \hat{V}_{GMV}) = 0$$

Considerando $x = V$ e $y = R$ si arriva a

$$\begin{aligned} x &= (y - \hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} - x(B \cdot \hat{s} - C) + (B \cdot \hat{s} \cdot D \cdot \hat{V}_{GMV}) - (C \cdot D \cdot \hat{V}_{GMV}) \\ &= \frac{(y - \hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} + (B \cdot \hat{s} \cdot D \cdot \hat{V}_{GMV}) - (C \cdot D \cdot \hat{V}_{GMV})}{(B \cdot \hat{s} - C)} \\ &= \frac{y^2 + y \cdot (-2\hat{R}_{GMV}) + (\hat{R}_{GMV})^2 - A \cdot \hat{V}_{GMV} + (B \cdot \hat{s} \cdot D \cdot \hat{V}_{GMV}) - (C \cdot D \cdot \hat{V}_{GMV})}{(B \cdot \hat{s} - C)} \end{aligned}$$

A questo punto si vede come la condizione posta sul verso della parabola ($a > 0$) possa essere riscritta come $B \cdot \hat{s} - C > 0$ da cui:

$$\hat{s} > \frac{C}{B} = \frac{(k-1)(n-1)}{n(n-k-1)}$$

Quindi per valori di \hat{s} maggiori della quantità appena trovata la parabola risulterà rivolta verso destra.

Mentre, la condizione posta sull'ascissa del vertice ($X_V > -\frac{\Delta}{4a}$), può essere riscritta come:

$$X_V = -\frac{\Delta}{4a} = \frac{-A \cdot \hat{V}_{GMV} + B \cdot \hat{s} \cdot D \cdot \hat{V}_{GMV} + C \cdot D \cdot \hat{V}_{GMV}}{B \cdot \hat{s} - C}$$

Dunque la positività del vertice si ottiene per:

- $Den > 0$ se $\hat{s} > \frac{C}{B}$
- $Num > 0$ se $\hat{s} > \frac{(C \cdot D) + A}{B \cdot D}$

Sostituendo i rispettivi valori dei parametri $A, B, C,$ e D si ottiene:

- $Den > 0$ se $\hat{s} > \frac{(k-1)(n-1)}{n(n-k-1)}$
- $Num > 0$ se $\hat{s} > \frac{k(n-1)}{n(n-k-2)}$

Siccome $\frac{k(n-1)}{n(n-k-2)} > \frac{(k-1)(n-1)}{n(n-k-1)} \forall k$ con $k > 2$ e $n > k$ ⁵ allora possiamo affermare che le condizioni affinché $X_V > 0$ saranno $\hat{s} > \frac{k(n-1)}{n(n-k-2)}$ oppure $\hat{s} < \frac{(k-1)(n-1)}{n(n-k-1)}$

Confrontando le due condizioni ($a > 0$ e $\frac{-\Delta}{4a} > 0$) otterremo allora 3 scenari possibili:

- se $\hat{s} > \frac{k(n-1)}{n(n-k-2)} = s_{lim2}$ la frontiera risulterà corretta, ovvero vertice positivo e rivolta verso destra.
- se $\frac{(k-1)(n-1)}{n(n-k-1)} < \hat{s} < \frac{k(n-1)}{n(n-k-2)}$ la frontiera risulterà con il vertice negativo ma rivolta verso destra.
- se $\hat{s} < \frac{(k-1)(n-1)}{n(n-k-1)} = s_{lim1}$ la frontiera risulterà distorta poichè avrà vertice positivo ma sarà rivolta verso sinistra.

Dalle formule dei limiti si nota come questi aumentino all'aumentare del numero di titoli nel portafoglio e diminuiscano al crescere del numero di osservazioni. Di seguito sono riportate le rappresentazioni grafiche dei 3 casi, dove la curva di colore blu identifica la frontiera efficiente non distorta à la Bodnar e Bodnar e la rossa la frontiera efficiente campionaria. Queste sono state ottenute tenendo costanti i valori di $\hat{\mu}, \hat{\Sigma}$ e k , facendo variare il numero di osservazioni n . Di conseguenza il valore di $\hat{s} = 0.1233$ rimarrà costante per tutte e tre le simulazioni, e varieranno solo gli intervalli per cui la frontiera risulterà accettabile. Con s_{lim2} indicheremo il limite oltre il quale la frontiera risulterà corretta, mentre con s_{lim1} quello per cui assume vertice positivo.

⁵ $k > 2$ e $n > k$ sono due condizioni poste da T.Bodnar e O.Bodnar nelle loro dimostrazioni

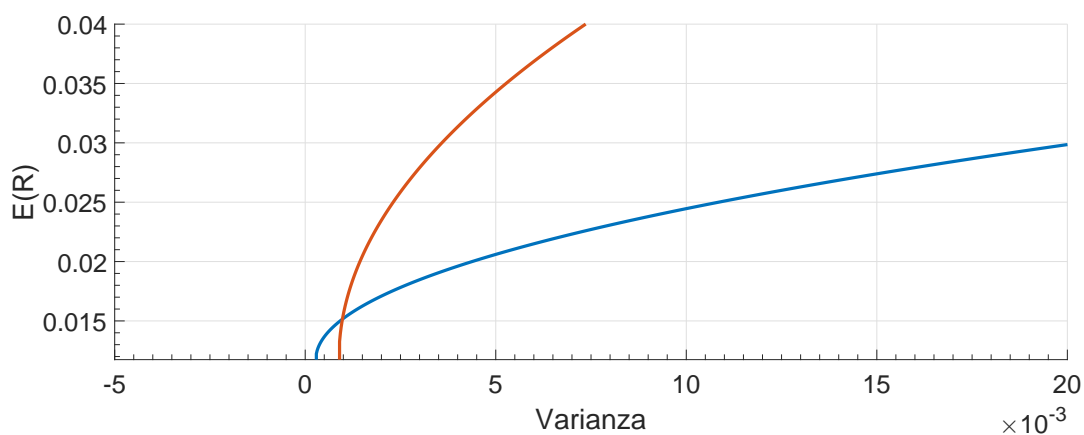


Figura 2.4: $n = 95$, $s_{lim2} = 0.1192$, $s_{lim1} = 0.1060$, Frontiera accettabile $\hat{s} > s_{lim2}$

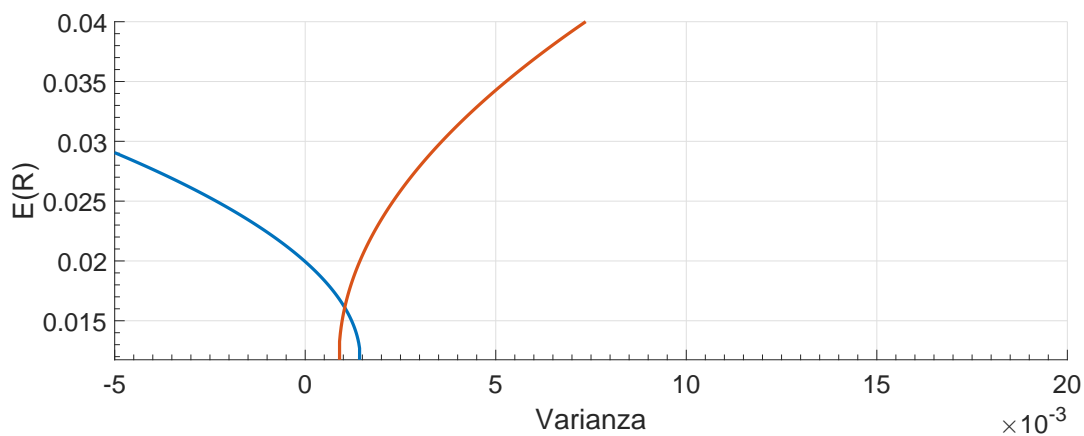


Figura 2.5: $n = 60$, $s_{lim2} = 0.2049$, $s_{lim1} = 0.1806$, Frontiera non accettabile $\hat{s} < s_{lim1}$

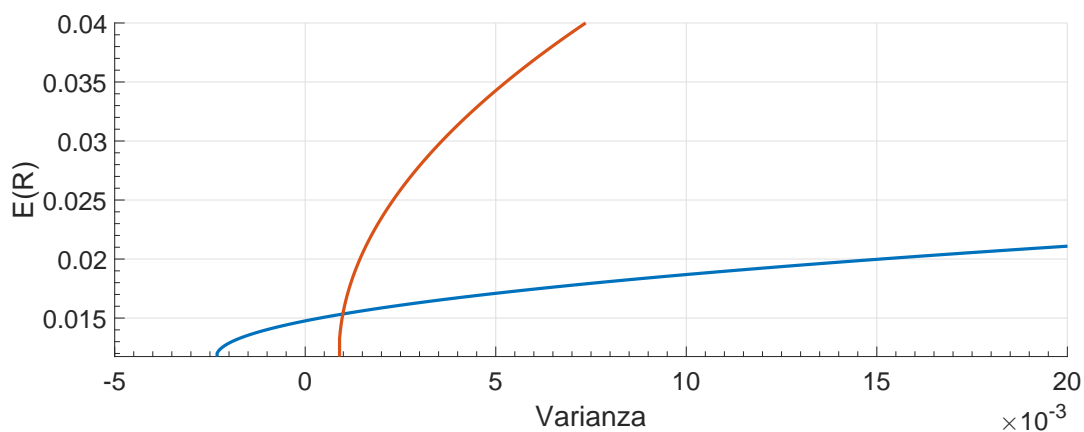


Figura 2.6: $n = 85$, $s_{lim2} = 0.1354$, $s_{lim1} = 0.1202$, Frontiera non accettabile $s_{lim1} < \hat{s} < s_{lim2}$

Capitolo 3

ANALISI DELLE COMPONENTI PRINCIPALI

3.1 Derivazione delle componenti principali

L'analisi delle componenti principali (detta anche PCA) è un metodo statistico che si basa sulla trasformazione lineare di una serie di p variabili osservate in un nuovo insieme q di variabili, dette appunto componenti principali. La trasformazione delle p variabili sarà ortogonale ed andrà a definire delle componenti principali caratterizzate dall'essere incorrelate tra loro. Il processo di estrazione prevede che la prima componente estratta sia quella con la varianza maggiore e tutte le successive siano caratterizzate da una varianza decrescente. Dunque, le componenti saranno estratte in ordine decrescente rispetto al contributo informativo dell'insieme delle p variabili osservate e la procedura di estrazione terminerà quando si arriverà ad un numero di componenti uguale al numero di variabili iniziali. L'insieme del contributo informativo delle q componenti principali sarà uguale a quello delle p variabili osservate. Lo scopo principale della PCA è quello di semplificare il processo di analisi, riducendo il numero di variabili e permettendo di controllare la perdita di informazione. Dunque possiamo riassumere le caratteristiche delle componenti principali:

- Le componenti principali risulteranno ortogonali fra di loro;

- Il contributo fornito dalle componenti principali è decrescente, ovvero la prima estratta garantirà il massimo contributo informativo rispetto alle p variabili iniziali, la seconda spiegherà il massimo dell'informazione rimanente.
- L'insieme delle componenti principali garantisce lo stesso contributo informativo delle variabili iniziali

La tecnica della componenti principali venne introdotta nel 1901 da Karl Pearson ¹, e implementata da Harold Hotelling nel 1933 ², e fra i suoi molti utilizzi troviamo l'analisi di data set e la stima di modelli produttivi. Il suo largo impiego nello studio dei data set è giustificato dal fatto che permette una riduzione dei dati da analizzare eliminando le variabili che presentano un contenuto informativo ridondante senza però perdere le informazioni del data set originale.

Per la scomposizione in componenti principali, soprattutto nel caso in cui le variabili osservate presentino unità di misura diverse, è preferibile attuare una standardizzazione. Quindi, date la matrice $X(n \times p)$ delle variabili osservate e $S(p \times p)$ la matrice delle covarianze, la prima componente sarà una combinazione lineare delle variabili x_j :

$$y_1 = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p$$

che possiamo riscrivere come:

$$y_1 = Xc_1$$

in cui:

- y_1 rappresenta il vettore di dimensioni $(n \times 1)$ dei valori della prima componente principale
- X rappresenta la matrice $(n \times p)$ delle variabili osservate
- c_1 rappresenta il vettore $(p \times 1)$ dei coefficienti che applicati alle variabili originali darà la prima componente principale.

¹PEARSON K., On Lines and Planes of Closet Fit to System of Points in Space, *Philosophical Magazine*, pp 559-572, 1901

²HOTTELING H., Analysis of a complex of statistical variables into principal component, *Journal of Educational Psychology*, pp 417-441, 1933

Poichè X , essendo la matrice dei dati originali, è nota bisognerà trovare c_1 tale che:

$$var(y_1) = max$$

sotto la condizione che:

$$c_1'c_1 = 1$$

La condizione si rende necessaria per evitare che il vettore c_1 dei coefficienti assuma valori tali per cui la varianza di y_1 diventi infinitamente grande. Dunque, bisognerà risolvere un problema di massimizzazione vincolata attraverso il metodo dei moltiplicatori di Lagrange. Dato che $var(y_1) = c_1'Sc_1$, allora:

$$\varphi = c_1'Sc_1 - \lambda_1(c_1'c_1 - 1)$$

Dove il moltiplicatore di Lagrange è definito da λ_1 . Ponendo la derivata di φ rispetto a c_1 uguale a zero avremo:

$$Sc_1 = \lambda_1c_1$$

Da quest'ultima si nota come il vettore c_1 che massimizza la varianza di y_1 è il primo autovettore di S , mentre λ_1 rappresenta la varianza della prima componente ed è data dal primo autovalore di S . Nel caso in cui le variabili utilizzate siano standardizzate la somma della loro varianza sarà uguale a p e di conseguenza il rapporto $\frac{\lambda_1}{p}$ indicherà la quota di varianza spiegata dalla prima componente.

Dopo aver calcolato la prima componente si prosegue andando ad estrarre la seconda che sarà data da:

$$y_2 = Xc_2$$

Come nel caso della prima componente si tratterà di massimizzare la varianza della componente.

$$Var(y_2) = max$$

I vincoli della seconda componente saranno:

$$c_2'c_2 = 1$$

$$cov(y_1y_2) = 0$$

Si nota come la prima condizione sia la stessa posta sulla prima componente, ed evita che la componente abbia varianza infinitamente grande, mentre il secondo vincolo impone che le due componenti siano incorrelate fra loro. Questo punto della scomposizione in componenti principali è fondamentale, poiché come già richiamato in precedenza, le componenti devono essere tutte ortogonali fra loro. Utilizzato anche per la seconda componente il metodo dei moltiplicatori di Lagrange avremo che c_2 è il secondo autovettore di S e λ_2 il corrispondente autovalore, mentre la varianza spiegata nel caso di variabili standardizzate sarà data da $\frac{\lambda_2}{p}$. Il procedimento di estrazione delle componenti continua fino a che non si arriva a definire un numero di componenti q uguale al numero delle variabili osservate. Di conseguenza alla fine del processo si otterrà una matrice Y di dimensioni $(n \times p)$ calcolata come:

$$Y = XC$$

dove:

- X rappresenta la matrice $(n \times p)$ delle variabili osservate.
- C rappresenta la matrice $(p \times p)$ degli autovettori della matrice delle covarianze di Z .

La matrice $Y(n \times p)$ sarà allora formata in modo tale che nella k -esima colonna avremo una componente principale y_k caratterizzata da una varianza pari a λ_k e poiché queste componenti sono ortogonali fra loro, la matrice delle covarianze di Y sarà una matrice diagonale così composta:

$$\begin{matrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \dots & \lambda_k \end{matrix}$$

Gli autovalori come detto in precedenza nel caso di variabili standardizzate sommano a p dunque si può scrivere che $\sum_{k=1,p} \lambda_k = p$ e di conseguenza ogni componente sarà in grado di spiegare una quota pari a $\frac{\lambda_k}{p}$ della varianza totale. Inoltre

gli autovalori verranno sempre estratti in ordine decrescente quindi $\lambda_1 \geq \lambda_2 \geq \lambda_k$. Di conseguenza possiamo sostenere sia che la prima componente sarà anche quella con il più alto contenuto informativo e l'ultima quella con il minore, sia che la totalità delle componenti riesce a spiegare in modo completo la varianza del fenomeno in esame, infatti si può affermare che $\sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \text{Var}(Y_i)$. Dunque si è visto come la tecnica delle componenti principali permetta la trasformazione di un data set composto da un grande numero di variabili interconnesse fra loro in un nuovo set di variabili incorrelate fra loro, dette componenti principali.

La PCA è stata spesso trattata come un caso particolare dell'analisi fattoriale, ma come spiegato da Joliffe (1986)³, le due tecniche si differenziano per alcune caratteristiche. Infatti, sebbene in entrambi i casi l'obiettivo sia quello di ridurre la dimensionalità delle variabili in esame, attraverso l'utilizzo della PCA si vanno a definire gli elementi diagonali, mentre nell'analisi fattoriale quelli non diagonali della matrice delle covarianze. Inoltre l'analisi delle componenti principali può essere svolta direttamente sulle variabili originali, a differenza dell'analisi fattoriale.

Una volta che si sono trasformati i dati originali in componenti principali bisognerà scegliere un numero di componenti, minore delle variabili iniziali, tali da riuscire a spiegare il fenomeno. Si possono utilizzare tre metodi differenti:

- Regola di Kaiser (o criterio degli autovalori): si sceglieranno solo le componenti che sono caratterizzate da un autovalore maggiore o uguale ad 1, o analogamente le componenti con varianza maggiore della varianza media.
- Criterio della percentuale di varianza spiegata: si considera un numero di componenti principali in grado di fornire un determinato valore di varianza spiegata. Una quota ritenuta soddisfacente è il 75%, ma spesso è accettata anche una quota inferiore.
- Criterio della rappresentazione grafica degli autovalori (o scree test): si utilizza il grafico degli autovalori detto appunto scree plot, definito dagli auto-

³JOLIFFE I., *Principal Component Analysis*, Springer, 1986

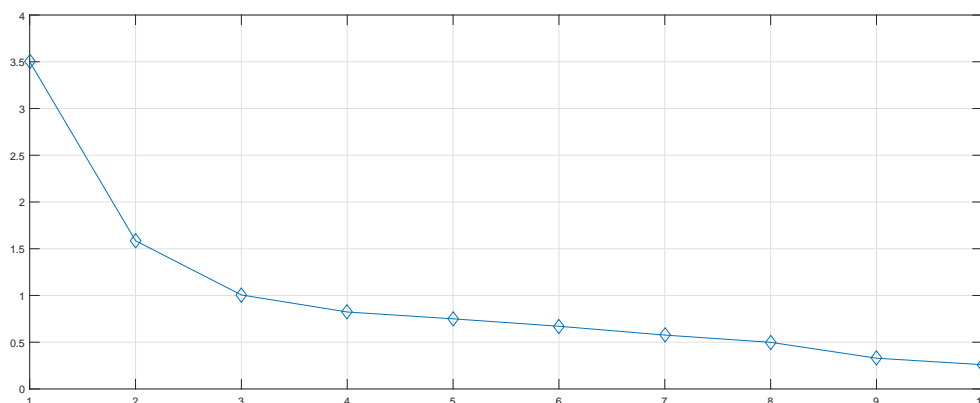


Figura 3.1: Esempio scree plot

valori sull'ordinata e dall'ordine di estrazione sulle ascisse, per determinare il numero di componenti da considerare. Si escluderanno le componenti per le quali la spezzata che unisce gli autovalori inizia ad essere quasi parallela all'asse delle ascisse. Un esempio di scree plot è fornito nel grafico (Fig.3.1).

3.2 Interpretazione grafica

Come abbiamo visto nel precedente paragrafo, ogni componente principale può essere espressa come:

$$y_1 = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p$$

Dove il coefficiente c_{ij} rappresenta il peso della variabile x_j nel determinare la componente principale y_j . Dal punto di vista geometrico l'analisi delle componenti principali può essere definita come la proiezione delle p variabili da uno spazio R^p ad un sottospazio R^k che deformi il meno possibile la distribuzione dei p punti. Il risultato di questa operazione è un nuovo spazio che avrà come assi le varie componenti principali in un ordine decrescente. Dunque il primo asse, ovvero quello della prima componente sarà anche quello con la massima variabilità, il secondo avrà variabilità inferiore ma comunque massima rispetto a quella rimasta.

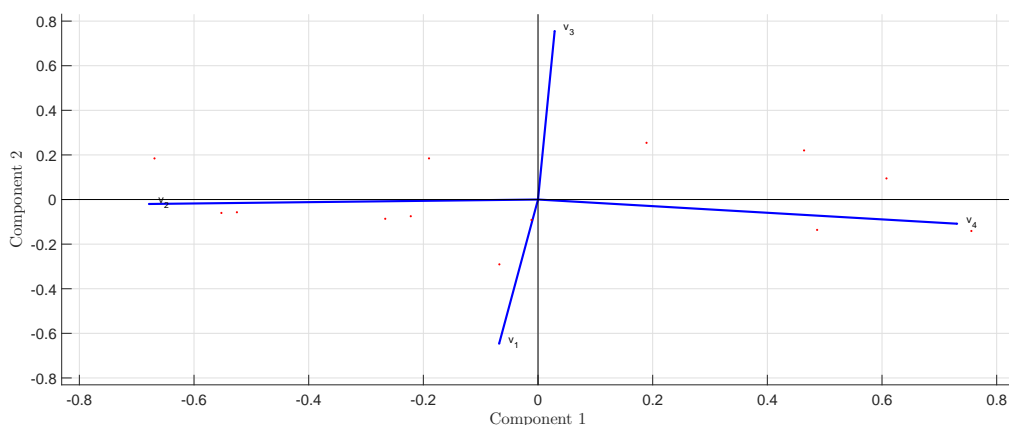


Figura 3.2: Proiezione sul piano R^2 (componente 1; componente 2)

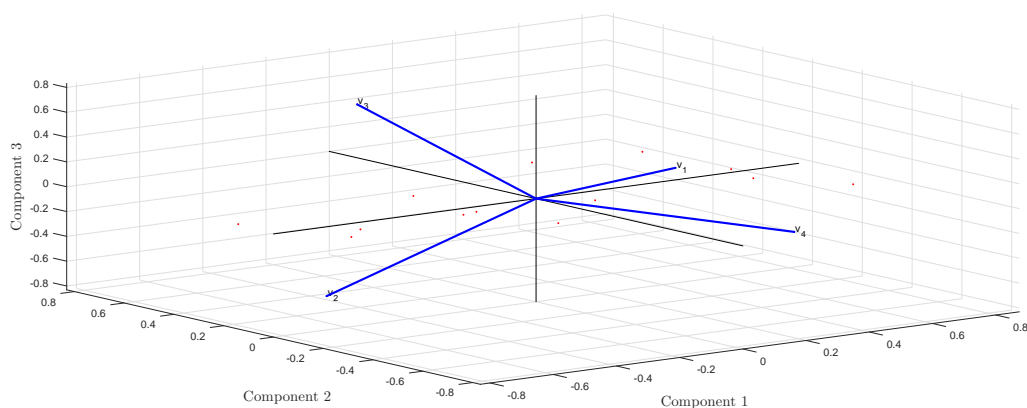


Figura 3.3: Proiezione sul piano R^3 (componente 1; componente 2; componente 3)

I grafici (Fig.3.2 e Fig.3.3) mostrano la trasformazione in componenti principali, in cui le quattro variabili della matrice dei dati originali sono rappresentate dai rispettivi vettori. La direzione e la lunghezza di ogni vettore indica quanto ogni variabile contribuisce alle componenti principali. Il primo grafico (Fig.3.2) mostra la proiezione in R^2 dei coefficienti delle prime due componenti principali, mentre il secondo (Fig.3.3) la proiezione in R^3 delle prime tre componenti principali. Dalla proiezione in R^2 si nota come la prima componente principale, che è rappresentata dall'asse orizzontale, ha un coefficiente positivo per la terza e la quarta variabile. In entrambi i grafici è inoltre possibile vedere le osservazioni e di conseguenza il loro contributo in termini di coefficienti nel determinare le componenti.

3.3 Applicazioni economiche

La tecnica delle componenti principali riesce a semplificare i problemi derivanti dall'eccessivo numero di variabili da analizzare ed è per questo che trova largo impiego in molti ambiti scientifici. Il suo utilizzo si è diffuso anche in campo economico infatti, è stata utilizzata per ricavare degli indici di mercato⁴ e per studiare il rischio sistemico del mercato⁵.

Nella nostra trattazione utilizzeremo la tecnica per sviluppare, partendo da dati reali, dei portafogli incorrelati che andranno a definire successivamente la frontiera efficiente.

Prima di fare ciò, sarà utile andare a vedere quali sono stati i maggiori contributi dell'applicazione della tecnica delle componenti principali per l'analisi di portafoglio. Partovi e Caputo (2004)⁶ sono stati i primi a considerare l'idea dell'utilizzo della PCA per la creazione di portafogli indipendenti, da loro definiti "principal portfolios". Gli autori partono dall'idea di trasformare i dati originali in un nuovo set di portafogli incorrelati tra loro, così da rendere il problema di selezione del portafoglio più semplice. Questo perchè non si dovrà più scegliere fra una serie di titoli correlati fra loro, ma fra una serie di portafogli incorrelati, garantendo una semplificazione nel processo di selezione. Il punto di partenza degli autori è la definizione della matrice di covarianza (σ) dei dati originali, che essendo simmetrica ($p \times p$) è anche ortogonalmente diagonalizzabile, e possiede un set completo di autovettori. Allora si può affermare che σ ammette p autovettori ortogonali e^μ , $\mu = 1, 2, \dots, p$, tali che $\sigma e^\mu = \lambda_\mu e^\mu$, dove $\lambda_\mu \geq 0$ sono gli autovalori della matrice di covarianza. Si può dunque riscrivere la matrice di covarianza in termini di autovettori e autovalori come $\sigma_{ij} = \sum_{\mu=1}^p \lambda_\mu e_i^\mu e_j^\mu$. I portafogli principali saranno definiti da $S_\mu = \sum_{i=1}^p e_i^\mu s_i / W_\mu$, dove $W_\mu = \sum_{i=1}^p e_i^\mu$. Dunque ogni portafoglio principale conterrà una quantità pari a e_i^μ / W_μ del titolo s_i . Va sottolineato come la quantità W_μ ovvero la quota di investimento relativa al μ -esimo autovettore, potrebbe risultare diversa da uno e addirittura negativa nel caso in cui le vendite

⁴HESTER D., Risk Aversion and Portfolio Choice, Wiley, 1967

⁵KRITZMAN M., LI Y., PAGE S. E RIGOBON R., Principal Component as a Measure of Systemic Risk, *The Journal of Portfolio Management*, 37(4), 2011

⁶PARTOVI M. E CAPUTO M., Principal Portfolios: Recasting the Efficient Frontier, *Economics Bulletin*, 7(3), pp 1-10, 2004

allo scoperto (short) dominino la costituzione del portafoglio. Possiamo definire a questo punto il rendimento atteso e la varianza dei "principal portfolios", come $R_\mu = \sum_{i=1}^p e_i^\mu r_i / W_\mu$ e $V_\mu = \lambda_\mu / W_\mu^2$.

Dall'intuizione di Partovi e Caputo si sono susseguiti diversi studi riguardanti la costruzione di "principal portfolios", fra cui quello di Meucci (2010)⁷. In quest'ultimo l'autore utilizza la PCA per derivare una distribuzione della diversificazione, uno strumento per analizzare la struttura della concentrazione di un portafoglio. La distribuzione è espressa come il rapporto tra la varianza di ogni "principal portfolio" e la varianza totale. Siccome tutti i portafogli principali sono incorrelati allora la varianza totale sarà data dalla loro somma. Il rapporto sarà di conseguenza un valore compreso tra 0 e 1 e sommerà sempre ad 1. Queste sono le stesse caratteristiche della probabilità, ovvero valori non negativi e che sommano ad 1. Secondo l'autore quindi, la massima diversificazione si ottiene quando la distribuzione di diversificazione è quasi uniforme. In altre parole, un portafoglio ben diversificato si ottiene quando si investe in modo uguale nelle componenti di rischio.

⁷MEUCCI A., *Managing Diversification, Bloomberg Education and Quantitative Research and Education Paper*, pp 74-79, 2010

Capitolo 4

ANALISI EMPIRICA

4.1 Analisi dei dati

In questa parte dell'elaborato ci si concentrerà sull'analisi empirica della frontiera efficiente campionaria e della sua stima proposta dal modello di Bodnar e Bodnar, di cui si è trattato nei precedenti capitoli dal punto di vista teorico. Inoltre, si analizzerà l'impatto dell'utilizzo della PCA sulla determinazione della frontiera efficiente corretta.

In una prima fase si andranno ad analizzare le caratteristiche dei dati che verranno utilizzati in seguito per la simulazione, effettuando una verifica di ipotesi sulla normalità dei rendimenti. Per svolgere questa operazione si farà ricorso al test di ipotesi Jarque-Bera, che si basa sulla misura di asimmetria e curtosi, in cui l'ipotesi nulla è data dalla normalità. Nel caso in cui il terzo o il quarto momento non risultino uguali ai valori teorici della distribuzione normale bisognerà rifiutare l'ipotesi nulla e quindi l'ipotesi di distribuzione normale dei rendimenti. In aggiunta si svolgerà il D'Agostino test per verificare la presenza di asimmetria e curtosi. Nella seconda fase si presenteranno le simulazioni effettuate con il software Matlab delle tre diverse frontiere efficienti, ovvero quella campionaria, quella risultante dal modello di Bodnar e Bodnar e quella ottenuta attraverso la tecnica delle componenti principali.

Gli 8 titoli, tutti quotati alla borsa di New York (NYSE), che verranno utilizzati per le successive simulazioni sono: Pfizer Inc(PFE), Coca-Cola(KO), Nike(NKE), J.P.Morgan(JPM), Goldman Sachs(GS), AT&T(T), General Electric(GE) e McDo-

nald's(MCD). I dati giornalieri e mensili sono stati ricavati dal sito web yahoo finance.

Per l'analisi si è scelto di utilizzare i rendimenti logaritmici ottenuti come:

$$r = \ln(P_t) - \ln(P_{t-1}).$$

Il rendimento logaritmico è un approssimazione lineare del rendimento percentuale e viene diffusamente utilizzato poichè gode della proprietà additiva che le altre tipologie di rendimenti non possiedono.

Per l'analisi grafica dei rendimenti si utilizzeranno due tipologie di grafici, un istogramma e uno di probabilità normale, che hanno lo scopo di confrontare la distribuzione empirica dei rendimenti con quella della normale. Il grafico di probabilità normale (normal probability plot) permette di visualizzare gli scostamenti dalla distribuzione normale, rendendo possibili identificare anche graficamente la presenza di asimmetria e curtosi.

4.1.1 Analisi della normalità

Come già trattato nel primo capitolo, l'assunzione della distribuzione normale dei rendimenti è un tema su cui si sono espressi numerosi studiosi. Tra i lavori maggiormente rilevanti in questo ambito troviamo quelli di Mandelbrot (1963)¹ e Fama (1965)², che attraverso un'analisi empirica mostrano come la distribuzione dei rendimenti non segua necessariamente la distribuzione normale. Dai loro studi emerge la tendenza dei rendimenti, soprattutto per quelli giornalieri, ad assumere una distribuzione leptocurtica, caratterizzata da una maggiore densità attorno al valore medio.

Per verificare la normalità dei rendimenti, sia giornalieri che mensili, si utilizzerà sia il Jarque-Bera test sia il D'Agostino test, con livello di significatività pari a $\alpha = 0.05$. Questi essendo entrambi test statistici per la verifica dell'ipotesi di normalità permetteranno di valutare se i dati siano distribuiti come una gaussiana, ed in particolare il D'Agostino test fornirà i p-value sia dell'asimmetria sia della

¹MANDELBROT B., The Variation of Certain Speculative Prices, *The Journal of Business*, 36(4), pp 394-419, 1963

²FAMA E., The Behavior of Stock-Market prices, *Journal of Business*, 38(1), pp 34-105, 1965

curtosi. Il terzo momento misura quanto la distribuzione sia simmetrica rispetto al rendimento medio, mentre il quarto dipende dalla probabilità con cui si possano verificare rendimenti distanti dal valore centrale. Quest'ultimo si può riscontrare graficamente dalla presenza di code pesanti.

Le formule analitiche per il calcolo dell'asimetria e della curtosi sono le seguenti:

$$S = \mathbb{E} \left[\left(\frac{R - \mu_R}{\sigma_R} \right)^3 \right] \quad (4.1)$$

$$K = \mathbb{E} \left[\left(\frac{R - \mu_R}{\sigma_R} \right)^4 \right]. \quad (4.2)$$

Nel caso in cui i rendimenti siano distribuiti come una normale si avrà $S(R) = 0$ e $K(R) = 3$.

Se i rendimenti logaritmici rispecchiano i valori della distribuzione normale allora la distribuzione Jarque Bera sarà del tipo χ^2 con due gradi di libertà. Il test sarà definito come:

$$JB = \frac{n}{6} (S^2 + \frac{1}{4}(K-3)^2)$$

dove n rappresenta il numero di osservazioni, ed S e K sono rispettivamente gli indici di asimmetria e curtosi calcolati sul campione. Nel caso di un livello di significatività del 5% l'ipotesi nulla sarà rifiutata per valori $JB > 6$. Se l'ipotesi nulla del test fosse verificata si potrà affermare che R sia distribuito come una normale. Infine, per valutare se la causa della distorsione dalla distribuzione normale sia dovuta all'asimetria o alla curtosi saranno forniti i p-value dei test sul terzo e il quarto momento ottenuti grazie al D'Agostino test.

4.1.2 Analisi per rendimenti mensili

Per valutare se la distribuzione dei rendimenti corrisponde a quella gaussiana sono stati calcolati i rendimenti mensili logaritmici per gli 8 titoli selezionati, con periodo di riferimento dal 1/01/2006 al 31/12/2017 e su questi sono stati effettuati

Tabella 4.1: Jarque Bera test e D'Agostino test per rendimenti mensili

Test normalità per rendimenti mensili				
Titolo	$p - value_{JB}$	JB	p-value simmetria	p-value curtosi
T	0.002	22.751	0.001	0.043
PFE	0.014	10.422	0.022	0.043
NKE	<0.001	3 130	0.243	<0.001
MCD	0.500	1.113	0.426	0.335
KO	0.002	25.002	0.024	0.002
JPM	0.004	18.513	0.003	0.014
GS	0.022	8.432	0.041	0.055
GE	<0.001	58.071	0.001	<0.001

il JB test e il D'Agostino test.

I risultati dei test e i rispettivi p-value sono mostrati nella tabella seguente (Tab.4.1)

Da questa si vede come l'unico titolo, tra quelli selezionati, che presenta un p-value e un valore di JB per cui non si rifiuta l'ipotesi di normalità è quello di McDonald's. Dunque tutti gli altri 7 titoli non presentano una distribuzione che si può definire normale.

Anche con il metodo grafico (Fig. da 4.1 a 4.4) si nota come le distribuzioni dei rendimenti mensili, per la quasi totalità dei titoli, si discostino in modo notevole dalla distribuzione normale. Inoltre i p-value ottenuti dal D'Agostino test ci indicano che solo i titoli MCD e GS non sono affetti da curtosi mentre solo MCD e NKE non presentano asimmetria con un livello di significatività pari a $\alpha = 0.05$. Dunque l'analisi porta a considerare la tendenza da parte della maggior parte dei titoli ad assumere una distribuzione leptocurtica, ovvero code più pesanti della normale e una frequenza maggiore attorno alla media.

Relativamente alla matrice di correlazione (Fig.4.5), si vede come la coppia che possiede il maggior coefficiente di correlazione è quella formata da Goldman Sachs e J.P.Morgan ($\rho = 0.6545$), situazione che è giustificata dal fatto che entrambe le società appartengono al settore bancario. Al contrario la coppia con il minor coefficiente di correlazione è quella composta da J.P.Morgan e AT&T ($\rho = 0.1360$).

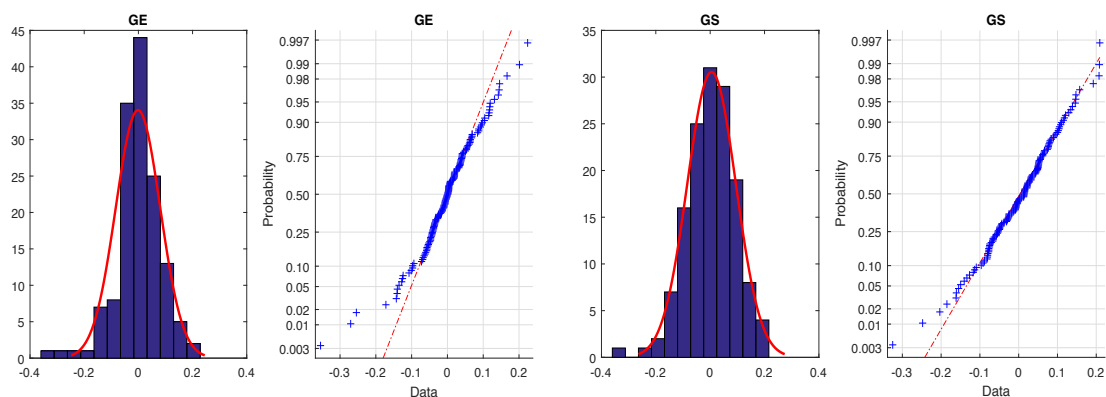


Figura 4.1: Istogramma e normal probability plot per i rendimenti mensili di GE e GS

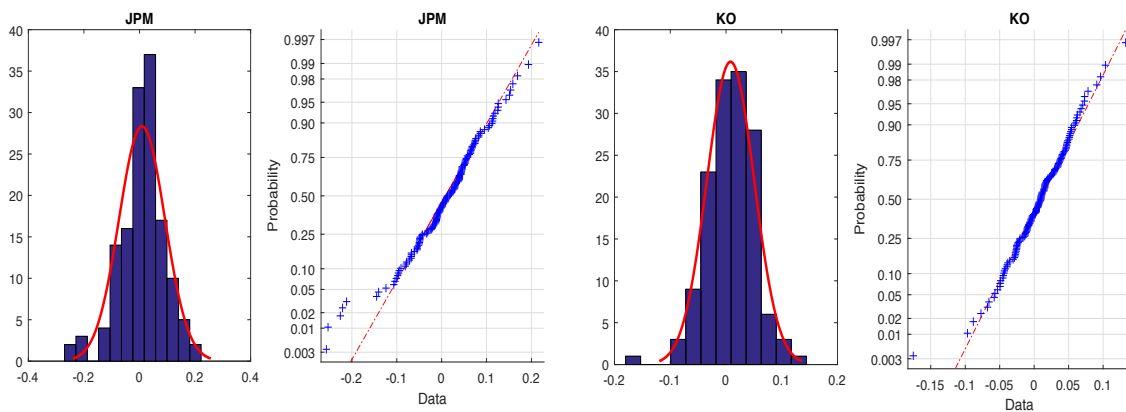


Figura 4.2: Istogramma e normal probability plot per i rendimenti mensili di JPM e KO

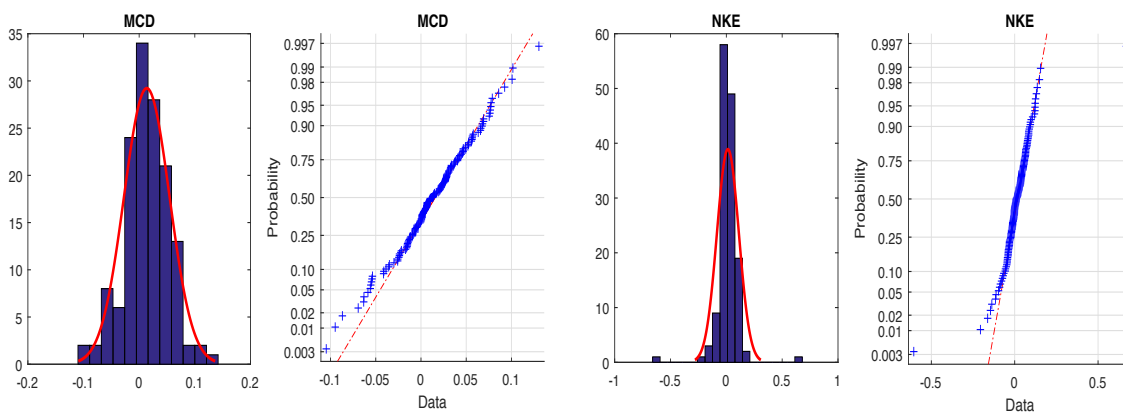


Figura 4.3: Istogramma e normal probability plot per i rendimenti mensili di MCD e NKE

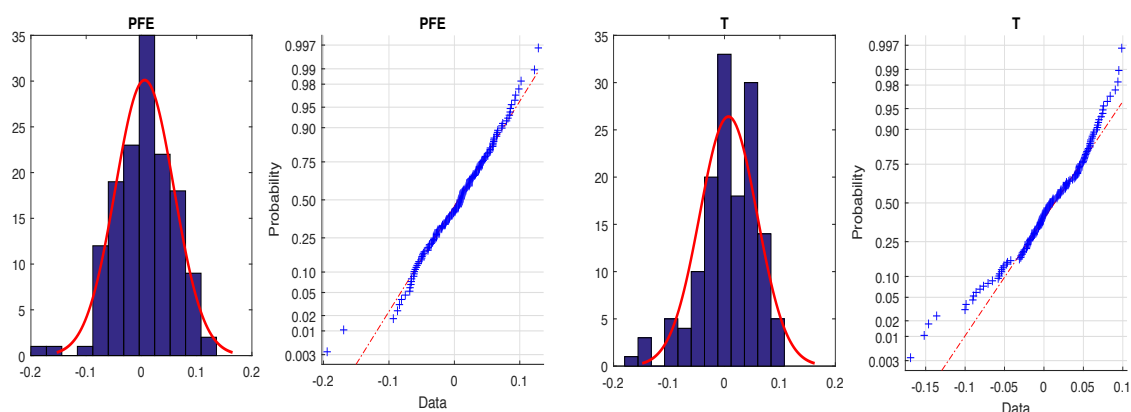


Figura 4.4: Istogramma e normal probability plot per i rendimenti mensili di PFE e T

	T	PFE	NKE	MCD	KO	JPM	GS	GE
T	1.0000	0.3529	0.1502	0.3311	0.4166	0.1360	0.1873	0.4110
PFE	0.3529	1.0000	0.1375	0.4404	0.3828	0.5186	0.2759	0.5192
NKE	0.1502	0.1375	1.0000	0.1885	0.1790	0.2553	0.2275	0.2741
MCD	0.3311	0.4404	0.1885	1.0000	0.5751	0.3012	0.2783	0.3665
KO	0.4166	0.3828	0.1790	0.5751	1.0000	0.2905	0.2960	0.5112
JPM	0.1360	0.5186	0.2553	0.3012	0.2905	1.0000	0.6545	0.5791
GS	0.1873	0.2759	0.2275	0.2783	0.2960	0.6545	1.0000	0.4420
GE	0.4110	0.5192	0.2741	0.3665	0.5112	0.5791	0.4420	1.0000

Figura 4.5: Matrice di correlazione per rendimenti mensili dal 1/1/2006 al 31/12/2017

4.1.3 Analisi per rendimenti giornalieri

Lo stesso procedimento è stato applicato ai dati giornalieri degli otto titoli selezionati per lo stesso arco temporale e i risultati sono mostrati in tabella (Tab.4.2).

Per tutti i titoli i valori del p-value del Jarque Bera test sono molto inferiori al livello di significatività, quindi per rendimenti giornalieri nessun titolo avrà una distribuzione normale. Considerando invece i risultati ottenuti dal test di D'Agostino si nota come i p-value siano molto piccoli, indicando che tutte le distribuzioni siano affette da asimmetria e curtosi; l'unica eccezione è data dal valore del p-value calcolato sull'asimmetria del titolo GE. Inoltre, confrontando i valori risultanti dai test d'ipotesi su dati giornalieri con quelli ottenuti sui rendimenti mensili si evidenzia un considerevole allontanamento dalla distribuzione normale, risultati che sono in linea con gli studi empirici di Fama e Mandelbrot.

Tabella 4.2: Jarque Bera test e D'Agostino test per rendimenti giornalieri

Test normalità per rendimenti giornalieri				
Titolo	$p - value_{JB}$	JB	p-value simmetria	p-value curtosi
T	<0.001	17 326	<0.001	<0.001
PFE	<0.001	70 008	0.001	<0.001
NKE	<0.001	8 956	<0.001	<0.001
MCD	<0.001	4 482	0.002	<0.001
KO	<0.001	22 560	<0.001	<0.001
JPM	<0.001	33 010	<0.001	<0.001
GS	<0.001	31 980	<0.001	<0.001
GE	<0.001	11 817	0.801	<0.001

Dai grafici che seguono (Fig. da 4.6 a 4.9) risulta una distribuzione leptocurtica per tutte le serie di rendimenti caratterizzati da una forte densità intorno al valore medio.

La matrice di correlazione (Fig.4.10) mostra come anche nel caso di osservazioni giornaliere la coppia che presenta maggior correlazione è quella composta da J.P.Morgan e Goldman Sachs ($\rho = 0.7471$), mentre quella con il coefficiente minore è McDonald's e Goldman Sachs ($\rho = 0.3785$).

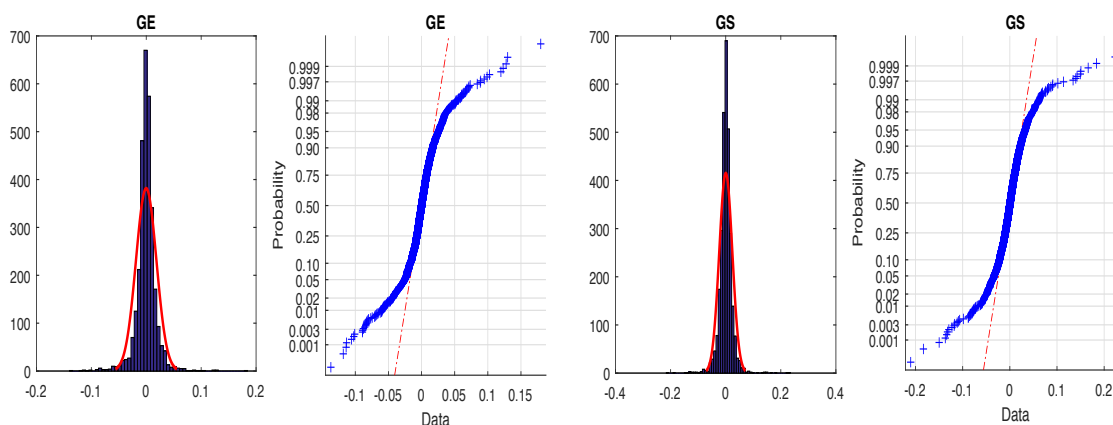


Figura 4.6: Istogramma e normal probability plot per i rendimenti giornalieri di GE e GS

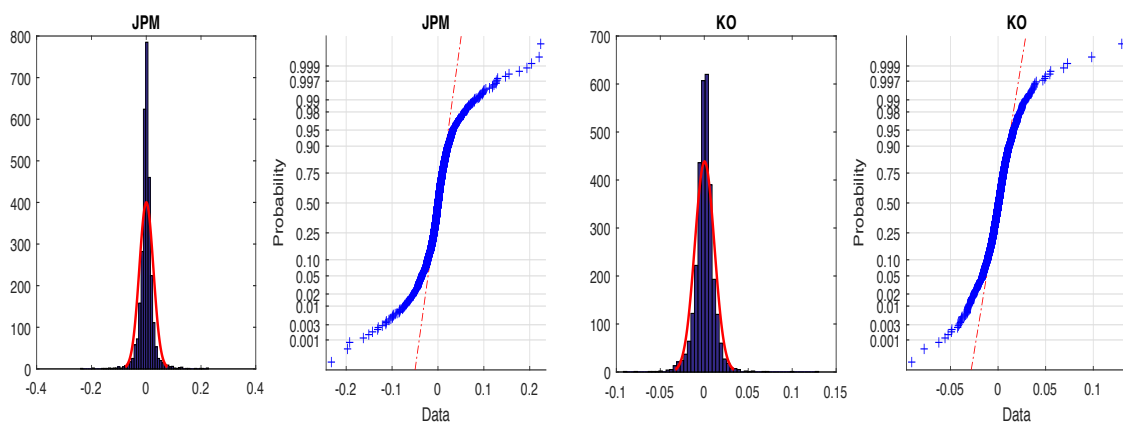


Figura 4.7: Istogramma e normal probability plot per i rendimenti giornalieri di JPM e KO

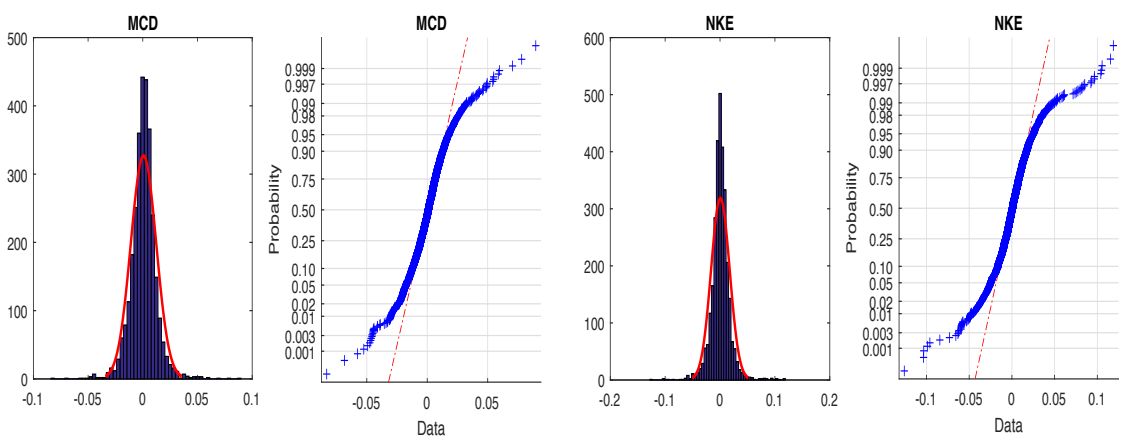


Figura 4.8: Istogramma e normal probability plot per i rendimenti giornalieri di MCD e NKE

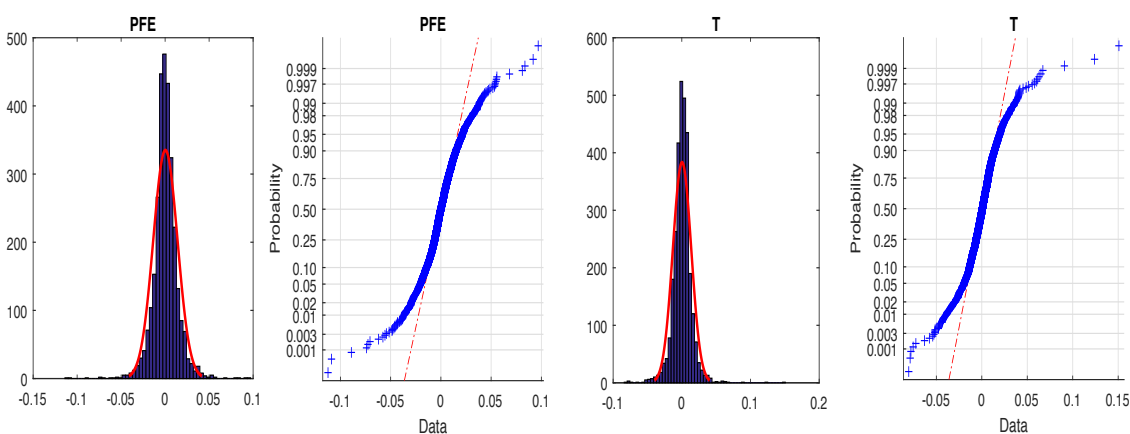


Figura 4.9: Istogramma e normal probability plot per i rendimenti giornalieri di PFE e T

	T	PFE	NKE	MCD	KO	JPM	GS	GE
T	1.0000	0.4898	0.4416	0.4197	0.4952	0.5014	0.5024	0.4913
PFE	0.4898	1.0000	0.3984	0.4224	0.4666	0.4781	0.4650	0.4981
NKE	0.4416	0.3984	1.0000	0.4616	0.4119	0.4752	0.4396	0.4738
MCD	0.4197	0.4224	0.4616	1.0000	0.4724	0.4182	0.3785	0.4278
KO	0.4952	0.4666	0.4119	0.4724	1.0000	0.3820	0.3800	0.4115
JPM	0.5014	0.4781	0.4752	0.4182	0.3820	1.0000	0.7471	0.6412
GS	0.5024	0.4650	0.4396	0.3785	0.3800	0.7471	1.0000	0.5355
GE	0.4913	0.4981	0.4738	0.4278	0.4115	0.6412	0.5355	1.0000

Figura 4.10: Matrice di correlazione per rendimenti giornalieri dal 1/1/2006 al 31/12/2017

L'analisi fin qui svolta sulla normalità dei rendimenti sia mensili che giornalieri, si allinea con i diversi studi che riportano una tendenza delle serie storiche dei rendimenti ad essere caratterizzate da code più pesanti rispetto a quelle della normale e ad una maggiore densità intorno allo zero. Questo significa che le probabilità che si verifichino rendimenti molto piccoli o molto grandi sono maggiori rispetto a quelle che si hanno con una distribuzione normale. Dalle matrici di correlazione emerge come tutte le coppie di titoli abbiano un coefficiente di correlazione positivo e che i titoli maggiormente correlati siano J.P.Morgan e Goldman Sachs.

4.2 Confronto fra frontiera efficiente Campionaria e frontiera efficiente à la Bodnar

In questa fase si svilupperà un'analisi sulle ipotesi di base del modello di Bodnar e Bodnar e un confronto fra la frontiera efficiente campionaria e la sua stima corretta, per evidenziarne le differenze.

4.2.1 Analisi sulle ipotesi di base di Bodnar e Bodnar

Prima di procedere con il confronto fra la frontiera efficiente campionaria e quella proposta da Bodnar e Bodnar, si andrà ad analizzare se gli assunti di base del

modello sui rendimenti siano verificati. Gli autori assumono che i rendimenti siano indipendenti e normalmente distribuiti con $k > 2$ e $n > k$, dove k ed n rappresentano rispettivamente il numero di titoli e le osservazioni. L'analisi sulla distribuzione normale dei rendimenti è stata svolta nel paragrafo precedente, da cui è risultato che solo il titolo McDonald's presenti per rendimenti mensili una distribuzione normale. Le ipotesi sui valori di k ed n sono verificate in quanto i titoli selezionati per l'analisi sono 8, e le osservazioni mensili dal 1/1/2006 al 31/12/2017 sono 143 e quelle giornaliere 3019.

Per quanto concerne l'autocorrelazione dei rendimenti, è stato dimostrato come la sua eventuale presenza non vada ad influenzare l'effettiva correzione della frontiera efficiente campionaria³. Va sottolineato comunque che se nel caso di rendimenti mensili sia accettabile l'ipotesi di mancanza di autocorrelazione questo non è sempre vero per osservazioni giornaliere in cui l'autocorrelazione può assumere valori non trascurabili.

Dall'analisi sui rendimenti sia giornalieri che mensili è emerso come gli assunti di base del modello di Bodnar e Bodnar non siano rispettati per i titoli selezionati, infatti il test sulla normalità ha evidenziato risultati non in linea con il modello. E' necessario ricordare che l'ipotesi sulla distribuzione gaussiana dei rendimenti abbia un ruolo fondamentale nella derivazione degli stimatori corretti per la frontiera campionaria di conseguenza un'assenza di questa caratteristica implica l'invalidazione del modello. Dunque si può affermare, dati i risultati empirici, che i titoli non godano di tutte le caratteristiche necessarie per soddisfare le ipotesi di base della teoria di Bodnar e Bodnar.

4.2.2 Confronto grafico dei modelli

Lo scopo del modello presentato da Bodnar e Bodnar è quello di fornire uno stimatore corretto della frontiera efficiente che corregga l'ottimismo della frontiera efficiente campionaria. Nel paragrafo precedente abbiamo verificato gli assunti di base del modello e nel secondo capitolo dell'elaborato è stato spiegato dal punto

³BRESOLIN E., Sulla distorsione della frontiera efficiente dei portafogli: analisi delle ipotesi di normalità ed indipendenza, *Università Ca' Foscari Venezia*, 2016-2017

di vista teorico il modello. In questa fase si confronteranno i risultati ottenuti dall'analisi delle due diverse tipologie di frontiere e si andrà a vedere se questi sono in linea con quelli ottenuti da Bodnar e Bodnar.

I grafici (Fig.4.11) mostrano le due frontiere ottenute con gli 8 titoli selezionati sia per rendimenti logaritmici mensili sia per quelli giornalieri. Per entrambe le figure la linea di colore blu corrisponde alla frontiera efficiente campionaria mentre quella rossa alla frontiera efficiente corretta. I risultati grafici sono in linea con quelli proposti dagli autori, infatti per valori elevati della varianza l'ottimismo della frontiera efficiente campionaria viene corretto dallo stimatore à la Bodnar.

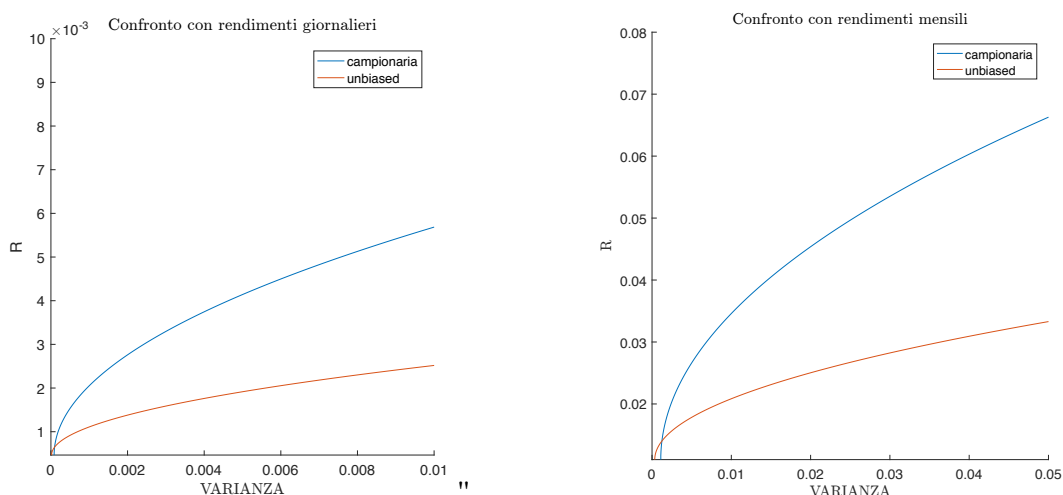


Figura 4.11: Confronto frontiere efficienti campionaria e corretta per rendimenti giornalieri e mensili

La frontiera per i rendimenti logaritmici giornalieri è ottenuta con $n = 3019$ e $k = 8$, da cui si è calcolato il portafoglio di minima varianza che è risultato pari a $\hat{R}_{GMV} = 4.6564e - 04$ e $\hat{V}_{GMV} = 8.3781e - 05$, mentre $\hat{s} = 0.0027$ è il valore assunto dal parametro che denota la pendenza della frontiera. E' stato inoltre calcolato il valore di γ , che seguendo l'analisi di Bodnar e Bodnar rappresenta il parametro per il quale la frontiera efficiente campionaria sovrastima la frontiera calcolata sulla popolazione. Il valore di γ è ottenuto nel seguente modo:

$$\gamma = \frac{2ns + (n - k)(k - 1) + n - 2}{kns + (n - k)(k - 1)} \quad (4.3)$$

Per i rendimenti giornalieri si ottiene $\gamma = 1.1403$ che indica che per $V < \gamma \hat{V}_{GMV} = 9.553e - 05$ si ha una sottostima della frontiera efficiente, mentre per $V > 9.553e - 05$

la frontiera efficiente campionaria sovrastima quella della popolazione. Come si può vedere dal confronto delle due frontiere, quella calcolata con il modello presentato da Bodnar e Bodnar corregge l'ottimismo di quella campionaria.

Lo stesso procedimento è stato svolto sui rendimenti logaritmici mensili, con $n = 143$ e $k = 8$, $\hat{R}_{GMV} = 0.011$, $\hat{V}_{GMV} = 0.0011$ e $\hat{s} = 0.0629$, da cui è risultato $\gamma = 1.0856$. Dunque per $V < 0.001185$ si avrà una sottostima della frontiera calcolata sulla popolazione, mentre per $V > 0.001185$ si avrà una sovrastima. Inoltre si può notare che per entrambe le simulazioni non si verifica una distorsione della frontiera efficiente, in quanto in entrambi i casi i valori \hat{s} sono maggiori dei valori che causano la distorsione.

Dunque nonostante non siano verificate le ipotesi di normalità e indipendenza dei rendimenti le simulazioni danno risultati in linea con la tesi di Bodnar e Bodnar. Infatti, dal punto di vista grafico si nota l'effetto di correzione sulla frontiera efficiente campionaria, che permette di correggerne l'ottimismo riscontrato da Siegel e Woodgate ⁴. Rimangono comunque numerosi dubbi sul modello poichè l'assunzione di normalità ed indipendenza dei rendimenti risultano, dal punto di vista matematico, vincolanti per la sua veridicità. Inoltre, come si analizzerà in seguito, il parametro \hat{s} può assumere valori che rendono il modello inutilizzabile, poichè restituisce una varianza negativa per lo stimatore della frontiera efficiente.

4.2.3 Analisi empirica sul parametro \hat{s}

Si è trattato nel secondo capitolo dei limiti sul parametro \hat{s} per il quale la frontiera efficiente à la Bodnar e Bodnar assume valori negativi per la varianza. Si è dimostrato che:

- se $\hat{s} > \frac{k(n-1)}{n(n-k-2)} = s_{lim2}$ la frontiera risulterà corretta, ovvero vertice positivo e rivolta verso destra.
- se $\frac{(k-1)(n-1)}{n(n-k-1)} < \hat{s} < \frac{k(n-1)}{n(n-k-2)}$ la frontiera risulterà con il vertice negativo ma rivolta verso destra.

⁴SIEGEL A. E WOODGATE A., Performance of Portfolios Optimized with Estimation Error, *Management Science*, 53(6), pp. 1005-1015, 2007

- se $\hat{s} < \frac{(k-1)(n-1)}{n(n-k-1)} = s_{lim1}$ la frontiera risulterà distorta poichè avrà vertice positivo ma sarà rivolta verso sinistra.

In questo paragrafo si analizzerà il comportamento della frontiera efficiente corretta al variare del numero di titoli e del numero di osservazioni sia giornaliere che mensili allo scopo di mostrare gli effetti che queste variazioni causano sul parametro \hat{s} e sulla frontiera. Nelle simulazioni effettuate sugli 8 titoli selezionati sia nel caso di rendimenti giornalieri sia per quelli mensili il valore di \hat{s} è risultato sufficientemente grande da non distorcere la frontiera. Infatti, nel caso dei rendimenti mensili si ha $\hat{s} = 0,0629$ con valore limite di $s_{lim2} = 0.0597$, mentre per rendimenti giornalieri $\hat{s} = 0,002748$ e valore limite di $s_{lim2} = 0,002657$. Nella prossima simulazione agli 8 titoli iniziali è stato aggiunto un ulteriore titolo, ovvero Ford (F). Le serie dei rendimenti del nuovo titolo corrispondono esattamente a quelle dei titoli iniziali.

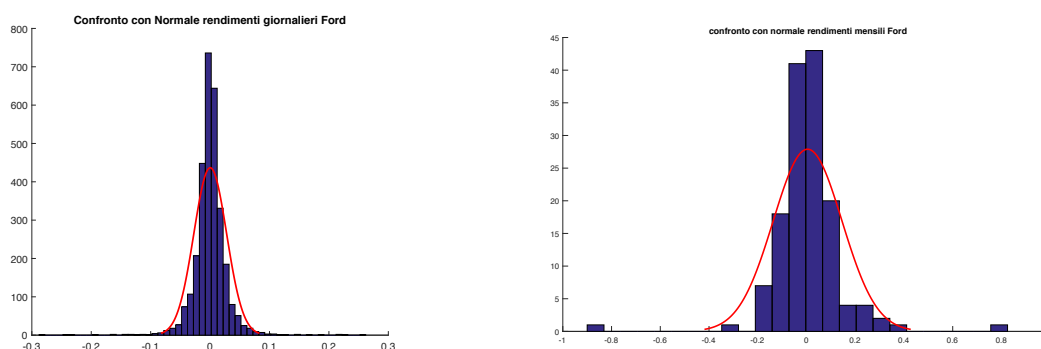


Figura 4.12: Confronto rendimenti giornalieri e mensili del titolo Ford con la normale

Il titolo Ford non presenta una distribuzione normale né per i rendimenti logaritmici mensili né per quelli giornalieri, infatti il JB test ha dato come risultati per dati giornalieri un $p\text{-value} < 1.000e-03$ e un valore del JB test pari a $JBstat = 4.426e+04$. Per i rendimenti mensili invece i risultati ottenuti sono: $p\text{-value} < 1,000e-03$ e $JBstat = 1.6073e+03$. Dai grafici (Fig.4.12) sulla distribuzione del titolo Ford si nota come questo sia contraddistinto da una distribuzione leptocurtica, soprattutto nel caso dei rendimenti giornalieri. Questo è in linea con l'analisi sulla normalità dei rendimenti svolta sugli 8 titoli iniziali.

Analizzata la distribuzione del titolo Ford, resta da vedere l'impatto di questo

sulla frontiera efficiente corretta e sul parametro \hat{s} .

Per i rendimenti giornalieri si riscontra $\hat{s} = 0.0027488$ valore per cui la frontiera efficiente corretta assume verso accettabile, ma vertice negativo. Infatti, il valore di \hat{s} risulta compreso fra i due limiti calcolati in precedenza, che in questo caso valgono rispettivamente $s_{lim1} = 0.0026657$ e $s_{lim2} = 0.0029910$. Questa degenerazione, come già ricordato, non ha alcun senso economico, in quanto valori della varianza minori di zero implicherebbero portafogli a rischio negativo.

Allo stesso modo per i rendimenti mensili si ottiene $\hat{s} = 0.0628$ dove $s_{lim1} = 0.0597$ e $s_{lim2} = 0.0677$, di conseguenza la degenerazione su entrambe le frontiere sarà dello stesso tipo; come si può vedere dai grafici (Fig.4.13).

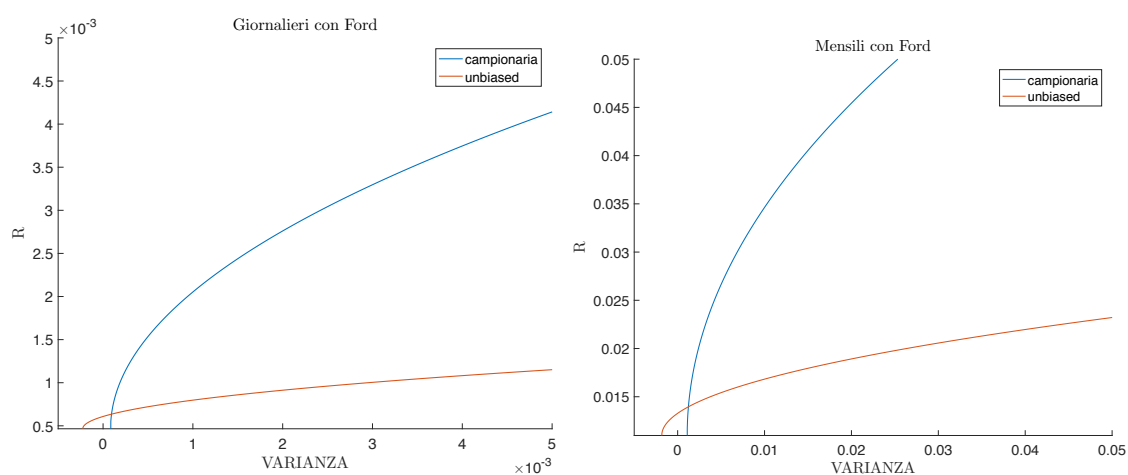


Figura 4.13: Frontiera efficiente campionaria e corretta per rendimenti giornalieri e mensili con il titolo Ford

L'aggiunta di un titolo nel portafoglio ha causato la degenerazione della frontiera, ma è interessante notare come questa aggiunta abbia causato in entrambi i casi un effetto quasi nullo sul valore di \hat{s} , mentre un aumento più significativo sui valori limiti.

Per continuare l'analisi si è scelto di suddividere entrambe le serie dei rendimenti iniziali in due sottoperiodi di uguale durata, e investigare quale sia l'effetto di una riduzione nel numero delle osservazioni. Si avranno allora due simulazioni per dati giornalieri rispettivamente una dal 1/03/2006 al 27/12/2011 e l'altra dal 28/12/2011 al 28/12/2017, ognuna delle quali con $n = 1508$. Lo stesso procedimento sarà eseguito sui rendimenti mensili e quindi si avranno due serie da 71 osservazioni. I limiti sul parametro \hat{s} per i quali la frontiera assume valori non accettabili

per le due simulazioni giornaliere saranno identici in quanto i valori limite dipendono unicamente da n e k che in questo caso valgono $n = 1508$ e $k = 8$. Allo stesso modo per le due simulazioni mensili i limiti saranno uguali poichè per entrambe si avrà $n = 71$ e $k = 8$.

Analizzando i risultati per i rendimenti giornalieri delle due simulazioni si nota che in entrambe si verifica la presenza di varianza negativa, ed essendo $s_{lim2} = 0.0053$ il limite al di sotto del quale la varianza assume comportamento non accettabile i risultati sono $\hat{s}_1 = 0.0040$ per la prima simulazione e $\hat{s}_2 = 0.0045$ per la seconda. Inoltre in tutt'e due si verifica un valore del parametro inferiore ad $s_{lim1} = 0.0046$, valore per il quale ci si aspetta che la frontiera efficiente assuma un verso opposto a quello corretto. I risultati delle due simulazioni giornaliere sono rappresentate nei grafici (Fig.4.14).

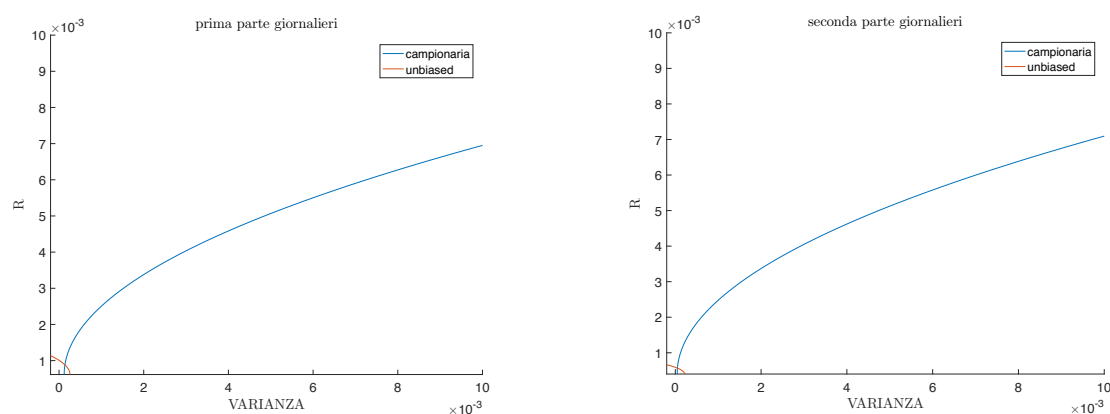


Figura 4.14: Confronto frontiere per rendimenti giornalieri

I risultati ottenuti per i rendimenti mensili mostrano una sostanziale differenza (Fig.4.15), in quanto la simulazione ottenuta sulla prima metà delle osservazioni produce una frontiera efficiente accettabile, mentre quella sulla seconda metà una frontiera non accettabile. I valori assunti dal parametro sono $\hat{s} = 0,1467$ per la prima metà e $\hat{s} = 0.0672$ per la seconda con i limiti $s_{lim1} = 0.1113$ e $s_{lim2} = 0.1293$. Dalla rappresentazione grafica della simulazione effettuata sulla prima metà dei dati si nota l'effetto correttivo dell'applicazione del modello alla Bodnar e Bodnar. Il valore del parametro \hat{s} , a differenza dei limiti, dipende sia dal valore della media $\hat{\mu}$ sia da quello della matrice di covarianza $\hat{\Sigma}$, dunque sarà utile confrontare i valori di $\hat{\mu}$ e $\hat{\Sigma}$ per le due diverse simulazioni.

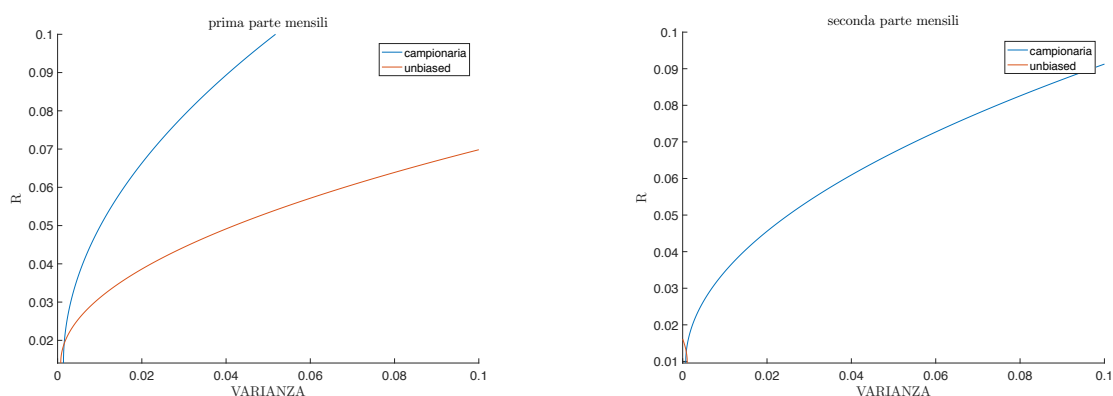


Figura 4.15: Confronto frontiere per rendimenti mensili

Si sono calcolate le medie dei rendimenti mensili per la prima e per la seconda parte, e i risultati sono mostrati in tabella (Tab.4.3).

Tabella 4.3: Confronto dei rendimenti mensili per prima e seconda metà

Confronto delle medie		
Titolo	1'parte	2'parte
T	0.0066	0.0083
PFE	0.0015	0.0104
NKE	0.0133	0.0135
MCD	0.0172	0.0104
KO	0.0099	0.0069
JPM	-0.0005	0.0171
GS	-0.0055	0.0128
GE	-0.0055	0.0018

Da questi sembra chiaro che nel secondo periodo i rendimenti mensili dei titoli siano nella maggior parte dei casi aumentati, infatti gli unici che mostrano una diminuzione dei rendimenti sono MacDonald's e Coca-Cola. I miglioramenti più significativi si hanno per i titoli J.P.Morgan, Goldman Sachs e General Electric che passano da avere un rendimento negativo ad uno positivo. Le matrici di covarianza per le due serie sono riportate di seguito (Fig.4.16 e 4.17), da queste si evidenzia una diminuzione sostanziale per i valori delle covarianze tra quelle calcolate sulla prima metà dei dati e quelle calcolate sulla seconda.

	T	PFE	NKE	MCD	KO	JPM	GS	GE
T	0.0033977	0.00144719	0.00144752	0.00116572	0.00129721	0.0012138	0.00194446	0.00249056
PFE	0.00144719	0.00395309	0.00142927	0.00144919	0.0012836	0.0034175	0.00168212	0.0038257
NKE	0.00144752	0.00142927	0.00471066	0.00142273	0.0018175	0.00369756	0.00310167	0.00490181
MCD	0.00116572	0.00144919	0.00142273	0.00200618	0.00128799	0.0015193	0.00165414	0.00205207
KO	0.00129721	0.0012836	0.0018175	0.00128799	0.00223705	0.00187679	0.00230227	0.00283074
JPM	0.0012138	0.0034175	0.00369756	0.0015193	0.00187679	0.00918802	0.0055751	0.00652567
GS	0.00194446	0.00168212	0.00310167	0.00165414	0.00230227	0.0055751	0.01114294	0.00497165
GE	0.00249056	0.0038257	0.00490181	0.00205207	0.00283074	0.00652567	0.00497165	0.01083296

Figura 4.16: Matrice di covarianza per rendimenti mensili dal 01/03/2006 al 27/12/2011

	T	PFE	NKE	MCD	KO	JPM	GS	GE
T	2.04E-03	5.05E-04	1.38E-04	0.00025429	0.0005488	5.01E-06	-9.08E-05	0.00107945
PFE	5.05E-04	1.71E-03	4.02E-05	0.00051807	0.00047289	1.13E-03	9.47E-04	0.0007411
NKE	1.38E-04	4.02E-05	1.48E-02	0.00014285	-0.0002568	4.00E-04	7.69E-04	-0.0004687
MCD	2.54E-04	5.18E-04	1.43E-04	0.00136905	0.00072152	6.49E-04	5.63E-04	0.00049645
KO	5.49E-04	4.73E-04	-2.57E-04	0.00072152	0.0014086	2.87E-04	1.45E-04	0.00083913
JPM	5.01E-06	1.13E-03	4.00E-04	0.00064925	0.00028677	4.31E-03	3.77E-03	0.00128466
GS	-9.08E-05	9.47E-04	7.69E-04	0.00056292	0.00014536	3.77E-03	4.40E-03	0.00140084
GE	1.08E-03	7.41E-04	-4.69E-04	0.00049645	0.00083913	1.28E-03	1.40E-03	0.00277015

Figura 4.17: Matrice di covarianza per rendimenti mensili dal 28/12/2011 al 28/12/2017

Dall'analisi svolta si evidenziano delle differenze nei valori di $\hat{\mu}$ e $\hat{\Sigma}$ tra le diverse simulazioni, ma più che i cambiamenti nell'uno o nell'altro parametro sembra essere il loro legame a determinare la distorsione della frontiera. Infatti, il valore di \hat{s} si ottiene attraverso il prodotto matriciale di elementi che dipendono sia dalla media campionaria sia dalla matrice di covarianza campionaria.

Dunque il parametro \hat{s} risulta molto sensibile alle variazioni sia in $\hat{\mu}$ sia in $\hat{\Sigma}$ e questo comporta un grosso limite per l'applicazione del modello proposto da Bodnar e Bodnar.

Per dimostrare l'effetto delle variazioni nei parametri $\hat{\mu}$ e $\hat{\Sigma}$, si è deciso di svolgere due simulazioni per il parametro \hat{s} . Siano $\hat{\mu}_1$ e \hat{R}_1 i parametri per il calcolo di \hat{s} per la prima metà dei rendimenti mensili logaritmici e $\hat{\mu}_2$ e \hat{R}_2 quelli per la seconda metà. Si ricorda che il valore di \hat{s} è dato da:

$$\hat{s} = \hat{\mu}' \hat{R} \hat{\mu}$$

dove

$$\hat{R} = \hat{\Sigma}^{-1} - \frac{\hat{\Sigma}^{-1} \mathbf{1} \mathbf{1}' \hat{\Sigma}^{-1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}}$$

Si è allora calcolato $\hat{s} = \hat{\mu}'_1 \hat{R}_2 \hat{\mu}_1 = 0.2307$ da cui emerge che applicando il vettore dei rendimenti attesi della prima metà dei dati nel calcolo di \hat{s} per la seconda metà, si verifica l'inversione della frontiera efficiente alla Bodnar da distorta ad accettabile. Al contrario calcolando $\hat{s} = \hat{\mu}'_2 \hat{R}_1 \hat{\mu}_2 = 0.0538$ si nota come il valore di \hat{s} per la prima metà dei dati a cui è stato applicato il vettore dei rendimenti attesi della seconda metà causa il passaggio della frontiera da accettabile a distorta. Da questi risultati sembra che per aumenti degli elementi di $\hat{\mu}$ si verifichino incrementi di \hat{s} .

Si è deciso di analizzare nel dettaglio l'effetto di variazioni percentuali positive e negative della matrice di covarianza sul parametro \hat{s} . I dati utilizzati per questa dimostrazione sono quelli dei rendimenti mensili logaritmici con $n = 143$ e $k = 8$. I risultati che si vedono in tabella (Tab.4.4) sono stati ottenuti tenendo costante la media dei rendimenti attesi e facendo variare la matrice di covarianza dei dati, moltiplicandone ogni elemento per uno scalare(α) che rappresenta la variazione percentuale dalla matrice di partenza. Si vede come per aumenti percentuali dei valori della matrice di covarianza si verifica una diminuzione del parametro \hat{s} , al contrario per riduzioni percentuali il valore del parametro aumenta. Questo è giustificato dal fatto che per determinare \hat{s} si utilizza \hat{R} , il quale dipende dalla matrice inversa di covarianza, di conseguenza a riduzioni dei valori della matrice di covarianza corrisponderanno aumenti di \hat{R} .

Tabella 4.4: Effetto delle variazioni della matrice di covarianza su \hat{s}

Effetto delle variazioni della matrice di covarianza su \hat{s}					
Variazione	\hat{s}_+	Var% ₊	Variazione	\hat{s}_-	Var% ₋
+5%	0.0595	-4.8%	-5%	0.0658	+5.2%
+10%	0.0568	-9.1%	-10%	0.0694	+11.0%
+15%	0.0543	-13.1%	-15%	0.0735	+17.6%
+20%	0.0521	-16.7%	-20%	0.0781	+24.92%
+80%	0.0347	-44.5%	-80%	0.3101	+399.7%

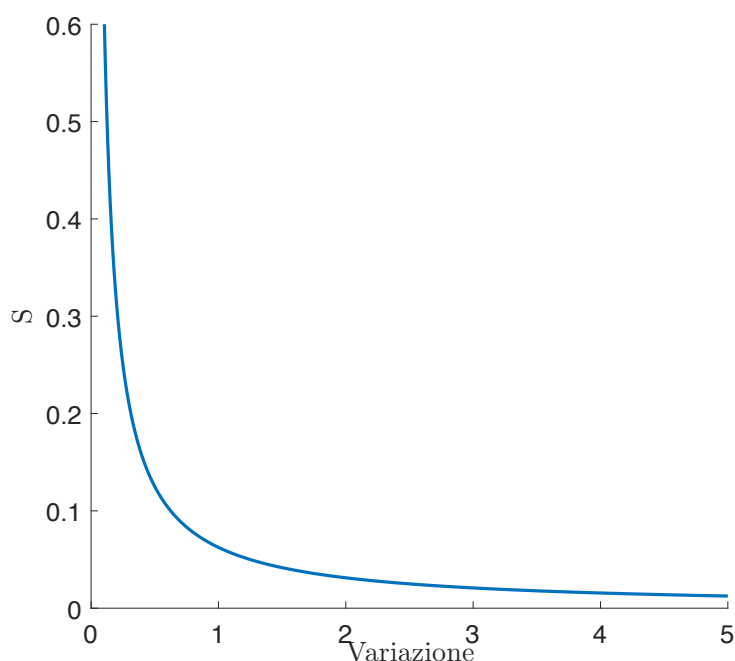


Figura 4.18: Rappresentazione di \hat{s} alle variazioni percentuali della matrice di covarianza

Graficamente la funzione di \hat{s} al variare della covarianza si presenta come un ramo d'iperbole con centro nell'origine degli assi (Fig.4.18), indicando che all'aumentare degli elementi della matrice di covarianza il valore di \hat{s} tende a zero, viceversa quando questi diminuiscono il parametro tende all'infinito. Il grafico mostra sull'asse delle ascisse il coefficiente che moltiplica la matrice di covarianza e sulle ordinate il valore di \hat{s} corrispondente. Quindi conoscendo la funzione di \hat{s} al variare di α si può ricavare di quanto debbano diminuire in percentuale i valori della matrice di covarianza affinché la frontiera alla Bodnar e Bodnar passi da distorta a corretta. Per esempio nel caso della seconda metà dei dati mensili, analizzato in precedenza, si era ricavato $\hat{s} = 0.0672$ con $s_{lim2} = 0.1293$, valori per cui la frontiera risultante era caratterizzata da varianza negativa. Di conseguenza per avere $\hat{s} > s_{lim2}$ c'è bisogno di un aumento di \hat{s} pari a $\approx 92\%$, che implica un valore di $\alpha \approx 0.51$ ovvero una diminuzione del 49% di ogni elemento della matrice di covarianza. Infatti, applicando $\alpha = 0.51$ alla matrice $\hat{\Sigma}$ dei dati iniziali e tenendo il vettore dei rendimenti attesi $\hat{\mu}$ costante si ottiene $\hat{s} = 0.1318$ che risulta maggiore del valore limite, di conseguenza la frontiera risulterebbe corretta.

L'analisi svolta ha evidenziato come il valore di \hat{s} risulti molto sensibile ai cambiamenti nel valore dei rendimenti attesi e della matrice di covarianza dei titoli comportando un grosso limite nell'applicazione del modello di Bodnar e Bodnar. In particolare è emerso come il parametro \hat{s} che indica la pendenza della frontiera sia inversamente proporzionale agli aumenti dei valori della matrice di covarianza.

4.2.4 Comportamento asintotico della frontiera efficiente corretta

Come descritto nel secondo capitolo la frontiera à la Bodnar nel caso di un aumento nel numero delle osservazioni tende alla frontiera efficiente campionaria. Questo perchè i parametri che determinano la correzione della frontiera al crescere di n tendono ad assumere valori tali per cui la frontiera campionaria e quella à la Bodnar e Bodnar coincidano. I rispettivi limiti sono:

$$\lim_{n \rightarrow +\infty} (A) = 0,$$

$$\lim_{n \rightarrow +\infty} (B) = 1,$$

$$\lim_{n \rightarrow +\infty} (C) = 0,$$

$$\lim_{n \rightarrow +\infty} (D) = 1.$$

Per quanto riguarda i risultati empirici per i rendimenti mensili degli 8 titoli selezionati si confermano i risultati teorici, infatti all'aumentare delle osservazioni si evidenzia che il valore assunto dai parametri tenda al loro valore limite. Gli effetti dell'aumento della numerosità campionaria per i parametri di correzione proposti da Bodnar e Bodnar sono riportati in tabella (Tab.4.5).

Dai risultati si evince come per una serie di 15 osservazioni l'effetto della correzione sia considerevole, in quanto i valori assunti da A,B,C e D differiscono in modo notevole dai loro valori limite, mentre già per $n=100$ le due frontiere risulteranno molto più simili. Gli stessi valori dei parametri si avranno nel caso di rendimenti giornalieri in quanto questi dipendono esclusivamente da n e k , e sono di conseguenza indipendenti dalle caratteristiche dei titoli. Questo comporta che per

Tabella 4.5: Effetto delle variazioni della numerosità campionaria sulla frontiera à la Bodnar e Bodnar

Comportamento della frontiera al variare di n			
Parametri	n=15	n=100	n=300
A	0.2889	0.0011	0.0035
B	0.4286	0.9192	0.9732
C	0.4667	0.0700	0.0233
D	1.6670	1.0868	1.0205

un'analisi sui rendimenti mensili un numero di osservazioni pari a 100 significa un arco temporale di poco più di 8 anni, ma nel caso di rendimenti mensili si tradurrebbe in meno di 3 mesi di osservazioni. Quindi nel caso di un'elevata numerosità campionaria l'utilizzo del modello di Bodnar e Bodnar non porta ad una correzione significativa della frontiera efficiente campionaria.

4.3 Utilizzo della PCA per stimare la frontiera à la Bodnar

In questa fase dell'analisi si utilizzerà la tecnica delle componenti principali per definire un nuovo set di variabili che siano per costruzione indipendenti fra loro. Le componenti principali, che sono una trasformazione lineare delle variabili iniziali, verranno in seguito utilizzate per stimare la frontiera efficiente à la Bodnar e Bodnar. Questo permetterà di misurare l'effetto sulla frontiera efficiente corretta dell'indipendenza fra le variabili.

4.3.1 Definizione delle componenti principali

La simulazione della frontiera efficiente corretta utilizzando l'analisi delle componenti principali sarà eseguita sia per i rendimenti giornalieri sia per quelli mensili. Le componenti principali sono state ottenute utilizzando l'apposito comando in matlab che restituisce la matrice dei coefficienti e gli autovalori. Si è poi eseguito il prodotto tra la matrice iniziale dei dati $X [n \times k]$ e la matrice dei coefficienti $C [k$

$n \times k$] da cui è risultata una nuova matrice Y [$n \times k$], le cui colonne rappresentano le componenti principali. Essendo per definizione le componenti indipendenti fra loro la matrice di covarianza calcolata su Y sarà diagonale, e di conseguenza si avrà che $Cov(Y_i, Y_j) = 0$, per $i \neq j$. Inoltre l'indipendenza comporta che la correlazione fra le diverse coppie di componenti sarà uguale a zero. Questo implica che la varianza totale calcolata su Y sarà data dalla somma delle varianze delle componenti principali e risulterà uguale alla varianza calcolata su X .

Come precedentemente esposto, le componenti principali possono essere espresse come combinazioni lineari delle variabili originali, e quindi una generica componente, ad esempio la prima, può essere riscritta come:

$$y_1 = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p$$

ed in forma compatta

$$y_1 = Xc_1$$

ogni elemento della matrice Y sarà allora un portafoglio in quanto combinazione lineare dei titoli selezionati. Il peso di ogni titolo sarà dato dal rapporto tra il coefficiente e la sommatoria dei valori del rispettivo autovettore. Va sottolineato che quest'ultima quantità non somma necessariamente ad uno e potrebbe assumere valore negativo nel caso in cui le vendite allo scoperto (short) dominino la costituzione del portafoglio. Ogni componente principale potrà essere considerata come un vero e proprio titolo, caratterizzata da un rendimento atteso e da una varianza, la quale corrisponderà all'autovalore della componente di riferimento. Attraverso l'applicazione del modello di Bodnar e Bodnar sulla nuova matrice dei dati Y si ricaveranno dunque dei portafogli efficienti caratterizzati dall'essere combinazione di componenti principali e quindi di attività indipendenti fra loro.

4.3.2 Applicazione della PCA per stimare modello à la Bodnar

L'applicazione della PCA al modello di Bodnar e Bodnar permette di ottenere una frontiera efficiente composta da elementi indipendenti fra loro. Questo implica che la correlazione e la covarianza per tutte le coppie di componenti principali sia

uguale a zero.

Data l'analisi sul parametro $\hat{\sigma}$, svolta in precedenza, da cui è emerso come la pendenza della frontiera efficiente corretta sia inversamente proporzionale agli aumenti della matrice di covarianza, con l'applicazione della PCA ci si aspetta un effetto positivo sul parametro.

Verranno proposte 3 diverse simulazioni per verificare gli effetti dell'applicazione della tecnica delle componenti principali sulla frontiera efficiente alla Bodnar e Bodnar: la prima sarà eseguito sulla totalità dei dati giornalieri, la seconda sulla totalità dei dati mensili e la terza sulle due metà dei dati mensili. Quest'ultima simulazione ci permetterà di valutare se l'effetto positivo sul parametro $\hat{\sigma}$ sarà in grado di correggere la distorsione della frontiera efficiente corretta.

Il grafico (4.19) mostra le due diverse frontiere ottenute su dati giornalieri, da cui si può notare che l'applicazione della PCA permette di ottenere una frontiera che domina per la quasi totalità la frontiera à la Bodnar. Infatti, solo per valori molto piccoli della varianza si verifica che la frontiera à la Bodnar e Bodnar classica assuma valore maggiore di quella calcolata con la PCA. Nella tabella (Tab.4.6) sono riportati i valori di rendimento atteso e varianza del portafoglio a minima varianza e del parametro $\hat{\sigma}$ ottenuti sulle due diverse frontiere.

Tabella 4.6: Confronto fra i valori di R_{GMV} , V_{GMV} e $\hat{\sigma}$ per dati giornalieri

Risultati per dati giornalieri		
Parametri	PCA	Bodnar
R_{GMV}	0.00029	0.00046
V_{GMV}	0.00001	0.00008
$\hat{\sigma}$	0.00531	0.00270

I risultati indicano che con l'applicazione della PCA si può ottenere un portafoglio di minima varianza caratterizzata da un minor rendimento e da minor rischio rispetto a quello ottenuto sui rendimenti iniziali. Inoltre si evidenzia come il valore del parametro $\hat{\sigma}$ sia influenzato positivamente dall'applicazione della PCA, infatti il suo valore passa da 0.00270 a 0.00531. Questo è giustificato dalle variazioni che si verificano sia nel parametro $\hat{\mu}$ dei rendimenti attesi sia sulla matrice di covarianza $\hat{\Sigma}$.

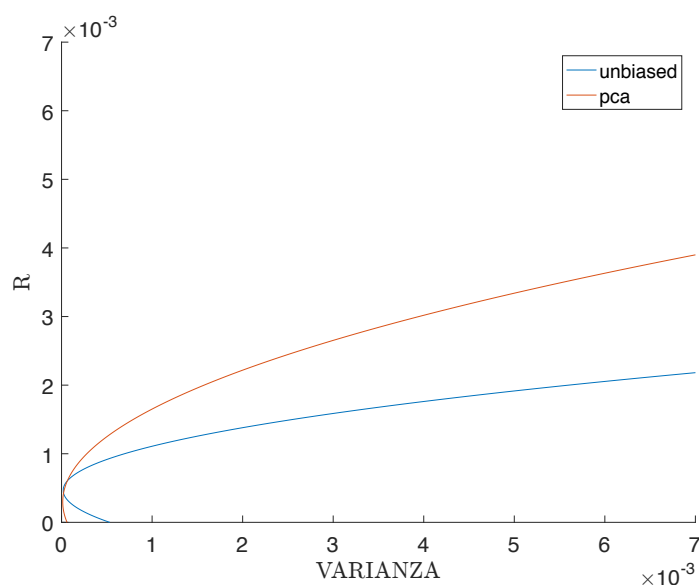


Figura 4.19: Confronto grafico fra il modello à la Bodnar e la sua stima con l'applicazione della PCA, per dati giornalieri

I risultati ottenuti dalla simulazione sui dati mensili sono in linea con quelli ottenuti per rendimenti giornalieri, la tabella (Tab.4.7) mostra infatti i valori assunti dai parametri.

Tabella 4.7: Confronto fra i valori di R_{GMV} , V_{GMV} e \hat{s} per dati mensili

Risultati per dati mensili		
Parametri	PCA	Bodnar
R_{GMV}	0.0049	0.0110
V_{GMV}	0.0002	0.0011
\hat{s}	0.0976	0.0629

Infatti, da questa si nota come anche per i dati mensili si ottengono valori del portafoglio a minima varianza inferiori e un aumento del valore di \hat{s} . Graficamente questa seconda simulazione si presenta come in figura (4.20).

Dalle due simulazioni emerge che l'applicazione della PCA al modello di Bodnar e Bodnar fornisce una nuova frontiera che domina per la quasi totalità quella ottenuta dai dati originali. Inoltre si rileva un impatto positivo sul parametro \hat{s} tale che in alcuni casi la frontiera ottenuta con la tecnica delle componenti principali riesce a correggere la distorsione di quella à la Bodnar e Bodnar. Per dare una prova di questa variazione nel parametro \hat{s} , si applicherà la PCA alle due serie

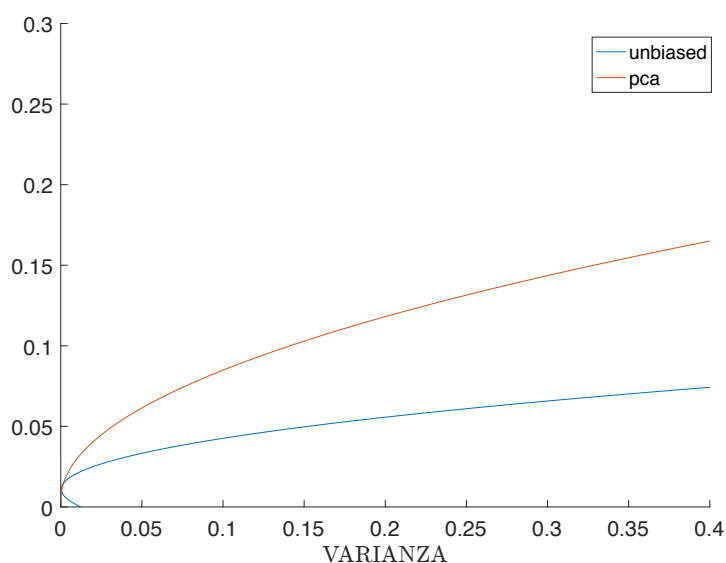


Figura 4.20: Confronto grafico fra il modello à la Bodnar e la sua stima con l'applicazione della pca, per dati mensili

di rendimenti mensili, ottenute dividendo a metà il campione iniziale, analizzate nel paragrafo 4.2.3 di questo capitolo. L'analisi iniziale aveva mostrato come la frontiera à la Bodnar e Bodnar ottenuta sulla prima metà dei dati risultasse corretta, mentre quella calcolata sulla seconda metà distorta.

I risultati delle due simulazioni sono rappresentati dai grafici (Fig.4.21), che mostrano come l'effetto positivo sul parametro \hat{s} nel caso della seconda simulazione sia tale da correggere la distorsione sulla frontiera. I valori dei limiti rimangono invariati per entrambe le simulazioni poichè dipendono unicamente da n e k , e valgono rispettivamente $s_{lim1} = 0.1113$ e $s_{lim2} = 0.1293$. Nel caso della prima metà dei dati si ottiene un valore di $\hat{s}_{pca} = 0.2658$ che risulta maggiore di quello derivante dalla frontiera à la Bodnar calcolata sui rendimenti iniziali $\hat{s} = 0.1467$. Si nota come il parametro sia influenzato positivamente dall'applicazione della PCA, sebbene in questo caso non sia necessaria alcuna correzione poichè la frontiera non presenta distorsione. Al contrario la frontiera calcolata sulla seconda metà dei dati risulta non accettabile in quanto il valore di $\hat{s} = 0.0672$ è inferiore a s_{lim2} , ma quella ottenuta dalle componenti principali risulta corretta poichè $\hat{s}_{pca} = 0.1845$.

Dunque dalle analisi svolte emerge che le modifiche nei valori di $\hat{\mu}$, e $\hat{\Sigma}$ derivanti dall'applicazione delle PCA comportano un aumento del parametro \hat{s} . Questo può comportare in alcuni casi la correzione della frontiera à la Bodnar, soprattutto

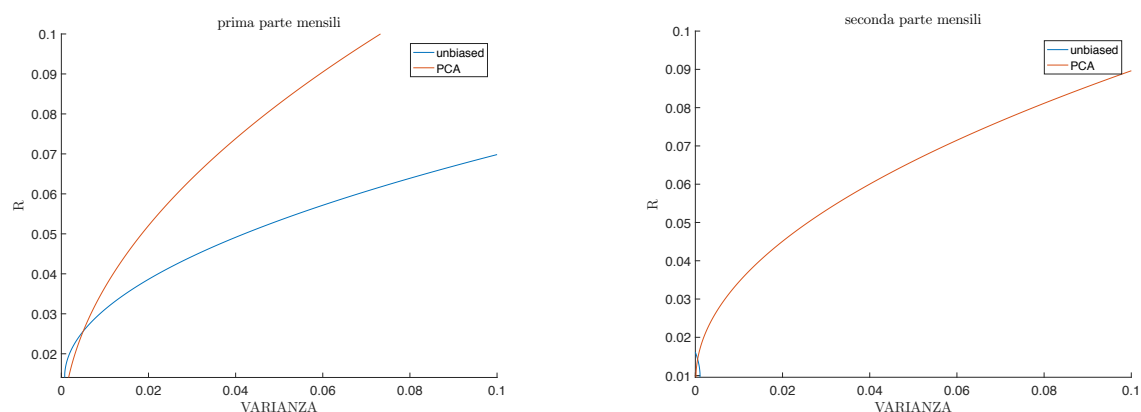


Figura 4.21: Confronto frontiera corretta à la Bodnar con quella ottenuta con l'applicazione della PCA per rendimenti mensili

quando il valore di \hat{s} risulta di poco inferiore ai valori limite per cui si verifica la distorsione. Inoltre, considerando il grafico per la prima metà delle osservazioni si vede come anche in questo caso la frontiera efficiente ottenuta con la PCA per la maggior parte dei valori di V domini quella à la Bodnar e Bodnar classica.

Concludendo, dati i risultati ottenuti dall'analisi sul parametro \hat{s} che indicano come questo sia inversamente proporzionale agli aumenti dei valori di $\hat{\Sigma}$ e quelli derivati dall'applicazione della PCA al modello di Bodnar e Bodnar si ritiene che non sia strettamente necessaria una trasformazione ortogonale, ma si possa prendere in considerazione un eventuale trasformazione obliqua. Questa si potrebbe ottenere attraverso un'analisi fattoriale e permetterebbe di ridurre le covarianze fra i titoli senza l'ipotesi di completa indipendenza.

CONCLUSIONI

L'analisi svolta ha messo in luce quelle che sono le caratteristiche principali della frontiera efficiente campionaria e della sua stima corretta proposta da T.Bodnar e O.Bodnar.

In particolare, si è visto come la distribuzione dei rendimenti logaritmici non segua necessariamente quella della normale; dall'analisi è risultato infatti che per rendimenti giornalieri la normalità non sia mai accettata mentre per quelli mensili non sia sempre verificata. Inoltre si è rilevata la tendenza dei rendimenti logaritmici ad assumere una distribuzione leptocurtica, che si accentua nel caso di rendimenti giornalieri. Le simulazioni hanno mostrato la correzione sulla frontiera efficiente campionaria derivante dall'applicazione del modello à la Bodnar e Bodnar, nonostante i rendimenti logaritmici dei titoli selezionati non fossero normalmente distribuiti. Va sottolineato che per la definizione dei parametri correttivi del modello gli autori assumono che i rendimenti siano distribuiti come una gaussiana e questo implica che non verificandosi la condizione di normalità, e quindi una delle basi su cui si fonda il modello, lo stesso perda di efficacia.

Inoltre, si è visto come la frontiera efficiente corretta all'aumentare del numero di osservazioni n tenda a quella campionaria classica, implicando una riduzione della distanza fra le due frontiere e pertanto un minor effetto derivante dalla correzione. Il numero di osservazioni n e il numero di titoli in portafoglio k sono inoltre gli unici parametri da cui dipendono i valori limite (\hat{s}_{lim1} e \hat{s}_{lim2}) per i quali la frontiera efficiente corretta assume varianza negativa e di conseguenza risulta inaccettabile. Dall'analisi è infatti emerso come per alcuni valori di \hat{s} che determina la pendenza della curva, la frontiera efficiente à la Bodnar e Bodnar assuma un comportamento anomalo.

Dalle diverse simulazione sui parametri che definiscono il valore di \hat{s} , cioè $\hat{\mu}$

CONCLUSIONI

vettore dei rendimenti attesi e $\hat{\Sigma}$ matrice di covarianza è risultato come \hat{s} sia molto sensibile a variazioni nei parametri che lo costituiscono. In particolare si è evidenziato come diminuzioni nei valori di $\hat{\Sigma}$ comportino un aumento del parametro \hat{s} .

Nell'ultima fase dell'analisi si è utilizzata la tecnica delle componenti principali (PCA) allo scopo di ottenere delle nuove variabili ortogonali fra loro per valutare l'effetto dell'indipendenza sulla frontiera efficiente corretta. I risultati hanno mostrato che la frontiera alla Bodnar e Bodnar calcolata sulle componenti principali offre migliori combinazioni di rischio rendimento rispetto a quella derivante dai rendimenti classici. In aggiunta l'applicazione della PCA ha comportato un impatto positivo sul parametro \hat{s} , tale che per alcune simulazioni la frontiera derivata dalle componenti principali risultasse accettabile mentre quella calcolata sui rendimenti distorta.

In conclusione, nel corso dell'elaborato si è visto come gli stimatori della frontiera efficiente campionaria classica siano effettivamente distorti e che l'applicazione degli stimatori proposti nel modello di T.Bodnar e O.Bodnar per dati reali non risulti una soluzione sempre accettabile. Questo perché il modello si basa sull'assunzione di distribuzione normale dei rendimenti e sulla loro indipendenza, ipotesi che alla luce delle analisi empiriche mal descrivono la realtà. Inoltre il suo utilizzo risulta viziato dalle problematiche derivanti dai possibili valori assunti da \hat{s} che potrebbero comportare una varianza negativa e di conseguenza l'inutilità del modello.

BIBLIOGRAFIA

- [1] BODNAR T. E SCHMID W., Econometrical Analysis of the Sample Efficient Frontier, *The European Journal of Finance*, 15(3), pp. 317-335, 2009
- [2] BODNAR T. E BODNAR O., Statistical inference procedure for the mean-variance efficient frontier with estimated parameters, *ASTA Advances in Statistical Analysis*, 93(3), pp. 295-306, 2009
- [3] BODNAR T. E BODNAR O., On the Unbiased Estimator of the Sample Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 13(7), pp. 1065-1073, 2010
- [4] BODIE Z., KANE A. E MARCUS A., Investment, 10th Edition, *McGraw-Hill*, 2014
- [5] BRACALENTE B., COSSIGNANI M. E MULAS A., Statistica Aziendale, *McGraw-Hill*, 2009
- [6] BRESOLIN E., Sulla distorsione della frontiera efficiente dei portafogli: analisi delle ipotesi di normalità ed indipendenza, *Università Ca' Foscari Venezia*, 2016-2017
- [7] ELTON E., GRUBER M., BROWN S. E GOETZMANN W., Teorie di portafoglio e analisi degli investimenti, *Apogeo Education*, 11-2017
- [8] FAMA E., The Behavior of Stock-Market prices, *Journal of Business*, 38(1), pp 34-105, 1965
- [9] HESTER D., Risk Aversion and Portfolio Choice, *Wiley*, 1967
- [10] HOTTELING H., Analysis of a complex of statistical variables into principal component, *Journal of Educational Psychology*, pp 417-441, 1933

BIBLIOGRAFIA

- [11] JOLIFFE I., *Principal Component Analysis*, Springer, 1986
- [12] KAHNEMAN D. E TVERSKY A., *Prospect Theory: An Analysis of Decision Under Risk*, *The Econometric Society*, 47(2), pp 263-292, 1979
- [13] KAN R. E SMITH D., *The Distribution of the Sample Minimum-Variance Frontier*, *Management Science*, 54(7), pp 1364-1380, 2008
- [14] KRITZMAN M., LI Y., PAGE S. E RIGOBON R., *Principal Component as a Measure of Systemic Risk*, *The Journal of Portfolio Management*, 37(4), 2011
- [15] MANDELBROT B., *The Variation of Certain Speculative Prices*, *The Journal of Business*, 36(4), pp 394-419, 1963
- [16] MARKOWITZ H., *Portfolio Selection*, *The Journal of Finance*, 7(1), pp. 77-91, 1952
- [17] MARASOVIC B., PIVAC S. E VUKASOVIC V., *The impact of transition costs on rebalancing an investment portfolio in portfolio optimization*, *World Academy of Science, Engineering and Technology*, 9(3), 2015
- [18] MEUCCI A., *Managing Diversification*, *Bloomberg Education and Quantitative Research and Education Paper*, pp 74-79, 2010
- [19] MICHIELI F. *Stimatori campionari non distorti della frontiera efficiente*, *Università Ca' Foscari Venezia*, 2013-2014
- [20] PARTOVI M. E CAPUTO M., *Principal Portfolios: Recasting the Efficient Frontier*, *Economics Bulletin*, 7(3), pp 1-10, 2004
- [21] PEARSON K., *On Lines and Planes of Closet Fit to System of Points in Space*, *Philosophical Magazine*, pp 559-572, 1901
- [22] SIEGEL A. E WOODGATE A., *Performance of Portfolios Optimized with Estimation Error*, *Management Science*, 53(6), pp. 1005-1015, 2007