



Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale
in Economia e Finanza
(*ordinamento ex D.M. 270/2004*)

Tesi di Laurea

Sulla distorsione della frontiera
efficiente dei portafogli:
analisi delle ipotesi di normalità
ed indipendenza

Relatore

Prof. Claudio Pizzi

Laureando

Enrico Bresolin

Matricola 844270

Anno Accademico

2016-2017

INDICE

Indice	i
Elenco delle tabelle	iii
Elenco delle figure	iv
Introduzione	1
1 Markowitz e la frontiera efficiente	3
1.1 I postulati alla base dell'approccio media-varianza	3
1.2 Derivazione analitica della frontiera efficiente	6
1.3 Fallacie nei presupposti della frontiera efficiente	9
2 Intervallo teorico per la Frontiera Efficiente	14
2.1 Le ipotesi di Bodnar e Schmid	14
2.2 La definizione della regione di confidenza di Bodnar e Schmid . . .	16
3 La distorsione della frontiera efficiente e la sua correzione	27
3.1 La Distorsione della frontiera efficiente	27
3.2 Test F sulla Frontiera Efficiente	33
3.3 Criticità della frontiera efficiente à la Bodnar e Bodnar	34
4 Bootstrap per l'analisi dei dati	39
4.1 Scopo dell'analisi	39
4.2 Il metodo bootstrap	40
4.3 Parametric bootstrap o bootstrap dei residui	41
4.4 Dati surrogati	43
4.5 Bootstrap a blocchi	45

5	Analisi della frontiera efficiente corretta	48
5.1	Dati utilizzati	48
5.2	Frontiera efficiente "à la Bodnar e Bodnar" per dati reali	56
5.3	Frontiera efficiente "à la Bodnar e Bodnar" per dati simulati	58
5.3.1	Simulazione di serie storiche normali multivariate e prive di autocorrelazione	58
5.3.2	Simulazione di serie storiche con funzione di distribuzione t di Student prive di autocorrelazione	61
5.3.3	Simulazione di serie storiche normali autocorrelate	63
5.3.4	Simulazione di serie storiche non normali e autocorrelate	66
6	Analisi esplorativa	68
6.1	Dati utilizzati	68
6.2	Analisi dei portafogli	69
	Conclusioni	81
A	Chiarimenti Modern Portfolio Theory	84
A.1	Funzione di utilità	84
A.2	Eguaglianza tra approccio semplificato e non nel processo di minimizzazione	85
A.3	Derivazione grafica della frontiera efficiente dei portafogli	86
B	Dimostrazione teorica dell'intervallo di confidenza	89
	Bibliografia	93

ELENCO DELLE TABELLE

5.1	Test di normalità	49
5.2	Test di normalità per serie DuPont	53
5.3	Parametri processi ARMA(1,1)	64
6.1	P-value dei test di linearità	69
6.2	P-value del test di normalità	70
6.3	Matrice di Varianza e Covarianza del Portafoglio A	73
6.4	Matrice di Varianza e Covarianza del Portafoglio B	73
6.5	Vettori delle medie dei portafogli A e B	74
6.6	Matrice di Varianza e Covarianza del Portafoglio A	77
6.7	Matrice di Varianza e Covarianza del Portafoglio C	77
6.8	Matrice di Correlazione del Portafoglio A	77
6.9	Matrice di Correlazione del Portafoglio C	78
6.10	Vettori delle medie dei portafogli A e C	78

ELENCO DELLE FIGURE

1.1	Rappresentazione della frontiera efficiente	11
2.1	Aree della regione di confidenza	22
2.2	Costruzione della regione di confidenza	24
2.3	Regione di confidenza per dati reali	25
3.1	Limite inferiore di k asset	30
3.2	Confronto tra le frontiere	32
5.1	Quantili distribuzioni empiriche mensili	49
5.2	Quantili distribuzioni empiriche settimanali	50
5.3	Quantili distribuzioni empiriche giornalieri	50
5.4	Distribuzioni mensili	51
5.5	Distribuzioni settimanali	52
5.6	Distribuzioni giornaliere	52
5.7	Box test per osservazioni giornaliere	53
5.8	Box test per osservazioni settimanali	54
5.9	Box test per osservazioni mensili	54
5.10	Frontiere efficienti a confronto	56
5.11	Intervallo bootstrap frontiera efficiente	57
5.12	Intervallo bootstrap serie normali	59
5.13	Frontiera efficiente corretta negativa	61
5.14	Frontiera efficiente per dati non normali	63
5.15	Frontiere efficienti per dati autocorrelati, 1424 osservazioni	65
5.16	Frontiere efficienti per dati autocorrelati, 120 osservazioni	65
5.17	Frontiere efficienti per dati autocorrelati	66

ELENCO DELLE FIGURE

6.1	Confronto tra distribuzioni empiriche e teoriche	71
6.2	Box Test Portafoglio A	72
6.3	Box Test Portafoglio B	72
6.4	Frontiere efficienti Portafoglio A e Portafoglio B	73
6.5	Frontiere efficienti Portafogli 1, 2, 3, 4	74
6.6	Frontiere efficienti Portafogli 1a, 2a, 3a, 4a	75
6.7	Frontiere efficienti Portafogli C, D, E, F	76
6.8	Frontiere efficienti Portafogli Alpha e Beta	79
A.1	Set di portafogli accettabili	87
A.2	Set di combinazioni (E,V) efficienti	88

*"Whenever we proceed from the known into the unknown we may hope to understand,
but we may have to learn at the same time a new meaning of the word understanding."*

W. Heisenberg

Alla mia famiglia

INTRODUZIONE

A metà del secolo scorso Harry Markowitz concettualizzando la Modern Portfolio Theory rivoluzionò completamente il mondo della finanza.

L'ipotesi innovativa di poter costruire un portafoglio di titoli basandosi esclusivamente sulle proprietà statistiche degli stessi costituì un importante snodo nell'approccio agli investimenti finanziari, inquadrando in una cornice più rigorosa ciò che agli operatori e agli economisti precedenti era comunque chiaro: la basilarietà della diversificazione.

È proprio attorno a questo concetto semplice ma fondamentale che Markowitz elaborò la sua teoria pur tuttavia introducendo postulati non del tutto attinenti alla realtà.

Proprio a causa di questa inesatta descrizione dell'ambiente in cui gli investitori si ritrovano ad agire, la Modern Portfolio Theory (MPT) non costituisce un modello utilizzabile per poter operare con sicurezza nel mercato finanziario.

Da qui è sorta la necessità di porre in essere miglioramenti alle assunzioni formulate nel 1952 ed economisti e statistici nel corso degli anni hanno proposto via via nuovi modelli correttivi, mantenendo però invariato il nocciolo della MPT.

L'aspetto comune a tutti questi lavori è nel riconoscere una distorsione della frontiera efficiente dei portafogli che porta gli operatori, qualora applicassero esclusivamente l'approccio *à la Markowitz*, a sottovalutare il rischio e a sopravvalutare le prestazioni dei titoli considerati.

Una delle ultime novità relative ad una possibile correzione della frontiera efficiente è stata proposta da Olha Bodnar e Taras Bodnar (2010), ma anche in questa formulazione sono presenti elementi di perplessità relativi ad alcuni postulati, specialmente riguardo l'ipotesi di assenza di autocorrelazione dei titoli presi in esame e la normalità della loro funzione di distribuzione.

Inoltre evidenze empiriche, come si provvederà ad illustrare, portano a sostenere che non sempre tale correzione è applicabile.

Con il presente lavoro di tesi quindi ci si propone di analizzare il comportamento di entrambe le frontiere (campionaria "à la Markowitz" e corretta "à la Bodnar e Bodnar"), verificandone la differenza statistica con la creazione di intervalli di confidenza per la frontiera efficiente campionaria ottenuti mediante metodo "block bootstrap". Si studieranno dataset con dati reali e con dati simulati, dapprima rispettando pienamente le principali assunzioni poste da Bodnar e Bodnar (2010), e poi violandole per osservare l'influenza nel risultato finale.

Si svolgerà anche un'analisi esplorativa su di un potenziale portafoglio azionario, al fine di individuare le probabili cause di errore riscontrate nel comportamento della frontiera efficiente corretta.

Il lavoro avrà la seguente struttura:

Capitolo primo: illustrazione della derivazione della MPT e i suoi fondamenti, evidenziandone alcune criticità.

Capitolo secondo: creazione di un possibile intervallo di confidenza teorico (Bodnar e Schmid, 2009) per la frontiera efficiente "à la Markowitz" e i motivi della sua scarsa efficacia applicativa.

Capitolo terzo: trattazione della distorsione della frontiera efficiente evidenziata da Bodnar e Bodnar (2010) e della correzione da questi proposta e focus sulle criticità emerse.

Capitolo quarto: spiegazione del metodo bootstrap utilizzato per le analisi svolte mediante codice R, al fine di costruire un intervallo di confidenza empirico per la frontiera "à la Markowitz"

Capitolo quinto: esposizione dei risultati ottenuti mediante le analisi svolte, sia per serie storiche reali sia per serie storiche simulate, in entrambi i casi di osservanza o meno delle assunzioni poste.

Capitolo sesto: breve analisi esplorativa di alcuni ipotetici portafogli al fine di rintracciare le cause di errore nella frontiera efficiente "à la Markowitz".

Conclusioni: conclusioni generali sugli aspetti analizzati.

Capitolo 1

MARKOWITZ E LA FRONTIERA

EFFICIENTE

1.1 I postulati alla base dell'approccio

media-varianza

La Modern Portfolio Theory costituisce il pilastro su cui si erge l'approccio odierno alla finanza. Harry Markowitz ebbe l'idea di analizzare gli investimenti guardando alle proprietà statistiche dei titoli presenti sul mercato: ipotizzò che un investitore fosse interessato solo a due aspetti dell'investimento, il rendimento e il rischio connesso allo stesso.

Richiamando le proprietà della funzione di distribuzione gaussiana, ovvero la possibilità di descrivere tutti i momenti della funzione in base ai primi due – media e varianza (per il secondo il momento centrato) – Harry Markowitz (1952) propose di utilizzare i dati storici dei rendimenti dei titoli per stimare questi due momenti, in quanto, almeno per osservazioni mensili, la distribuzione delle stesse è di tipo normale.¹

¹In realtà successivi lavori di Fama e al. hanno evidenziato come anche le osservazioni di rendimenti mensili, che dovrebbero godere di maggiore indipendenza da fattori di perturbazione del mercato, presentano problemi di leptocurtosi. Tuttavia Tu e Zhou (2004) evidenziano anche che non vi sono vantaggi evidenti nell'utilizzare altre funzioni di distribuzione in sostituzione di quella normale: se ne conclude quindi che l'utilizzo di una distribuzione gaussiana è accettabile in questo contesto, come affermato tempo addietro anche da Fama (1976)

Postulò inoltre altre assunzioni relative al contesto in cui le osservazioni si verificano, che seppur necessarie per inquadrare il problema in un ambiente di riferimento, a causa delle frizioni con la realtà riducono l'operatività e l'efficacia della sua teorizzazione:

- tutti gli operatori sul mercato valutano ogni titolo in modo analogo, con la stessa funzione di distribuzione e lo stesso arco temporale di riferimento;
- il mercato è perfetto (e di conseguenza vi è assenza di tassazione e di costi di transazione);
- tutti gli operatori misurano il rischio in modo proporzionale alla volatilità del titolo (sia essa espressa in forma di varianza o di deviazione standard);
- la funzione di utilità degli investitori U è proporzionale (direttamente) al rendimento e (inversamente) alla variabilità del titolo;
- vi è perfetta divisibilità di un titolo (è possibile acquistare anche una frazione di azione);
- gli operatori sono "price takers", ovvero non possono influenzare il mercato con le loro azioni;

Un aspetto chiave nella teorizzazione di Markowitz è l'approccio media-varianza che descrive la razionalità degli investitori: essi per un dato livello di rischio prediligono il maggior rendimento possibile, mentre per un dato livello di rendimento scelgono il titolo che presenta il minor grado di rischio.

Questa formulazione è fondamentale per descrivere l'approccio di Markowitz proposto nel 1952 riguardo alla selezione dei titoli da inserire nel portafoglio: si inizia con lo stimare i parametri di interesse (media e varianza dei rendimenti, ovvero le misure di aspettativa di rendimento e grado di rischio) a partire dai dati storici, poi sono scartati in prima battuta i titoli inefficienti, ovvero dominati in base al criterio media-varianza e da ultimo si selezionano i titoli da inserire nel portafoglio, rispettando i postulati inerenti agli investitori di cui sopra, ovvero realizzando un portafoglio che nel complesso rispetti le volontà dell'investitore in termini di rischio.

Si inserisce a questo punto un altro elemento cruciale nella scelta fra titoli: la correlazione tra gli stessi. Infatti, considerando i titoli come variabili aleatorie, qualora si prenda in esame un loro insieme si è interessati anche alla relazione

che vi è tra i diversi elementi che si stanno considerando: questa relazione è misurata dalla correlazione, e l'approccio markovitziano alla diversificazione si basa proprio su questo indice, il quale consente di misurare il grado di dipendenza lineare esistente tra le variabili considerate.

Si può così definire la frontiera efficiente dei portafogli: visto che ogni portafoglio è composto da più titoli, se questi hanno correlazioni minori di 1 allora il portafoglio totale sarà caratterizzato da un rendimento pari alla media -pesata- dei rendimenti dei titoli e da una varianza inferiore alla media pesata delle varianze.

Infatti se x e y sono variabili aleatorie, caratterizzate dall'aver la propria media e la propria varianza, e z è la somma di tali variabili, allora la variabile aleatoria risultante sarà $z = x + y$ e sarà caratterizzata dall'aver come valore atteso la somma dei valori attesi delle variabili di cui è costituita, ovvero $E(z) = E(x) + E(y)$ e come varianza una somma di tre addendi: non solo la somma delle varianze delle due variabili, ma anche di un terzo elemento, la covarianza, rappresentante la misura di dipendenza stocastica delle variabili coinvolte nella costituzione di z , ovvero $V(z) = V(x) + V(y) + 2Cov(x, y)$.

Si inserisce così in un contesto squisitamente analitico la presenza del beneficio della diversificazione: più basso è il coefficiente di correlazione, maggior diversificazione si ottiene nel portafoglio, ovvero una diminuzione del rischio dello stesso, misurato dalla sua varianza.

Risulta così fondamentale per la composizione di portafogli a varianza ridotta inserire, se possibile, all'interno degli stessi titoli negativamente correlati.

Per definire in modo rigoroso l'approccio è però necessario calcolare la composizione ottima del portafoglio. Tale composizione permette di massimizzare il rendimento dato un certo livello di rischiosità, oppure di minimizzare il rischio dato un certo livello di rendimento.

1.2 Derivazione analitica della frontiera efficiente

Harry Markowitz propose una derivazione grafica della frontiera efficiente dei portafogli, mentre una formulazione analitica la si deve al lavoro di Robert Merton (1972).²

La derivazione di Merton prevede la risoluzione di un problema di minimizzazione, così riassumibile: dati n titoli, e tutti i possibili portafogli creati come combinazione di quei titoli, si è interessati a selezionare solo quei portafogli che a parità di valore atteso posseggono la varianza minore.

In termini più formali: siano considerati n titoli diversi fra loro e siano note le seguenti quantità:

μ : vettore dei rendimenti attesi dei diversi titoli, di dimensione $n \times 1$;

Σ : matrice di varianza e covarianza dei rendimenti dei titoli, di dimensione $n \times n$;

ω : vettore dei pesi da assegnare ad ogni titolo considerato (ogni peso ha valore compreso tra 0 e 1), di dimensione $n \times 1$;

$\mathbf{1}$: vettore di 1, di dimensione $n \times 1$.

Si considerano inoltre i seguenti vincoli:

$\mu_p = \omega' \mu$, ovvero il rendimento di un generico portafoglio è pari alla media ponderata dei rendimenti attesi dei titoli che lo compongono;

$1 = \omega' \mathbf{1}$, ovvero la somma dei pesi deve essere pari a 1 (l'investitore crea un portafoglio utilizzando tutta la sua ricchezza).

Ciò che si intende risolvere è la minimizzazione vincolata di $\omega' \Sigma \omega$, tenendo conto dei vincoli esposti sopra, ovvero:

$$\begin{array}{ll} \min_{\omega} & (\omega' \Sigma \omega) \\ \text{sub to} & \begin{cases} \mu_p = \omega' \mu \\ 1 = \omega' \mathbf{1} \end{cases} \end{array}$$

²Nel presente lavoro, al fine di concentrarsi maggiormente sugli sviluppi successivi del modello proposto da Markowitz, si è scelto di riportare solo in Appendice A la derivazione grafica della frontiera, in quanto l'approccio analitico di Merton fornisce assai più informazioni e costituisce la base per i modelli successivi.

Il procedimento di ottimizzazione consiste nella creazione del Lagrangiano \mathcal{L} , la sua successiva derivazione rispetto alla variabile di interesse ω e la ricerca dei valori nulli di tale funzione.

La funzione Lagrangiana, che incorporando al suo interno i vincoli della funzione precedente risulta essere nel complesso una funzione non vincolata, è la seguente:

$$\mathcal{L}(\omega, \lambda, \gamma) = \frac{1}{2} \omega' \Sigma \omega + \lambda (\mu_p - \omega' \mu) + \gamma (1 - \omega' \mathbf{1}) \quad (1.1)$$

In questo caso per praticità si è scelto di inserire $\frac{1}{2}$ all'interno della formula di \mathcal{L} per agevolare i conti.³

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \omega'} &: \Sigma \omega + \lambda \mu + \gamma \mathbf{1} \\ \frac{\partial \mathcal{L}}{\partial \omega'} = 0 &\implies \omega = \lambda \Sigma^{-1} \mu + \gamma \Sigma^{-1} \mathbf{1} \end{aligned}$$

È necessario ora ricercare i valori dei pesi associati ai vincoli nella funzione \mathcal{L} : la soluzione di tali incognite può essere ottenuta attraverso la risoluzione del seguente sistema di equazioni:

$$\begin{cases} \mu_p = \mu' \omega \\ \mathbf{1} = \mathbf{1}' \omega \end{cases}$$

nel quale sostituendo ad ω quanto prima calcolato si ottiene:

$$\begin{cases} \mu_p = \lambda \mu' \Sigma^{-1} \mu + \gamma \mu' \Sigma^{-1} \mathbf{1} \\ \mathbf{1} = \lambda \mathbf{1}' \Sigma^{-1} \mu + \gamma \mathbf{1}' \Sigma^{-1} \mathbf{1} \end{cases}$$

Operando delle sostituzioni, rispettivamente:

- $a = \mu' \Sigma^{-1} \mu$,
- $b = \mu' \Sigma^{-1} \mathbf{1}$ e
- $c = \mathbf{1}' \Sigma^{-1} \mathbf{1}$,

è possibile riscrivere il precedente sistema in forma compatta.

Esso diventa:

$$\begin{pmatrix} \mu_p \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix}$$

³In Appendice A si è provveduto ad illustrare l'equivalenza dei due procedimenti

ed isolando a sinistra la componente di interesse:

$$\begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \frac{1}{(ca-b^2)} \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} \mu_p \\ 1 \end{pmatrix}$$

Risolviendo tale sistema si ottengono i seguenti risultati per i pesi dei due vincoli:

- $\lambda = \frac{\mu_p c - b}{ca - b^2}$
- $\gamma = \frac{a - \mu_p b}{ca - b^2}$

E sostituendo nel risultato del processo di ottimizzazione si può così ottenere il valore di ω , che risulta essere

$$\omega = \frac{(\mu_p c - b)}{(ca - b^2)} \Sigma^{-1} \mu + \frac{(ca - \mu_p b)}{(ca - b^2)} \Sigma^{-1} \mathbf{1}$$

Si può così risolvere il problema di ottimizzazione della varianza del portafoglio:

$$\begin{aligned} \omega' \Sigma \omega &= (\omega' \Sigma) (\lambda \Sigma^{-1} \mu + \gamma \Sigma^{-1} \mathbf{1}) \\ &= \lambda \omega' \Sigma^{-1} \Sigma \mu + \gamma \omega' \Sigma^{-1} \Sigma \mathbf{1} \\ &= \lambda \omega' \mu + \gamma \omega' \mathbf{1} \\ &= \lambda \mu_p + \gamma \end{aligned}$$

Sostituendo alle quantità λ e γ quanto precedentemente trovato si ottiene il valore di σ_p (ovvero $\omega' \Sigma \omega$):

$$\begin{aligned} \sigma_p^2 &= \frac{(c\mu_p^2 - b\mu_p)}{ca - b^2} + \frac{(a - \mu_p b)}{(ca - b^2)} \\ &= \frac{c\mu_p^2 - 2b\mu_p + a}{ca - b^2} \end{aligned}$$

e riorganizzando si può notare come la funzione di σ_p^2 non è altro che una parabola, dipendente dai valori di μ_p , e di conseguenza la frontiera efficiente dei portafogli teorizzata da Markowitz risulta essere il ramo superiore di tale parabola:

$$\sigma_p^2 = \frac{c}{ca - b^2} \mu_p^2 - \frac{2b}{ca - b^2} \mu_p + \frac{a}{ca - b^2} \quad (1.2)$$

È interessante notare come tale rappresentazione della frontiera efficiente dei portafogli consenta, oltre che una rappresentazione grafica efficace, di isolare

facilmente una quantità fondamentale, ovvero il portafoglio di minima varianza (MGV “minimum global variance portfolio”), quel portafoglio che giace sulla frontiera efficiente e che è caratterizzato dall’aver il minor grado di rischio possibile.

È infatti possibile ricavare tale portafoglio minimizzando la funzione di σ_p^2 :

$$V_{MGV} : \frac{\partial \sigma_p}{\partial \mu_p} = 0$$
$$\mu_p = \frac{b}{c}$$

Ottenendo così il valore minimo di μ_p che sostituito all’interno della funzione σ_p^2 fornisce la V_{MGV} :

$$V_{MGV} = \frac{c}{ca - b^2} \frac{b^2}{c^2} - \frac{2b}{ca - b^2} \frac{b}{c} + \frac{a}{ca - b^2}$$
$$V_{MGV} = \frac{1}{c}$$

Stando a questa formulazione, un investitore è così virtualmente in grado di costruire un portafoglio le cui caratteristiche di media e varianza dominano nettamente quelle dei singoli titoli.

1.3 Fallacie nei presupposti della frontiera efficiente

Quanto fin qui illustrato è derivato dai lavori di Markowitz (1952) e Merton (1972), i quali si basano su un insieme di ipotesi fortemente limitativo, originato nell’iniziale lavoro di Markowitz citato in precedenza.

I fondamenti su cui tale formulazione si basa non sono però così solidi se confrontati con la realtà in cui gli operatori e gli investitori si trovano ad agire.

La tanto distribuzione normale dei rendimenti, basilare per la costruzione di una frontiera efficiente, non è quasi mai presente per serie storiche reali di titoli azionari, in quanto quasi tutte presentano una assodata evidenza di leptocurtosi.

Inoltre il corredo di postulati circa le condizioni di funzionamento di mercato (assenza costi di transazione, titoli perfettamente divisibili, orizzonte uniperiodale per l'investimento) produce un'ulteriore lontananza tra mondo reale e condizioni teoriche ipotetiche.

Studi successivi, che si richiameranno nei prossimi capitoli, hanno inoltre dimostrato come la formulazione di Merton della frontiera efficiente risulta affetta da distorsione: essa non è così come ci appare graficamente in base all'equazione (1.2) – il ramo superiore di una parabola descritta in uno spazio rendimento atteso - deviazione standard – ma presenta una inclinazione meno accentuata, con un coefficiente di slope minore rispetto a quello precedentemente calcolato.

Questo comporta quindi che per un investitore che basa la costruzione del proprio portafoglio sul criterio media-varianza, il rischio che fronteggia per ottenere il rendimento desiderato differisce sensibilmente da quello calcolato tramite l'utilizzo della MPT.

Essa è eccessivamente ottimista circa la deviazione standard del portafoglio a parità di rendimento atteso, sottovalutando in maniera sostanziale il rischio vero connesso all'investimento.

Questo aspetto diventa ancor più fondamentale nelle ipotesi di investimenti a leva: il passo successivo nella teorizzazione di Markowitz circa la costruzione di un portafoglio consiste infatti, dato il tasso di rendimento di un titolo privo di rischio (*risk free*, generalmente un titolo di stato), nell'individuare il portafoglio ottimo, punto di tangenza tra la retta originata a partire dal valore di tale tasso e la frontiera efficiente, e comporre un investimento creato ad hoc per l'investitore in base alle sua avversione al rischio e alla sua funzione di utilità⁴ realizzando una combinazione tra investimento in titolo free risk e investimento nel portafoglio ottimo.

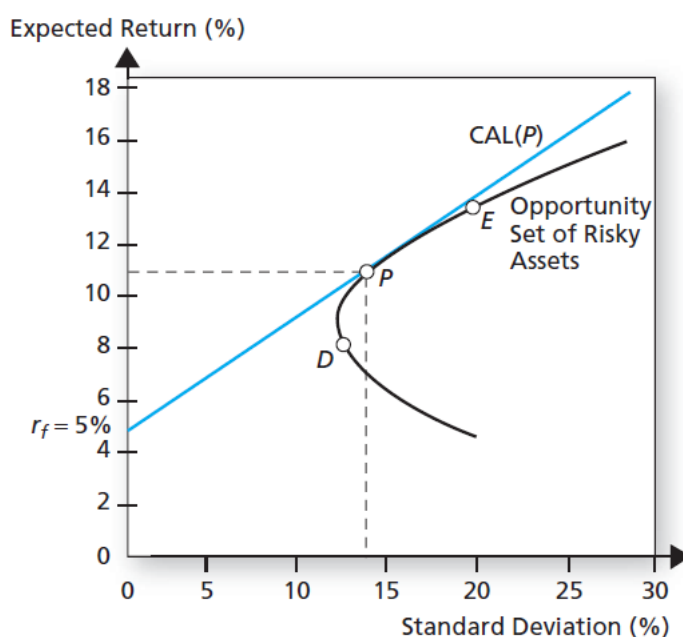
Tutti i portafogli così creati giacciono sulla retta tangente la frontiera efficiente avente per intercetta il rendimento del titolo privo di rischio e per coefficiente

⁴La funzione di utilità lega il rendimento atteso dell'investimento e il suo rischio al gradimento che l'investitore ottiene dalla sua realizzazione. È quindi evidente che tale funzione è basata su parametri soggettivi che cambiano da investitore ad investitore, come ad esempio il suo grado di avversione al rischio. Una spiegazione della funzione di utilità sarà data nella sezione Appendice A di questo lavoro

angolare l'indice di Sharpe.⁵

Risulta così evidente che in base alla diversa avversione al rischio degli investitori la composizione del portafoglio in cui si investe (mix tra risk free e portafoglio ottimo) cambia. Investitori con scarsa avversione al rischio possono infatti aprire una posizione corta sul titolo privo di rischio per finanziare l'acquisto di una maggior quantità di portafoglio ottimo.

Figura 1.1: Rappresentazione della frontiera efficiente con portafoglio ottimo e *risk free*



Fonte: Bodie Z., Kane A., Marcus A.J., *Investments*, McGraw-Hill Education, 2014

Tanto più si sfrutta l'effetto leva nell'investimento, tanto più la distorsione risulta affliggere la bontà dello stesso, aggiungendo un ulteriore elemento di rischiosità, originato dalla fiducia riposta in un modello almeno parzialmente sbagliato.

Come già si accennava in precedenza più studiosi si sono dedicati all'analisi di questo problema. I lavori più recenti che hanno fornito una spiegazione di questo fenomeno e un possibile metodo correttivo sono quelli di Siegel and Woodgate (2007), Bodnar e Schmid (2009) e di Bodnar e Bodnar (2010).

In questo ultimo lavoro gli autori derivano in modo analitico la distribuzio-

⁵L'indice di Sharpe è calcolato come rapporto tra differenza di rendimento tra titolo rischioso e titolo risk free e deviazione standard del titolo rischioso, ovvero $\frac{E(r)-r_f}{\sigma}$. Indica la remunerazione aggiuntiva che un investitore si attende (premio per il rischio atteso) per unità di rischio assunto.

ne statistica delle quantità coinvolte nel calcolo della frontiera alla Markowitz, evidenziandone così la distorsione: il valore atteso della frontiera efficiente campionaria differisce dal valore della frontiera efficiente teorica e di conseguenza la condizione di correttezza non è verificata.

Da qui gli autori riescono quindi a ricavare l'entità della distorsione e a calcolare un fattore correttivo da inserire nella formulazione della frontiera efficiente dei portafogli fornita dall'equazione (1.2).

Tuttavia per giungere a questo risultato gli stessi autori formulano ipotesi circa la funzione di distribuzione dei rendimenti azionari e il grado di dipendenza tra diversi titoli che necessitano di essere verificate.

Scopo di questo lavoro di tesi, come annunciato già nelle sezioni introduttive, è studiare il comportamento della frontiera efficiente à la Bodnar e Bodnar qualora vengano meno alcuni postulati, in particolare quelli di normalità distributiva dei rendimenti dei titoli e di indipendenza temporale dei rendimenti logaritmici degli stessi.

L'ipotesi di indipendenza statistica infatti non appare molto fondata, tenuto conto delle evidenze empiriche riguardanti la presenza di autocorrelazione nelle serie storiche dei rendimenti dei titoli azionari, specialmente nel caso di osservazioni giornaliere. Infatti la soluzione riguardante l'utilizzo di serie storiche con osservazioni mensili, seppur garantendo alcune proprietà a livello teorico, a livello pratico è penalizzante circa l'operatività: serie storiche con osservazioni giornaliere o settimanali forniscono informazioni maggiori e più accurate per l'operatore, soprattutto per quanto riguarda l'analisi della relazione tra diversi titoli, in quanto consentono di catturare meglio il rischio fronteggiato per tutta la durata dell'investimento.

Inoltre, come evidenziato da Corazza e Pizzi, il fattore correttivo applicato da Bodnar e Bodnar usando dati reali con osservazioni giornaliere (essi tuttavia non escludono che lo stesso possa avvenire anche per dati caratterizzati da osservazioni mensili) può produrre una deformazione inaccettabile nella frontiera, portando la varianza ad assumere valori negativi (si ricorda la sua impossibilità, considerando che la varianza è originata dalla sviluppo di un quadrato, e quindi sempre positiva per definizione).

Quanto fin qui affermato rende necessaria quindi la verifica della correttezza del metodo proposto da Bodnar e Bodnar che si provvederà ad illustrare nel capitolo terzo.

Si richiameranno nel corso del seguente capitolo invece anche lavori precedenti che hanno posto le basi su cui si è costituita successivamente la formulazione della distorsione e la conseguente correzione sopra menzionate.

Capitolo 2

INTERVALLO TEORICO PER LA FRONTIERA EFFICIENTE

2.1 Le ipotesi di Bodnar e Schmid

Questo capitolo tratterà la formulazione di Bodnar e Schmid circa la possibilità di costruire un intervallo di confidenza, basandosi sulla distribuzione della frontiera efficiente, in modo tale da individuare una possibile area nel piano media-varianza nella quale si è ragionevolmente sicuri giaccia la frontiera efficiente.

Tale approfondimento risulta necessario in quanto esso ha consentito di svolgere un primo passo verso la successiva formulazione di distorsione della frontiera, accennata a conclusione del capitolo precedente. Utilizzando come base di partenza i lavori precedenti di Jobson and Korkie (1980) relativamente alla distribuzione asintotica della frontiera, gli autori sviluppano una distribuzione della stessa, basata sulle seguenti ipotesi:

- i rendimenti dei titoli sono distribuiti normalmente,
- l'orizzonte temporale di riferimento per l'investimento è uniperiodale,
- il mercato non presenta costi di transazione e tutti i partecipanti sono considerati price taker,
- non vi è presenza di autocorrelazione tra le serie storiche dei rendimenti dei titoli azionari.

Sfruttando la definizione analitica di Merton, $\sigma_p^2 = \frac{c\mu_p^2 - 2b\mu_p + a}{ca - b^2}$ è possibile riorganizzare il tutto, riscrivendo l'equazione della frontiera efficiente utilizzando tre componenti fondamentali:

- R_{GMV} , ovvero il rendimento atteso del portafoglio a minima varianza
- V_{GMV} , ovvero la varianza del portafoglio a minima varianza
- s , ovvero il coefficiente di slope della frontiera efficiente, con $s = \boldsymbol{\mu}' \mathbf{R} \boldsymbol{\mu}$ e con

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

L'equazione della frontiera efficiente dei portafogli diventa:

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

E sostituendo successivamente alle quantità ignote i rispettivi stimatori:

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

Nel succitato articolo gli autori utilizzano questa riformulazione della frontiera efficiente in quanto consente di evidenziare tutte e tre le quantità coinvolte nel calcolo della stessa, per le quali è possibile formulare intervalli di confidenza teorici.

Si è inoltre scelto di seguire quanto svolto nell'articolo di Bodnar e Schmid (2009) e di rinominare con g_p la frontiera efficiente teorica (e quindi calcolata sulla popolazione) e con g_s quella stimata (ovvero calcolata sul campione).

Grazie ai lavori precedenti di Kan e Smith (2008) e Bodnar e Schmid (2009), che derivano indipendentemente la distribuzione congiunta di \hat{R}_{GMV} , \hat{V}_{GMV} e \hat{s} , è possibile calcolare delle statistiche test (T_V , T_R e T_S) la cui distribuzione sotto ipotesi preliminari è nota.

Risulta quindi utile richiamare i lavori di Kan e Smith (2008) e di Bodnar e Schmid (2009).

2.2 La definizione della regione di confidenza di Bodnar e Schmid

Per giungere ad una rappresentazione efficace di tale area nel piano media-varianza, gli autori come già esposto derivano le esatte distribuzioni delle quantità coinvolte nel calcolo della frontiera, assieme alla distribuzione congiunta delle stesse.

Il Lemma sotto richiamato espone proprio le funzioni di densità di probabilità degli elementi succitati.

Lemma 1:

Dati k titoli, e n osservazioni dei rendimenti mensili di ciascun titolo, siano $\mathbf{X}_1, \dots, \mathbf{X}_n$ vettori aleatori indipendenti, e sia $\mathbf{X}_i \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Si assume inoltre che la matrice di varianza e covarianza $\boldsymbol{\Sigma}$ sia definita positiva, e che il numero di osservazioni n sia maggiore del numero dei titoli considerati k .

Ne deriva che:

- a) \hat{V}_{GMV} è indipendente da \hat{R}_{GMV} e da \hat{s}
- b) $(n-1)\hat{V}_{GMV}$ e $V_{GMV} \sim \chi_{n-k}^2$
- c) $\frac{n(n-k+1)}{(n-1)(k-1)}\hat{s} \sim F_{k-1, n-k+1, ns}$
- d) $\hat{R}_{GMV}|\hat{s} = y \sim N(R_{GMV}, \frac{1+\frac{n-1}{n}y}{n}V_{GMV})$
- e) la funzione di densità congiunta è data da:

$$f_{\hat{R}_{GMV}, \hat{V}_{GMV}, \hat{s}}(x, y, z) = \frac{n(n-k+1)}{(k-1)V_{GMV}} f_{\chi_{n-k}^2} \left(\frac{n-1}{V_{GMV}} z \right) f_{N(R_{GMV}, \frac{1+\frac{n-1}{n}y}{n}V_{GMV})}(x)$$

$$f_{F_{k-1, n-k+1, ns}} \left(\frac{n(n-k+1)}{(n-1)(k-1)} y \right)$$

Questa distribuzione congiunta è stata derivata indipendentemente da Bodnar e Schmid (2009) e da Kan e Smith (2008).

Da questa distribuzione gli autori hanno potuto costruire le statistiche test sopramenzionate che hanno prodotto un notevole miglioramento rispetto all'approccio passato: le analisi precedenti infatti, vista l'impossibilità di costruzione di test statistici si limitavano all'osservazione del comportamento della frontiera nei casi in cui i suoi componenti raggiungessero valori asintotici, mentre conoscendo

la distribuzione delle quantità coinvolte nel calcolo della stessa è ora possibile analizzare il comportamento e l'influenza di ogni fattore anche per numerosità campionarie finite. Ciò potrebbe avere rilevanza pratica perché, a differenza di una metodologia basata su analisi asintotiche, consentirebbe di individuare l'intervallo di confidenza anche in caso di frequenti riallocazioni dei pesi del portafoglio (ipotesi molto reale considerando il mondo del risparmio gestito in modo attivo).

Come si illustrerà più avanti questo metodo ha permesso di individuare una regione di confidenza per la frontiera efficiente dei portafogli, consentendo un miglioramento rispetto all'approccio di tipo asintotico formulato da Jobson (1991). Un test congiunto su tutte e tre le quantità coinvolte permetterebbe infatti di capire quanto ampio sia il margine di errore connesso alla stima della frontiera efficiente, ovvero offrirebbe una rappresentazione grafica della dispersione della frontiera efficiente campionaria.

Più semplicemente questo approccio può portare ad avere una migliore idea di quanto ci si possa sbagliare nello stimare la frontiera efficiente, e quindi di riflesso quanto possano essere differenti da quanto stimato quantità quali rendimento atteso e rischio, elementi fondamentali per operatori ed investitori, meno interessati agli aspetti di natura statistica ma decisamente volti ad un miglioramento pratico nella gestione del risparmio.

Si procede quindi all'illustrazione delle statistiche coinvolte nel calcolo dell'intervallo di confidenza. Si presenterà in modo esaustivo il caso in cui si voglia eseguire un test congiunto su tutte e tre le quantità coinvolte, considerato di maggior interesse, e si illustrerà brevemente solo per conoscenza i casi in cui si provveda a calcolare un intervallo per ogni singola variabile¹.

Da quanto precedentemente illustrato circa la distribuzione delle quantità coinvolte nella stima della frontiera efficiente, gli autori ricavano una serie di proposizioni e di interessanti risultati.

La prima quantità ad essere esaminata è la \hat{V}_{GMV} , ovvero la varianza stimata del portafoglio a minima varianza. Il test eseguito su tale valore ha la seguente struttura di ipotesi:

¹Per approfondimenti si rimanda a Bodnar e Schmid (2009)

$$\begin{cases} H_0 : V_{GMV} = V_0 \\ H_1 : V_{GMV} = V_1 \neq V_0 \end{cases}$$

e per questo sistema di ipotesi è calcolata la seguente statistica test:

$$T_V = (n-1) \frac{\hat{V}_{GMV}}{V_0}$$

Sostanzialmente quindi si vuole verificare se la varianza del GMVportfolio sia uguale o meno ad un dato valore, chiamato V_0 . Questa è l'eventualità contemplata dall'ipotesi nulla, mentre l'ipotesi alternativa presuppone che la quantità considerata sia invece eguale ad un valore V_1 , diverso dal precedente V_0 .

Bodnar e Schmid hanno dimostrato che sotto le assunzioni precedenti la quantità $\frac{V_0 T_V}{V_1}$ si distribuisce come una χ^2 con $n-k$ gradi di libertà. T_V si distribuisce quindi come una χ^2 con $n-k$ ovvero $T_V \sim \chi_{n-k}^2$.

È quindi possibile calcolare un intervallo di confidenza bilaterale con livello di confidenza $(1-\alpha)$ così definito:

$$\left[\frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k; 1-\alpha/2}^2}, \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k; \alpha/2}^2} \right]$$

nel quale la quantità $\chi_{n-k; \alpha/2}^2$ indica il quantile $\alpha/2$ della distribuzione χ^2 con $n-k$ gradi i libertà.

Allo stesso modo è possibile definire un intervallo di confidenza per il rendimento atteso del portafoglio a minima varianza, utilizzando un sistema di ipotesi del tutto simile a quello precedente:

$$\begin{cases} H_0 : R_{GMV} = R_0 \\ H_1 : R_{GMV} = R_1 \neq R_0 \end{cases}$$

in cui si testa l'uguaglianza della quantità R_{GMV} ad un valore dato R_0 .

La statistica test è definita come segue:

$$T_R = \sqrt{n\mathbf{1}'\Sigma^{-1}\mathbf{1}} \frac{\sqrt{n-k}}{\sqrt{n-1}} \frac{\hat{R}_{GMV} - R_0}{\sqrt{1 + \frac{n}{n-1} \hat{\boldsymbol{\mu}}' \mathbf{R} \hat{\boldsymbol{\mu}}}}$$

La sua distribuzione è derivata come segue dalla seguente proposizione:

Proposizione 1

Si assume che $\mathbf{X}_1 \dots \mathbf{X}_n$ siano vettori indipendenti di variabili casuali, che $\mathbf{X}_i \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, e che la matrice $\boldsymbol{\Sigma}$ sia definita positiva.

Assumendo inoltre che $n > k$, ne deriva che:

a) La densità di T_R è data da:

$$F_{TR}(x) = \frac{n(n-k+1)}{(k-1)(n-1)} \int_0^\infty f_{t_{n-k, \delta(y)}}(x) f_{F_{k-1, n-k+1, ns}}\left(\frac{n(n-k+1)}{(k-1)(n-1)}y\right) dy$$

con $\delta(y) = \sqrt{n}\lambda_1 / \sqrt{1 + \frac{n}{n-1}y}$, $\lambda_1 = (R_1 - R_0) / \sqrt{V_1}$ e con $s = \boldsymbol{\mu}' \mathbf{R} \boldsymbol{\mu}$: δ e ns indicano i parametri di non centralità per le distribuzioni t di Student e F di Fisher rispettivamente.

b) sotto l'ipotesi nulla consegue che T_R è indipendente da \hat{s} e si distribuisce come una t di Student con $n-k$ gradi di libertà, ovvero $T_R \sim t_{n-k}$.

La prova della distribuzione della statistica T_R è ricavabile dal precedente Lemma 1, in quanto è possibile ricavare che $T_R | \hat{s} = y \sim t_{n-k, \delta(y)}$, e per altri enunciati nello stesso Lemma 1 consegue che la distribuzione di $T_R | \hat{s}$ risulta essere una distribuzione t di Student centrale e indipendente da y , quindi equivalente alla distribuzione non condizionale della statistica T_R .

L'intervallo di confidenza bilaterale per la quantità R_{GMV} è così definito:

$$\left[\hat{R}_{GMV} - \frac{\sqrt{n-1}}{\sqrt{n-k}} \sqrt{1 + \frac{n}{n-1} \hat{s}} \frac{t_{n-k; 1-\alpha/2}}{\sqrt{n \mathbf{1}' \boldsymbol{\Sigma} \mathbf{1}}}; \hat{R}_{GMV} + \frac{\sqrt{n-1}}{\sqrt{n-k}} \sqrt{1 + \frac{n}{n-1} \hat{s}} \frac{t_{n-k; 1-\alpha/2}}{\sqrt{n \mathbf{1}' \boldsymbol{\Sigma} \mathbf{1}}} \right]$$

in cui il valore $t_{p; \beta}$ esprime il quantile di ordine β della distribuzione centrale t di Student con p gradi di libertà.

Da notare soprattutto come nella funzione di densità di T_R la dipendenza in termini di λ_1 e s possa essere totalmente ricompresa nella sola dipendenza da $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$.

Se λ_1 misura la non-centralità della funzione di distribuzione, il parametro s assume un valore più importante: esso è la vera misura del coefficiente di slope della parabola che descrive la frontiera efficiente dei portafogli.

Da ultimo si analizza la statistica test per la quantità s , il parametro di pendenza della parabola.

Il sistema di ipotesi ha la medesima forma dei precedenti:

$$\begin{cases} H_0 : s = s_0 \\ H_1 : s = s_1 \neq s_0 \end{cases}$$

La statistica test T_S ha invece la seguente struttura:

$$T_S = \frac{n(n-k+1)}{(k-1)(n-1)} \hat{\boldsymbol{\mu}}' \hat{\mathbf{R}} \hat{\boldsymbol{\mu}}$$

Ne deriva che la statistica test ha due distribuzioni differenti sotto le due diverse ipotesi, nel caso più specifico: $T_S \sim F_{k-1, n-k+1; ns_0}$ sotto l'ipotesi H_0 , e $T_S \sim F_{k-1, n-k+1; ns_1}$ sotto l'ipotesi H_1 .

Ciò è equivalente al domandarsi se il parametro di non centralità sia s_0 piuttosto che s_1 .

Si fa notare come quest'ultimo aspetto riguardante il parametro di non centralità s_0 sia già stato trattato in letteratura in lavori precedenti – fu oggetto di studio da parte di Lam (1987) – e come il metodo utilizzato per la creazione dell'intervallo si basi sulla ricerca delle soluzioni dell'equazione: $F_{p,q,\lambda_\beta}(x) = \beta$ per qualsiasi realizzazione della variabile casuale considerata, distribuita come una F di Fisher con p e q gradi di libertà. Lam (1987) propone anche una procedura ricorsiva per il calcolo, non inserita in questo lavoro in quanto esula dallo scopo di questa tesi.

Si può quindi procedere analizzando il caso più interessante: la creazione di un intervallo di confidenza che esamini il comportamento congiunto di tutti e tre gli stimatori coinvolti nel calcolo della frontiera efficiente.

In questo modo è possibile individuare l'area in cui giace la frontiera efficiente reale (non campionaria) e avere quindi uno strumento capace di misurare il potenziale errore di stima commesso – legato in modo ineluttabile alla procedura di inferenza delle quantità ignote necessarie per il calcolo della frontiera – nell'utilizzo della frontiera efficiente campionaria.

Per questo specifico caso il sistema di ipotesi deve essere riformulato affinché tenga conto contemporaneamente di tutte e tre le quantità in gioco:

$$\begin{cases} H_0 : V_{GMV} = V_0; R_{GMV} = R_0; s = s_0; \\ H_1 : V_{GMV} = V_1 \neq V_0; R_{GMV} = R_1 \neq R_0; s = s_1 \neq s_0; \end{cases}$$

La statistica test utilizzata per saggiare questo sistema di ipotesi è $T = (T_R^*, T_V, T_S)'$ in quanto T_V è indipendente da T_R^* e da T_S , con $T_R^* = \sqrt{n} \frac{\hat{R}_{GMV} - R_0}{\sqrt{V_0} \sqrt{1 + \frac{n}{n-1} \hat{\mu}' R \hat{\mu}}}$. L'uso di T_R^* al posto di T_R è motivato proprio dalla sua indipendenza con la statistica test T_V , condizione non soddisfatta invece dalla statistica precedentemente utilizzata. Dal Lemma 1. precedentemente trattato si può ricavare la distribuzione di $T_R^* | \hat{s}$ sotto l'ipotesi nulla, ovvero: $T_R^* | \hat{s} = y \sim N(0, 1)$.

Si rendono necessarie a questo punto delle ulteriori considerazioni, enunciate dagli autori nella seguente proposizione:

Proposizione 2

Siano X_1, \dots, X_n vettori aleatori di variabili casuali e sia $X_i \sim N(\mu, \Sigma)$ per ogni $i=1, \dots, n$. Si assume inoltre che Σ sia definita positiva e che $n > k$.

Allora:

a) la funzione di densità di T è data da:

$$f_T(x, y, z) = \eta^{-1} f_{\chi_{n-k}^2} \left(\frac{z}{\eta} \right) f_{N(\sqrt{\eta} \tilde{\delta}(y), \eta)}(x) f_{F_{k-1, n-k+1, ns}}(y) \quad (2.1)$$

con $\tilde{\delta}(y) = \sqrt{n} \lambda_1 / \sqrt{1 + \frac{k-1}{n-k+1} y}$ e con $\eta = V_1 / V_0$. λ_1 ed s sono le medesime quantità già descritte.

b) vista l'indipendenza tra le variabili che compongono T , allora sotto l'ipotesi nulla la densità di T può essere riscritta come:

$$f_T(x, y, z) = f_{\chi_{n-k}^2}(z) f_{N(0,1)}(x) f_{F_{k-1, n-k+1, ns_0}}(y)$$

Grazie ai postulati della Proposizione 2. è possibile costruire l'insieme di caratteristiche per tracciare l'intervallo di confidenza. Esso si basa sull'indipendenza delle statistiche T_R^* , T_V e T_S e per un livello di confidenza $(1 - \alpha)$ comprende tutti i punti (R_{GMV}, V_{GMV}, s) che soddisfano le seguenti condizioni:

$$(R_{GMV} - \hat{R}_{GMV})^2 \leq z_{1-\alpha^*/2}^2 \left(\frac{1}{n} + \frac{\hat{s}}{n-1} \right) V_{GMV}, \quad (2.2)$$

$$V_{GMV} \in \left[\frac{(n-1)V_{GMV}}{\chi_{n-k; 1-\alpha^*/2}^2}; \frac{(n-1)V_{GMV}}{\chi_{n-k; \alpha^*/2}^2} \right], \quad (2.3)$$

$$\hat{s}_{\alpha^*/2} \leq s \leq \hat{s}_{1-\alpha^*/2} \quad (2.4)$$

dove $F_{k-1, n-k+1, n\hat{s}_\beta}(T_S) = \beta$ e $\alpha^* = 1 - \sqrt[3]{1-\alpha}$

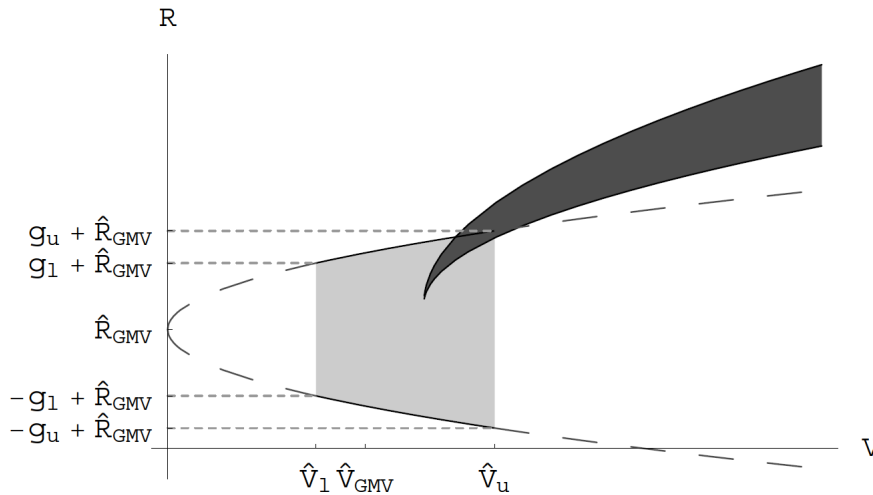
Se si rinomina \mathcal{A} il set di caratteristiche appena enunciate, una procedura per monitorare l'intera frontiera efficiente si basa sulla costruzione di un set di equazioni \mathcal{B} definite nello spazio media-varianza. \mathcal{B} ha la seguente struttura:

$$\mathcal{B} = \left\{ (R - R_{GMV})^2 = s(V - V_{GMV}) : (R_{GMV}, V_{GMV}, s) \in \mathcal{A} \right\}$$

Analizzando la struttura geometrica di \mathcal{B} è possibile considerare \mathcal{B} in due modi distinti: o come set di parabole oppure come set collocato nello spazio media-varianza. Il set \mathcal{B} interpretato in quest'ultimo modo è stato rinominato \mathcal{B}^* . \mathcal{B} risulta quindi essere composto da tutte le parabole del tipo: $(R - R_{GMV}^*)^2 = s^*(V - V_{GMV}^*)$ con $R^* \in \mathbb{R}$ e con $V^*, s^* > 0$.

La seguente figura illustra questo fatto:

Figura 2.1: Aree della regione di confidenza in base alle differenti condizioni



Fonte: Bodnar T. e Schmid W., *Econometrical Analysis of the Sample Efficient Frontier*, *The European Journal of Finance*, 2009

Il grafico presenta due aree distinte, rappresentanti situazioni diverse: la parte più chiara rappresenta tutti i punti nel piano media-varianza per cui le quantità R_{GMV} e V_{GMV} rispettano i vincoli posti dalle prime due asserzioni contenute nel set \mathcal{A} . La parte più scura invece è determinata come intervallo sul valore di slope della frontiera efficiente (terza condizione del set \mathcal{A}) per una frontiera con

vertice giacente nello spazio individuato dall'area grigio chiaro (ovvero per un portafoglio di minima varianza che soddisfa le prime due condizioni del del set \mathcal{A}).

Appare così scontato che la regione di confidenza totale per la frontiera efficiente è data dalla combinazione delle due aree: la si ottiene tracciando un intervallo di confidenza per il valore s per ogni punto che soddisfa le prime due condizioni di \mathcal{A} .

Per una trattazione analitica si prosegue come di seguito indicato.

Si definiscano le quantità:

$$g_l = z_{1-\alpha^*/2} \sqrt{1 + \frac{n\hat{s}}{n-1} \frac{\sqrt{n-1} \sqrt{\hat{V}_{GMV}}}{\sqrt{n\chi_{n-k;1-\alpha^*/2}^2}}}, \quad g_u = z_{1-\alpha^*/2} \sqrt{1 + \frac{n\hat{s}}{n-1} \frac{\sqrt{n-1} \sqrt{\hat{V}_{GMV}}}{\sqrt{n\chi_{n-k;\alpha^*/2}^2}}} \quad (2.5)$$

Sapendo che la costruzione della regione di confidenza per la frontiera avviene come sopra indicato, gli autori dimostrano che la medesima regione è individuata da cinque parabole.

Per necessità di concisione si riporta solo il risultato finale.

Analiticamente quindi la regione di confidenza è così costruita:

$$\left\{ \begin{array}{ll} V \leq \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;1-\alpha^*/2}^2} & \text{con } R \in \mathbb{R} \\ V \leq z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} (R - \hat{R}_{GMV})^2 & \text{con } R \in \mathbb{I}_1^l \\ V \geq \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha^*/2}^2} + \hat{s}_{\alpha^*/2}^{-1} (R - \hat{R}_{GMV} + g_u)^2 & \text{con } R \in \mathbb{I}_2^l \\ V \leq \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;1-\alpha^*/2}^2} \hat{s}_{1-\alpha^*/2}^{-1} (R - \hat{R}_{GMV} - g_l)^2 & \text{con } R \in \mathbb{I}_1^u \\ V \leq \frac{z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1}}{1 + z_{1-\alpha^*/2}^{-2} \hat{s}_{1-\alpha^*/2} \frac{n(n-1)}{n(1+\hat{s})-1}} (R - \hat{R}_{GMV})^2 & \text{con } R \in \mathbb{I}_2^u \\ V \leq \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha^*/2}^2 + \hat{s}_{1-\alpha^*/2}^{-1}} (R - \hat{R}_{GMV} - g_u)^2 & \text{con } R \in \mathbb{I}_3^u \end{array} \right.$$

$$\text{con } \mathbb{I}_1^u = [\hat{R}_{GMV} + g_l; \hat{R}_{GMV} + g_l/t], \quad \mathbb{I}_2^u = [\hat{R}_{GMV} + g_l/t; \hat{R}_{GMV} + g_u/t], \quad \mathbb{I}_3^u = (\hat{R}_{GMV} +$$

$$g_u; \infty), \mathbb{I}_1^l = [\hat{R}_{GMV} - g_u; \hat{R}_{GMV} - g_l], \mathbb{I}_2^l = [\hat{R}_{GMV} - g_u; +\infty)$$

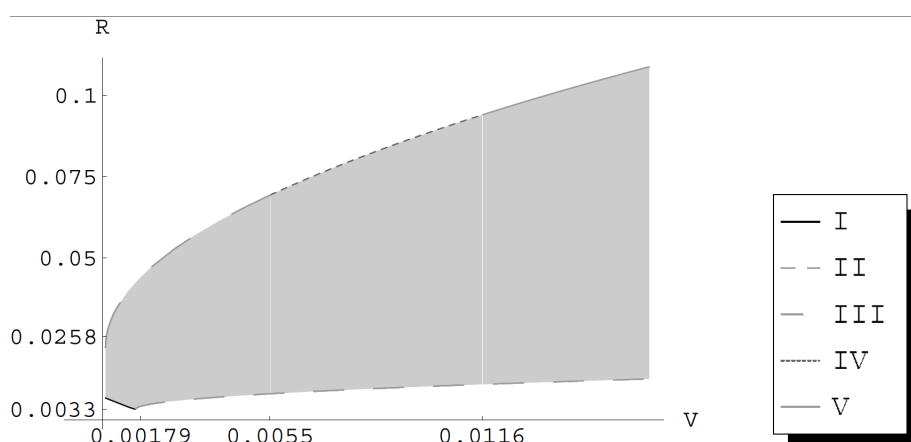
$$e \text{ con } t = \frac{1}{1 + z_{1-\alpha^*/2}^2 \hat{s}^{1-\alpha^*/2} \frac{n(n-1)}{n(1+\hat{s})-1}} \leq 1$$

La trattazione teorica di questa regione di confidenza è riportata in Appendice B.

Si può però affermare che la spiegazione trae le sue origini dall'approccio grafico esposto in precedenza.

Nell'esposizione dell'argomento gli autori proseguono replicando sul piano operativo quanto illustrato su quello teorico. Per il loro test utilizzano però dati mensili dei rendimenti di indici di mercato azionari, più specificamente di paesi quali: Canada, UK, USA, Germania e Svizzera, per il periodo compreso tra luglio 1994 e giugno 1999. La regione di confidenza estrapolata a partire da tali dati, per la quale sono stati utilizzati i seguenti parametri: $k = 5$, $n = 60$, $\alpha = 0.15$, $\hat{R}_{GMV} = 0.0145664$, $\hat{V}_{GMV} = 0.0010165$, e $\hat{s} = 0.227957$, è illustrata in figura:

Figura 2.2: Costruzione della regione di confidenza



Fonte: Bodnar T. e Schmid W., *Econometrical Analysis of the Sample Efficient Frontier*, *The European Journal of Finance*, 2009

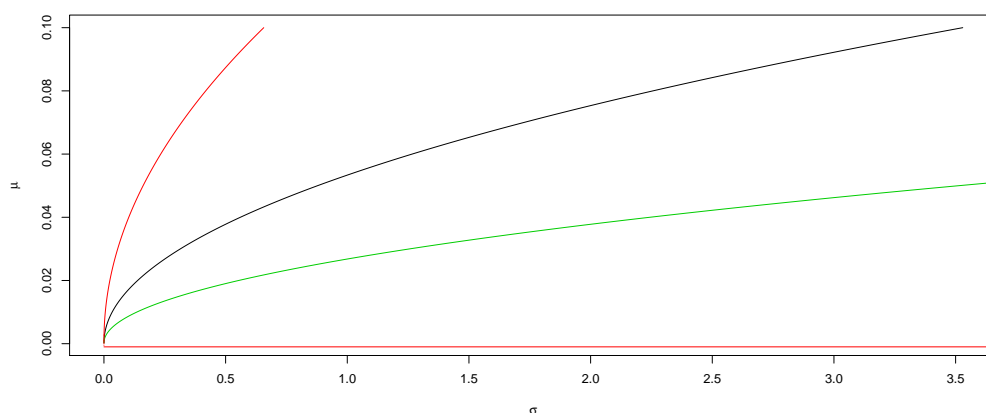
È interessante notare come viene sottolineata la scelta di un valore di α relativamente basso: per valori più elevati non è detto che esista un limite inferiore e in tal caso viene suggerito da parte degli autori di porre come valore limite l'asse delle ascisse.

Concludendo questo capitolo è necessaria una precisazione per quanto si tratterà più avanti: una composizione così fatta per la regione di confidenza porta all'individuazione di un'area troppo vasta (la quasi totalità del primo quadrante

del piano). Se questo aspetto può anche non essere di grande rilevanza per coloro che sono interessati alla sola "misurazione" del potenziale errore compiuto nel processo di stima, per i fini di questo lavoro di tesi non risulta però applicabile: l'ampiezza della regione non permette di verificare se la correzione per la distorsione della frontiera proposta (che si illustrerà nel prossimo capitolo) sia statisticamente differente dalla frontiera efficiente.

Il grafico riportato di seguito fornisce una rappresentazione di questo problema: sono stati selezionati i rendimenti logaritmici giornalieri per quattro titoli (DuPont, ExxonMobil, General Electric e Goldman Sachs, per il periodo 01/01/2012 - 31/08/2017), sono state calcolate le due frontiere (in nero quella campionaria e in verde quella corretta) e l'intervallo di confidenza per la frontiera "à la Markowitz" (riportato in rosso).

Figura 2.3: Intervallo di confidenza teorico per un portafoglio costituito da quattro titoli, utilizzando rendimenti logaritmici giornalieri



periodo 01/01/2012 - 31/08/2017

Per i motivi sopra menzionati, risulta così evidente l'inefficacia di un simile intervallo di confidenza dal punto di vista operativo.

Per risolvere tale problema si è quindi pensato di utilizzare un metodo bootstrap per la creazione dell'intervallo di interesse.

Si tiene quindi a far notare che, sebbene la costruzione di un intervallo teorico per la frontiera efficiente sia di notevole rilevanza, esso perde di efficacia qualora

risultati troppo vasto: la sua utilità in questi casi è infatti nulla, poiché non aggiunge alcuna informazione all'insieme conoscitivo di cui l'investitore è già in possesso.

Capitolo 3

LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

3.1 La Distorsione della frontiera efficiente

Come si ha avuto modo di spiegare in precedenza, il presente lavoro di tesi ha come obiettivo la verifica delle ipotesi poste da Bodnar e Bodnar (2010) per derivare la distorsione della frontiera efficiente dei portafogli.

Risulta quindi indispensabile spiegare dettagliatamente la distorsione calcolata dai due autori, e le criticità che conseguono a questo approccio, che come molte formulazioni teoriche non è esente da critiche e perplessità.

Nell'articolo "On the unbiased estimator of efficient frontier" – Bodnar e Bodnar (2010) – gli autori riprendono l'aggiustamento della formulazione analitica già presentato nel capitolo precedente e sviluppano attraverso una serie di postulati, assunzioni e successive conseguenze la distribuzione della frontiera efficiente stimata.

Innanzitutto si ricorda che si ha a che fare con quantità ignote che necessitano di essere stimate: questo come esplicitato nel capitolo che precede porta ad una differenza tra frontiera teorica (o calcolata sulla popolazione) e frontiera campionaria (ovvero stimata a partire dal campione).

In questo capitolo si tratterà esclusivamente della frontiera efficiente campionaria,

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

pertanto si richiama la definizione analitica di partenza:

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

Il problema principale riguardante l'utilizzo della frontiera efficiente campionaria è l'eccessivo ottimismo che la caratterizza: Siegel e Woodgate (2007) infatti dimostrano come la frontiera campionaria sovrastimi costantemente quella teorica.

Conoscendo l'esatta distribuzione delle quantità coinvolte nel calcolo della frontiera campionaria – \hat{V}_{GMV} , \hat{R}_{GMV} , \hat{s} – illustrata al Lemma 1 del capitolo precedente, è possibile evidenziare la distorsione presente nella stessa.

Si necessita del seguente lemma contenuto nel lavoro di Bodnar e Bodnar (2010):

Lemma 2: siano $X_1 \dots X_n$ vettori di variabili casuali dipendenti e sia $X_i \sim N_k(\boldsymbol{\mu}; \boldsymbol{\Sigma})$ per $i = 1 \dots n$. Si assume quindi che $\boldsymbol{\Sigma}$ sia definita positiva, che il numero di titoli considerati sia $k > 2$ e quello di osservazioni per ogni titolo sia $n > k$.

Ne consegue che:

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

La prova di queste identità deriva direttamente dalla distribuzione delle suddette quantità e dalle proprietà della distribuzione F non centrale.

Ne consegue che:

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

e applicando il Lemma 1 ne consegue che:

$$\begin{aligned} Var(\hat{R}_{GMV}) &= Var(E(\hat{R}_{GMV}|\hat{s})) + E(Var(\hat{R}_{GMV}|\hat{s})) \\ &= Var(R_{GMV}) + \left(\frac{1}{n} + \frac{1}{n-1}E(\hat{s})\right)V_{GMV} \\ &= \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s\right)V_{GMV} \end{aligned} \tag{3.1}$$

Gli altri enunciati del Lemma 1 infine contribuiscono a condurre al risultato sopra illustrato, specialmente sfruttando le proprietà della distribuzione χ^2 .

Ciò che è rilevante sottolineare – al di là della dimostrazione tecnica – è che i risultati del Lemma 2 portano ad affermare che $E(g_s(R, V)) \neq 0$ per $(R, V) \in \mathbb{R} \times \mathbb{R}_+$

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

in modo che $E(g_p(R, V)) = 0$.

Non resta quindi che calcolare il valore atteso della frontiera efficiente campionaria, in modo da misurarne la distorsione.

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

L'ultima equivalenza deriva direttamente dal Lemma 1 a) – riguardante l'indipendenza tra le diverse quantità – e applicando i risultati del Lemma 2 si ottiene:

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

e di conseguenza:

$$\begin{aligned} E(g_s(R, V)) &= g_p(R, V) - \left(\frac{k}{n-k-1}s + \frac{(n-1)(k-1)}{n(n-k-1)} \right) V \\ &\quad + \left(\frac{2}{n-k-1}s + \frac{(n-k)(k-1) + n-2}{n(n-k-1)} \right) V_{GMV} \end{aligned} \quad (3.2)$$

Risulta così evidente che le quantità coinvolte nel calcolo di $g_s(R, V)$ producono una distorsione considerevole, portando il valore atteso di quello che è possibile definire come stimatore della frontiera efficiente a differire sensibilmente dal suo valore teorico calcolato a partire dalla popolazione e non dal campione.

Infatti $E(g_s(R, V))$ non è pari a $g_p(R, V)$ (ipotesi di non distorsione di g_s) ma alla somma tra tale quantità e un'altra, data dalla combinazione di valori quali s , V , V_{GMV} , k e n .

L'ultima equazione porta a concludere che la frontiera campionaria è uno stimatore corretto solo per il portafoglio con varianza $V = \gamma V_{GMV}$. Infatti conferendo a γ il seguente valore:

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

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

se la varianza del portafoglio considerato è maggiore del valore γV_{GMV} allora la frontiera efficiente campionaria sovrastima quella teorica, nel caso di varianza minore della quantità indicata si è invece in presenza di un caso di sottostima.

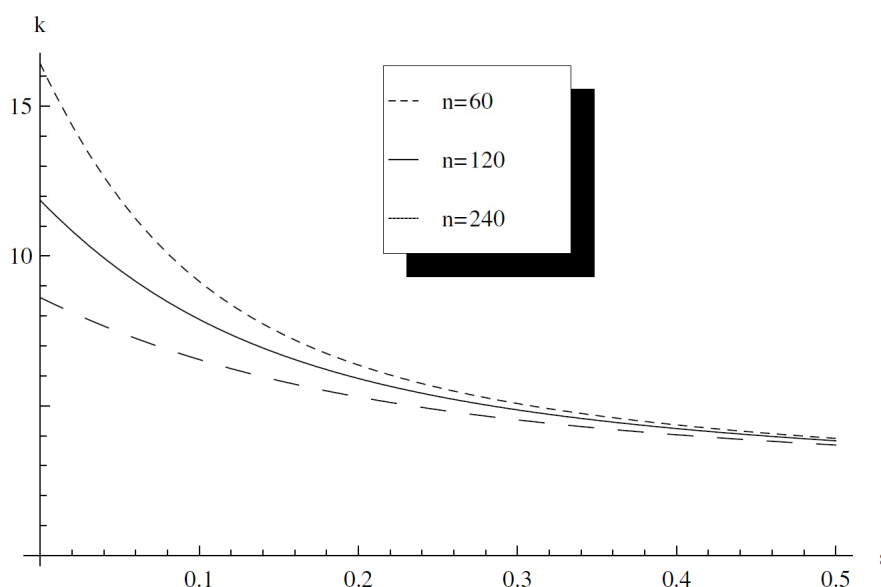
Vi è un caso che si ritiene importante segnalare, ovvero quello in cui è rispettata la seguente condizione:

$$1 + \frac{-ns \sqrt{(ns+2)^2 + 4n - 12}}{2} < k < n$$

se tale condizione risulta rispettata allora il valore γ risulterà essere minore di 1, e la varianza di ogni portafoglio che giace sulla frontiera campionaria sarà sempre maggiore di γV_{GMV} , e di conseguenza sovrastimerà costantemente la frontiera efficiente reale.

Si riporta di seguito un utile grafico illustrante il rapporto tra k e s circa il limite inferiore per il quale il portafoglio stimato presenta un eccessivo ottimismo rispetto a quello calcolato sulla popolazione.

Figura 3.1: Limite inferiore di k asset da includere nel portafoglio per un dato valore s per il quale il portafoglio campionario presenta una sovrastima del portafoglio teorico



Fonte: Bodnar T. e Bodnar O., On the Unbiased Estimator of the Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 2010

Si può notare come, per valori minimi di s , il numero di asset da includere diffe-

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

risce molto a seconda della numerosità campionaria: più questa aumenta, meno asset sono necessari affinché la frontiera efficiente campionaria risulti affetta da distorsione, sovrastimando la frontiera reale g_p . Per valori sempre crescenti di s invece il limite tende asintoticamente ad un valore k prossimo a 4.

Da quanto illustrato si evince che la quantità k risulta essere una funzione decrescente in s . Si nota inoltre che per ogni portafoglio con asset $k < 10$ la sovrastima è comunque presente per tutta la frontiera, se il valore s risulta essere maggiore di 0.2.

Come si ha avuto modo di affermare nei capitoli precedenti, questo fatto è perfettamente in linea con le scoperte di Siegel e Woodgate (2007) riguardanti l'eccessivo ottimismo che caratterizza la frontiera efficiente dei portafogli e la reale locazione della stessa nello spazio media-varianza.

A questo punto si può procedere illustrando il metodo che consente di giungere alla formulazione di uno stimatore corretto per la frontiera efficiente dei portafogli.

Bodnar e Bodnar (2010) espongono al riguardo un teorema, e le prove che ne determinano la validità:

Teorema:

Siano $X_1 \dots X_n$ vettori aleatori indipendenti di variabili casuali, e sia $X_i \sim \mathcal{N}_k(\mu, \Sigma)$. Sia inoltre Σ definita positiva e siano poste le seguenti condizioni su n e k : $k > 2$ e $n > k$.

Allora si può affermare che:

$$g_u(R, V) = (R - \hat{R}_{GMV})^2 - \frac{(n-2)(n-1)}{n(n-k)(n-k-1)} \hat{V}_{GMV} - \left(\frac{n-k-1}{n-1} \hat{s} - \frac{k-1}{n} \right) \left(V - \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)} \hat{V}_{GMV} \right). \quad (3.3)$$

in cui la quantità $g_u(R, V)$ rappresenta la frontiera efficiente corretta, ovvero modificata affinché non sia più presente la distorsione provocata dall'utilizzo nel procedimento di calcolo delle quantità stimate \hat{V}_{GMV} e \hat{s} .

La prova di questo teorema è data di seguito, e si fonda sulla dimostrazione che il valore atteso di $g_u(R, V)$ è pari a 0 per ogni punto $(R, V) \in \mathbb{R} \times \mathbb{R}_+$ tale che

$$g_p(R, V) = 0.$$

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

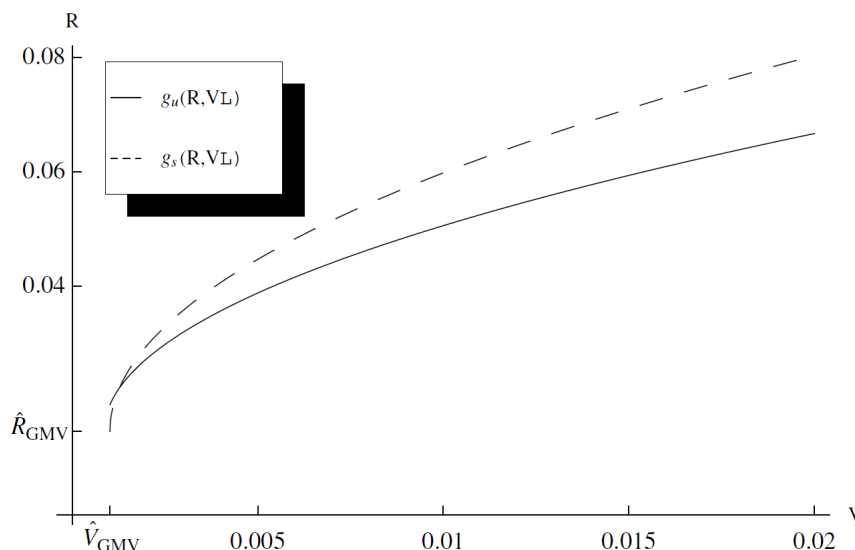
Applicando gli enunciati del Lemma 1 e del Lemma 2 si può giungere ad affermare che:

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

Un esempio pratico può aiutare nell'esposizione.

Come illustrato in Bodnar e Bodnar (2010), facendo riferimento a studi precedenti di Bodnar e Schmid (2009), è possibile tracciare in un grafico media varianza i due diversi tipi di frontiera efficiente: quella campionaria (indicata con $g_s(R, V)$) e quella corretta (indicata con $g_u(R, V)$).

Figura 3.2: Confronto tra le frontiere $g_s(R, V)$ e $g_u(R, V)$ nel piano media-varianza



Fonte: Bodnar T. e Bodnar O., On the Unbiased Estimator of the Efficient Frontier, *The International Journal of Theoretical and Applied Finance*, 2010

Come già accennato, i dati per la costruzione delle due frontiere sono stati ricavati da un articolo precedente e per completezza vengono di seguito presentati:

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

$n = 60$, $k = 5$, $\hat{s} = 0.221457$, $\hat{R}_{GMV} = 0.0145664$ e infine $\hat{V}_{GMV} = 0.0010337$.

Di conseguenza il valore soglia di γ risulta essere $\gamma = 1.00707$ e quindi si ha sovrastima per $V < \gamma V_{GMV}$ cioè per $V < 0.00104101$, mentre per valori maggiori si è in presenza di sottostima.

Emerge chiaramente, da un'analisi grafica anche superficiale, come sia evidente il fenomeno di eccessivo ottimismo scoperto da Siegel e Woodgate (2005): la frontiera campionaria domina per tutti i valori di $V > \gamma V_{GMV}$ la rispettiva frontiera corretta.

Questo implica che la formulazione ufficiale della frontiera efficiente remunera eccessivamente il rischio assunto, portando l'investitore ad aspettarsi rendimenti più elevati per il rischio assunto oppure – il che è equivalente, dal punto di vista tecnico, ma può condurre a perdite più elevate – a sottovalutare il rischio insito in un investimento con un dato rendimento atteso.

3.2 Test F sulla Frontiera Efficiente

Gli autori nell'articolo in esame presentano inoltre un test statistico per verificare la profittabilità, e quindi l'efficienza, di ogni portafoglio sulla frontiera efficiente. Essa come si ha avuto modo di illustrare già numerose volte, dipende crucialmente dal parametro s , ovvero dalla curvatura che assume la parabola che la descrive. Nel caso ipotetico in cui tale quantità assuma valore nullo, allora la frontiera efficiente degenera, assumendo una forma retta parallela all'asse delle ascisse. In questo caso limite allora vi è un unico portafoglio efficiente, quello a minima varianza.

Il test F proposto in Bodnar e Bodnar (2010) è basato esattamente su questa assunzione, e pertanto la struttura del test di ipotesi è così definita:

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

Il test ha quindi una interpretazione strettamente finanziaria: se si accetta l'ipotesi nulla allora non ha senso detenere un portafoglio rischioso differente dal portafoglio di minima varianza, poiché tutti gli altri portafogli (dati dalla combinazione

dei k assets) sarebbero caratterizzati dal medesimo rendimento ma da una varianza più elevata.

La statistica test proposta per verificare questa ipotesi ha la seguente forma:

$$T_S = \frac{n(n-k+1)}{(n-1)(k-1)} \hat{s}$$

e visto che qualora $s = 0$ il parametro di non centralità della distribuzione F non centrale diventa nullo (e di conseguenza la distribuzione diventa centrale), si rifiuta l'ipotesi nulla quando:

$$T_S > F_{k-1, n-k+1; \alpha}$$

Dove $F_{k-1, n-k+1; \alpha}$ denota il quantile di ordine α della distribuzione F con $k-1$ e $n-k+1$ gradi di libertà.

Come illustrato dagli autori, si sottolinea che il medesimo risultato è ottenibile se si esegue un test che verifichi l'uguaglianza di ogni μ_k : se tutti i titoli hanno il medesimo rendimento atteso allora anche tutti i portafogli da questi originati hanno il medesimo rendimento atteso, e, ancora una volta, l'unico portafoglio selezionabile risulta essere il portafoglio di minima varianza.

3.3 Criticità della frontiera efficiente à la Bodnar e Bodnar

A questo punto risulta necessario fare un po' di chiarezza circa questo nuovo approccio che sembrerebbe consentire una operatività migliore.

Anche già da una prima occhiata emerge chiaramente che questo nuovo approccio all'ottimizzazione di portafoglio non elimina alcune pesanti assunzioni poste da Markowitz, sia di natura statistica sia di natura economica.

Bodnar e Bodnar infatti nulla dicono circa il superamento di presupposti quali: perfetta divisibilità dei titoli, assenza di costi di transazione, assenza di operatori *market maker*, titoli perfettamente liquidi.

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

Inoltre essi preservano le medesime assunzione circa la distribuzione dei rendimenti e l'assenza di autocorrelazione nelle relative serie storiche.

È soprattutto quest'ultimo aspetto che necessita di indagini più approfondite: se infatti uno sfaldamento degli assunti economici circa il mercato ha un impatto finale considerevole ma non eccessivo (la liquidità degli investimenti impatta sul valore finale realizzato dall'investimento, non su quello teorico calcolato) una modifica delle assunzioni statistiche potrebbe portare non solo all'accumularsi di un ulteriore impatto negativo (rendimento atteso inferiore a cui si cumula una riduzione di valore dovuta ad illiquidità, costi di transazione ecc.) ma anche all'inapplicabilità dell'approccio stesso di ottimizzazione per come è stato illustrato. Il presente lavoro di tesi quindi analizzerà in primo luogo la presenza di differenza statistica tra frontiera efficiente campionaria e frontiera efficiente corretta, proponendo un intervallo di confidenza per la frontiera campionaria calcolato mediante una metodologia bootstrap¹, e successivamente si provvederà a stressare le ipotesi di normalità e assenza di autocorrelazione.

Questo è giustificato da uno sguardo alla operatività nel mondo del risparmio gestito: serie storiche mensili dei rendimenti dei titoli azionari, se da una parte concedono il vantaggio di una distribuzione di tipo normale e dell'assenza di autocorrelazione, dall'altra non permettono di estrarre informazioni qualitativamente equiparabili a quelle ottenute da serie storiche giornaliere. E queste ultime, come assodato dalla teoria e dalla pratica, non rispecchiano le ipotesi di normalità e indipendenza temporale delle osservazioni.

A conclusione di questo capitolo però si ritiene necessario illustrare ulteriori criticità della frontiera efficiente corretta: esse sono emerse dall'analisi di dati reali giornalieri (Corazza e Pizzi, 2014) che ha evidenziato in alcuni casi una varianza negativa per i portafogli della frontiera efficiente e un comportamento asintotico dello stimatore che può considerevolmente ridurre l'efficacia della correzione.

Relativamente alla negatività della varianza, tale possibilità non è ovviamente minimamente contemplata² e ha probabile origine dall'introduzione delle quantità correttive.

¹Si illustrerà nei prossimi capitoli in dettaglio la metodologia utilizzata e quelle considerate non adeguate

²Si ricorda che la varianza è lo sviluppo di una quantità al quadrato, e quindi necessariamente sempre positiva

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

Per meglio chiarire questo punto si ricorre nuovamente ad una riscrittura della formulazione della frontiera efficiente, in questo caso dello stimatore della stessa proposto in Bodnar e Bodnar (2010):

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

con rispettivamente:

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

Dall'analisi di Corazza e Pizzi (2014) emerge inoltre un'ulteriore criticità sopra accennata, relativa al comportamento asintotico dello stimatore in esame: per una numerosità campionaria tendente ad infinito lo stimatore della frontiera à la Bodnar e Bodnar converge alla frontiera efficiente campionaria, ovvero allo stimatore distorto.

Questo può essere agevolmente dimostrato prendendo in esame la Proposizione 1 di Corazza e Pizzi (2014) e la prova che offrono a suo sostegno.

Proposizione 1 *Lo stimatore corretto della frontiera efficiente proposto da Bodnar e Bodnar (2010), al crescere della numerosità n del campione tende asintoticamente alla frontiera efficiente campionaria:*

$$\lim_{n \rightarrow \infty} (\text{Frontiera efficiente corretta}) = \text{Frontiera efficiente campionaria}$$

La prova è fornita di seguito:

è infatti possibile riscrivere le quantità A, B, C e D con le loro approssimazioni:

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

CAPITOLO 3. LA DISTORSIONE DELLA FRONTIERA EFFICIENTE E LA SUA CORREZIONE

È ora facilmente dimostrabile che se la numerosità campionaria cresce si ottengono i seguenti risultati:

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

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

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

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

e di conseguenza:

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

Il comportamento al limite delle quantità A, B, C e D mostra come la correzione proposta in Bodnar e Bodnar (2010) cambi al variare della numerosità delle serie incluse nel procedimento di stima della frontiera efficiente: più tale numerosità è elevata, più accurata è la stima della frontiera campionaria, e quindi minore è l'entità della correzione. Viceversa, per una stima eseguita su campioni con poche osservazioni, la distorsione della frontiera efficiente campionaria risulterà più marcata, e di conseguenza risulterà maggiore l'impatto della quantità correttiva. È possibile quindi affermare che per la stima della frontiera efficiente di "lungo periodo" (quindi con elevata quantità di dati disponibili per la stima) le due frontiere efficienti si equivalgono, essendo la correzione proposta precedentemente poco influente - tendendo ad un valore nullo mano a mano che si incrementa la durata del periodo considerato - , mentre per stime di "breve periodo" (caratterizzate quindi dall'impiego di un numero contenuto i dati) esse possono essere sensibilmente differenti.

Si continua ora con l'analisi del problema di negatività della varianza. Corazza e Pizzi (2014) hanno dimostrato che tale fenomeno si riscontra qualora le quantità in gioco assumono determinati valori, più precisamente quando si verifica la seguente condizione:

$$\hat{s} \leq \frac{(n-1)k}{n(n-k-2)}$$

con

$$\hat{s} = \sum_{i=1}^k \frac{\hat{\mu}_i^2}{\hat{\sigma}_i^2} (1 - \hat{p}_i) \sum_{i=1}^k \sum_{j=1 \neq i}^k \frac{\hat{\mu}_i \hat{\mu}_j}{\hat{\sigma}_i^2} \hat{p}_j,$$

$$\hat{p}_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^k \frac{1}{\sigma_j^2}} \in (0, 1).$$

Per valori di \hat{s} superiori a quello indicato la frontiera efficiente non mostra forme anomale dovute alla negatività della sua misura di rischio.

Come si è già affermato in precedenza, avere un portafoglio caratterizzato da una varianza minore di 0 è non solo un controsenso matematico, ma anche economico: l'affermare che un investimento è caratterizzato rischio negativo è una asserzione priva di qualsivoglia significato, in quanto il minimo rischio assumibile è il rischio nullo – $V = 0$, ovvero l'assenza totale di rischio –. Corazza e Pizzi hanno riscontrato questo effetto per dati giornalieri di rendimenti logaritmici caratterizzati da una elevata numerosità: non escludono tuttavia che il medesimo problema possa verificarsi per dati mensili.

Capitolo 4

BOOTSTRAP PER L'ANALISI DEI DATI

4.1 Scopo dell'analisi

Come si ha già avuto modo di affermare nel corso di questo lavoro, terminata la parte relativa alla trattazione teorica del problema affrontato, ci si focalizza ora sulla sua verifica empirica.

Nel corso del secondo capitolo si è avuto modo di riportare l'intervallo teorico per la frontiera efficiente campionaria sviluppato da Bodnar e Schmid. Tale intervallo però, vista l'ampiezza, risulta di scarsa utilità pratica.

Si è quindi pensato di ricercare un intervallo di confidenza al 95% tramite una metodologia bootstrap: sono state generate 1000 repliche differenti per ognuna delle k serie di dati, ed è stata calcolata la frontiera efficiente campionaria in 100 punti diversi. Per creare l'intervallo di confidenza si è selezionato il venticinquesimo e il novecentosettantacinquesimo valore di rendimento atteso a parità di varianza, per ognuno dei cento punti costituenti la frontiera. In questo modo si può costruire il limite inferiore e superiore dell'intervallo.

Se lo stimatore corretto della frontiera cade al di fuori dell'area così determinata, allora esso è statisticamente differente dalla frontiera campionaria su cui l'intervallo è costruito.

Necessario però a questo punto è illustrare le diverse tipologie di bootstrap: per giungere al risultato finale infatti, si sono sperimentati diversi metodi bootstrap, ognuno con i propri pregi e difetti.

Si illustreranno brevemente quali sono e in cosa consistono i diversi tipi di boo-

tstrap (se ne sono selezionati solo alcuni tra quelli più utilizzati in letteratura), quali sono i principali problemi legati all'applicazione su serie storiche dei rendimenti azionari e quali sono state le difficoltà riscontrate. Si farà riferimento soprattutto a quanto illustrato da Kreiss e Lahiri (2012), i quali forniscono un'adeguata esposizione dei vari metodi esistenti.

4.2 Il metodo bootstrap

Per bootstrap si intende un metodo per creare una serie di dati a partire da una serie di cui si ha già disponibilità. Trova particolare applicazione nella ricerca della distribuzione di alcuni stimatori (ad esempio la media).

Esso si fonda sulla seguente idea di base: si finga di avere a disposizione un campione di 100 osservazioni della statura di altrettante persone, dello stesso sesso e della stessa età e di essere interessati alla stima della statura media. Si procederà quindi calcolando la media delle 100 osservazioni, visto che la media è uno stimatore corretto. Ma come si può fare per avere un'idea un po' più precisa dell'incertezza dello stimatore?

L'applicazione di un metodo bootstrap può portare alla soluzione. A partire dalle 100 osservazioni di partenza è possibile costruire un nuovo campione, in cui ogni osservazione è generata da un'estrazione casuale con ripetizione dal campione precedente. In questo modo si è ottenuto un campione *sosia* del precedente, il quale ha la medesima media teorica (i dati provengono dallo stesso modello generatore dei dati), ma diversa media campionaria (Efron, 1979).

Calcolando la media del *campione sosia* e ripetendo il procedimento per un numero sufficiente di volte si possono così ottenere diverse osservazioni della media campionaria, ed è possibile quindi calcolarne la distribuzione.

Questo metodo però si rivela inutilizzabile per alcune tipologie di dati: è il caso dei dati caratterizzati dalla presenza di autocorrelazione, come ad esempio i dati dei rendimenti giornalieri dei titoli azionari, o comunque dati di tipo time-series. Una siffatta procedura, infatti, causa la rottura della struttura di autocorrelazione, alterando quindi i dati e rompendo la logica del legame temporale tra le diverse osservazioni.

È quindi necessario trovare un metodo alternativo, che consenta di preservare la struttura di partenza. Inoltre quasi tutti i modelli in letteratura sono presentati e applicati per campioni univariati, e si è di conseguenza dovuto modificare l'approccio di tutti questi metodi per renderlo valido nel caso multivariato. Un bootstrap valido per una sola time series necessita infatti di essere riadattato qualora lo si applichi a più serie storiche contemporaneamente. Questo in quanto la struttura di correlazione tra diverse serie (ciò che rende efficace la costruzione di un portafoglio di titoli) deve assolutamente essere preservata. Nel caso non lo fosse, ogni replica bootstrap per la matrice contenente tutte le serie analizzate, anche se correttamente eseguita per ogni singola serie storica, porterebbe inevitabilmente alla distorsione del valore di correlazione tra le medesime, alterando quindi la matrice di varianza e covarianza Σ menzionata più e più volte nei capitoli precedenti.

Tenendo conto di quanto appena illustrato si prosegue esaminando le tipologie bootstrap adeguate nel caso di serie storiche.

4.3 Parametric bootstrap o bootstrap dei residui

Questo metodo bootstrap è alquanto interessante e versatile: data una serie storica di partenza si inizia stimando un modello per la media e uno per la volatilità più accuratamente possibile.

Stimati i due modelli e applicandoli alle osservazioni disponibili, se sono stati inclusi tutti i regressori e non si sono commessi errori di sorta si otterrà una serie di residui, tra di loro incorrelati e distribuiti identicamente, con media nulla e varianza unitaria (si assume solitamente una distribuzione normale standardizzata o t di Student standardizzata).

A questo punto è possibile generare una replica bootstrap dei residui (oppure, cosa pressoché equivalente, generare un campione di n osservazioni necessarie alla stima, estratte casualmente da una distribuzione analoga a quella dei residui), e a partire da quest'ultimi ricostruire, ponendoli come innovazioni e utilizzando

le stime ottenute dai modelli per la media e per la volatilità, una serie di dati che abbia le medesime caratteristiche di quella di partenza (Efron e Tibshirani, 1986). Considerato che per il presente lavoro di tesi si ha necessariamente a che fare con dati caratterizzati sia da correlazione tra diverse serie storiche sia da autocorrelazione, è stato quindi necessario aggiustare la metodologia bootstrap, in quanto né la stima del modello per la media né quella del modello per la volatilità consentono di catturare la struttura di correlazione tra le diverse serie di dati.

Si è quindi pensato di applicare un bootstrap non per le singole colonne della matrice dei dati, ma riga per riga, in modo che il residuo al tempo t_0 per la serie x fosse sempre contemporaneo al residuo t_0 della serie y . Tale accortezza non viene considerata nella trattazione del bootstrap sui residui, in quanto si tratta di un semplice aggiustamento che dipende non solo dai dati che si utilizzano ma anche dalle necessità del ricercatore.

In questo modo quindi è stato possibile preservare per ogni replica bootstrap la medesima struttura di autocorrelazione e di dipendenza tra diverse serie.

Similmente al metodo parametrico appena illustrato si ritiene opportuno citare brevemente anche il metodo *sieve bootstrap*: anch'esso infatti si basa sulla stima di un modello per i dati di input, un modello autoregressivo di ordine p con un numero di ritardi inseriti dipendenti dalla lunghezza della serie analizzata. In questo modo è possibile filtrare la serie, ottenere una nuova serie di residui, e similmente a sopra eseguire il bootstrap su questi ultimi e poi ricostruire la serie a partire dal modello usato per filtrare i dati.

Bühlmann (1995) espone le proprietà non parametriche di questo bootstrap, evidenziandone l'affidabilità rispetto ad altri metodi: non si tiene conto però della spesa computazionale di questo metodo, che per consentire un buon risultato necessita di stimare un modello AR tendente ad infinito. Si è pertanto deciso di abbandonare tale metodologia.

Anche per quanto riguarda il bootstrap sui residui si è riscontrato un problema analogo e di conseguenza non è stato utilizzato. Se infatti sembra particolarmente efficace, esso tuttavia ha un costo computazionalmente e temporalmente elevato: oltre al tempo necessario affinché si ottenga la replica bootstrap dei residui, si deve necessariamente aggiungere il tempo speso per la stima dei modelli, che diventa assai rilevante.

È proprio con riferimento a tale problema che si è ritenuto opportuno abbandonare questa metodologia preferendone una alternativa. Si è infatti inizialmente provato ad eseguire un bootstrap sui residui con un procedimento di stima automatico dei modelli per media e volatilità, ma ciò impone necessariamente dei limiti sulle variabili da includere nel modello affinché un simile procedimento sia computazionalmente sostenibile. L'inserimento di questi limiti comporta in modo inequivocabile la perdita di accuratezza nel procedimento di stima e, di conseguenza, nella ricostruzione dei dati da utilizzare per le analisi successive. Concludendo questa breve illustrazione del bootstrap sui residui si può affermare che esso sia un metodo molto efficace e sicuro, a patto che si disponga di una potenza di calcolo elevata e di modelli per media e volatilità assai accurati e affidabili. In caso contrario questo metodo è sconsigliato, a causa delle conseguenze dell'applicazione di modelli inesatti.

4.4 Dati surrogati

Questo metodo è un'alternativa ai bootstrap sopra indicati. Esso non è propriamente una tipologia di bootstrap, in quanto non ricampiona da un insieme di dati, ma crea i dati alternativi a partire da trasformazioni dei dati di partenza. Tale approccio si basa sul calcolo della trasformata di Fourier¹ per il campione di dati disponibili: la trasformata permette di rappresentare nel dominio delle frequenze una funzione che è caratterizzata da dipendenza temporale e per ottenere questo risultato essa applica una decomposizione della funzione stessa nella base di funzioni esponenziali e prodotto scalare. La seguente formula è quella utilizzata per il calcolo della trasformata e prende il nome di *Trasformata di Fourier*

¹La prima versione è contenuta nel trattato *Théorie analytique de la chaleur* ad opera dello stesso Fourier

Veloce:

$$X_q = \sum_{k=0}^{n-1} x_k e^{-i \frac{2\pi}{n} k q} = 0 \ 1 \ \dots \ n-1$$

in cui X_q indica la q -esima osservazione della serie trasformata, mentre n invece indica la numerosità campionaria.

Si nota fin da subito che questa applicazione scompone i dati in ingresso in numeri complessi, dotati quindi di parte reale e di parte immaginaria: per ottenere i dati surrogati è sufficiente operare una randomizzazione della fase della trasformata, elevando la stessa nel seguente modo:

$$X_q r = X_q^{z_q * i}$$

tenuto conto che è possibile riscrivere X_q come:

$$X_q = A(f) * e^{i\phi(f)}$$

ovvero evidenziando ampiezza A e fase ϕ .

$X_q r$ rappresenta invece la trasformata con fase randomizzata e z un numero casuale compreso tra 0 e 2π .

Questo metodo esposto in Theiler e al. (1991), Theiler e al. (1992), Prichard e Theiler (1994), Chan (1997), Schreiber (1998) e Schreiber e Schmitz (2000) consente di ottenere quindi una trasformata randomizzata simile alla precedente, ma che una volta trasformata inversamente e quindi riportata nel dominio del tempo produce un insieme di nuovi dati con le medesime caratteristiche di quelli di partenza ma con valori differenti: in questo modo si preserva l'autocorrelazione, ottenendo repliche leggermente differenti del campione di partenza.

Affinché siano inoltre preservate anche le correlazioni tra le diverse serie storiche è sufficiente randomizzare ogni serie con la stessa sequenza di numeri casuali z , in modo tale che Y_q e X_q siano elevate sempre al medesimo z_q : così facendo ogni serie viene randomizzata nella stessa misura delle altre.

Tuttavia questo approccio ha un difetto di fondo: i dati generati – surrogati per l'appunto – non rispettano pienamente tutte le caratteristiche dei dati originari in quanto da campione a campione si rileva un notevole cambio di ordine di grandezza degli stessi.

Questo ovviamente non permette di utilizzarlo per lo scopo di questo lavoro di tesi.

Si è quindi testata l'applicabilità di un approccio differente, basato sul riordinamento del campione in seguito ai valori della trasformata di Fourier discreta randomizzata nella fase: questo metodo consente di preservare l'ordine di grandezza, tramite dati surrogati che non sono quindi nuovi dati generati, ma sono semplicemente i dati già disponibili, riorganizzati in un ordine temporale differente ma tale da preservare le caratteristiche di autocorrelazione. Anche in questo caso è stato necessario riadattare il codice R affinché permettesse di svolgere tale trasformazione per dati multivariati.

Tuttavia questo metodo, sebbene colmi alcune lacune del precedente, lascia il fianco scoperto sotto altri aspetti: esso reimpiega la totalità dei dati di partenza, così che si ottiene un nuovo campione caratterizzato da media e varianza uguali al campione iniziale.

Ciò non è quindi di alcuna utilità per i fini del presente lavoro in quanto non si otterrebbero differenze di sorta tra la serie surrogata e quella reale.

4.5 Bootstrap a blocchi

Da ultimo ci si concentra sul metodo bootstrap utilizzato: il bootstrap a blocchi. L'idea di fondo di tale metodo consiste nel dividere il campione di dati disponibili in blocchi sufficientemente grandi in modo che all'interno di ogni blocco sia preservata la struttura di autocorrelazione e di eseguire un bootstrap sull'insieme dei blocchi (Künsch, 1989).

Questo metodo ha anche un notevole vantaggio qualora applicato a dati multivariati: è sufficiente dividere a blocchi l'intera matrice dei dati, così come si opera la suddivisione sulla singola serie. In questo modo è preservata non solo la correlazione tra serie, ma in misura sufficiente anche l'autocorrelazione, in misura dipendente dalla dimensione del blocco stesso, come esposto da Hall et al. (1995).

Esistono in letteratura differenti metodi per la realizzazione di un bootstrap a blocchi: a blocchi fissi, a blocchi mobili, a blocchi sovrapposti.

In ogni caso la validità di un metodo dipende soprattutto da un fattore: la dimen-

sione del blocco. Più essa è elevata, tanto maggiore è la preservazione dell'autocorrelazione, più la dimensione è ridotta meno tale caratteristica è mantenuta. Tuttavia blocchi grandi comportano numerosi problemi e difficoltà, prima fra tutte la scarsa differenza tra le diverse serie replicate e la serie originale. Ipotizzando infatti un blocco con numerosità notevole di osservazioni al suo interno, si ottiene un numero ridotto di blocchi complessivi nel quale il campione di partenza è diviso. In questo modo il numero di permutazioni con ripetizioni disponibili si abbassa e si incorre nel rischio (assai concreto per elevati valori di repliche) di ottenere più serie fra loro equivalenti. Nello script utilizzato, si è applicato un metodo di formazione del blocco di tipo probabilistico, fissando a priori un valore denominato *gprob*, compreso tra 0 e 1, rappresentante la probabilità di transizione da un blocco al successivo.

Si seleziona un valore casuale tra quelli nella serie storica dei dati: esso sarà il primo valore della replica bootstrap. Poi si estrae un numero casuale compreso tra 0 e 1 da una distribuzione uniforme e si procede in questo modo: se il valore estratto è maggiore di *gprob* allora si continua il blocco, altrimenti se ne inizia uno di nuovo.

Un esempio può aiutare a chiarire il procedimento.

Si immagini di avere a disposizione 1000 dati, e di fissare un valore $gprob = 0.5$. Estrae casualmente dai 1000 dati si ottiene l'osservazione numero 850: questa sarà la prima osservazione della replica. Si continua poi estraendo un numero casuale compreso tra 0 e 1, ad esempio 0.657. Essendo maggiore di *gprob* si continua inserendo nella seconda posizione della serie replica l'osservazione 851 del campione di partenza. Si ipotizzi di estrarre altri 9 numeri casuali superiori a 0.5: le prime dieci osservazioni della serie replicata saranno quindi le osservazioni 850-859 del campione di partenza. Se il numero casuale successivo è minore di *gprob* allora si estrae una nuova osservazione dal campione e un nuovo blocco ha inizio.

Questo metodo consente di coniugare alcuni vantaggi di altri bootstrap a blocchi: la dimensione del blocco è decisa in modo probabilistico (a minore *gprob* corrisponde minore probabilità che il blocco abbia numerosità campionaria ridotta) così come la sovrapposizione o meno dei blocchi (essa dipende solamente dall'osservazione iniziale di ogni blocco estratta casualmente dall'insieme dei dati

di partenza).

Si è scelto questo metodo in quanto è quello caratterizzato dall'aver minori controindicazioni nel suo utilizzo e di conseguenza minori effetti distorsivi: la quasi totalità dei metodi precedentemente illustrati, per una pluralità eterogenea di motivi hanno portato nell'applicazione pratica a risultati inconcludenti se non ampiamente e chiaramente erranei.

Capitolo 5

ANALISI DELLA FRONTIERA EFFICIENTE

CORRETTA

5.1 Dati utilizzati

Dopo aver esposto il metodo bootstrap utilizzato per generare le repliche delle serie storiche, si prosegue ora illustrando i dati utilizzati per eseguire le differenti analisi.

Vista la condizione più volte richiamata nel presente lavoro che pone $k > 2$, per simulare un portafoglio azionario sono stati selezionati quattro differenti titoli quotati alla NYSE (*New York Stock Exchange*) e facenti parte dell'indice borsistico *Dow Jones Industrial Average*, per il periodo compreso dal 01/01/2012 al 31/08/2017: DuPont (*ticker symbol* DD), ExxonMobil (*ticker symbol* XOM), General Electric (*ticker symbol* GE) e Goldman Sachs (*ticker symbol* GS).

Si sono inoltre calcolati i rendimenti logaritmici di ogni titolo per diversi archi temporali: rendimenti giornalieri, rendimenti settimanali e rendimenti mensili.

Per ogni serie così ottenuta si è svolto il test D'Agostino sulla distribuzione, per verificare la presenza o meno di normalità.

Nella tabella alla pagina seguente sono esposti i *p-value* delle statistiche test per ogni serie storica:

Tabella 5.1: Test di normalità

	Mensile	Settimanale	Giornaliero
DD	0.0000000	0.000000e+00	0.000000e+00
XOM	0.9777286	1.386493e-04	0.000000e+00
GE	0.2485197	5.846952e-09	0.000000e+00
GS	0.5156155	2.647238e-03	5.551115e-16

Si nota da subito come le serie dei rendimenti del titolo DuPont non presentino mai una distribuzione normale, nemmeno per osservazioni mensili. Ciò ha portato alla necessità di cercare di individuare il motivo di tale comportamento: si è proceduto quindi con l'operare un confronto tra quantili della distribuzione empirica e quantili teorici della distribuzione normale. Si riportano di seguito i grafici illustranti la relazione tra i suddetti quantili. Per agevolare il confronto si è scelto di illustrare i grafici di tutte le serie storiche, suddividendoli a seconda del tipo di osservazione.

Figura 5.1: Confronto tra quantili della distribuzione normale e quantili della distribuzione empirica per le serie dei rendimenti mensili

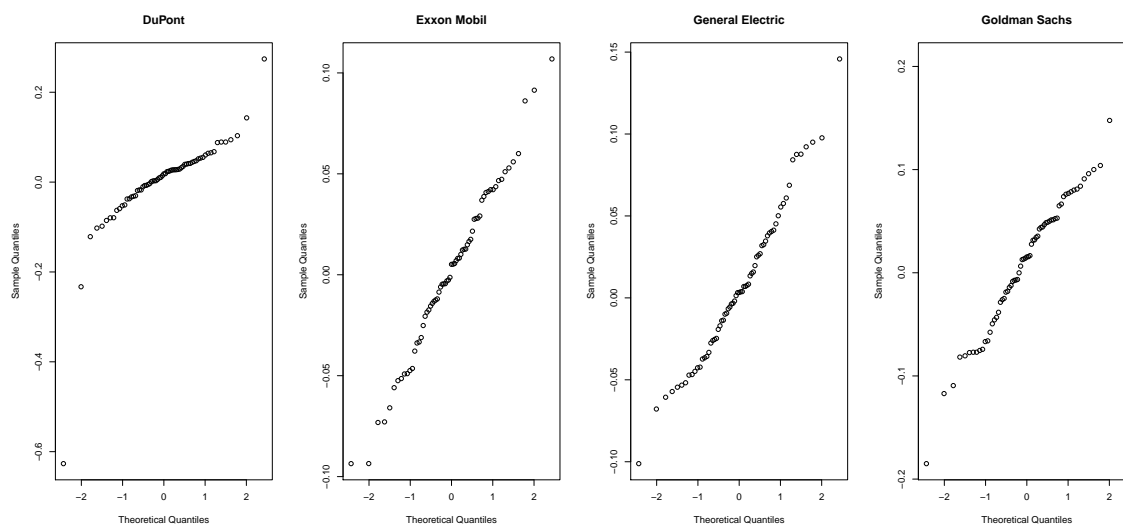


Figura 5.2: Confronto tra quantili della distribuzione normale e quantili della distribuzione empirica per le serie dei rendimenti settimanali

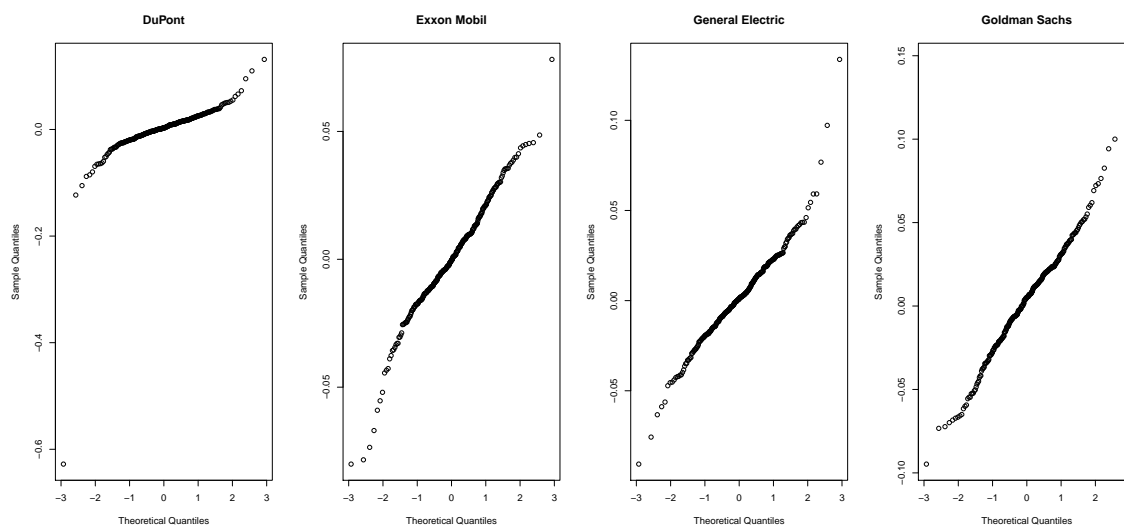
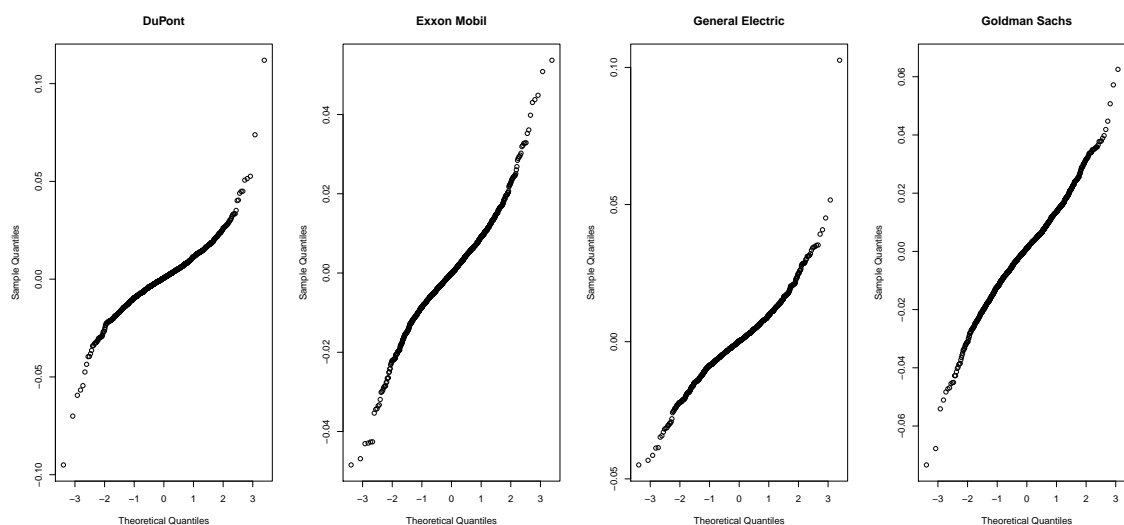


Figura 5.3: Confronto tra quantili della distribuzione normale e quantili della distribuzione empirica per le serie dei rendimenti giornalieri



Dall'analisi di tali grafici è possibile notare come per rendimenti mensili i quantili evidenzino un comportamento normale per tutte le serie eccetto quella dei rendimenti del titolo DuPont: tale serie risente infatti della presenza anomala di valori situati sulle code della funzione di densità di probabilità, rendendo di fatto più probabili osservazioni estreme rispetto a quanto accadrebbe per una distribuzione gaussiana.

Il medesimo comportamento riscontrato per i rendimenti mensili di tale titolo è

presente anche nei rendimenti giornalieri e settimanali dello stesso. Inoltre si fa notare come per serie storiche di questo tipo (giornaliere e settimanali) anche le distribuzioni dei rendimenti degli altri titoli deviano da quella di tipo normale, evidenziando quanto appena illustrato per il titolo DuPont: eventi estremi più probabili causano l'ispessimento delle code della funzione di distribuzione e pertanto la presenza di *leptocurtosi* impedisce la normalità della distribuzione dei rendimenti.

Quanto appena affermato è supportato anche dall'analisi del confronto tra distribuzione empirica e distribuzione normale (riportato nei grafici a seguire).

Come in precedenza, si illustrano per ogni tipologia di rendimento (mensile, settimanale e giornaliero), i grafici per ognuna delle quattro serie storiche utilizzate. La curva tracciata in rosso identifica la distribuzione empirica dei rendimenti, mentre la curva in grigio rappresenta la distribuzione normale con parametri di media e varianza pari a quelli del campione utilizzato: in questo modo è possibile avere un riscontro visivo di quanto le due curve si distanzino tra loro e quindi di quanto l'ipotesi di normalità risulta errata o accettabile.

Figura 5.4: Confronto tra distribuzione empirica e distribuzione normale per le serie dei rendimenti mensili

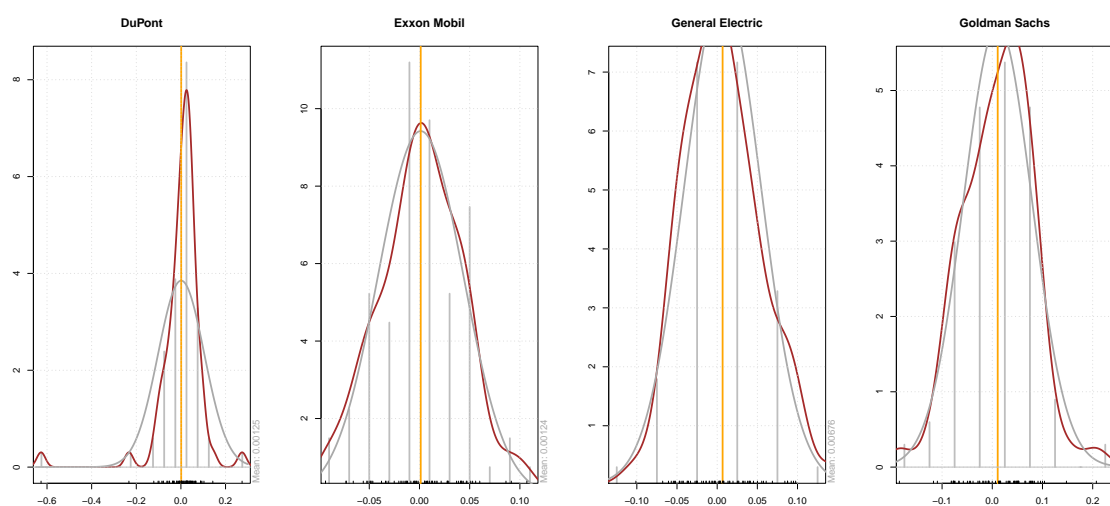


Figura 5.5: Confronto tra distribuzione empirica e distribuzione normale per le serie dei rendimenti settimanali

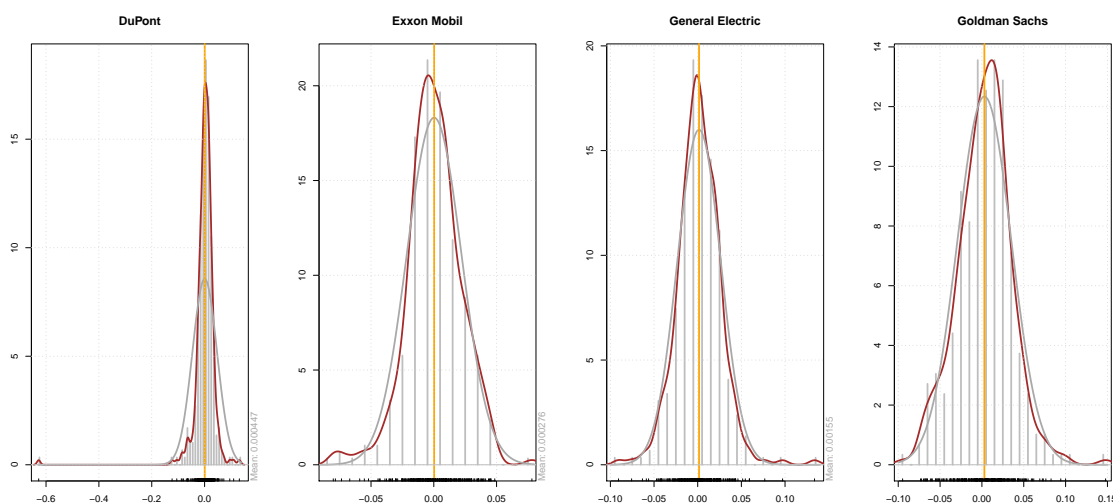
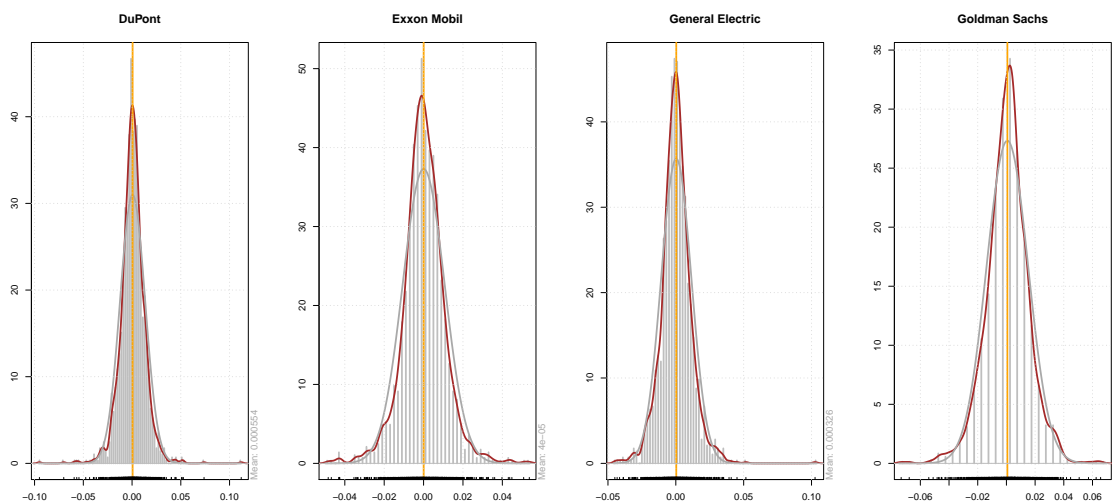


Figura 5.6: Confronto tra distribuzione empirica e distribuzione normale per le serie dei rendimenti giornalieri



Si può osservare che per le serie di rendimenti mensili la distribuzione empirica segue con buona approssimazione la distribuzione teorica. Come evidenziato dall'analisi precedente anche in questo caso per il titolo DuPont si riscontrano forti anomalie rispetto all'ipotesi di normalità: valori estremi portano non solo al succitato problema di leptocurtosi, ma inoltre evidenziano una sostanziale asimmetria della distribuzione.

A conferma di ciò si è eseguito un test D'Agostino sul titolo DuPont, che ha

prodotto il seguente risultato:

Tabella 5.2: Test di normalità per i rendimenti mensili del titolo DuPont

	p-value
Omnibus Test	< 2.2e-16
Skewness Test	1.475e-11
Kurtosis Test	5.953e-09

I risultati esposti in 5.2 confermano quanto evidenziato dall'analisi grafica svolta in precedenza.

Per avere un quadro un po' più completo del comportamento delle serie storiche utilizzate si è deciso di verificare anche la presenza di autocorrelazione fra le osservazioni delle stesse. Si è quindi predisposto un Box test per un numero di ritardi pari a 30, e si sono rappresentati graficamente i p-value ottenuti dal test statistico. Si ricorda che per questo test l'ipotesi nulla è l'assenza di autocorrelazione, e per agevolare il riconoscimento di p-value inferiori al livello di significatività $\alpha = 0.05$ è stata inserita nel grafico una retta rossa, indicando proprio il livello discriminante tra accettazione e rifiuto dell'ipotesi nulla.

Figura 5.7: Box test per osservazioni giornaliere delle quattro serie storiche in esame

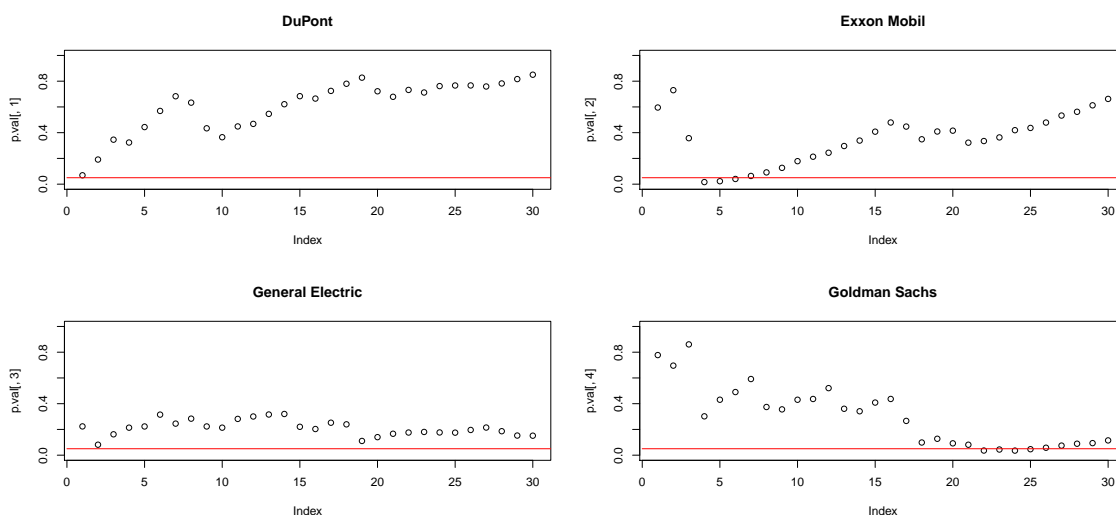


Figura 5.8: Box test per osservazioni settimanali delle quattro serie storiche in esame

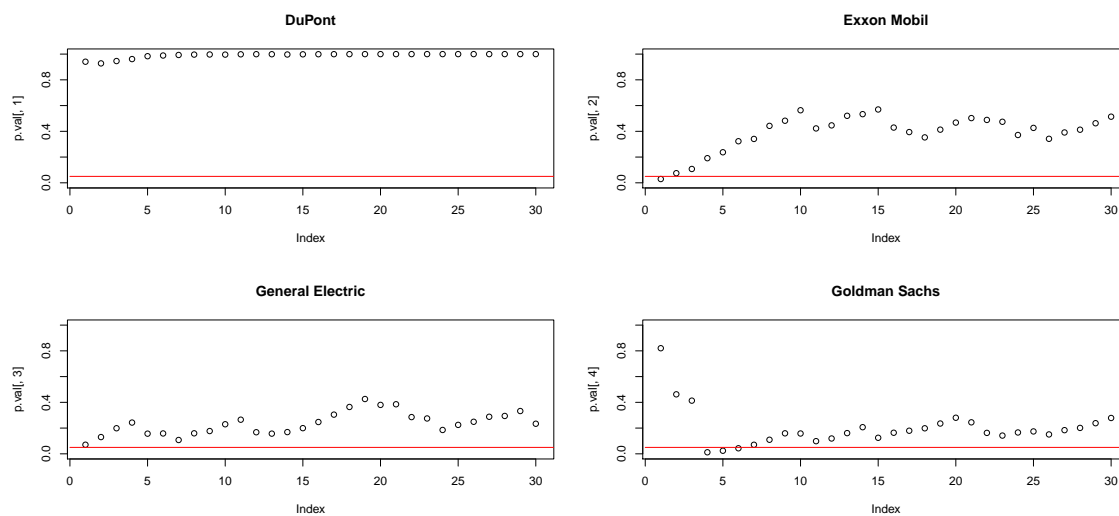
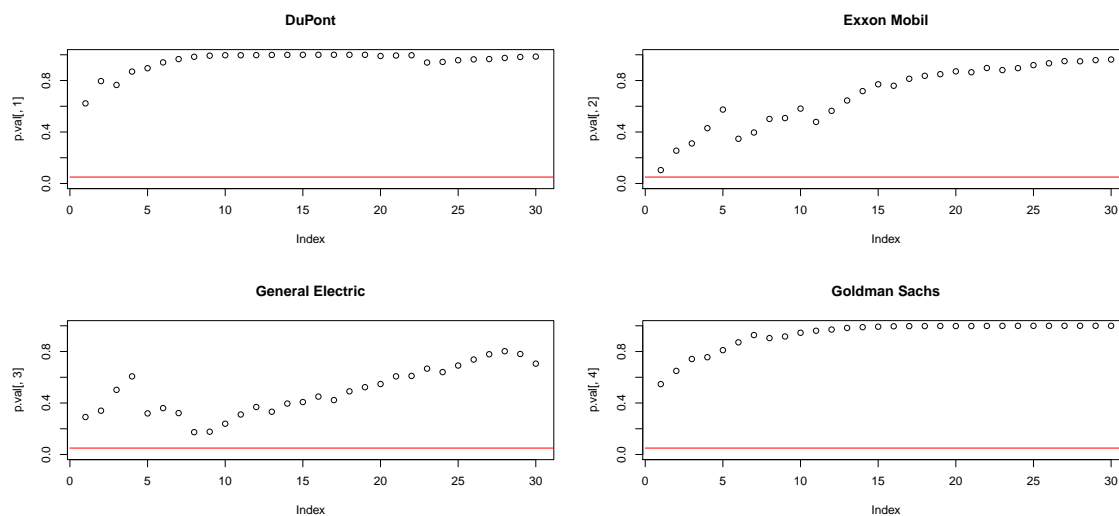


Figura 5.9: Box test per osservazioni mensili delle quattro serie storiche in esame



Come si osserva dai grafici sopra riportati, il test è stato svolto per tutte le serie storiche di ogni titolo (osservazioni mensili, settimanali e giornaliere).

È interessante notare come il comportamento della serie cambi a seconda della tipologia di osservazioni utilizzate: per osservazioni giornaliere le serie storiche dei rendimenti di Exxon Mobil e Goldman Sachs riscontrano evidenza di auto-correlazione (ad eccezione di quella del titolo DuPont, il cui p-value per l'unico ritardo dubbio è pari a 0.06895, e di General Electric), mentre per osservazioni

mensili il test consente di accettare l'ipotesi di osservazioni non autocorrelate per tutte le serie storiche. Questo risultato è in linea con il metodo proposto da Markowitz (1952) e Bodnar e Bodnar (2010) circa l'utilizzo di osservazioni mensili al fine di lavorare con serie storiche non autocorrelate.

Come però si ha avuto modo di illustrare nei capitoli precedenti l'informazione ridotta dei rendimenti mensili dei titoli azionari porta alla necessità di utilizzare serie storiche giornaliere e quindi possibili osservazioni autocorrelate.

Infine per quanto riguarda le osservazioni settimanali non si riscontra un comportamento univoco: i rendimenti di alcuni titoli (DuPont e General Electric) si caratterizzano per la non autocorrelazione, mentre quelli di altri (Exxon Mobil e Goldman Sachs) preservano tracce di autocorrelazione nelle osservazioni, similmente a quanto esposto per le serie di rendimenti giornalieri.

Questo consente di ipotizzare che in periodi di mercato non anomali (bolle speculative o crisi economiche) i prezzi giornalieri e settimanali dei titoli (e quindi i rendimenti giornalieri e settimanali degli stessi) sono influenzati da fattori esterni in misura maggiore rispetto a quanto avviene per i prezzi mensili.

Riassumendo, la distribuzione normale dei rendimenti è accettabile solo per le serie storiche mensili, ad eccezione del titolo DuPont, poiché la distribuzione delle altre serie analizzate (rendimenti giornalieri e settimanali di tutti i titoli) riproduce evidenze di *asimmetria* e *leptocurtosi* che portano a rifiutare la normalità della distribuzione.

Per quanto riguarda invece l'autocorrelazione è possibile affermare che non è mai presente per le diverse serie dei rendimenti del titolo DuPont, mentre, per gli altri titoli inseriti nel portafoglio, la presenza o meno di autocorrelazione dipende dal tipo di osservazione considerata: assenza di autocorrelazione per rendimenti mensili, presenza di autocorrelazione per rendimenti giornalieri e settimanali.

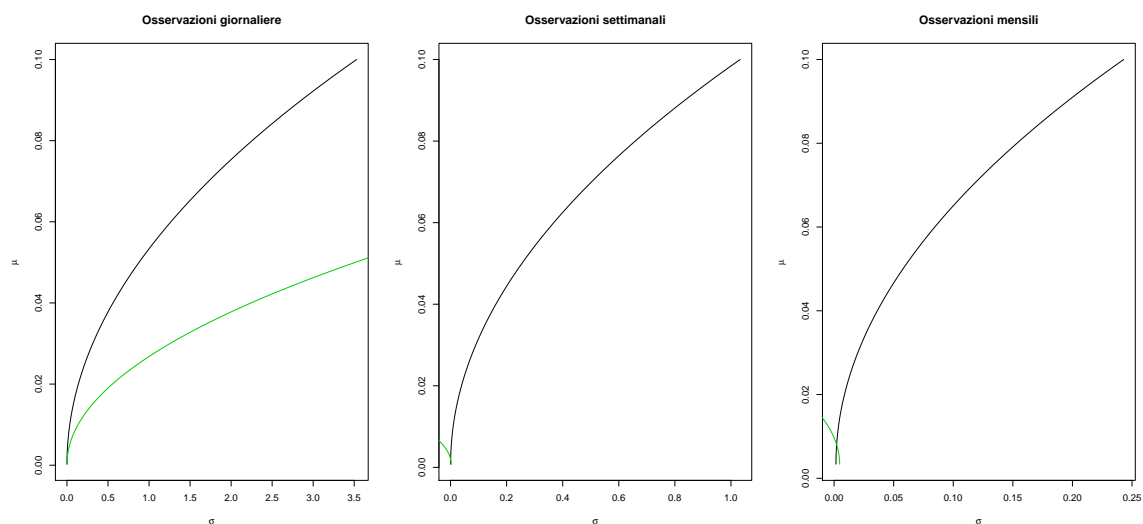
5.2 Frontiera efficiente "à la Bodnar e Bodnar" per dati reali

Effettuate le analisi preliminari appena descritte per capire il comportamento delle serie considerate, si è proceduto con la stima delle due frontiere efficienti presentate nei precedenti capitoli di questo lavoro: la frontiera efficiente campionaria (o "à la Markowitz") e la frontiera efficiente corretta (o "à la Bodnar e Bodnar").

La stima per entrambe le frontiere è stata ottenuta con le procedure descritte nei capitoli 1 e 3. Ottenuti i valori dei pesi del portafoglio di minima varianza è stato poi possibile calcolare un insieme di 100 punti della frontiera (ogni punto è funzione di R_{GMV} , V_{GMV} e R - come illustrato nel corso del secondo capitolo - e quindi fissando un dato valore di R per ogni punto di stima della frontiera è possibile risalire al rispettivo valore di V).

I risultati grafici per le tre tipologie di serie storiche (osservazioni giornaliere, settimanali e mensili) sono di seguito forniti:

Figura 5.10: Confronto tra le due diverse frontiere efficienti proposte: in nero la frontiera efficiente campionaria, in verde la frontiera efficiente corretta



Si nota subito che nel caso di osservazioni mensili e settimanali la frontiera efficiente "à la Bodnar e Bodnar" risulta inaccettabile in quanto presenta una varianza negativa, come evidenziato già a chiusura del capitolo terzo.

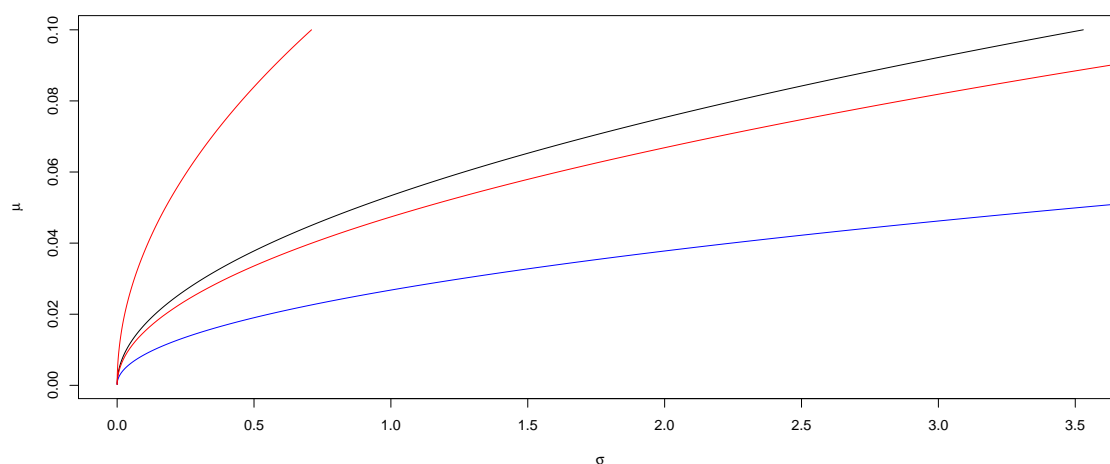
Ciò verifica, anche per serie storiche con questo tipo di osservazioni, la validità di quanto evidenziato da Corazza e Pizzi sul comportamento della frontiera riscontrato in alcuni casi di serie storiche di rendimenti giornalieri.

A tal riguardo si è provveduto con il verificare la diversità statistica tra le due frontiere solo per il portafoglio creato a partire da serie di rendimenti giornalieri, visto il risultato inconcludente ottenuto per gli altri due portafogli.

È stato applicato quindi il metodo bootstrap a blocchi (con parametro probabilistico pari a 0.5) per la generazione di mille repliche delle osservazioni di partenza, generando mille alternativi portafogli e stimando la frontiera efficiente campionaria per ognuno di questi, in modo da ottenere un vasto numero di osservazioni sulle quali calcolare l'intervallo.

Tale intervallo empirico risulta evidenziare come le due frontiere nel caso considerato risultino statisticamente differenti: si provvede a fornire di seguito il grafico.

Figura 5.11: Intervallo empirico mediante "block bootstrap" per la frontiera efficiente campionaria



Si nota infatti come i due limiti dell'intervallo di confidenza al livello di significatività $\alpha = 5\%$ rappresentati sul grafico dai due rami di parabola in rosso, racchiudano al loro interno solamente la frontiera efficiente campionaria (rappresentata in nero), mentre la frontiera efficiente corretta (colorata in blu) si colloca nettamente al di sotto.

È quindi possibile affermare che anche nel caso di serie storiche di rendimenti giornalieri, i quali violano alcune delle assunzioni alla base della quantità correttiva calcolata da Bodnar e Bodnar (2010), le due frontiere efficienti stimate risultino statisticamente differenti.

Si deve ovviamente tener conto però che le distribuzioni delle serie storiche utilizzate non sono bene approssimate da distribuzioni normali. Ciò comporta che l'applicazione sia del procedimento di stima della frontiera efficiente campionaria sia della quantità correttiva introdotta da Bodnar e Bodnar possa portare a risultati parzialmente sbagliati.

5.3 Frontiera efficiente "à la Bodnar e Bodnar" per dati simulati

A fronte dei risultati ottenuti dall'utilizzo di dati reali, non si è in grado di determinare quale sia il fattore che influenza il risultato ottenuto, nel caso specifico la diversità tra le due frontiere efficienti.

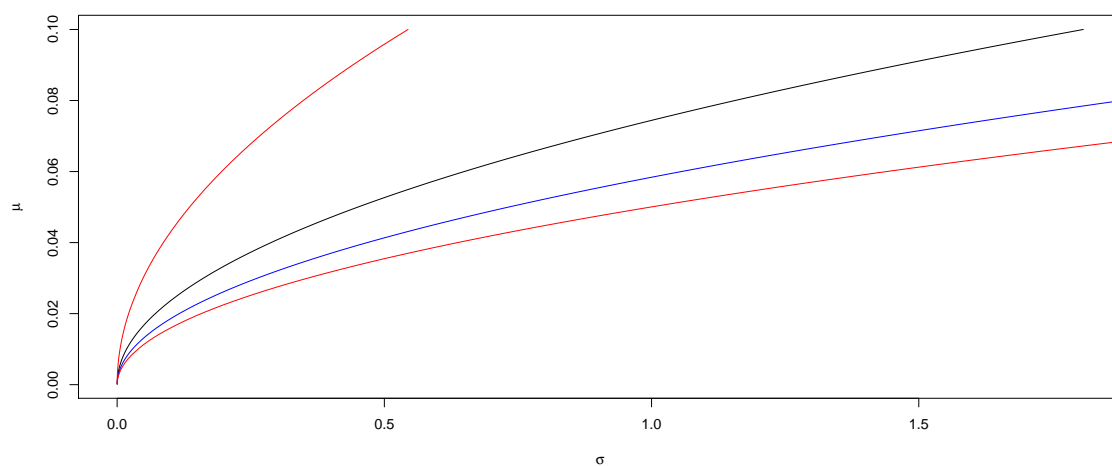
Tenuto conto che l'analisi è stata svolta per osservazioni giornaliere si è deciso di sottoporre a stress una per volta le ipotesi base della correzione formulata da Bodnar e Bodnar (2010): normalità dei rendimenti e assenza di autocorrelazione nelle serie storiche degli stessi.

5.3.1 Simulazione di serie storiche normali multivariate e prive di autocorrelazione

Per prima cosa è stato simulato un dataset perfettamente aderente alle ipotesi teoriche, estraendo le osservazioni tempo per tempo da una distribuzione normale multivariata, con media pari alla media campionaria delle serie storiche reali e con varianza pari alla matrice di varianza e covarianza delle medesime. In tal modo si è cercato di ottenere dati di input per la stima della frontiera, la cui unica differenza rispetto a quelli reali fosse nella funzione di densità di probabilità e nella presenza o meno di autocorrelazione.

È stato generato quindi un nuovo dataset, sul quale è stata verificata la differenza statistica tra le due frontiere con la metodologia bootstrap già illustrata. Sono state simulate quattro serie storiche, ognuna con un numero di osservazioni pari a 1424 (come per i dati reali), aspettandosi che le due frontiere, come evidenziato da Corazza e Pizzi, per numerosità così elevata ricadano nel medesimo intervallo.

Figura 5.12: Intervallo empirico per la frontiera efficiente campionaria con dati simulati provenienti da una distribuzione normale multivariata



Come nel grafico riportato, entrambe le frontiere giacciono all'interno del medesimo intervallo di confidenza, portando la loro differenza ad essere statisticamente nulla.

Si riporta che tale analisi è stata svolta per 100 dataset simulati a partire dalla stessa distribuzione normale multivariata: nella maggioranza dei casi (69) si è riscontrata l'uguaglianza statistica, mentre solo una minoranza di 31 casi ha riportato una diversità tra le due frontiere.

Ciò necessita di un migliore approfondimento: la metodologia utilizzata, ovvero la creazione dell'intervallo di confidenza empirico attraverso repliche delle osservazioni ottenute con un "block bootstrap", è robusta: per ognuno dei cento dataset il metodo ha fornito un intervallo pressoché costante. Ciò che ha prodotto la differenza nell'esito è il procedimento di stima delle frontiere. Analizzando la matrice di varianza e covarianza e il vettore delle medie dei rendimenti dei titoli in portafoglio per ogni dataset si è notato come minime differenze nei valori degli

stessi producano elevate diversità nel comportamento sia della frontiera efficiente campionaria, sia della frontiera efficiente corretta.

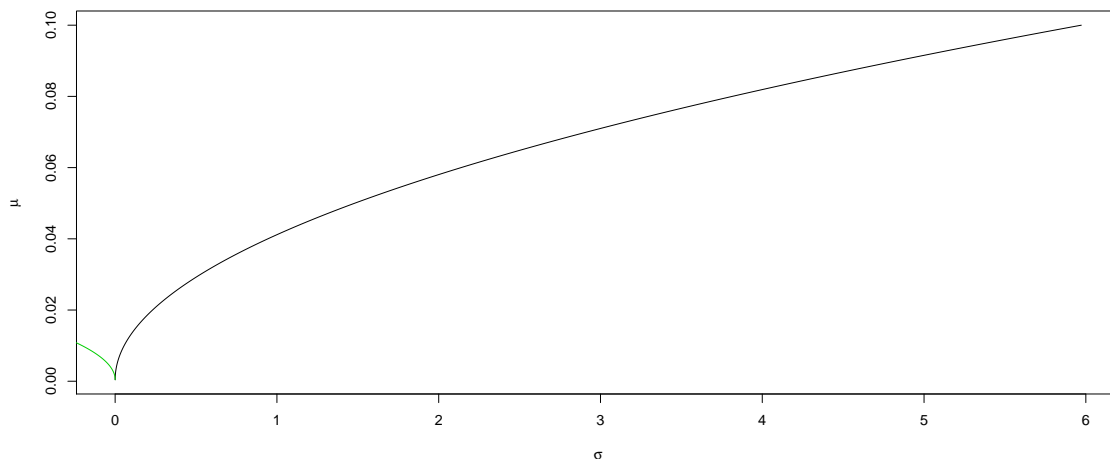
Per tale motivo si è ritenuto congruo eseguire il procedimento un numero sufficiente di volte tale per cui la maggioranza degli esiti costituisca un risultato affidabile.

Il medesimo procedimento è stato inoltre eseguito, utilizzando lo stesso numero di dataset, per serie storiche distribuite come le precedenti (si è utilizzata la stessa funzione di distribuzione normale multivariata) aventi però un numero inferiore di osservazioni, specificamente pari a 120. In questo modo si è voluto verificare il comportamento di entrambe le frontiere qualora non vi sia presenza di un numero di osservazioni sufficientemente elevato da costituire una buona approssimazione di una serie di lunghezza tendente ad infinito. I risultati hanno confermato le attese: nella maggioranza dei casi (77 casi su 100) le due frontiere risultano statisticamente differenti, mentre solo una minoranza di 23 dataset porta ad affermare l'equivalenza tra le due frontiere. Anche in questo caso i due limiti dell'intervallo ottenuti con il metodo bootstrap risultano affidabili: al variare del dataset, cui corrisponde - come già accennato - una grande variabilità nella forma assunta dalle frontiere efficienti, essi preservano le loro caratteristiche, subendo minime variazioni (dovute soprattutto alla diversità del campione). Per svolgere questa analisi però si è dovuto tener conto della ridotta dimensione del campione da cui ricavare le repliche bootstrap e quindi è stato modificato il valore del parametro probabilistico che regola la dimensione dei blocchi, elevandolo a 0.8 dal precedente 0.5. In questo modo si ha una probabilità più elevata di generare blocchi diversi.

Come ultimo aspetto connesso a questa verifica empirica, si fa notare che per alcuni dataset, si è ottenuta una frontiera efficiente corretta con varianza negativa. Ovviamente tali dataset non sono stati inclusi nel conteggio per l'ovvia inaccettabilità del risultato e conseguente impossibilità di confronto.

Alla pagina seguente, si riporta a titolo esemplificativo un grafico che evidenzia tale problema.

Figura 5.13: Frontiera efficiente "à la Bodnar e Bodnar" utilizzando dati simulati da una distribuzione normale multivariata



Considerando che la distribuzione è rimasta invariata rispetto al caso illustrato dal grafico precedente, l'unica diversità la si riscontra solamente nella matrice delle stime delle varianze e covarianze e nel vettore delle medie empiriche. La spiegazione di tale fenomeno giace quindi nei valori assunti dalle stime contenute nella matrice e nel vettore appena citati.

Tutte le altre condizioni, oltre ad essere invariate rispetto al caso riportato per la creazione dell'intervallo, rispettano fedelmente le assunzioni di normalità ed assenza di autocorrelazione richieste dalla teoria.

5.3.2 Simulazione di serie storiche con funzione di distribuzione t di Student prive di autocorrelazione

Avendo esposto la verifica del comportamento delle frontiere efficienti nel caso in cui i presupposti di normalità e assenza di autocorrelazione siano pienamente soddisfatti si procede ora con il sottoporre a stress la correzione proposta da Bodnar e Bodnar (2010) simulando serie storiche con osservazioni estratte non da una distribuzione normale, ma da una distribuzione t di Student con 4 gradi di libertà.

In tal modo si intende verificare se la diversità tra le due frontiere permane oppure

se esse risultano essere statisticamente equivalenti.

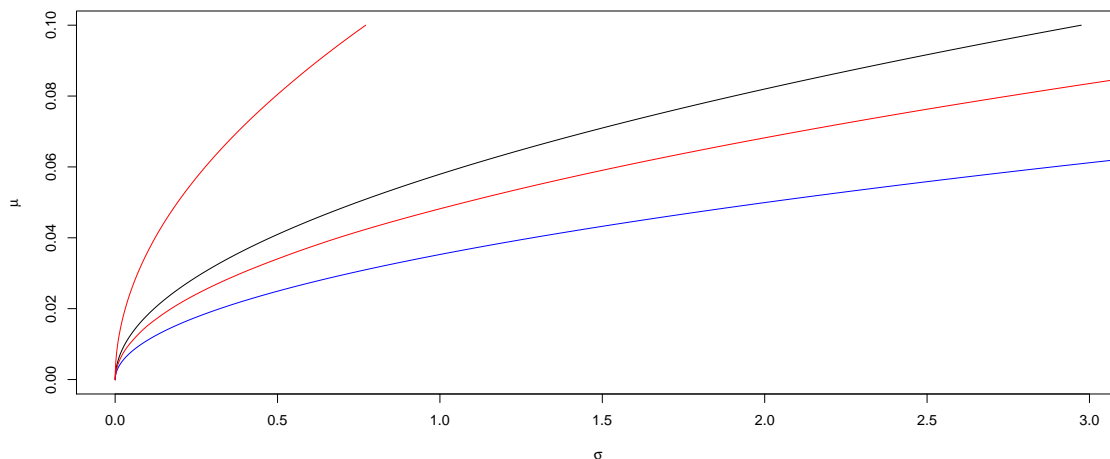
Come si ha già avuto modo di esporre in precedenza, affinché il risultato fosse maggiormente attendibile si è eseguita la simulazione su cento dataset differenti, utilizzando serie storiche con lunghezza pari a 1424 osservazioni. Nella maggioranza dei casi (72 dataset su 100 rispetto a 28 dataset su 100), le due frontiere risultano essere statisticamente differenti.

È quindi possibile affermare che anche per serie storiche non normali (almeno per quanto riguarda quelle costituite da osservazioni distribuite come t di Student) la frontiera efficiente dei portafogli proposta da Bodnar e Bodnar risulta essere diversa da quella campionaria. In aggiunta, tale frontiera per dati distribuiti in modo non normale non risente della convergenza alla frontiera campionaria all'aumentare della numerosità, ovvero per $n \rightarrow \infty$. Le serie storiche simulate hanno infatti uguale lunghezza delle corrispettive reali: a differenza della simulazione di perfetta aderenza alle assunzioni di Bodnar e Bodnar (dati normali e non autocorrelati, che porta ad una convergenza della frontiera corretta a quella campionaria) - come illustrato nel paragrafo precedente- , utilizzando un egual numero di osservazioni per serie con diversa funzione di distribuzione, tale comportamento asintotico viene meno. Plausibilmente questo comportamento è dovuto al fatto che per serie storiche con distribuzioni che non rispettano le assunzioni teoriche la componente correttiva applicata è errata, in quanto deriva direttamente dall'applicazione di valori attesi calcolati sulla base di una distribuzione normale: quando ciò non si verifica si riproduce quindi una frontiera efficiente modificata, il cui valore atteso non è necessariamente uguale alla frontiera efficiente teorica ($E(g_u(R, V)) \neq g_p(R, V)$).

Il risultato certo e rilevante è la diversità tra le due frontiere anche in presenza di dati distribuiti in modo non normale.

È di seguito fornito un grafico che illustra la differenza statistica delle due frontiere:

Figura 5.14: Intervallo di confidenza per la frontiera efficiente campionaria utilizzando dati simulati da una distribuzione t di Student multivariata



Si ritiene altresì che la sensibilità della stima delle frontiere efficienti (campionaria e corretta) ai valori contenuti nel vettore μ dei rendimenti medi e nella matrice Σ di varianza e covarianza non giustifichi l'assenza di convergenza, in quanto i risultati favorevoli sono in numero assai ridotto rispetto a quanto avviene per dati con distribuzione normale. Si è quindi propensi a rintracciare la causa di tale comportamento nella non normalità della funzione di distribuzione e nella conseguente applicazione erronea della frontiera efficiente corretta.

5.3.3 Simulazione di serie storiche normali autocorrelate

Si procede ora con la verifica della frontiera efficiente corretta nel caso l'ipotesi di assenza di autocorrelazione non sia presente, preservando però la distribuzione normale dei rendimenti.

A tal fine si è pensato di equiparare le serie storiche a realizzazioni di un processo ARMA (processo autoregressivo a media mobile), in modo che vi fosse presenza di autocorrelazione tra le osservazioni delle diverse serie storiche. Si sono quindi fissati in principio i parametri della componente autoregressiva e di quella a media mobile, ipotizzando un modello ARMA (1,1) così definito:

$$y_t = \phi_1 y_{t-1} + \Theta_1 \epsilon_{t-1} + \epsilon_t$$

con $\epsilon_t \sim iid(0, \sigma_{\epsilon_t}^2)$

Ogni serie è la realizzazione di un processo ARMA(1,1) differente.

Nella seguente tabella si riportano i parametri dei processi utilizzati:

Tabella 5.3: Parametri processi ARMA(1,1)

	Serie 1	Serie 2	Serie 3	Serie 4
ϕ_1	0.823	-0.487	-0.399	0.613
Θ_1	-0.115	0.243	-0.322	-0.342

Si sono quindi generati 100 dataset, ognuno composto da quattro serie con lunghezza pari a 1424 osservazioni, ricavate dalla generazione di innovazioni distribuite normalmente con media nulla e varianza paria a quella campionaria della serie di riferimento ridotta di un ordine di grandezza. Per ogni dataset si è poi provveduto a testare la differenza statistica tra le due frontiere: ne è emerso che le due differenti stime della frontiera efficiente dei portafogli risultano essere uguali nella maggioranza dei casi (54 casi contro 46).

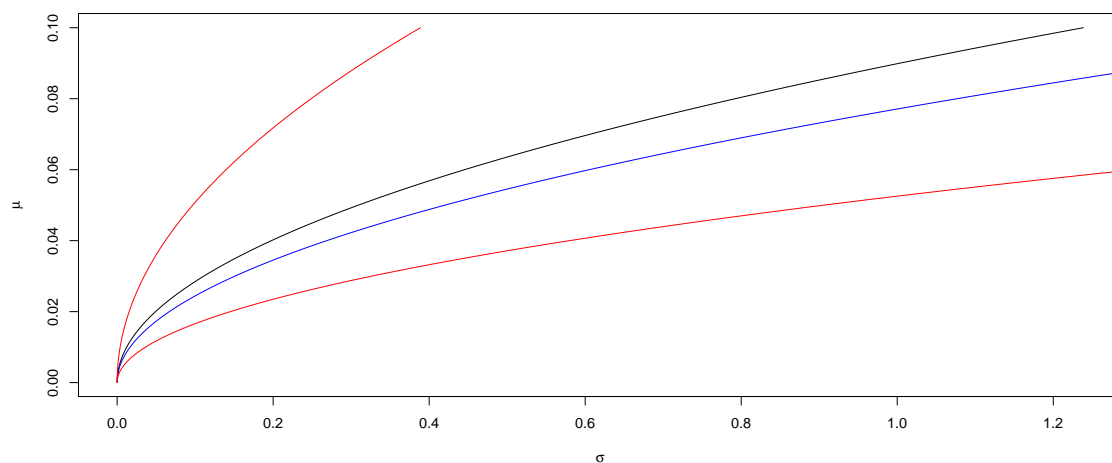
Anche in questo caso, tale risultato può essere dovuto al comportamento asintotico della frontiera efficiente "à la Bodnar e Bodnar" già menzionato. Per verificare ciò si è ripetuto il procedimento, cambiando solo la dimensione del campione: la lunghezza di ogni serie storica è stata posta uguale a 120 osservazioni.

I risultati sembrerebbero confermare l'influenza del comportamento asintotico anche per serie di rendimenti autocorrelati: infatti usando campioni di dimensioni ridotte (ma parti dello stesso processo stocastico che ha generato i dataset precedenti) l'affidabilità del risultato cambia nettamente. In 68 prove su 100 le due frontiere risultano essere statisticamente differenti.

I casi in cui le due frontiere ricadono all'interno dello stesso intervallo di confidenza sono probabilmente dovuti, come nelle prove precedenti, alla sensibilità della stima rispetto alla matrice di varianza e covarianza e al vettore dei rendimenti medi: minime variazioni in entrambi possono provocare forti diversità nel comportamento delle due frontiere.

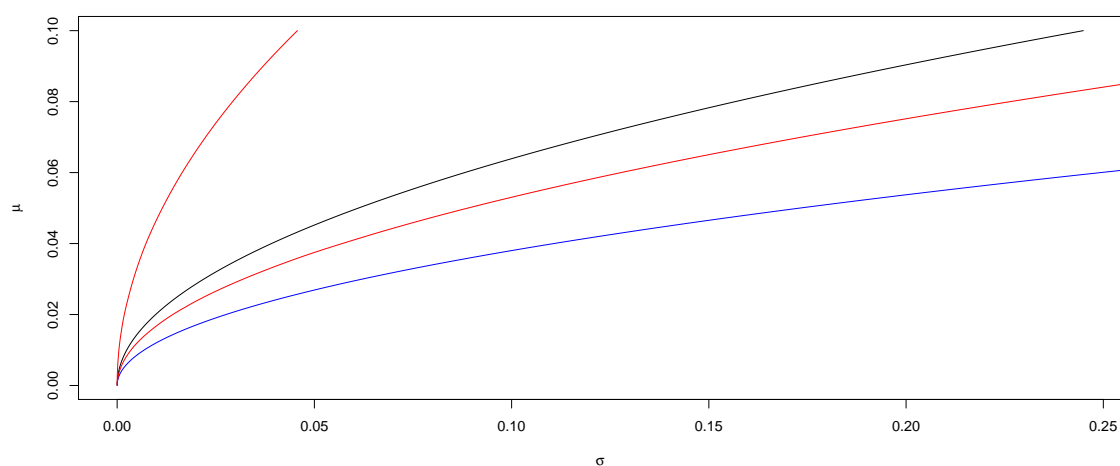
Di seguito si riportano entrambi i grafici:

Figura 5.15: Intervallo di confidenza per la frontiera efficiente campionaria utilizzando dati simulati mediante un processo ARMA(1,1) con innovazioni distribuite normalmente



Serie storiche con lunghezza pari a 1424 osservazioni

Figura 5.16: Intervallo di confidenza per la frontiera efficiente campionaria utilizzando dati simulati mediante un processo ARMA(1,1) con innovazioni distribuite normalmente



Serie storiche con lunghezza pari a 120 osservazioni

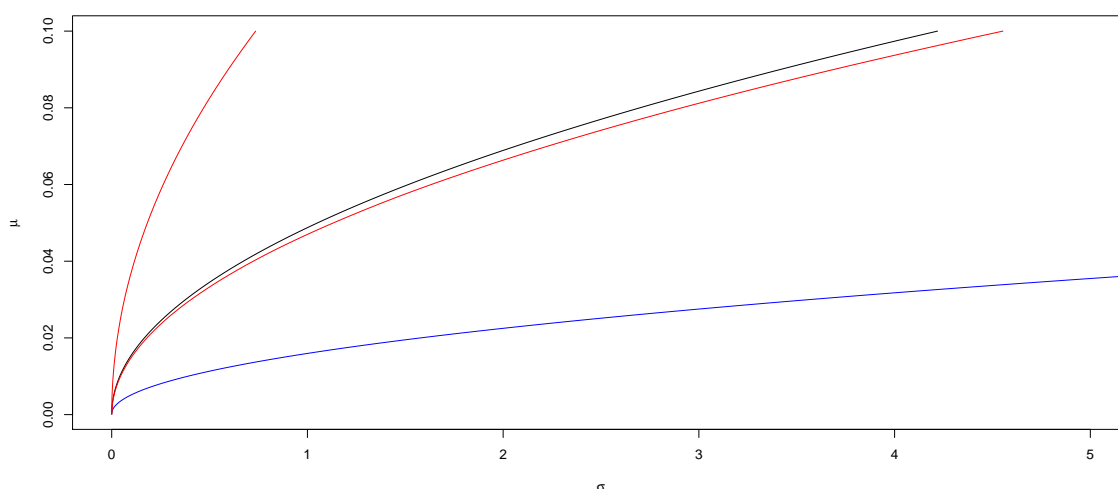
5.3.4 Simulazione di serie storiche non normali e autocorrelate

Con questa ultima analisi di dati simulati si è cercato di osservare il comportamento della frontiera efficiente corretta nei confronti di quella campionaria qualora siano violate le due assunzioni cruciali più e più volte richiamate. Conoscere se i dati rispettino o meno gli assiomi di normalità e assenza di autocorrelazione permette di ottenere un risultato la cui possibile differenza dai casi precedenti (analizzati nelle sezioni del presente capitolo) è unicamente imputabile al rispetto o meno delle suddette condizioni, visto che le altre caratteristiche delle serie sono mantenute invariate.

Questa analisi, sebbene eseguita attraverso l'uso di innovazioni distribuite secondo una t di Student per tutte le serie storiche, cerca di imitare il più possibile il comportamento di dati reali, per i quali - come si ha già avuto modo di spiegare in precedenza - distribuzione normale e assenza di autocorrelazione sono ipotesi in molti casi irrealistiche.

Come illustrato per lo studio del caso di serie autocorrelate e distribuite normalmente, si è provveduto ad esaminare serie storiche di lunghezza pari a 1424 osservazioni e si fornisce di seguito una rappresentazione grafica delle frontiere efficienti (campionaria e corretta) da queste stimate.

Figura 5.17: Intervallo di confidenza per la frontiera efficiente campionaria utilizzando dati simulati mediante un processo ARMA(1,1) con innovazioni $\epsilon \sim T(4)$



Visti i risultati ottenuti non si è ritenuto necessario verificare il comportamento del portafoglio per un numero di osservazioni inferiore. Infatti, anche utilizzando un dataset di dimensioni analoghe a quello dei dati reali non si è riscontrata la presenza di un comportamento asintotico della frontiera efficiente corretta, a differenza di quanto avviene per dati distribuiti normalmente: la maggioranza dei casi esaminati (58 su 100) riporta la presenza di diversità statistica tra le due frontiere, analogamente a quanto avviene per dataset caratterizzati da assenza di autocorrelazione e distribuzione non normale.

Concludendo il presente capitolo è possibile affermare quindi che la diversità statistica tra le due frontiere proposte (campionaria e corretta) è sempre presente per portafogli composti da serie storiche con scarso numero di osservazioni. Al crescere di quest'ultimo invece si riscontra una diversità di comportamento a seconda del tipo di distribuzione che caratterizza le serie storiche: se distribuite normalmente si verifica il comportamento asintotico che porta la quantità correttiva contenuta nella frontiera "à la Bodnar e Bodnar" ad avere valore nullo, se distribuite in modo differente invece tale comportamento viene meno. Si ritiene che questo fenomeno sia imputabile alla quantità correttiva proposta da Bodnar e Bodnar, il cui calcolo è fondato sull'ipotesi di normalità distributiva, e qualora tale assunzione viene meno anche la correzione stessa risulta perdere di affidabilità. Cambiando infatti la distribuzione sottostante, anche le quantità stimate necessarie per la costruzione della frontiera efficiente¹ risultano avere una funzione di distribuzione diversa da quella ricavata da Bodnar e Schmid (2007), ed è quindi plausibile che la quantità posta a correzione della frontiera campionaria risulti inadeguata a garantire la correttezza della frontiera efficiente "à la Bodnar e Bodnar".

Da ultimo per quanto riguarda l'influenza dell'autocorrelazione delle serie sulla stima della frontiera efficiente corretta è possibile affermare che non si è riscontrato alcun effetto che possa modificare il comportamento di questa nei confronti di quella campionaria indipendentemente dal tipo di distribuzione, confermando la differenza statistica tra le due.

¹La trattazione in dettaglio dell'argomento è presente al capitolo secondo e terzo.

Capitolo 6

ANALISI ESPLORATIVA

6.1 Dati utilizzati

La stima della frontiera efficiente "à la Bodnar e Bodnar", come si ha già avuto modo di esporre in precedenza, produce in determinati casi una frontiera efficiente inaccettabile, caratterizzata da una misura di rischio (varianza) negativa, priva quindi sia di significato economico sia di validità matematica.

In questo capitolo si cercherà di individuare i fattori che influiscono sul processo che porta ad ottenere stime di questo tipo, analizzando due differenti portafogli, e monitorando il comportamento degli stessi e delle serie storiche che ne determinano la composizione.

Si è proceduto come segue: si sono selezionati due portafogli (entrambi composti da quattro differenti titoli) per i quali la stima della frontiera efficiente corretta sia in un caso accettabile e nell'altro no.

Si sono analizzate le serie storiche - linearità, normalità della distribuzione, presenza di autocorrelazione - ricercando possibili motivi di disturbo nel processo di stima, analizzando anche i vettori delle medie e le matrici di varianza e covarianza.

Poi è stata variata la composizione dei portafogli, aggiungendo titoli provenienti dal portafoglio alternativo e si è verificato il comportamento della frontiera efficiente corretta.

Si è anche provveduto a mantenere invariato il numero di titoli presenti nel portafoglio e a variarne ugualmente la composizione, al fine di verificare se la stima della frontiera efficiente variasse (da inaccettabile ad accettabile e viceversa).

In questo modo, monitorando i nuovi titoli aggiunti al portafoglio (o sostituiti ai precedenti) e quelli eliminati dallo stesso è possibile verificare se il comportamento della stima è influenzato da caratteristiche intrinseche alle serie storiche che compongono il portafoglio stesso o se dipende invece da fattori differenti, come ad esempio la struttura di dipendenza tra i diversi titoli utilizzati.

Si procede ora con l'esposizione della composizione dei portafogli e delle serie storiche che li costituiscono.

6.2 Analisi dei portafogli

Sono stati selezionati due portafogli di partenza composti da quattro titoli ciascuno, facenti parte dell'indice Dow Jones Industrial Average, e per ognuno di essi si sono calcolati i rendimenti logaritmici per il periodo 01/01/2012 - 31/08/2017.

I due portafogli risultano così costruiti:

Portafoglio A: Caterpillar (*ticker symbol "CAT"*), Chevron Corporation (*ticker symbol "CVX"*), Cisco Systems (*ticker symbol "CSCO"*) e Coca-Cola Company (*ticker symbol "KO"*)

Portafoglio B: Home Depot (*ticker symbol "HD"*), Intel (*ticker symbol "INTC"*), IBM (*ticker symbol "IBM"*) e Johnson & Johnson (*ticker symbol "JNJ"*).

Per tutti e otto i titoli si è verificata la presenza di linearità attraverso il test Terasvirta e il test di White. I risultati sono di seguito esposti:

Tabella 6.1: P-value dei test di linearità

Titolo	Terasvirta Test	White Test
CAT	0.34401750	0.42095790
CVX	0.00348080	0.02287521
CSCO	0.24165032	0.07080268
KO	0.25161300	0.37848190
HD	0.02298258	0.32765747
IBM	0.02853256	0.03513779
INTC	0.02198053	0.01388302
JNJ	0.10064200	0.33525600

Si può notare come i due test statistici non sempre siano concordi relativamente all'accettazione dell'ipotesi nulla di linearità. In casi di risultati divergenti si è considerato il solo test Terasvirta. Si può quindi affermare che le serie dei rendimenti dei titoli Caterpillar, Cisco Systems, Coca-Cola e Johnson & Johnson siano lineari, mentre quelle dei titoli rimanenti presentano un comportamento non lineare.

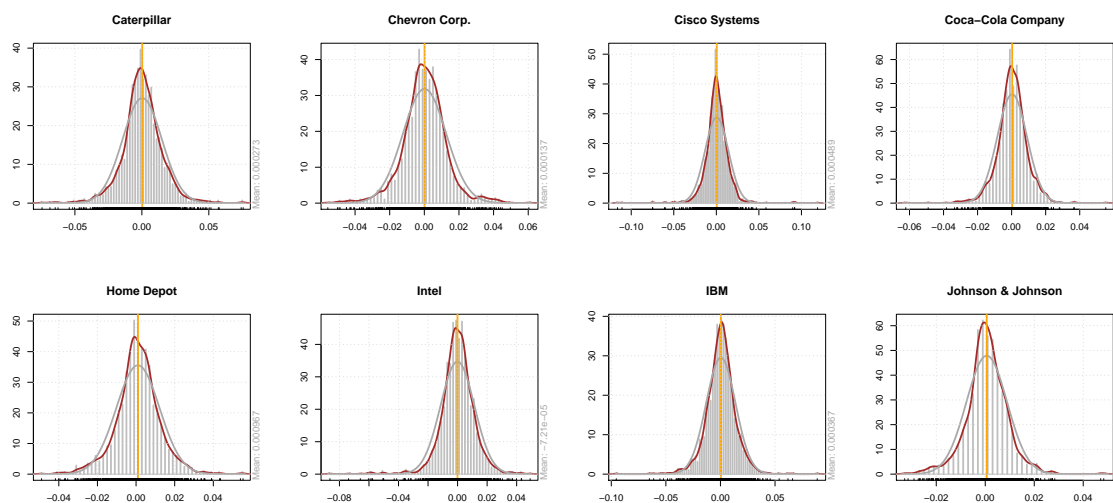
Allo stesso modo si è proseguito per la verifica di normalità distributiva, eseguita mediante il test D'Agostino. I risultati del test, presentati più avanti, hanno evidenziato l'assenza di distribuzione di tipo normale per tutte e otto le serie considerate.

Come già discusso nei capitoli precedenti, la presenza di leptocurtosi nelle serie storiche dei rendimenti giornalieri porta a rifiutare l'ipotesi nulla di normalità. Ancora una volta, per poter appurare il motivo della distanza della distribuzione empirica da quella normale, è stato necessario stimare la prima (stima eseguita con metodo Kernel) e confrontarla con la seconda. In questo modo è possibile verificare la presenza o meno di masse di probabilità sulle code delle distribuzioni empiriche (che portano quindi ad una maggior probabilità di eventi *estremi* rispetto a quanto avviene per una distribuzione normale): se presenti allora è verificato il fenomeno di leptocurtosi. Di seguito si riportano i risultati per il test D'Agostino e i grafici di confronto tra distribuzione empirica e teorica.

Tabella 6.2: P-value del test di normalità

Titolo	D'Agostino test
CAT	0.000000e+00
CVX	0.000000e+00
CSCO	0.000000e+00
KO	0.000000e+00
HD	1.998401e-15
IBM	0.000000e+00
INTC	0.000000e+00
JNJ	0.000000e+00

Figura 6.1: Confronto tra distribuzioni empiriche e teoriche per i titoli dei portafogli A e B



È facile notare come tutte le serie evidenzino la presenza di un numero maggiore di osservazioni sulle code della distribuzione, confermando difatti l'ipotesi di distribuzioni leptocurtiche, le quali, unitamente all'asimmetria di alcune serie (e.g. Intel, Johnson & Johnson), provocano un allontanamento dalla distribuzione teorica (riportata in grigio), e quindi dall'ipotesi di normalità distributiva.

Appurata la non normalità della distribuzione dei rendimenti dei titoli oggetto di studio, si è indagata la presenza o meno di autocorrelazione, svolgendo il test Box per ognuna delle otto serie. Come spiegato nel capitolo precedente, per ogni grafico si è evidenziata con una retta rossa la soglia di accettabilità dell'ipotesi nulla.

Si può agevolmente notare come per il Portafoglio A nessuna serie mostri evidenza di autocorrelazione, a differenza di quanto avviene per il Portafoglio B, la cui unica serie non autocorrelata è quella dei rendimenti di Intel.

I risultati sono esposti nella pagina seguente.

Figura 6.2: Box test per le serie dei titoli inclusi nel Portafoglio A

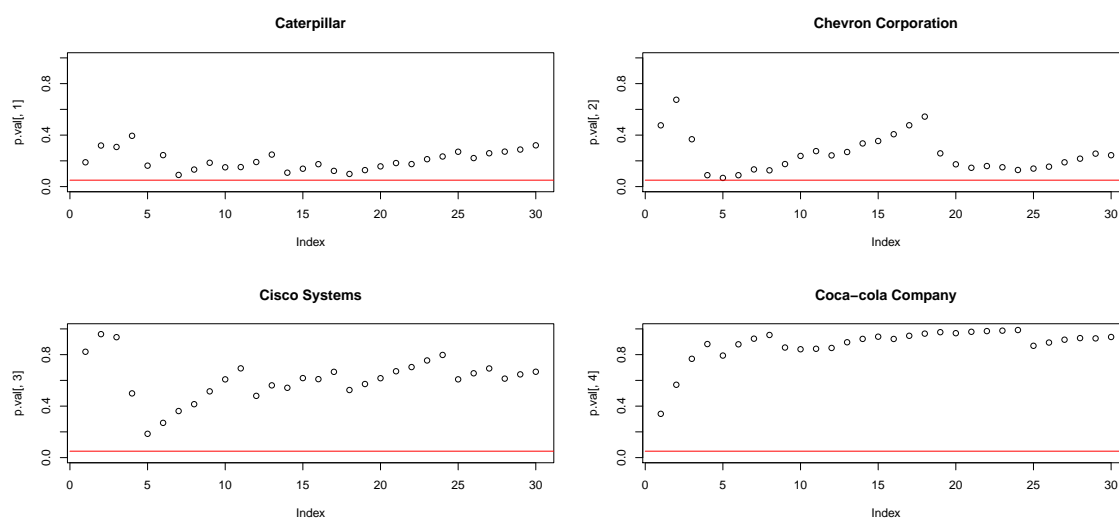
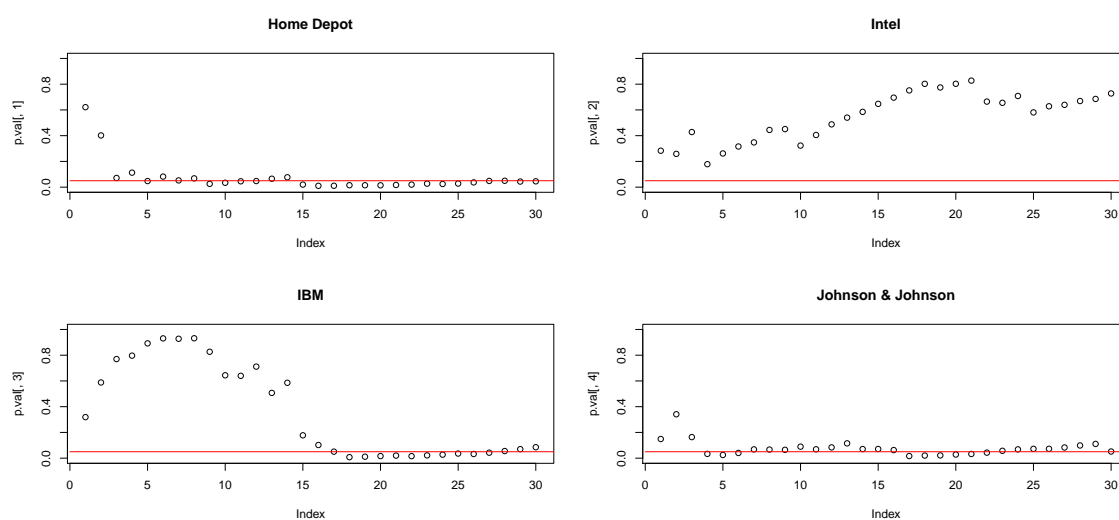


Figura 6.3: Box test per le serie dei titoli inclusi nel Portafoglio B

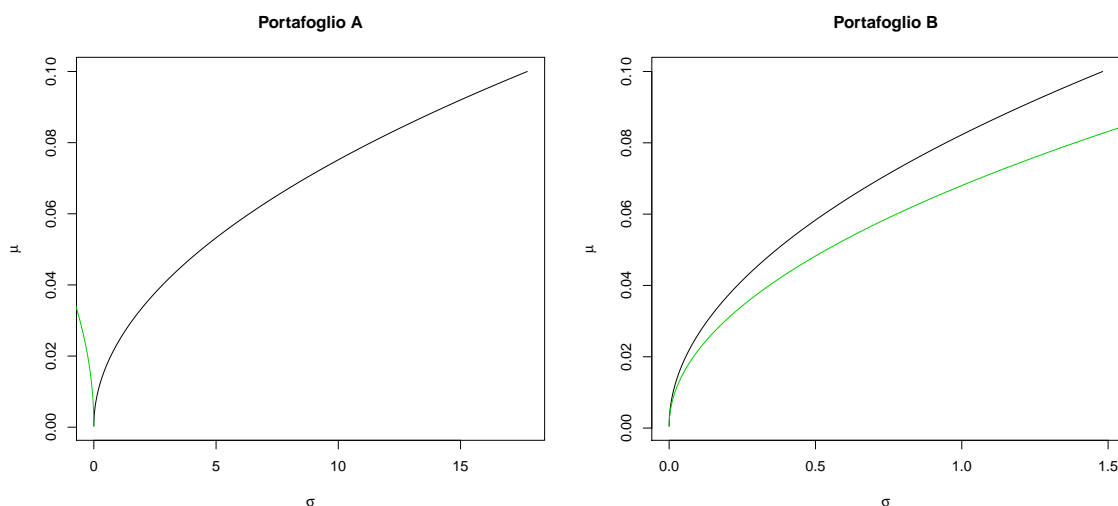


Terminata l'analisi preliminare del comportamento delle singole serie, si passa ora ad analizzare le due frontiere efficienti ("à la Markowitz" e "à la Bodnar e Bodnar") per entrambi i portafogli.

Come accennato ad inizio capitolo la frontiera efficiente corretta del Portafoglio A presenta il già citato problema di negatività della varianza, a differenza di quella del Portafoglio B.

Si procede quindi con le stime delle frontiere per entrambi i portafogli, le cui rappresentazioni grafiche sono di seguito fornite.

Figura 6.4: Rappresentazione grafica delle frontiere efficienti per il Portafoglio A e per il Portafoglio B



Le matrici di varianza e covarianza di entrambi i portafogli illustrano la presenza di titoli positivamente correlati, con valori del medesimo ordine di grandezza: si riportano entrambe di seguito.

Tabella 6.3: Matrice di Varianza e Covarianza del Portafoglio A

	CAT	CVX	CSCO	KO
CAT	2.162795e-04	9.679962e-05	7.417240e-05	3.143570e-05
CVX	9.679962e-05	1.571198e-04	6.235280e-05	4.045650e-05
CSCO	7.417240e-05	6.235280e-05	1.919076e-04	3.705898e-05
KO	3.143570e-05	4.045650e-05	3.705898e-05	7.673902e-05

Tabella 6.4: Matrice di Varianza e Covarianza del Portafoglio B

	HD	IBM	INTC	JNJ
HD	1.259772e-04	4.476723e-05	5.270212e-05	3.810879e-05
IBM	4.476723e-05	1.321365e-04	6.527840e-05	3.654679e-05
INTC	5.270212e-05	6.527840e-05	1.825251e-04	4.079604e-05
JNJ	3.810879e-05	3.654679e-05	4.079604e-05	6.946423e-05

Inoltre il determinante di entrambe le matrici è caratterizzato dall'aver il medesimo ordine di grandezza, con valore 2.491003e-16 per il Portafoglio A e 1.072809e-

16 per il Portafoglio B. I vettori delle medie invece, di seguito riportati, evidenziano una sostanziale uguaglianza nell'ordine di grandezza dei valori in essi contenuti, ma anche importanti differenze tra gli stessi: il rendimento medio del titolo IBM ad esempio ha segno negativo, ma ciò non influisce sull'accettabilità della frontiera efficiente corretta, a differenza di quanto accade per il portafoglio A, per il quale il vettore delle medie è costituito da soli elementi positivi.

Tabella 6.5: Vettori delle medie dei portafogli A e B

	Titolo 1	Titolo 2	Titolo 3	Titolo 4
Portafoglio A	0.0002727890	0.0001366708	0.0004888353	0.0003054939
Portafoglio B	9.669509e-04	-7.210281e-05	3.670478e-04	6.138196e-04

Per verificare se il comportamento anomalo della frontiera efficiente corretta dipende dalla struttura di dipendenza tra i diversi titoli o dalle caratteristiche dei titoli stessi, si è provveduto ad aggiungere al portafoglio A uno alla volta i titoli del portafoglio B. Si sono quindi ottenuti i seguenti portafogli:

Portafoglio 1: CAT, CVX, CSCO, KO, HD

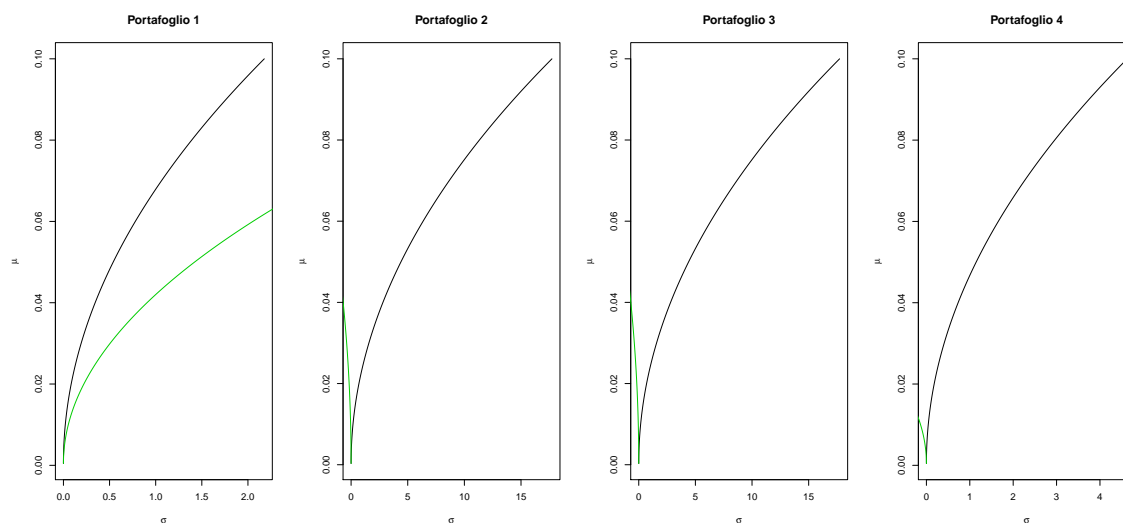
Portafoglio 2: CAT, CVX, CSCO, KO, IBM

Portafoglio 3: CAT, CVX, CSCO, KO, INTC

Portafoglio 4: CAT, CVX, CSCO, KO, JNJ

e di seguito si illustrano le rispettive frontiere efficienti corrette:

Figura 6.5: Rappresentazione grafica delle frontiere efficienti per i portafogli 1, 2, 3 e 4



È evidente che l'unico portafoglio che registra un miglioramento è il Portafoglio 1, creato dal portafoglio A aggiungendo il titolo Home Depot. A tal riguardo è interessante notare come la serie storica dei rendimenti di tale titolo abbia alcune caratteristiche rilevanti: non è lineare (p-value del test Terasvirta inferiore a 0.05), non è distribuita normalmente ed è autocorrelata. Ciò porta quindi a propendere per l'idea che l'errore insito nel calcolo della frontiera efficiente corretta sia imputabile alla matrice di varianza e covarianza Σ (struttura di dipendenza tra i titoli) o a quella del vettore dei rendimenti medi μ . Infatti, i portafogli 1a, 2a, 3a e 4a, rispettivamente così composti:

Portafoglio 1a: IBM, INTC, JNJ

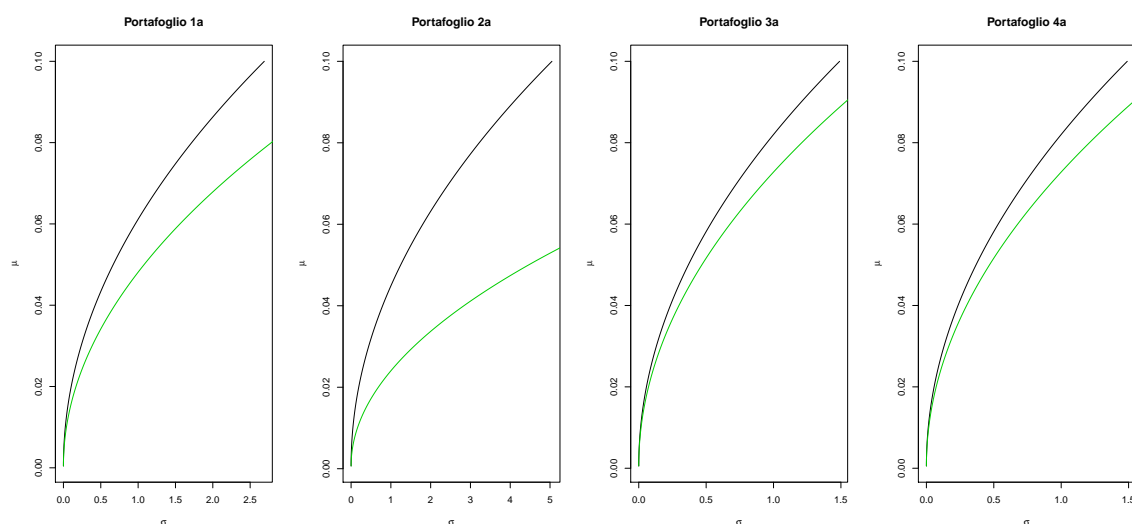
Portafoglio 2a: HD, INTC, JNJ

Portafoglio 3a: HD, IBM, JNJ

Portafoglio 4a: HD, IBM, INTC

mostrano comunque una frontiera efficiente corretta accettabile (come illustrato nel grafico sottostante), con curvatura diversa da caso a caso, dovute a variazioni in Σ e μ .

Figura 6.6: Rappresentazione grafica delle frontiere efficienti per i portafogli 1a, 2a, 3a e 4a



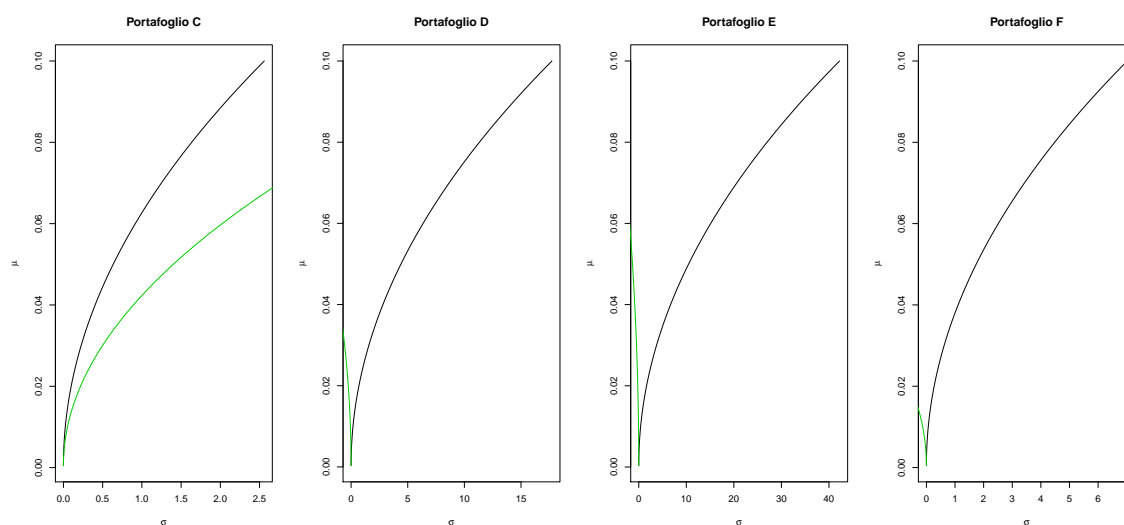
Di notevole rilevanza è il portafoglio 2a, composto interamente da serie storiche autocorrelate e non normali, e da una sola serie lineare, quella dei rendimenti del titolo Johnson & Johnson, il quale produce ugualmente una frontiera efficiente

corretta accettabile, supportando l'ipotesi che l'anomalia nella stima della stessa dipenda dalle quantità μ e Σ .

Da ultimo si presentano i portafogli creati eliminando la serie più problematica (autocorrelata, non lineare e non normale) dal portafoglio A, e sostituendola di volta in volta con una serie del portafoglio B, ottenendo i risultati presentati di seguito.

Appare chiaro che anche in questo caso non sembra esserci legame tra l'anomalia della frontiera e le caratteristiche analizzate del titolo: infatti la frontiera corretta mostra un miglioramento solamente per il caso in cui la serie *Chevron Corporation* sia sostituita con quella di *Home Depot*. Tuttavia, come fatto notare sopra, anche la serie di Home Depot presenta le medesime caratteristiche di quella eliminata.

Figura 6.7: Rappresentazione grafica delle frontiere efficienti per i portafogli C, D, E e F



Portafoglio C: CAT, CSCO, KO, HD; Portafoglio D: CAT, CSCO, KO, IBM; Portafoglio E: CAT, CSCO, KO, INTC; Portafoglio F: CAT, CSCO, KO, JNJ

Assodato quindi che l'assenza di normalità e di linearità e la presenza di autocorrelazione non sono problematiche per il calcolo della frontiera efficiente corretta, resta da analizzare se il comportamento anomalo dipenda dalla stima del vettore delle medie dei titoli o da quella della matrice di varianza e covarianza, oppure da una relazione tra le due quantità. Partendo dalla seconda ipotesi si riporta un confronto tra la matrice di varianza e covarianza del Portafoglio A e quella

del portafoglio C (ottenuto dalla sostituzione del titolo Chevron Corporation con Home Depot):

Tabella 6.6: Matrice di Varianza e Covarianza del Portafoglio A

	CAT	CVX	CSCO	KO
CAT	2.162795e-04	9.679962e-05	7.417240e-05	3.143570e-05
CVX	9.679962e-05	1.571198e-04	6.235280e-05	4.045650e-05
CSCO	7.417240e-05	6.235280e-05	1.919076e-04	3.705898e-05
KO	3.143570e-05	4.045650e-05	3.705898e-05	7.673902e-05

Tabella 6.7: Matrice di Varianza e Covarianza del Portafoglio C

	CAT	HD	CSCO	KO
CAT	2.162795e-04	5.461505e-05	7.417240e-05	3.143570e-05
HD	5.461505e-05	1.259772e-04	5.028327e-05	3.354556e-05
CSCO	7.417240e-05	5.028327e-05	1.919076e-04	3.705898e-05
KO	3.143570e-05	3.354556e-05	3.705898e-05	7.673902e-05

Interessante notare come le due matrici non abbiano differenze sensibili circa l'ordine di grandezza o il segno degli elementi in esse contenuti: l'aver sostituito un titolo con un altro non ha provocato un effetto evidente. Per esaminare meglio l'entità della variazione, si è considerata anche l'analisi delle matrici di correlazione: essendo la misura di correlazione un indicatore adimensionale si può quindi valutare in modo più accurato il cambiamento negli elementi costituenti la matrice stessa.

Tabella 6.8: Matrice di Correlazione del Portafoglio A

	CAT	CVX	CSCO	KO
CAT	1.0000000	0.5251103	0.3640732	0.2440099
CVX	0.5251103	1.0000000	0.3590826	0.3684383
CSCO	0.3640732	0.3590826	1.0000000	0.3053791
KO	0.2440099	0.3684383	0.3053791	1.0000000

Tabella 6.9: Matrice di Correlazione del Portafoglio C

	CAT	HD	CSCO	KO
CAT	1.0000000	0.3308710	0.3640732	0.2440099
HD	0.3308710	1.0000000	0.3233936	0.3411780
CSCO	0.3640732	0.3233936	1.0000000	0.3053791
KO	0.2440099	0.3411780	0.3053791	1.0000000

Dall'analisi dei componenti di queste matrici è possibile notare come la dipendenza tra i titoli Caterpillar e Chevron Corporation sia più marcata rispetto a quella riscontrata per le serie Caterpillar e Home Depot. La relazione tra Home Depot e gli altri due titoli non sembra essere di molto diversa da quella del titolo Chevron Corporation nei confronti degli stessi.

Visto che l'osservazione dei singoli elementi delle matrici non ha prodotto risultati di nota si è valutato il determinante di entrambe, riscontrando una stretta somiglianza. Infatti indicando con Σ_A e Σ_C le matrici di varianza e covarianza dei rispettivi portafogli, i determinanti calcolati risultano essere: $\det(\Sigma_A) = 2.491003e-16$ e $\det(\Sigma_C) = 2.447849e-16$.

Si è successivamente analizzato il vettore delle medie dei rendimenti delle serie storiche facenti parte dei due portafogli: si è potuto constatare che il rendimento medio di Home Depot è nettamente maggiore di quello di Chevron Corporation. Si riportano di seguito i vettori delle medie di entrambi i portafogli:

Tabella 6.10: Vettori delle medie dei portafogli A e C

	Titolo 1	Titolo 2	Titolo 3	Titolo 4
Portafoglio A	0.0002727890	0.0001366708	0.0004888353	0.0003054939
Portafoglio C	0.0002727890	0.0009669509	0.0004888353	0.0003054939

L'aver sostituito il secondo titolo del portafoglio ha portato all'innalzamento della media del vettore, portandola da 0.0003009472 a 0.0005085173. Infatti si nota come il rendimento medio di Home Depot (Titolo 2 del Portafoglio C) sia quasi un ordine di grandezza più elevato di quella di Chevron Corporation (Titolo 2 del Portafoglio A).

Per chiarire quindi se il comportamento anomalo della frontiera efficiente corretta dipenda dalla stima della matrice di varianza e covarianza, da quella del

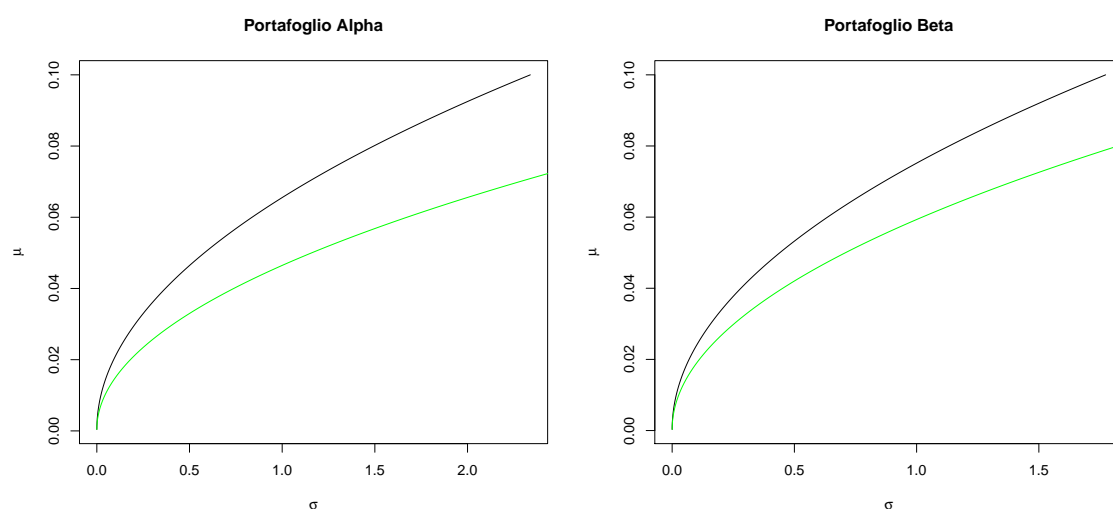
vettore dei rendimenti medi oppure dalla relazione che lega entrambi, si sono costruiti due portafogli virtuali, uno con vettore delle medie μ pari a quello del Portafoglio C e con matrice di varianza e covarianza pari a quella del Portafoglio A, denominato Portafoglio Alpha; l'altro, denominato Portafoglio Beta, è sostanzialmente una modifica del Portafoglio A, con vettore μ pari a quello del portafoglio di origine ma con matrice Σ ridotta di un ordine di grandezza. Questo perché considerando che, come esposto da Corazza e Pizzi (2014), il comportamento anomalo della frontiera efficiente corretta dipende dal fattore s^1 , e che tale valore contiene sia il vettore μ sia la matrice Σ allora non si è in grado di determinare se il problema nasca unicamente dai valori contenuti in μ , da quelli in Σ oppure se dipenda anche dalla relazione tra i due.

Il risultato ottenuto mostra per entrambi i casi un miglioramento della frontiera efficiente "à la Bonar e Bodnar", la cui varianza passa da negativa a positiva, e quindi da inaccettabile ad accettabile.

Questo risultato fa propendere per l'ipotesi che sia rilevante non tanto i valori assunti da μ o da Σ , quanto la relazione che lega entrambi e che si concretizza nel prodotto vettoriale che dà origine al parametro s .

Di seguito sono illustrate entrambe le frontiere.

Figura 6.8: Rappresentazione grafica delle frontiere efficienti per il Portafoglio Alpha e per il Portafoglio Beta



¹Per le condizioni specifiche che devono essere rispettate si rimanda al paragrafo 3.3 del capitolo terzo.

Concludendo questo capitolo si può quindi affermare che il problema relativo all'inaccettabilità della frontiera efficiente "à la Bodnar e Bodnar" è originato dalla relazione esistente tra i rendimenti medi dei titoli inseriti nel portafoglio e le rispettive varianze e covarianze. Variazioni anche minime nelle stime di μ e di Σ producono una forte alternanza nei risultati, portando la frontiera efficiente corretta ad assumere valori negativi qualora non siano rispettate le condizioni poste sul parametro s derivate da Corazza e Pizzi (2014) e qui illustrate nel capitolo terzo.

CONCLUSIONI

In questo ultimo capitolo si fornisce un riepilogo generale delle analisi svolte nel presente lavoro.

La frontiera efficiente corretta fornita in Bodnar e Bodnar (2010) risulta essere valida e statisticamente differente da quella campionaria (Markowitz, 1952 e Merton, 1972) sia per i portafogli con rendimenti dei titoli distribuiti normalmente sia per quelli composti da serie storiche caratterizzate da una distribuzione diversa da quella normale. Per i primi tuttavia si evidenzia il comportamento asintotico della frontiera corretta, la quale al crescere delle osservazioni dei rendimenti per ogni titolo, tende al valore della frontiera campionaria. Tale comportamento non è stato riscontrato per i dati distribuiti secondo una *t* di Student con quattro gradi di libertà.

L'analisi svolta sui dati autocorrelati porta a rifiutare l'ipotesi di una qualsiasi influenza sul risultato finale dovuto all'autocorrelazione. Infatti, ipotizzando un modello ARMA(1,1), le frontiere in esame risultano diverse sia per i processi con residui distribuiti normalmente sia per quelli con residui distribuiti secondo una *t* di Student (ipotesi di distribuzione non normale). Anche in questo caso però, per i portafogli composti da serie con residui normalmente distribuiti si è riscontrato il medesimo comportamento asintotico illustrato in precedenza, affievolendosi di intensità mano a mano che si riduce il numero di osservazioni, in linea con quanto affermato a conclusione del capitolo terzo.

Relativamente all'influenza del tipo di distribuzione seguita dalle serie dei rendimenti dei titoli costituenti il portafoglio emerge che l'utilizzo di dati non normali inficia notevolmente la quantità correttiva, portando la frontiera efficiente "à la Bodnar e Bodnar" a non essere necessariamente una frontiera statisticamente corretta: infatti le quantità correttive introdotte dai due autori sono derivate sulla

base dell'assunzione di normalità distributiva dei rendimenti, e venendo meno la stessa, la correzione applicata non è più valida.

L'effetto di tale invalidità lo si riscontra nell'assenza del comportamento asintotico della frontiera corretta (la quale, come illustrato al capitolo terzo, all'aumentare della numerosità del campione disponibile per ogni serie tende al valore della frontiera efficiente campionaria) comportando o una assenza totale di convergenza oppure la necessità di un numero di osservazioni maggiori per il suo raggiungimento. L'ipotesi relativa alla distribuzione dei rendimenti impiegati per la stima della frontiera efficiente corretta risulta quindi essenziale affinché essa sia uno stimatore corretto della vera frontiera efficiente dei portafogli, inficiandone quindi la robustezza.

Infine, per quanto riguarda la criticità della frontiera efficiente corretta relativa alla negatività della varianza riscontrata in alcuni portafogli, si può affermare che essa dipenda, da quanto è emerso dalle analisi esplorative svolte nel capitolo sesto, in misura ampiamente rilevante non da caratteristiche intrinseche nelle serie storiche (assenza di linearità, distribuzioni non normali, presenza di autocorrelazione) ma dalla relazione esistente tra la matrice di varianze e covarianze del portafoglio e il vettore dei rendimenti medi dei titoli dello stesso, il cui prodotto è parte costituente del parametro s di pendenza della frontiera, come illustrato nei capitoli secondo e terzo.

Concludendo, si può affermare che la correzione proposta da Bodnar e Bodnar (2010) è applicabile per dati con distribuzione normale con presenza o meno di autocorrelazione; per serie storiche caratterizzate da non normalità, invece, l'assenza di convergenza per elevati numeri di osservazioni è indice di mancanza di validità della componente correttiva, e pertanto il suo utilizzo non consente di eliminare la distorsione della frontiera campionaria. Un utilizzo pratico della stessa risulta inoltre inficiato dal già discusso problema di negatività della varianza: una frontiera efficiente corretta di tale forma infatti non è di alcuna utilità per operatori ed investitori, in quanto non fornisce nessun miglioramento informativo, producendo solamente un nonsenso matematico ed economico. Si può quindi affermare che se per serie storiche di rendimenti mensili tale problema è meno frequente, per osservazioni giornaliere non è poi così raro: visto il netto incremento di informazione che fornisce un'analisi di dati giornalieri rispetto a

CONCLUSIONI

quelli mensili, l'applicazione pratica della correzione esaminata risulta sì attuabile, ma non costituisce una procedura affidabile e sostituta della classica frontiera efficiente "markowitziana", come invece avviene per osservazioni mensili.

Appendice A

RICHIAMI SULLA DERIVAZIONE DELLA FRONTIERA EFFICIENTE "À LA MARKOWITZ"

A.1 Funzione di utilità

In economia un ruolo fondamentale è svolto dalle preferenze dei soggetti circa la scelta di beni/investimenti. Essi infatti realizzano scelte personali di gestione delle proprie risorse -necessariamente limitate e non infinite-, decidendo quindi su quali beni convogliarle.

Nel caso specifico di interesse di questo lavoro di tesi, ovvero la scelta tra diversi investimenti, la funzione di utilità permette di comprendere e di misurare la soddisfazione che ogni investitore trae dall'investimento sottoscritto, soddisfazione che dipende sia da parametri soggettivi (avversione al rischio) sia da parametri oggettivi (deviazione standard e rischio dell'investimento).

In tal caso quindi si necessita di uno strumento capace di raccogliere e misurare questi diversi aspetti.

La funzione di utilità risulta così descritta:

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

In tale rappresentazione le quantità indicate hanno il seguente significato:

$E(r)$ rappresenta il rendimento atteso dell'investimento, σ^2 invece rappresenta il rischio dello strumento finanziario considerato, mentre A rappresenta il grado di avversione al rischio dell'investitore.

L'utilità quindi risulta essere direttamente proporzionale al rendimento atteso e inversamente proporzionale al rischio e alla avversione al rischio. Più l'aspetto soggettivo di avversione al rischio è pronunciato, più l'utilità ricavata dallo strumento è bassa.

A.2 Eguaglianza tra approccio semplificato e non nel processo di minimizzazione

Dalla risoluzione di del problema di ottimizzazione di \mathcal{L} , senza moltiplicare $\omega' \Sigma \omega$ per $\frac{1}{2}$ si ottiene il seguente valore di ω :

$$\omega = \frac{\lambda}{2} \Sigma^{-1} \mathbf{r} + \frac{\gamma}{2} \Sigma^{-1} \mathbf{1}$$

Il che porta, dopo aver operato le sostituzioni enunciate nel capitolo, alla risoluzione del seguente sistema:

$$\frac{1}{2} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \frac{1}{(ca - b^2)} \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} \mu_p \\ 1 \end{pmatrix}$$

Le soluzioni per λ e γ risultano essere le seguenti:

$$\gamma = \frac{2(a - \mu_p b)}{ca - b^2} \quad \lambda = \frac{2(\mu_p c - b)}{ca - b^2}$$

Operata la sostituzione dei valori λ e γ all'interno del risultato ω si ottiene:

$$\begin{aligned} \omega &= \frac{2(\mu_p c - b)}{2(ca - b^2)} \Sigma^{-1} \mathbf{1} + \frac{2(a - \mu_p b)}{2(ca - b^2)} \Sigma^{-1} \mathbf{r} \\ &= \frac{(\mu_p c - b)}{(ca - b^2)} \Sigma^{-1} \mathbf{1} + \frac{(a - \mu_p b)}{(ca - b^2)} \Sigma^{-1} \mathbf{r} \end{aligned}$$

Ovvero il medesimo risultato ottenuto mediante l'iniziale semplificazione dei conti.

A.3 Derivazione grafica della frontiera efficiente dei portafogli

La prima spiegazione riguardante la frontiera efficiente dei portafogli fu fornita da Markowitz (1952) e si basa esclusivamente su una trattazione grafica della stessa.

Per prima cosa si inizia con lo spiegare la differenza tra portafogli efficienti ed inefficienti, ovvero tra quei portafogli per i quali è impossibile aumentare il rendimento atteso senza un conseguente aumento del rischio assunto (portafogli efficienti, sono già in una condizione ottimale) e quei portafogli che invece consentono, a parità di rendimento di ridurre il rischio, oppure a parità di rischio di aumentare il rendimento atteso (portafogli inefficienti, ovvero migliorabili).

Un portafoglio efficiente è quindi caratterizzato dall'aver il minimo rischio assumibile per il rendimento atteso che offre, oppure, equivalentemente, il massimo rendimento atteso per il livello di rischio assunto.

Markowitz, non deriva il set di portafogli efficienti da un punto di vista matematico, ma procede per una via differente, preferendo un approccio grafico al problema.

Dapprima egli analizza il caso di un portafoglio formato da tre titoli:

$$\begin{aligned}
 E &= \sum_{i=1}^3 X_i \mu_i \\
 V &= \sum_{i=1}^3 \sum_{j=1}^3 X_i X_j \sigma_{ij} \\
 \sum_{i=1}^3 X_i &= 1 \\
 X_i &\geq 0 \quad \text{for } i = 1, 2, 3
 \end{aligned}$$

e di conseguenza si ottiene che

$$X_3 = 1 - X_1 - X_2$$

e quindi sia il rendimento atteso E , sia la varianza V , possono essere riscritti in funzione dei soli pesi X_1 e X_2 :

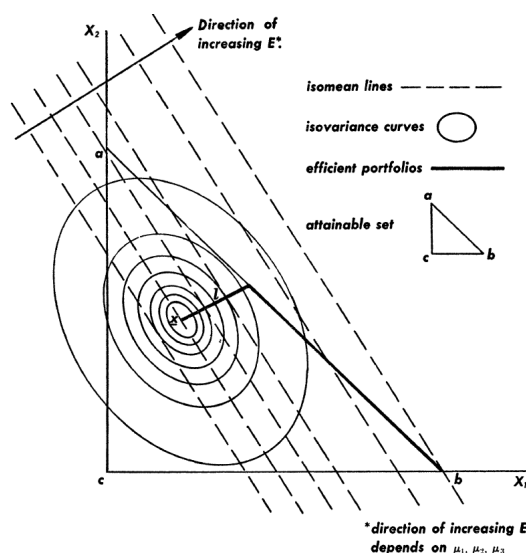
$$E = E(X_1, X_2)$$

$$V = V(X_1, X_2)$$

$$X_1 \geq 0, \quad X_2 \geq 0, \quad 1 - X_1 - X_2 \geq 0$$

Ovviamente il vincolo sui pesi deve essere rispettato: la somma dei pesi deve essere pari a 1 e nessuno peso può avere valore negativo (cioè non sono ammesse vendite allo scoperto). In tal modo Markowitz ricava due set di curve: un set di curve aventi la stessa media, *isomean curve* (il medesimo valore atteso) e un altro di curve aventi la stessa varianza, *isovariance curve*. Di seguito si riporta la rappresentazione grafica fornita dallo stesso Markowitz:

Figura A.1: Set di portafogli accettabili



Markowitz H., *Portfolio Selection*, The Journal of Finance, 1952

Le *isomean curves* consistono in un sistema di rette parallele, mentre le *isovariance curves* formano un set di ellissi concentriche, il cui centro rappresenta il punto a varianza minima. Più ci si allontana dal centro, più la varianza aumenta: devono però essere rispettati i vincoli predetti, così che un portafoglio può giacere solo all'interno dell'area rappresentata dal triangolo, i cui vertici sono nominati: *a*, *b* e *c*.

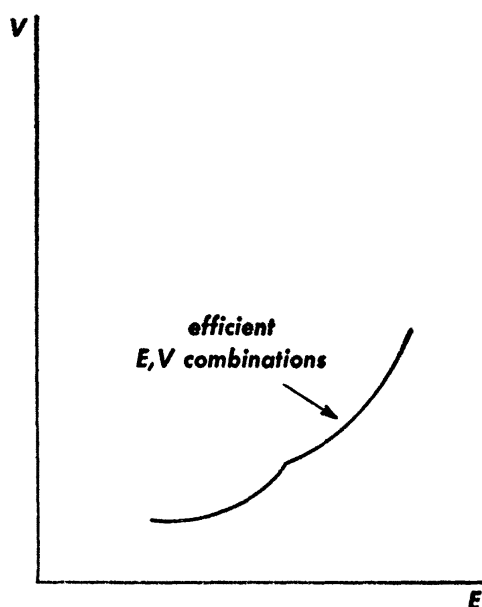
Dall'analisi di Markowitz emerge come il punto *X*, cioè il centro dell'ellisse, possa giacere anche all'esterno del triangolo sopra definito.

Relativamente alle curve a medesima media, è intuitivo notare come per una stessa retta vi siano più punti di intersezione con il set di isovariance curve. Un portafoglio efficiente tra quelli rispettanti i vincoli, è quindi il portafoglio che, a

parità di valore atteso, giace il più vicino possibile al centro del sistema di ellissi. In tal modo la varianza ne risulta minimizzata e il rendimento resta invariato.

Il set delle coppie efficienti (E,V) risulta così ricavabile dal caso in cui si considerano solo tre titoli (per numerosità di titoli più elevata il ragionamento resta comunque valido). Infatti in tal caso, E risulta essere un piano e V un paraboloidale: il set di combinazioni efficienti è quindi dato da un insieme di segmenti, che rappresentati in un piano (E,V) , producono la seguente rappresentazione grafica:

Figura A.2: Set di combinazioni (E,V) efficienti



Markowitz H., *Portfolio Selection*, The Journal of Finance, 1952

Tale insieme di combinazioni efficienti per la coppia di variabili (E,V) è la prima definizione e rappresentazioni della frontiera efficiente dei portafogli.

Appendice B

APPROFONDIMENTO SULL'INTERVALLO DI CONFIDENZA PER LA FRONTIERA EFFICIENTE E DIMOSTRAZIONE TEORICA

In questa sezione si fornisce la spiegazione dettagliata del lavoro di Bodnar e Schmid (2008) in cui derivano esattamente l'intervallo di confidenza per un livello di significatività α relativo alla frontiera efficiente dei portafogli.

Considerato che dalla condizione (2.4) del capitolo secondo si ha a che fare con un vertice fissato e definito da V_{GMV} e da R_{GMV} , il limite superiore è dato da:

$$(R - R_{GMV})^2 = \hat{s}_{1-\alpha^*/2}(V - V_{GMV}) \quad (B.1)$$

Il vertice invece può assumere valori definiti dalle condizioni (2.3) e (2.2).

Per costruire l'intervallo prima di tutto gli autori calcolano un limite superiore per ogni frontiera efficiente: tale set consiste in differenti parabole, da scegliere per diversi valori di R .

Visto che la localizzazione della parabola nel piano è determinata dal vertice (e tutte le parabole hanno il medesimo coefficiente di slope) ne deriva che R_{GMV} e V_{GMV} sono in un qualche modo funzioni di R . Risolvendo quindi la (B.1) rispetto

**APPENDICE B. DIMOSTRAZIONE TEORICA DELL'INTERVALLO DI
CONFIDENZA**

a V si ottengono i punti giacenti sulla parabola:

$$V = V_{GMV} + \frac{1}{\hat{s}_{1-\alpha^*/2}} (R - R_{GMV})^2$$

$$z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} (R_{GMV} - \hat{R}_{GMV})^2 + \frac{1}{\hat{s}_{1-\alpha^*/2}} (R - R_{GMV})^2 \quad (\text{B.2})$$

La quantità $V = V(R_{GMV})$ è minimizzata rispetto a $R_{GMV} \in [\hat{R}_{GMV} + g_l, \hat{R}_{GMV} + g_u]$, e g_l e g_u sono le quantità già definite nel secondo capitolo dalla (2.5). La soluzione di tale minimizzazione (ottenuta ponendo la derivata prima della (B.2) uguale a zero) è la seguente:

$$R_{GMV;m} = tR + (1-t)\hat{R}_{GMV} \quad (\text{B.3})$$

per la quale t assume il valore già richiamato dall'equazione $t = \frac{1}{1 + z_{1-\alpha^*/2}^{-2} \frac{\hat{s}_{1-\alpha^*/2} \frac{n(n-1)}{n(1+\hat{s})-1}} \leq 1$ nel secondo capitolo.

Dato che la derivata secondo dell'equazione (B.2) è:

$$V''(R_{GMV}) = 2z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} + \frac{2}{\hat{s}_{1-\alpha^*/2}} > 0 \quad (\text{B.4})$$

ne deriva che per un dato valore di R la parabola che delinea il limite dell'intervallo ha vertice $(tR + (1-t)\hat{R}_{GMV}, z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} t^2 (R - \hat{R}_{GMV})^2)$ considerato che $V_{GMV;m} = z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} (R_{GMV;m} - \hat{R}_{GMV})^2$.

Si fa notare come la quantità $tR + (1-t)\hat{R}_{GMV} \in [\hat{R}_{GMV} + g_l, \hat{R}_{GMV} + g_u]$ se e solo se $R \in \mathbb{I}_2^u = [\hat{R}_{GMV} + g_l/t, \hat{R}_{GMV} + g_u/t]$.

Sostituendo l'equazione (B.3) nella (B.2) si ottiene:

$$V = z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} t(R - \hat{R}_{GMV})^2 = \frac{z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1}}{1 + z_{1-\alpha^*/2}^{-2} \hat{s}_{1-\alpha^*/2} \frac{n(n-1)}{n(1+\hat{s})-1}} (R - \hat{R}_{GMV})^2. \quad (\text{B.5})$$

Nel caso in cui $R \in \mathbb{I}_1^u = [\hat{R}_{GMV} + g_l, \hat{R}_{GMV} + g_l/t]$ il minimo di V si ottiene se si pone $R_{GMV;m} = \hat{R}_{GMV} + g_l$ e di conseguenza:

$$V = \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;1-\alpha^*/2}^2} + \hat{s}_{1-\alpha^*/2}^{-1} (R - \hat{R}_{GMV} - g_l)^2 \quad (\text{B.6})$$

APPENDICE B. DIMOSTRAZIONE TEORICA DELL'INTERVALLO DI CONFIDENZA

Per $R \in \mathbb{I}_3^u = (\hat{R}_{GMV} + g_u/y, \infty)$ ovvero $R_{GMV;m} = \hat{R}_{GMV} + g_u$, si ottiene che:

$$V = \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha^*}^2} + \hat{s}_{1-\alpha^*/2}^{-1}(R - \hat{R}_{GMV} - g_u)^2 \quad (\text{B.7})$$

Unendo tutte le precedenti condizioni, il limite superiore per la regione di confidenza è formato da tre parabole:

$$V = \begin{cases} \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;1-\alpha^*/2}^2} \hat{s}_{1-\alpha^*/2}^{-1}(R - \hat{R}_{GMV} - g_l)^2 & \text{con } R \in \mathbb{I}_1^u \\ \frac{z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1}}{1 + z_{1-\alpha^*/2}^{-2} \hat{s}_{1-\alpha^*/2} \frac{n(n-1)}{n(1+\hat{s})-1}} (R - \hat{R}_{GMV})^2 & \text{con } R \in \mathbb{I}_2^u \\ \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha^*/2}^2 + \hat{s}_{1-\alpha^*/2}^{-1}} (R - \hat{R}_{GMV} - g_u)^2 & \text{con } R \in \mathbb{I}_3^u \end{cases} \quad (\text{B.8})$$

In modo simile è possibile ricavare anche il limite inferiore per la regione di confidenza, anche se in tal caso si è interessati alla curva che giace al di sotto di tutte le frontiere efficienti con vertice situato nella parte inferiore della parabola definita dalla condizione (2.2).

Considerato che $\hat{R}_{GMV} > R_{GMV}$ e che $R_{GMV} < R$ allora:

$$V(R_{GMV})' = 2z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} (R_{GMV} - \hat{R}_{GMV}) + \frac{2}{\hat{s}_{\alpha^*/2}} (R_{GMV} - R) < 0$$

e il massimo della funzione

$$\begin{aligned} V(R_{GMV}) &= V_{GMV} + \frac{1}{\hat{s}_{\alpha^*/2}} (R - R_{GMV})^2 \\ &= z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{s})-1} (R_{GMV} - \hat{R}_{GMV})^2 + \frac{1}{\hat{s}_{\alpha^*/2}} (R - R_{GMV})^2 \end{aligned} \quad (\text{B.9})$$

è raggiunto in $(\hat{R}_{GMV} - g_u, \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha/2}^2})$.

Ponendo $\mathbb{I}_1^l = [\hat{R}_{GMV} - g_u, \hat{r}_{GMV} - g_l]$ e $\mathbb{I}_2^l = [\hat{R}_{GMV} - g_u, +\infty)$ il limite inferiore

*APPENDICE B. DIMOSTRAZIONE TEORICA DELL'INTERVALLO DI
CONFIDENZA*

della regione di confidenza è definito dalle seguenti due parabole:

$$V = \begin{cases} V \leq z_{1-\alpha^*/2}^{-2} \frac{n(n-1)}{n(1+\hat{\delta})-1} (R - \hat{R}_{GMV})^2 & \text{con } R \in \mathbb{I}_1^l \\ V \geq \frac{(n-1)\hat{V}_{GMV}}{\chi_{n-k;\alpha^*/2}^2} + \hat{\delta}_{\alpha^*/2}^{-1} (R - \hat{R}_{GMV} + g_u)^2 & \text{con } R \in \mathbb{I}_2^l \end{cases} \quad (\text{B.10})$$

Il risultato del teorema enunciato nel secondo capitolo è così dimostrato.

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] BÜHLMANN P., Sieve bootstrap for time series, Technical report No. 431, Dept. of Statistics University of California, Berkeley, 1995
- [5] CHAN K.S., On the validity of the method of surrogate data, *Fields Institute Communications*, 11, pp. 77-97, 1997
- [6] CORAZZA M. E PIZZI C., Comunicazione a Mathematical and Statistical Methods for Actuarial Sciences and Finance, 2014
- [7] EFRON B., Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, 7, pp. 1-26, 1979
- [8] EFRON B., TIBSHIRANI R., Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, 1(1), pp. 54, 75, 1986
- [9] FAMA E., *Foundations of finance*, New York: Basic Books, 1976
- [10] FOURIER J.B.J., *Théorie analytique de la chaleur*, 1822

BIBLIOGRAFIA

- [11] FREEDMAN D.A., Bootstrapping regression models, *The Annals of Statistics*, 9, pp. 1218-1228, 1984
- [12] HALL P., HOROWITZ J.L. E JING B.-Y, On blocking rules for the bootstrap with dependent data, *Biometrika*, 82(3), pp. 561-574, 1995
- [13] JOBSON J.D., Confidence regions for the mean-variance efficient set: An alternative approach to estimation risk, *Review of Quantitative Finance and Accounting*, 1, pp. 235-257, 1991
- [14] JOBSON J., KORKIE B., Estimation for Markowitz Efficient Portfolios, *Journal of the American Statistical Association*, 75(371), p. 544, 1991
- [15] KAN R. E SMITH D., The Distribution of the Sample Minimum-Variance Frontier, *Management Science*, 54(7), pp. 1364-1380, 2008
- [16] KREISS J.P., FRANKE J., Bootstrapping stationary ARMA models, *Journal of Time Series Analysis*, 13(4), pp. 297-317, 1992
- [17] KREISS J.P. E LAHIRI S.N., Bootstrap Method for time series, in *Time series analysis: methods and Applications*, Elsevier, Amsterdam, 2012
- [18] KÜNSCH H. R., The jackknife and the bootstrap for general stationary observation, *The Annals of Statistics*, 17(3), pp. 1217-1241, 1989
- [19] LAM Y.-M., Confidence Limits for Non-Centrality Parameters of Non-central Chi-Squared and F Distributions, *ASA Proceedings of the Statistical Computing Section*, pp. 441-443, 1987
- [20] MARKOWITZ H., Portfolio Selection, *The Journal of Finance*, 7(1), pp. 77-91, 1952
- [21] MARKOWITZ H., *Portfolio Selection*, Cambridge, Mass.: B. Blackwell, 1991
- [22] MERTON R., An Analytic Derivation of Efficient Portfolio Frontier, *The Journal of Finance and Quantitative Analysis*, 7(4), pp. 1851-1872, 1972
- [23] PRICHARD D., THEILER J., Generating surrogate data for time series with several simultaneously measured variables, *Physical Review Letters*, 73(7), 1994

BIBLIOGRAFIA

- [24] SCHREIBER T., Interdisciplinary application of nonlinear time series methods, *Physics Reports*, 308, pp. 11-64, 1999
- [25] SCHREIBE T., SCHMITZ A., Surrogate time series, *Physica D*, 142, pp. 346-382, 2000
- [26] SIEGEL A. E WOODGATE A., Performance of Portfolios Optimized with Estimation Error, *Management Science*, 53(6), pp. 1005-1015, 2007
- [27] SWANEPOEL J.W., VAN WYK J.W.Y., The bootstrap applied to power spectral density function estimation, *Biometrika*, 73, pp. 135 -141, 1986
- [28] THEILER J., EUBANK S., LONGTIN A., GALDRIKIAN B., DOYNE FARMER J., Testing for non linearity in time series: the method of surrogate data, *Physica D*, 58, pp. 77-94, 1992
- [29] THEILER J., EUBANK S., LONGTIN A., GALDRIKIAN B., DOYNE FARMER J., Using surrogate data to detect non linearity in time series, in *Nonlinear Modelling and Forecasting*, Eds. M. Casdagli and S. Eubank, Santa Fe Institute, Addison-Wesley, pp. 163-188, 1992
- [30] TU J. E ZHOU G., Data-generating process uncertainty: What difference does it make in portfolio decisions?, *Journal of Financial Economics*, 72(2), pp. 385-421, 2004
- [31] TU J. E ZHOU G., Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics*, 99(1), pp. 204-215, 2011
- [32] YIN Y., SHANG P., Modified cross sample entropy and surrogate data analysis method for financial time series, *Physica A*, 433, pp. 17-25, 2015