



UNIVERSITA' DEGLI STUDI DI PADOVA
FACOLTA' DI SCIENZE STATISTICHE
CORSO DI LAUREA IN
SCIENZE STATISTICHE, ECONOMICHE,
FINANZIARIE E AZIENDALI

TESI DI LAUREA SPECIALISTICA

CLUSTERING DI FONDI
COMUNI D'INVESTIMENTO ED EFFETTI
SULLA VALUTAZIONE DELLA PERFORMANCE

RELATORE: CH.MO PROF. MASSIMILIANO CAPORIN

LAUREANDA: KLODIANA NUSHI

MARTICOLA: 531074_SEA

ANNO ACCADEMICO 2007/2008

INDICE

Introduzione	3
Capitolo 1	
Fondi Comuni d’Investimento e Rendimenti	9
1.1 Il Risparmio Gestito	10
1.2 I Fondi Comuni d’Investimento	12
1.2.1 Classificazione dei fondi comuni d’investimento.....	14
1.2.2 Struttura dei fondi comuni e costi.....	16
1.3 Nozioni sui Mercati Finanziari	21
1.3.1 Strumenti finanziari e indici di mercato.....	23
1.4 Alcuni concetti sui Rendimenti	26
1.4.1 Risk-free e extrarendimenti.....	29
Capitolo 2	
CML,CAPM e Indici di Performance	31
2.1 CML, Modello di Markowitz	32
2.1.1 Costruzione della CML e FE.....	33
2.2 CAPM: Regressioni in Serie Storica	38
2.3 Strategie di Gestione del Portafoglio	43
2.4 Definizione della Performance	49
2.4.1 Scelta del benchmark.....	51
2.5 Misure di Performance sui Rendimenti	54
2.5.1 Misure di performance basate su modelli econometrici.....	55
2.5.1.1 Misura Alpha di Jensen.....	55
2.5.1.2 Misura Beta del portafoglio.....	57
2.5.2 Misure di RAP – Risk Adjusted Performance.....	58
2.5.2.1 Indice di Sharpe.....	59

2.5.2.2 Indice di Treynor.....	60
2.5.2.3 Information Ratio.....	62
2.5.2.4 Indice di Sortino.....	65
Capitolo 3	
Cluster Analysis.....	67
3.1 Prossimità e Distanze.....	68
3.1.1 Distanza Euclidea.....	69
3.1.2 Indice di Similarità.....	73
3.1.3 Metodo del Multidimensional Scaling.....	76
3.2 Analisi di Raggruppamento.....	77
3.2.1 Metodi gerarchici di classificazione.....	80
3.2.1.1 Valutazione della classificazione del metodo gerarchico.....	84
3.2.2 Metodi non gerarchici di classificazione.....	87
3.4 Approccio di Sharpe, Brown e Goetzman.....	89
Capitolo 4	
Analisi dei fondi comuni d'investimento USA.....	91
4.1 I Fondi Comuni Statunitensi.....	91
4.2 Caratteristiche del dataset.....	95
4.2.1 Analisi prezzi e rendimenti.....	97
4.2.2 Performance del portafoglio.....	103
4.3 Cluster analysis dei fondi comuni USA.....	109
4.4 Effetti sulla performance dopo il clustering.....	117
4.4.1 Valutazioni dal Clustering sugli indici.....	132
Conclusioni.....	133
Appendice.....	137
Bibliografia.....	143
Webgrafia.....	145

INTRODUZIONE

Dataset di grandi dimensioni sui dati finanziari sono stati resi disponibili agli investitori soltanto recentemente. L'insieme delle informazioni grezze contenute in questi dataset è un modo abbastanza complesso e lento, per capire e valutare gli strumenti finanziari a cui si è interessati, in quanto non informativi dal punto di vista economico-finanziario. Di conseguenza, si ha la necessità di elaborare l'informazione grezza mediante una corretta analisi modellistica per poter ottenere il massimo delle informazioni senza errori grossolani. Proprio la ricchezza delle informazioni contenute in dataset a dimensioni mutevoli, hanno motivato e continuano tutt'ora le ricerche in analisi quantitativa finanziaria, di conseguenza hanno portato a individuare strumenti nuovi per la loro analisi, cosa che è sempre in continuo progresso. Su queste basi trovano sostegno la teoria statistica classica e software sofisticati, che uniti agli algoritmi genetici e reti neurali artificiali tendono a rappresentare un nuovo modo di concepire e affrontare "l'analisi finanziaria", sia da persone altamente qualificate e sia da tutti coloro che sono semplicemente appassionati dell'analisi finanziaria.

Alla luce delle varie ricerche e scoperte in campo finanziario, uno degli obiettivi di questa tesi è quello di ricavare informazioni realistiche, rappresentative ed empiricamente verificate per i fondi comuni d'investimento statunitensi. Le informazioni ottenute devono essere bene interpretate e sufficienti per valutare la performance dei fondi durante l'intera permanenza nel mercato statunitense e per guidare l'investitore nelle sue valutazioni e scelte di investimento. Infatti, i fondi verranno classificati e raggruppati in modo da estrarre maggiore informazione possibile per avere valutazioni più accurate della performance e imparare a interpretare il loro comportamento rispetto al mercato. In quanto, sia nelle fasi propizie dei mercati e sia in quelle avverse, in ambito finanziario non bisogna smettere di raggruppare e selezionare, in quanto il mercato è in continuo movimento e molto spesso è soggetto a forti ribassi o rialzi. I cambiamenti

inaspettati del mercato sono molto spesso dovuti, ad effetti della politica predominante di un paese oppure conseguenza delle decisioni prese dalla Commissione Europea.

I fondi comuni rappresentano un portafoglio finanziario manovrabile da investitori professionisti e da tutti gli appassionati dell'analisi finanziaria. Se nel passato le banche offrivano prevalentemente fondi monomarca agli investitori, oramai tutto questo è cambiato, in quanto i promotori finanziari sono sempre più competitivi e le loro scelte sono legate alla qualità dei fondi, fattore principale questo ultimo per poter ottenere massimo profitto. I clienti nella maggior parte dei casi non cercano di conoscere a fondo ciò che acquistano ma, su affidamento delle informazioni date dalla banca¹, che collocano il loro denaro su alcuni fondi comprando parte delle quote e acconsentono all'investimento, come deciso dal contratto precedentemente concordato con il promotore finanziario. Solo una volta decisa la ripartizione del portafoglio tra azionario, obbligazionario e liquidità in modo coerente con i propri bisogni finanziari e in base alla propria tolleranza al rischio, l'investitore con l'aiuto del promotore può procedere alla scelta dei fondi con cui realizzare concretamente *l'asset allocation*.

Lo stile d'investimento del promotore finanziario è molto importante, in quanto generalmente valuta dove il fondo deve essere impiegato e verifica che tipo di strategia d'investimento seguire. In particolare, gli **strumenti azionari** sono molto difficili da valutare (bisogna conoscere il profilo della società e avere informazioni riguardanti i reali valori economici), di conseguenza, affidarsi a professionisti può effettivamente creare notevoli differenze sotto il profilo rischio/rendimento. Questi prodotti si distinguono anche per la maggiore esposizione alle avversità dei mercati rispetto ad altre categorie di investimento. Con i **fondi bilanciati** è cruciale la capacità del gestore di impostare un'asset allocation vincente tra azioni e obbligazioni. Per i **fondi obbligazionari** sono, invece importanti il posizionamento del gestore sulla curva dei tassi e la scelta tra bond societari e governativi. La complessità di gestione diminuisce con i **fondi di liquidità**, che hanno un raggio di azione molto ridotto, dal momento che i gestori possono diversificare il patrimonio solo in obbligazioni di breve termine. Una volta definito il portafoglio iniziale di fondi occorre sempre monitorare i propri investimenti. Periodicamente

¹ Le quali devono essere veritiere, secondo Basilea 2;

l'investitore dovrebbe accertarsi che i proprio prodotti scelti siano in linea con le proprie esigenze, che nel tempo possono mutare. Tali modifiche possono riguardare anche le caratteristiche e la qualità del fondo stesso, nonché l'eventualità di un possibile cambio del gestore. L'elevato rating di un fondo è dovuto all'abilità del money manager, o del team di gestione, che nell'arco della vita lavorativa può cambiare. Per avere investitori preparati sul mercato dei fondi comuni, analisti specializzati rilasciano periodicamente classificazioni sui fondi esistenti a seconda degli obiettivi d'investimento prefissati dai managers.

Negli Stati Uniti nel 1984 è stata fondata *Morningstar*, società indipendente, leader nell'analisi e nella valutazione dei fondi comuni e dei prodotti di risparmio gestito (non solo fondi, ma anche hedge fund, ETF, fondi chiusi ecc.). Dal 18 aprile 2001 è presente anche in Italia e offre dati e analisi su circa 6.000 fondi e servizi informativi a tutte le primarie istituzioni finanziarie. Secondo Morningstar Rating, per valutare un fondo non basata guardare i rendimenti passati, ma bisogna tenere in considerazione il rischio assunto dal gestore e i costi sostenuti dall'investitore.

In Italia, così come in tutti gli altri paesi Europei, *l'associazione nazionale dei managers* sui fondi comuni, *Assogestioni dal 1984*, pubblica la propria classificazione basata su osservazioni periodiche dei fondi di “*portafolio holdings*”. Ogni *classe* di Assogestioni è definita da limitazioni specifiche per ogni *asset allocation* e queste devono essere soddisfatte da tutti i fondi, *con almeno tre anni di vita*, che lo compongono. La classificazione si articola innanzitutto in cinque macro categorie: *Azionario, Bilanciati, Obbligazionario, Liquidità e Flessibili*, che meglio definiremo in seguito. Ogni macro categoria si suddivide a sua volta in diverse categorie definite sulla base dei fattori di rischio che le caratterizzano e i managers dichiarano la *classe* che vogliono attribuire ai loro fondi, senza violare lo stato delle limitazioni dell'asset allocation, specificate da una regolamentazione interna all'associazione. Per la regolamentazione interna si ha anche il controllo del comportamento dei managers, cosa che si effettua periodicamente e in caso di violazioni, il manager è obbligato a rivedere la composizione del portafoglio oppure a scegliere diverse classi per i fondi. Nel caso in cui i managers non obbediscano allora si ha una altra classificazione obbligatoria da parte della stessa Assogestioni.

Le società estere o italiane che siano, per valutare delle *classificazioni statistiche* si basano sui *rendimenti passati* dei fondi e soprattutto sono soluzioni a “basso costo”. Le serie storiche dei rendimenti sono facili e meno costosi anche per la loro reperibilità, si riferiscono a osservazioni giornaliere derivate da risorse pubbliche e sono variabili nel lungo periodo. Alla luce di questo, l’obiettivo della tesi è quello di confrontare la classificazione a priori sui rendimenti mensili dei 50 (cinquanta) fondi comuni d’investimento definita in base alla capitalizzazione, agli stili di gestione e al Morningstar Rating, rispetto agli risultati ottenibili dal clustering. Infatti, la nuova classificazione che ci proponiamo di ottenere è generata dall’elaborazione dell’informazione contenuta nei rendimenti e negli indici di performance ricavati per i fondi, cioè mediante una cluster analysis.

Il portafoglio in analisi è abbastanza diversificato, in quanto è composto da 50 fondi comuni d’investimento statunitensi e valutati in un arco temporale di 5 (cinque) anni, *dal 01 Gennaio 2002 fino al 31 Dicembre del 2006*. I fondi comuni d’investimento azionari sono stati scaricati dal portale www.finance.yahoo.com e appartengono a *sette (7) diverse categorie Morningstar*, sia per capitalizzazione e sia per stile di gestione, e quindi hanno anche a diverse misure di Morningstar Rating. In particolare sono:

- *11 fondi comuni ‘Small Growth’;*
- *9 fondi ‘Small Value’;*
- *10 fondi ‘Small Blend’;*
- *5 fondi ‘Large Growth’;*
- *5 fondi ‘Large Value’;*
- *5 fondi ‘Large Blend’;*
- *e 5 fondi ‘Mid-Cap-Growth’.*

La scelta di strutturare in questo modo il portafoglio non è casuale, l’intenzione è quella di analizzare gruppi di origine abbastanza eterogenea (classificazione questa definita a priori, nota) e in base al clustering verificare come muta l’informazione attorno alla performance del portafoglio. Le classificazioni che si ottengono dal clustering (classificazione a posteriori) saranno chiave di analisi e di una serie di confronti tra le informazioni a priori e a posteriori che si hanno sulla performance dei fondi. Lo scopo è quello di valutare se alla luce dei risultati ottenuti con la cluster analysis sugli indici di performance, otteniamo gruppi maggiormente

informativi. Quindi, le classificazioni a priori e quelle a posteriori in cosa differiscono, il clustering è stato utile per la valutazione della performance dei fondi comuni oppure no?

Il lavoro è organizzato nel seguente modo:

- nel capitolo 1 vengono definiti gli strumenti base per l'analisi finanziaria e alcune conoscenze generali legate al mercato finanziario;
- in seguito, nel capitolo 2 si considerano definizioni importanti legate al *CML* (*Capital o Security Market Line*), al *CAPM* (*Capital Asset Pricing Model*) e agli indici di performance di tipo *RAP* (*Risk Adjusted Performance*);
- nel capitolo 3 verranno affrontate definizioni e proprietà legate alla Classificazione e alla Cluster analysis;
- infine, nel capitolo 4 verrà presentata l'analisi di clustering del dataset, costituito dai rendimenti mensili dei 50 fondi comuni d'investimento statunitensi.

L'analisi, oltre a valutare la performance in base ai risultati ottenuti dal CAPM, mira a considerare le opportunità di guadagno e a rivalutare la classificazione dei fondi comuni mediante l'applicazione di un clustering sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi in analisi e sugli indici di performance calcolati a priori². Vedremo che a posteriori si otterranno classificazioni analizzabili mediante confronti grafici e tabulari per i rispettivi clusters (gruppi) e in seguito verranno considerati gli effetti sulla valutazione della performance del portafoglio dei fondi comuni, confrontandoli con le informazioni risalenti alla classificazione a priori.

² prima dell'applicazione del clustering;

CAPITOLO 1

FONDI COMUNI D'INVESTIMENTO E RENDIMENTI

Nel corso degli anni il mercato finanziario è cambiato, si è trasformato diventando più trasparente, meno rigido e a “misura” del risparmiatore. Oggi esiste l’opportunità di investimento per ogni tipo d’investitore, in base al grado di rischio a cui si vuole esporre e al capitale iniziale in possesso, quindi si riesce a trovare il più adeguato strumento d’investimento personalizzandolo.

In particolare, negli ultimi decenni il risparmiatore, a causa anche della politica economica regnante in Italia, non si è mai fermato nella ricerca della migliore fonte alternativa, rispetto a un reddito fisso, di guadagno. Infatti, sono tanti coloro che decidono di investire i loro risparmi in strumenti finanziari (asset classes) efficaci ed efficienti, cioè tali da garantire massimo rendimento a fronte di un rischio minimo. Ovviamente, bisogna tenere presente che ogni risparmiatore ha una quota di investimento diversa, il grado di rischio da sopportare è diverso e le conoscenze su come funziona, veramente, il mondo dei mercati finanziari sono differenti. Di recente, l’interesse dei risparmiatori si è puntato su organismi d’investimento collettivo, quindi il risparmiatore non gestisce direttamente il proprio capitale, ma lo affida alla gestione di operatori qualificati per poter ottenere una razionale diversificazione del portafoglio investito, difficilmente replicata dall’investitore stesso. Vedremo in questo capitolo, il perché intraprendere tale scelta e come i fondi comuni si distinguono rispetto agli altri strumenti finanziari presenti nel mercato finanziario.

1.1 IL RISPARMIO GESTITO

Investire i propri risparmi, a capitale iniziale variabile, fa parte della cultura di molti paesi, in quanto è visto come una necessità per garantire profitti, utilizzabili per i progetti che si vogliono intraprendere nel futuro e non solo. A tale scopo, i mercati finanziari propongono molti strumenti, i quali sono diversi l'uno dall'altro in base alle loro caratteristiche per ogni tipologia (vedremo queste differenze in modo approfondito nei paragrafi che seguono) e alla luce di due elementi, rischio e rendimento, vengono studiati da gestori altamente qualificati.

Ultimamente sono molti i risparmiatori che scelgono di investire parte dei loro risparmi in strumenti finanziari “sicuri” e a *facile reperibilità*, parliamo infatti dei *fondi comuni d'investimento*. I fondi comuni sono vantaggiosi proprio per la loro “*accessibilità*”, in quanto sono disponibili anche ai risparmiatori con capitale iniziale di poche migliaia di euro. I primi investimenti di questa tipologia sono originari negli Stati Uniti, poi si sono diffusi in Europa centrale, per poi arrivare anche in Italia verso i primi anni '90. Da sempre, i risparmiatori si distinguono l'uno dall'altro per la loro *avversione al rischio* e la *quota di risparmio* da investire quindi, anche gli strumenti finanziari a cui fare d'appoggio sono variabili. Di recente “l'arma vincente” per molti risparmiatori (quota di investimento medio-bassa e medio-alta) sono proprio i fondi comuni d'investimento.

In generale, si può definire come *risparmio gestito* la quota di accantonamento che ogni risparmiatore detiene e affida ad uno o più gestori professionali, i quali provvedono ad amministrarlo. Oggi giorno quando si parla del risparmio gestito si fa riferimento alle seguenti attività:

- i) *attività di gestione professionale del risparmio operate dai fondi comuni di investimento mobiliare e dalle SICAV*, distinte per la raccolta di depositi e l'erogazione di impieghi;
- ii) *attività di gestione di patrimoni mobiliari individuali (GPM)* effettuata da banche e da società d'intermediazione mobiliare (*SIM*), cioè in tal caso si ha una completa traslazione sugli investitori del rischio proveniente dalle oscillazioni del valore dei titoli detenuti in portafoglio;

iii) *attività d'investimento per conto dei risparmiatori operate dai fondi pensione e dalle compagnie di assicurazione nell'ambito della cosiddetta previdenza complementare.*

Invece, nelle statistiche internazionali il risparmio gestito comprende: *le gestioni di patrimoni individuali, gli OICR (fondi comuni di investimento e SICAV), le assicurazioni sulla vita e i fondi pensione.* Molto spesso però, ci si richiama ad una nozione più significativa per il mercato italiano comprendente le sole ***gestioni di portafogli:***

1. ***collettivi:*** sono da intendersi le gestioni di fondi comuni di investimento immobiliare e le SICAV,
2. ***individuali:*** ci si riferisce alla gestione mobiliare e del patrimonio.

Principalmente, *gli investitori qualificati* sono rappresentati delle seguenti *categorie di soggetti:*

- I) *le imprese di investimento* comprendenti le SIM e imprese di investimento estere, le banche, gli agenti di cambio, le società di gestione del risparmio, le SICAV e i fondi pensione, le imprese di assicurazione, le società finanziarie capogruppo di gruppi bancari;
- II) *i soggetti esteri autorizzati* a svolgere, in forza della normativa in vigore nel proprio Paese d'origine, le medesime attività;
- III) *le fondazioni bancarie;*
- IV) *le persone fisiche e giuridiche e gli altri enti* in possesso di specifica competenza ed esperienza.

Segue che, la *rilevanza economico-sociale* attribuita a questi soggetti fa riferimento al ruolo da essi svolto nell'ambito del mercato finanziario, quindi *le principali funzioni* sono:

- a) *funzione di intermediazione:* collegamento tra soggetti in surplus e quelli in deficit finanziario;
- b) *funzione di investimento:* il risparmio raccolto viene investito selezionando le opportunità di investimento più adeguate alle necessità del soggetto finanziato, nel rispetto di determinati livelli di rischio e di rendimento;
- c) *funzione di gestione collettiva delle risorse:* i cui vantaggi consistono principalmente nella possibilità di realizzare, una più ampia diversificazione

degli investimenti con un minor grado di rischio ed un più contenuto costo di gestione del risparmio;

- d) *funzione di consulenza*: sia nei confronti dei risparmiatori e anche a favore dei soggetti finanziati, bisogna quindi avere le capacità di incoraggiarne i piani di investimento in relazione alle loro specifiche necessità.

Attualmente, dal punto di vista normativo, *il risparmio gestito* è disciplinato dal **TUIF** (Testo Unico delle disposizioni in materia di Intermediazione Finanziaria) e dai relativi regolamenti emanati dalla Consob e dalla Banca d'Italia.

1.2 I FONDI COMUNI D'INVESTIMENTO

Negli ultimi anni i fondi comuni di investimento sono stati definiti anche come, *patrimonio autonomo*, in quanto suddiviso in quote di appartenenza di una pluralità di risparmiatori e viene gestito dagli OICR. I fondi vengono definiti "**comuni**", perché la somma versata dal singolo risparmiatore cessa di essere direttamente riferibile ad esso in quanto, diventa parte di un patrimonio unico gestito a monte e sul quale ciascun risparmiatore vanta i medesimi diritti (proporzionalmente all'entità del capitale sborsato). Ma oltre a questo, sono definiti "**d'investimento**", perché le somme raccolte devono necessariamente essere investite con le modalità stabilite dalla legge e dal regolamento del fondo.

Un fondo comune di investimento può essere visto come un organismo che, mediante la sottoscrizione di apposite quote di partecipazione, raccoglie risorse finanziarie presso i risparmiatori al fine di gestirle in modo collettivo attraverso il loro investimento in valori mobiliari o altri beni, con lo scopo di realizzare, nel tempo, un incremento del valore dei capitali investiti. In particolare, i fondi comuni sono suddivisi in quote e in base a chi sottoscrive tale investimento, acquista un numero di quote in proporzione all'importo versato.

Uno dei principali vantaggi offerti dai fondi comuni (accanto alla competenza professionale del gestore ed alla *riduzione dei costi di transazione*, grazie agli elevati volumi di titoli scambiati) è la possibilità, anche per il risparmiatore con

disponibilità finanziarie piuttosto limitate, di orientarsi verso quella tipologia di strumenti finanziari che maggiormente preferisce e di *diversificare* il proprio investimento tra i *vari mercati*:

► *nel monetario*;

► *nell'obbligazionario (MOT)*: titoli di credito, cioè prestiti contratti dallo Stato, da banche, da società. In questo caso, tranne nei casi di insolvenza, il possessore ha diritto alla restituzione, a scadenza, del valore nominale nonché al pagamento degli interessi;

► *nell'azionario (MAT)*: titoli che rappresentano quote di partecipazione al capitale di un'azienda e il possessore ha diritto ai dividendi (se previsti). In caso di liquidazione non è garantito il valore nominale e si può anche perdere tutta la somma investita.

Le scelte e le decisioni non si basano sull'influenza esercitata dalle scelte del gestore, ma semplicemente scegliendo o *ripartendo le disponibilità da investire* tra i vari prodotti, investendo ingenti patrimoni su una vastissima gamma di titoli, riuscendo così a ridurre al minimo il rischio di perdite finanziarie.

Un altro vantaggio legato all'investimento in fondi è rappresentato dalla *gestione professionale*, o meglio i fondi sono gestiti da operatori qualificati che quotidianamente analizzano mercati e settori, selezionando i titoli migliori da inserire nel portafoglio. Inoltre, si possono elencare altri vantaggi, quali:

- I *Trasparenza*: il valore della quota dei fondi è pubblicato giornalmente sui principali quotidiani nazionali e sui siti finanziari;
- II *Affidabilità*: i fondi comuni d'investimento operano in base alle leggi imposte da numerosi organismi istituzionali cioè, in riferimento a un quadro legislativo rigoroso e nello stesso tempo trasparente;
- III *Flessibilità*: questo consente di poter trasferire l'investimento da un fondo a un altro in tempi molto rapidi e con costi contenuti;
- IV Infine, *liquidità*: le quote dei fondi possono essere vendute facilmente in qualsiasi momento attraverso il riscatto, senza poi l'applicazione di nessuna penalità.

La valutazione dei fondi comuni deve essere molto accurata, in quanto sono molti gli elementi da considerare. Prima di tutto è estremamente utile confrontare la *performance*, o meglio l'andamento del fondo in un determinato arco di tempo,

con il proprio **benchmark** di riferimento, utile questo in quanto si può dedurre lo stile di gestione del fondo. In Italia la Consob definisce il benchmark come: “un parametro oggettivo di riferimento costruito facendo riferimento a indicatori finanziari elaborati da soggetti terzi e di comune utilizzo, e soprattutto coerente con i rischi connessi alla gestione del fondo, con il quale confrontare il rendimento del fondo stesso”.

1.2.1 CLASSIFICAZIONE DEI FONDI COMUNI

Gli elementi di specializzazione, considerando il ciclo vitale del fondo comune d'investimento, tendono ad accrescersi nel tempo. Quando entra nel mercato un prodotto è poco o per nulla conosciuto da potenziali acquirenti e lo sforzo delle società di gestione è quello di promuoverlo proponendolo alle fasce di utenza interessate che, in questa fase, sono di solito abbastanza ristrette. In seguito, si passa ad ulteriori suddivisioni in funzione di determinate specializzazioni assunte dai fondi di investimento.

Il concetto di fondo comune è molto ampio e comprende una serie di tipologie, le quali sono definite in funzione alle differenti forme tecniche e in base alla specializzazione.

- 1) Un primo criterio di classificazione può essere rappresentato dal **grado di diversificazione del portafoglio** e si hanno:
 - i) **fondi diversificati**: investono in titoli di società che appartengono a settori economici diversi e/o a Paesi diversi;
 - ii) **fondi specializzati**: investono in titoli di società appartenenti ad un solo settore (specializzazione economica) o ad un solo Paese o ad un'unica area geografica (specializzazione geografica).
- 2) Sulla base della **natura dell'investimento preferenziale**, operato dalla società di gestione e quindi del tipo di valore mobiliare prevalente in portafoglio, è possibile distinguere:
 - a) **Fondi comuni azionari**: sono caratterizzati da un prevalente investimento in **titoli azionari**. La continuità legata alla compravendita di azioni,

consente elevati rendimenti uniti a un elevato rischio, in questo modo si soddisfano anche le esigenze di crescita patrimoniale più rapide esponendo l'investitore ad un maggior livello di rischio.

- b) **Fondi comuni bilanciati**: sono definiti dalla ricerca di un tendenziale equilibrio tra investimenti in *titoli a reddito fisso* e quelli a *rendimento variabile*. L'investimento riguarda sia i titoli azionari che i titoli obbligazionari, la prevalenza dell'una o dell'altra forma di investimento è in funzione delle politiche di gestione perseguite e dell'andamento congiunturale dei rispettivi mercati. Inoltre, questi fondi devono avere una quota d'investimenti azionari non superiori al 70% e non inferiori al 20% rispetto al totale.
- c) **Fondi comuni obbligazionari**: sono distinti da un orizzonte d'investimento a medio e lungo termine, da minori livelli di volatilità e di rischio per il partecipante, in quanto si investe in titoli obbligazionari. Ad esempio: Buoni del Tesoro, titoli di debito emessi da imprese, ecc.
- d) **Fondi monetari e di liquidità**: orientati ad *investimenti nel breve termine*. Ad esempio: BOT, pronti contro termine, certificati di deposito, ecc. Sono adatti a convogliare disponibilità in attesa di un impiego più duraturo, caratterizzati da un livello di rischio molto basso.
- e) **Fondi flessibili**: distinti dall'assenza di vincoli sia sull'asset allocation di base (azioni/obbligazioni), sia all'interno di ciascuna categoria (durata dei titoli obbligazionari e/o del mercato monetario, area geografica o settore per le azioni).
- f) **Fondi pensione**: le fonti costitutive possono avere la forma di contratti e accordi collettivi, ovvero di accordi fra gli stessi lavoratori, o ancora di regolamenti di enti o aziende. Sono utilizzate prevalentemente da istituzioni finanziarie appositamente costituite per garantire, sia a lavoratori dipendenti sia a lavoratori autonomi, un adeguato trattamento previdenziale al raggiungimento dell'età pensionabile. La finalità principale è l'erogazione di un trattamento pensionistico, complementare rispetto a quello pubblico che è obbligatorio.
- g) **Fondi di Fondi**: si investe il patrimonio degli investitori in quote di altri fondi e/o Sicav. Una Sicav funziona in modo simile a un fondo comune

d'investimento dal punto di vista pratico dell'investitore. La differenza tra questi due prodotti finanziari è *giuridica*, o meglio il sottoscrittore di una Sicav non compra "quote" ma "azioni" acquisendo anche i diritti tipici di un azionista. In generale, la gran parte dei prodotti distribuiti in Italia dai grandi gruppi stranieri sono Sicav.

3) A seconda delle **modalità di sottoscrizione** avremo:

- I) *Versamento in unica soluzione*, che comunque potrà essere integrato mediante eventuali versamenti successivi (*Piano di investimento in contanti, Pic*);
- II) *Piano d'accumulo*: sistema di investimento mediante il quale il sottoscrittore entra nel fondo comune con una cifra iniziale molto bassa e programma contemporaneamente di versare periodicamente una certa cifra (fissa o variabile a seconda dei casi) per un certo periodo di tempo. Rappresenta una *forma di risparmio molto flessibile* poiché è possibile, secondo modalità e condizioni che variano da società a società, variare l'importo delle rate, effettuare versamenti aggiuntivi o prolungare il piano, sospendere o interrompere i versamenti in qualsiasi momento. Il *Piano di accumulazione del capitale (Pac)* prevede il versamento di un prefissato ammontare mensile per un certo periodo di tempo, sviluppando in maniera automatica il *cost averaging*³. Un *Piano ad accumulazione variabile (Pav)* è simile ad un Pac, con la differenza che l'ammontare del versamento mensile può variare, sviluppando il *value averaging*⁴.

Il valore della quota del proprio investimento indipendentemente dalla modalità di sottoscrizione del fondo, vedremo in seguito che è molto semplice da ottenere.

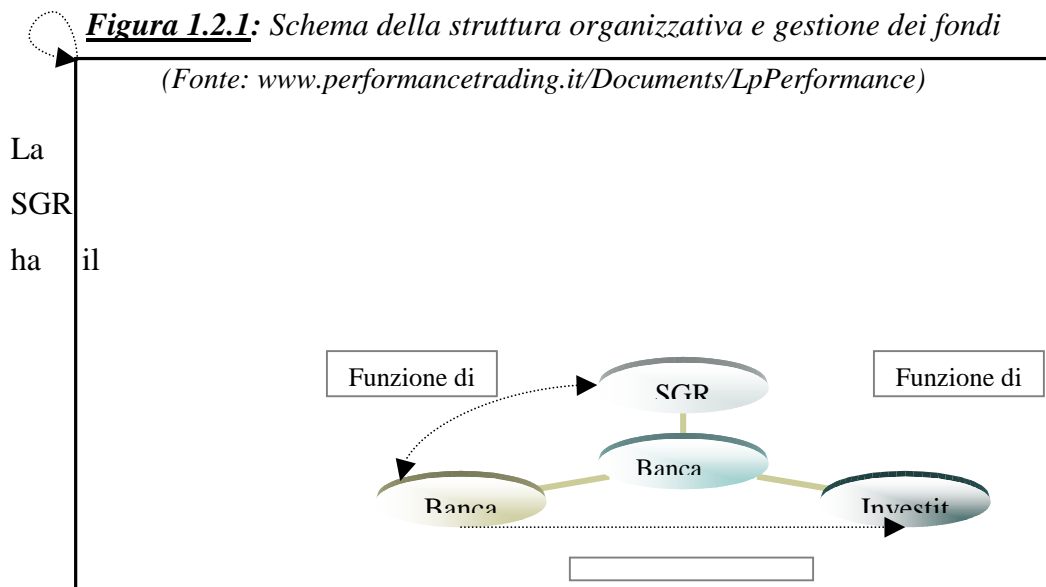
1.2.2 STRUTTURA DEI FONDI COMUNI E COSTI

La struttura organizzativa (vedi Figura 1.2.1) dei fondi comuni d'investimento si articola in: società di gestione del risparmio (SGR), banca depositaria e l'insieme

³ Costi medi in riferimento al Pac;

⁴ valore medio aggiunto previsti dal Pav;

dei partecipanti (gli investitori). Sul complesso delle attività che svolgono questi tre soggetti è previsto il controllo da parte della Banca d'Italia e della Consob. I capitali versati dai sottoscrittori formano il *compendio patrimoniale* autonomo (*fondo comune*) sul quale la SGR esercita le funzioni di amministrazione e la banca depositaria quelle, di custodia e di controllo sull'attività svolta dalla società di gestione del risparmio, assumendosene le relative responsabilità.



compito di impiegare i capitali dei sottoscrittori e di farli rendere nel miglior modo possibile, ossia decide la politica di investimento e segue l'evoluzione del portafoglio in modo da mantenerlo a un livello qualitativo ottimale. Il rischio di insolvenza delle SGR è molto basso, in quanto il patrimonio gestito è legato da quello della società e quindi essere rimborsato in ogni caso dagli investitori. I sottoscrittori non hanno diritto a influire sull'attività di gestione del patrimonio del fondo, si limitano solo ad aderire ad una proposta contrattuale predefinita comprendendo così anche una fonte di tutela nel controllo amministrativo dell'organo di vigilanza sul contenuto (e sul rispetto) del regolamento contrattuale. I fondi comuni d'investimento vengono quindi, gestiti direttamente dalle società di gestione del risparmio (SGR), costituite su iniziativa di banche, compagnie di assicurazione oppure grandi società finanziarie. Tali società sono autorizzate dal Ministero del Tesoro e controllate continuamente dalla Consob, mentre l'approvazione del quadro informativo, che riguarda gli aspetti gestionali del fondo e delle società stesse, è approvato dalla Banca d'Italia.

L'adozione di tale *modello organizzativo* offre una maggiore tutela ai partecipanti in quanto la netta separazione delle diverse funzioni consente una alta professionalità tecnica nell'ambito della società di gestione del risparmio e una precisa individuazione della sua attività. Dal punto di vista strutturale si osserva che, i fondi comuni in base alla *forma del capitale detenuta* si distinguono in tre principali classi:

1.) *Il fondo aperto* o meglio definito a capitale variabile, rappresenta la forma più comunemente utilizzata nel campo dei fondi comuni di investimento. Il suo *patrimonio varia* continuamente (sia nella composizione e sia nel valore) in relazione agli *acquisti e alle vendite* realizzate sul mercato a opera dei gestori del fondo, e all'*andamento del saldo* netto tra nuova raccolta e riscatti chiesti dai partecipanti. Il patrimonio è suddiviso in "*quote*" e il valore (o prezzo) di ciascuna quota si determina come:

$$V_q = \frac{\text{Totale attività nette del fondo}}{\text{Numero delle quote circolanti}} \quad (1.2.1)$$

I partecipanti al fondo aperto hanno diritto di chiedere, in qualsiasi tempo, il rimborso delle quote secondo le modalità previste dalle regole di funzionamento del fondo. Inoltre, in ogni momento il sottoscrittore può verificare l'andamento del prezzo della quota pubblicato quotidianamente sui principali quotidiani, Televideo e siti Internet. Molto spesso si parla del *fondo comune d'investimento mobiliare aperto*, il quale rappresenta *un patrimonio destinato all'investimento*, generato dalle assegnazioni di una massa di risparmiatori. Il fondo è "*mobiliare*" poiché, il suo patrimonio è impiegato esclusivamente in strumenti e prodotti finanziari ed è "*aperto*" in quanto il sottoscrittore è libero di entrare ed uscire dal fondo in qualsiasi momento, fondi che non rispettano uno o più di tali vincoli si dicono "*non armonizzati*". Investendo in tali fondi, un risparmiatore affida una determinata somma di denaro a società che svolgono professionalmente la gestione (in forma collettiva) di attività finanziarie, o meglio alle SGR.

2.) *Il fondo chiuso*, è uno strumento finanziario che raccoglie capitali presso investitori istituzionali (quali banche, fondazioni, compagnie assicurative, fondi pensione) e presso privati, vincolandoli a medio-lungo termine per investirli in imprese non quotate ad alto potenziale di sviluppo. Essi sono gestiti da società specializzate nell'investimento istituzionale nel capitale di rischio e rappresentano

uno strumento che, permette di allargare enormemente il potenziale operativo di questa attività. L'istituzione di fondi chiusi è consentita alle società di gestione del risparmio e ciascun fondo chiuso deve disporre di un proprio regolamento (*ad esempio*: in Italia è approvato dalla Banca d'Italia) che individua le caratteristiche del fondo e stabilisce le modalità attraverso le quali la società di gestione è abilitata ad operare. La caratteristica di chiusura si basa sul fatto che non sia possibile, a differenza dei fondi aperti, né entrare, né uscire, in qualsiasi momento e si possono suddividere in:

- a) *fondi mobiliari chiusi*: nel caso in cui la società di gestione intenda aumentare il volume della massa di mezzi finanziari a disposizione del fondo, dovrà stabilire il collocamento di nuove quote, secondo una procedura in un certo modo paragonabile a quella prevista per gli aumenti di capitale delle società per azioni;
- b) *fondi immobiliari chiusi*: la loro gestione è affidata ad una società di gestione del risparmio e il patrimonio del fondo deve essere raccolto mediante un'unica emissione di quote, di eguale valore unitario, che devono essere sottoscritte entro il termine massimo di un anno a decorrere dalla data del *nulla osta* da parte della Consob alla pubblicazione del prospetto informativo. La durata del fondo non può essere inferiore a dieci anni e superiore a trenta.

3.) I fondi speculativi o hedge funds: sono previsti dal regolamento del Ministero del Tesoro e hanno una struttura diversa rispetto sia alla tradizionale dicotomia fondi aperti/chiusi sia all'oggetto di investimento. Si tratta di *fondi non armonizzati con soglia minima di sottoscrizione pari ad un milione di euro*, il cui gestore, in deroga alle norme prudenziali sul contenimento del rischio previste dalla normativa. Tra questi fondi rientrano numerosi prodotti esteri e, da qualche tempo, i non armonizzati di diritto italiano, cioè *i fondi riservati*. I fondi non armonizzati possono essere aperti o chiusi, anch'essi disciplinati da recenti disposizioni e alla cui sottoscrizione possono accedere unicamente investitori qualificati, che derogano alle norme prudenziali di limitazione e frazionamento del rischio. Lo scopo è l'ottenimento dei risultati positivi indipendentemente dall'andamento dei mercati in cui si opera utilizzando sofisticate tecniche di investimento, cercando di migliorare le performance e proteggendo il portafoglio

da eventuali ribassi del mercato ricorrendo all'apertura di posizioni speculative mediante vendite allo scoperto.

In generale, i fondi offrono ai risparmiatori la possibilità di *diversificare i propri investimenti in un portafoglio più ampio*, ottenendo una gestione professionale del risparmio, in campi di non facile accesso all'investitore individuale, ma sottoscrivendo un fondo comune non si devono dimenticare i costi che devono essere affrontati. ***I costi di commissione*** sono diversi, in quanto:

I) Alcune sono *a carico degli investitori*:

- *Commissioni d'ingresso*: sono richieste nel momento in cui si acquistano le quote del fondo all'atto del primo versamento e di quelli successivi. Possono essere fisse o variabili, e se tali decrescono al crescere della somma investita. Sono definiti *Fondi No Load* quelli che non prevedono commissioni d'ingresso, però i costi di gestione possono essere più elevati.
- *Commissioni di uscita*: richieste al momento in cui si decide il disinvestimento delle quote possedute. Molto spesso decrescono all'aumentare degli anni di permanenza nel fondo e in alcuni casi sono un'alternativa alle commissioni d'ingresso.
- *Commissioni di switch*: previste quando si effettua uno switch, o meglio si decide di spostare il proprio investimento da un fondo all'altro.

II) E altre *a carico del fondo*:

- *Commissione di gestione*: non è altro che la remunerazione fissa del lavoro svolto dal gestore del fondo e vengono calcolati quotidianamente sul patrimonio netto del fondo.
- *Commissione di performance*: rappresenta la remunerazione prevista nel caso in cui il fondo riesca a battere il cosiddetto "benchmark".
- *Costi d'intermediazione*: include le spese che il fondo sostiene per la compravendita, *ad esempio* dei valori mobiliari.

Inoltre, esiste anche il *trattamento fiscale* riservato all'investimento in fondi comuni, il quale differisce in base al tipo di fondo. Per *i fondi di diritto Italiano* è prevista una *tassazione del 12,5% sul rendimento positivo* ottenuto ogni anno. In

tal caso l'investitore non deve provvedere ad alcun adempimento fiscale perché le imposte sono prelevate direttamente dal fondo, cioè la quota viene in seguito calcolata al netto dell'imposta. Per quanto riguarda, *i fondi di diritto estero comunitari* è prevista una tassazione a carico dell'investitore, a viene applicata una *ritenuta a titolo d'imposta del 12,5% al momento dell'incasso* del provento erogato del fondo entro 12 mesi dall'investimento.

1.3 NOZIONI SUI MERCATI FINANZIARI

Il mercato finanziario è definito come l'insieme delle *negoziazioni di strumenti finanziari* ed opera come collegamento tra *settori in surplus* (caratterizzati dagli operatori che presentano un rimanenza finanziaria) e *settori in deficit* (cioè, evidenziano un disavanzo finanziario) canalizzando in questo modo le risorse disponibili verso forme adeguate di impiego. In generale, *il mercato deve essere in grado di offrire* ai risparmiatori ed agli utilizzatori finali *risorse finanziarie* in grado di rispondere alle loro esigenze future in termini contrattuali (scadenza) e di un adeguato rapporto di rendimento/rischio. Possibili classificazioni dei mercati si basano sul tipo e durata di vita dell'attività finanziaria scambiata, per data di consegna e per struttura organizzativa. *Un attività finanziaria* è definita come un contratto stipulato tra due controparti, rispettivamente *il venditore e l'acquirente*, che stabilisce per ciascuna data futura e stato del mondo, la quantità di moneta o contratto finanziario che il venditore deve trasferire al compratore. In particolare, in base alla *struttura organizzativa* si determina la *tipologia di informazione disponibile sulle transazioni*, infatti i mercati si distinguono in:

- D) ***Mercati regolamentati***: sono fondati e gestiti da società per azioni (*dette anche società di gestione*), responsabili dell'organizzazione e del controllo delle transazioni, dell'ammissione degli strumenti da negoziare e degli operatori. In Italia l'esercizio dei mercati regolamentati è autorizzato dalla Consob. Infatti, i *mercati Italiani* autorizzati sono iscritti in un elenco tenuto

dalla Consob⁵, invece i *mercati esteri* sono riconosciuti ai sensi dell'ordinamento comunitario e sono iscritti in un apposita sezione⁶ e altri non inclusi nell'elenco sono riconosciuti dalla Consob sulla base di accordi stipulati con le corrispondenti autorità⁷. Sia i mercati esteri e sia i mercati italiani funzionano di continuo, le transazioni avvengono in qualsiasi momento durante l'apertura e il prezzo varia in base agli ordini oppure alle "grida", gli ordini vengono raggruppati ed eseguiti contemporaneamente nel corso di un'asta che si svolge una o più volte al giorno.

II) ***Mercati over the counter (OTC)***: ci si riferisce a operatori collegati da reti telematiche che negoziano titoli in maniera essenzialmente non regolamentata, o meglio rappresenta tutte quelle operazioni di compravendita di titoli che non figurano nel listino di borsa. Le negoziazioni, si svolgono al di fuori dei circuiti *Borsistici Ufficiali* e di un mercato organizzato come lo è la *Borsa Valori*, e sono effettuate tra intermediari ed investitori.

I dati usati dagli analisti finanziari possono essere, *serie storiche di prezzi*, cioè campione di osservazioni relative allo stesso soggetto e a date diverse, mentre quando le osservazioni si riferiscono alla stessa data per diversi soggetti si ha un campione in *cross-section*. Nei mercati in cui esiste un intermediario incaricato di compiti ***market maker*** (responsabile della determinazione del prezzo di equilibrio di un titolo attraverso l'offerta e la domanda), i prezzi quotati sono diversi se si riferiscono all'acquisto o alla vendita del titolo, di solito come prezzo di riferimento si usa quello di *chiusura aggiornato*. Proprietà importante dei prezzi è la *liquidità del mercato* da cui provengono. Inoltre, *l'ampiezza del margine* definito dalla differenza tra il prezzo effettivamente venduto e quello richiesto, cioè *denaro-lettera (bid-ask spread)* è una misura della liquidità di un titolo finanziario. Se è bassa, l'intermediario applica un margine ampio come remunerazione del rischio legato all'essere in possesso di un titolo per il quale sarà difficile trovare un acquirente. Ma dall'altra parte, non si deve dimenticare che tale *spread* è un segnale del potere di mercato degli intermediari. Le quotazioni

⁵ secondo l'art. 63, comma 2, D.Lgs. n.58/1998;

⁶ dell'elenco secondo l'art.67, comma 1, D.Lgs. n.58/'98;

⁷ secondo l'art.67, comma 2, D.Lgs. n.58/'98;

disponibili non corrispondono sempre ai prezzi ai quali avvengono le transazioni, in quanto esistono anche i costi di transazione, commissioni e imposte.

1.3.1 STRUMENTI FINANZIARI E INDICI DI MERCATO

Negli ultimi tempi possedere un piccolo risparmio e avere la voglia di incrementarlo, sono la spinta più frequente per conoscere le varie opportunità offerte dal mercato finanziario. La chiave del successo (creare guadagno), come vedremo, stà nell'individuare la ripartizione del portafoglio ottimale fra i diversi strumenti finanziari (*assets classes*) quali azioni, obbligazioni e aree di investimento quindi, individuare l'*asset allocation* ottimale che significa, raggiungere la diversificazione nei propri investimenti in relazione al grado di rischio che si vuole sopportare. Le prime attenzioni vanno a focalizzarsi sul significato economico e finanziario che hanno i vari strumenti finanziari, partendo dai *contratti finanziari* (o *attività finanziarie*) finalizzati al trasferimento di moneta o merci a diverse date di esigibilità (o scadenze) subordinante alla realizzazione di diversi stati del mondo. Di conseguenza si ha che, la gestione di *assets classes* lo si può anche descrivere come una serie di attività svolte al fine di costruire un portafoglio finanziario tale da massimizzare i rendimenti e minimizzando la differenza tra i risultati attesi e quelli effettivi. Riepilogando possiamo dire che, il punto di partenza prima di effettuare un investimento è quello di *scegliere l'asset allocation, cioè la ripartizione del portafoglio.*

In finanza è molto importante fare riferimento a un indice (*parametro oggettivo*) oppure a una composizione di indici di mercato calcolati da soggetti terzi rispetto alla società di gestione promotrice del fondo, utile per ogni tipo di analisi.

Un indice è il *prezzo di un portafoglio di titoli* e possono essere tutti quelli quotati su un mercato oppure essere scelti fra i più rappresentativi di un particolare segmento. La composizione del portafoglio può variare nel tempo in seguito al declino economico di alcune società (o all'emergere di altre). Il valore dell'indice è spesso normalizzato fissandone il livello (molto importante per la valutazione

dell'indice nel futuro, prezzo base) a una data di riferimento e con prezzi dei titoli che vengono sommati in modo ponderato con la capitalizzazione della società, ossia il numero delle azioni emesse per il loro prezzo. Per ogni *parametro oggettivo* composto da due o più indici di mercato, il periodo di ribilanciamento dei pesi delle singole componenti del parametro è stabilito dalla società di gestione e non può eccedere i tre mesi. La periodicità del ribilanciamento deve essere esplicitata nel prospetto e il sistema più adeguato è ritenuto quello a *proporzioni costanti*, secondo il quale i pesi degli indici che compongono il parametro di riferimento vengono periodicamente ribilanciati al fine di mantenere costante la proporzione di ogni componente ad inizio periodo. In caso di interruzione del calcolo o della pubblicazione di uno o più indici indicati nel parametro oggettivo di riferimento, la società di gestione potrà fare automaticamente riferimento ad un *parametro sostitutivo*. Ad esempio: Assogestioni ne definisce uno per ogni categoria del sistema di classificazione in dotazione. Segue che, la qualità di gestione non è misurata esclusivamente dalle capacità di offrire un extra-rendimento rispetto al *benchmark* (indice utilizzato per misurare il rischio sistematico dei portafogli passivi) ma anche dalla capacità di sceglierlo in modo appropriato. Le scelte più corrette sono quelle che si riferiscono a degli *indici* che hanno la capacità di riassumere in un unico numero l'evoluzione di tutto il mercato.

Gli Indici Ufficiali delle principali Borse sono *definiti* dalla Banca Centrale e Istituti Governativi a cui i mercati fanno riferimento. La *costruzione degli indici* e la diffusione dei dati relativi al loro andamento vengono effettuate da società chiamate *Index Provider* e generalmente appartengono a gruppi finanziari o a società di gestione di mercati regolamentati, o ancora anche da gruppi editoriali (Reuters, Bloomberg, ecc.) e si riferiscono a metodologie indipendenti che però, tuttavia presentano tratti comuni. Gli indici che caratterizzano *il mercato finanziario globale* sono definiti per aerea di origine e sviluppo.

In Italia il **MibTel** rappresenta l'indice generale della Borsa di Milano e sintetizza l'andamento di tutti i titoli quotati a Piazza Affari. L'attuale valore di riferimento è 10.000 punti, fissato il 3 gennaio '94. Gli altri indici Italiani sono: **S&P/Mib**, **MIBTEL**, **MIDEX**, **MIB**, **All Stars**, **MEX (Expandi)** e **D.J.Eurostoxx**

50, distinti per tipologia di capitalizzazione e altre caratteristiche delle società di riferimento.

In Europa e all'estero, gli indici principali⁸ sono: il CAC40 (Parigi), il Ftse 100 (Londra), Dax Xetra (Francoforte), Bel30 (Bruxelles), Aex (Amsterdam), Ibex 35 (Madrid), Smi (Zurigo), Hang Seng (Hong Kong), AllOrd (Sydney) e Nikkei (Tokio).

Mentre, negli Stati Uniti gli indici più importanti sono il Dow Jones (New York), lo Standard & Poor's 500 (rappresenta le prime 500 società a media capitalizzazione) e Nasdaq (rappresenta le società tecnologiche della cosiddetta "new economy"). Il Dow Jones è anche definito l'indice di Borsa più antico del mondo, in quanto risale al 1896. Il paniere che lo compone è rappresentato da 30 società americane a grande capitalizzazione che sono leader di mercato nei rispettivi settori. Una particolarità del Dow Jones è quella di essere una semplice media aritmetica dei prezzi delle 30 azioni che lo compongono, senza ponderare il peso dei singoli titoli a seconda della loro capitalizzazione. La conseguenza è che, nel calcolo dell'indice, una variazione dei titoli a prezzo più alto ha un impatto superiore a quella dei titoli a prezzi più bassi.

*Gli indici più noti sono probabilmente quelli del mercato azionario, ma non di meno sono importanti gli indici obbligazionari. Infatti, da recenti indagini si ha che, oltre l'80% del patrimonio dei fondi italiani armonizzati e riservati è gestito con riferimento agli indici di quattro **index provider**: J.P. Morgan, MSCI, MTS SPA e Merrill Lynch. Si nota che, per i fondi azionari risulta ancora più evidente l'elevato grado di concentrazione, oltre il 62% del loro patrimonio è gestito in riferimento agli indici del provider. Nel comparto obbligazionario invece, il mercato appare *più diversificato*, nonostante la prevalenza degli indici J.P. Morgan, diversi provider concorrono a definire i benchmark. Oltre tutto, gli indici di Borsa sono un riferimento base per le obbligazioni indicizzate e per i contratti derivati, come: *futures, opzioni e warrant*. A volte nell'analisi finanziaria, così come anche per i fondi comuni che sono oggetto di studio e analisi di questa tesi, si fa riferimento ai fondi comuni d'investimento USA, quindi ai rendimenti che presentano nel tempo. In modo naturale ci si chiede, cos'è il rendimento e che proprietà deve avere perché sia ben definito?*

⁸ simili al MibTel italiano;

1.4 ALCUNI CONCETTI SUI RENDIMENTI

I rendimenti misurano la variazione percentuale dei prezzi in un certo intervallo. Esistono due modi leggermente diversi di calcolare i rendimenti, in entrambi i casi si fa riferimento ai prezzi del titolo, però l'utilizzo è diverso.

I) *Capitalizzazione semplice*: (o percentuali) il rendimento di un titolo fra due date ($t-1$) e t è definito come:

$$r_t = \frac{p_t + d_t}{p_{t-1}} - 1 \quad (1.4.1)$$

$$r_t + 1 = \frac{p_t}{p_{t-1}} \quad \text{se} \quad d_t = 0 \quad (1.4.2)$$

dove $p(t)$ e $p(t-1)$ sono i prezzi del titolo alle date t e $t-1$, invece $d(t)$ si riferisce al cash flow generato dal titolo fra le due date. La somma $1+r(t)$ è definito come *fattore di capitalizzazione*. Per costruzione $r(t) \geq -1$, quindi i rendimenti semplici non possono essere generati da una distribuzione normale. Due rendimenti che si riferiscono a intervalli di lunghezza diversa possono essere confrontati dopo essere stati trasformati in modo da corrispondere allo stesso intervallo, di solito annuale, chiamato anche normalizzazione. Con la capitalizzazione semplice è facile ricavare la relazione fra rendimenti dei titoli e quello di un portafoglio. Dati I -titoli, $i = 1, 2, \dots, I$; i prezzi alle date $t-1$ e t di un portafoglio p composto da una quantità di q_i di ciascuno di essi sono pari a:

$$p_{pt-1} = \sum_{i=1}^I q_i p_{it-1} \quad \text{e} \quad p_{pt} = \sum_{i=1}^I q_i p_{it} \quad (1.4.3)$$

se il costo del portafoglio è diverso da 0, possiamo dividere per p_{pt-1} :

$$\frac{p_{pt}}{p_{pt-1}} = \frac{\sum_{i=1}^I q_i p_{it}}{p_{pt-1}} \quad (1.4.4)$$

ottenendo i rendimenti:
$$r_{pt} = \sum_{i=1}^I w_i r_{it} \quad (1.4.5)$$

e la quota di ricchezza investita nel titolo i è:
$$w_i = \frac{q_i p_{it-1}}{p_{pt-1}} \quad (1.4.6)$$

inoltre si nota che, per costruzione le quote sommano a 1. Il tutto può essere rappresentato anche matricialmente e il vincolo di somma unitaria delle quote possono essere riformulate come segue:

$$r_{pt} = w' r_t \quad (1.4.7)$$

e poi:
$$t'w = 1 \quad (1.4.8)$$

II) La definizione alternativa del rendimento è basata sulla **capitalizzazione continua**, secondo la quale i rendimenti sono dati da:

$$r_t^* = \ln(p_t + d_t) - \ln(p_{t-1}) = \ln\left(\frac{p_t + d_t}{p_{t-1}}\right) = \ln(1 + r_t) \quad (1.4.9)$$

a differenza del caso precedente, tale rendimento non è limitato inferiormente, ed è quindi compatibile con l'ipotesi di distribuzione normale. *In assenza di dividendi*, questa regola semplifica notevolmente l'aggregazione nel tempo dei rendimenti e per annualizzarlo è sufficiente dividerlo per i periodi su cui è osservato. Tuttavia l'ipotesi di assenza di flussi in entrata e in uscita nel periodo considerato ai fini del calcolo del rendimento è del tutto irrealistica. Se si prende in considerazione il patrimonio complessivo del fondo, questo è continuamente soggetto alla movimentazione legata alla sottoscrizione e al rimborso di quote, al prelievo fiscale, all'eventuale distribuzione di proventi; nell'ottica del singolo investitore. Inoltre, è possibile che la somma inizialmente versata sia stata integrata da ulteriori conferimenti, che siano stati effettuati dis_investimenti parziali, che siano stati percepiti proventi. Da notare è il fatto che, il rendimento continuo di un portafoglio non è dato dalla media ponderata dei rendimenti dei titoli.

Le definizioni precedenti sul rendimento conducono in realtà a risultati molto simili. Tale proprietà è verificata espandendo il *logaritmo dei prezzi in serie di Taylor al secondo ordine* rispetto a $p(t)$ in un intorno di $p(t-1)$. Si può notare inoltre che: $r_t^* \leq r_t$ e che la differenza tra le due definizioni è in generale piccola, data l'unità di misura dei rendimenti. Le proprietà valide per i rendimenti semplici sono approssimativamente valide anche per i rendimenti continui, e viceversa. Di conseguenza la scelta di quale tipo di rendimento utilizzare dipende strettamente dalla tipologia dell'analisi condotta. Se l'attenzione è concentrata su una *serie storica*, la capitalizzazione continua tende a essere preferita per la facilità con la quale è possibile aggregare rendimenti consecutivi. Se invece, il campione in

analisi è di *cross-section*, allora è più preferibile utilizzare i rendimenti semplici. Da notare è soltanto la differenza che: *i rendimenti percentuali (o semplici) ci portano ad avere i rendimenti di un portafoglio come somma ponderata dei rendimenti dei singoli titoli, invece per i rendimenti logaritmici (o continui) questo non è valido, in quanto il logaritmo di una sommatoria non è pari alla sommatoria dei logaritmi e la somma di log-normali non è log_normale.*

L'evoluzione del prezzo di un'azione nel tempo è rappresentato da un processo stocastico, cioè successione di variabili casuali indicizzate, chiamato **Random Walk**.

$$\ln p_t = \mu + \ln p_{t-1} + u_t \quad (1.4.10)$$

E si verifica che:
$$E(u_t | p_{t-1}, p_{t-2} \dots) = 0 \quad (1.4.11)$$

Si nota che **il prezzo è somma di tre componenti**: una deterministico (la costante μ), il prezzo precedente e infine il termine casuale $u(t)$, indipendente dai valori passati dei prezzi. L'analisi statistica di una serie storica generata da un processo random walk è complicata, in quanto i prezzi non risultano stazionari in senso debole, cioè il valore atteso condizionato è noto, in quanto uguale a quello in $(t-1)$ e anche la varianza condizionata è nota. Quello che cambia nel RW sono il valore atteso e la varianza non condizionate, così come la covarianza non condizionata che dipendono dal tempo t , questo fa sì che la varianza diventi molto grande.

La *stazionarietà* del processo che ha generato la serie storica può essere verificata mediante vari test di stazionarietà, che vedremo in seguito. Tale proprietà è tipica delle quotazioni ed è gestibile soltanto se si considerano i rendimenti, i quali aiutano a capire cos'è avvenuto e a valutare la performance del portafoglio a cui si è interessati. Come vedremo nel paragrafo successivo, sono i rendimenti lo strumento migliore per ogni analisi che si voglia fare sui titoli o portafogli, in quanto verificano tutte le proprietà necessarie, come la stazionarietà.

Il problema della misurazione del rendimento di un *investimento finanziario* è un problema di scelta e di *calcolo di una media*. In generale, una media viene definita come utile "riassunto", o "caratteristica sintetica" di qualcosa di più complesso, in una logica funzionale, e di "rispondere a un dato scopo". Quindi, non si può dire (in assoluto) se una media è migliore di un'altra, se non nel senso che risponde meglio dell'altra allo scopo dichiarato. Per la definizione dell'indice

di rendimento vale la stessa logica (come visto prima), quindi non è possibile definire un indice come il migliore in tutti i sensi, a prescindere dalle finalità specifiche. In particolare *l'indice di rendimento time-weighted* (dei rendimenti giornalieri, o delle quote) è adeguato al fine di confrontare il rendimento di un fondo con il rendimento del benchmark, prescindendo dall'effetto dei volumi investiti (abilità del gestore); *l'indice money-weighted lineare* consente invece, di misurare appropriatamente il rendimento di periodo, tenendo conto delle movimentazioni, nella logica del tasso interno.

1.4.1 RISK-FREE E EXTRARENDIMENTI

Per le analisi che vogliamo condurre è necessario un titolo non rischioso oltre a quelli rischiosi. Il rendimento risk-free è definito dalla data di acquisto, anche se in realtà tale titolo viene di solito identificato mediante un deposito bancario o un'obbligazione a breve termine. Un risparmiatore, ovviamente, può decidere di scegliere ad investire una parte della propria ricchezza nel titolo non rischioso, tenendo presente le quote di ricchezza investite nel titolo risk-free e quelle nei titoli rischiosi, allora possiamo scrivere:

$$r_p = w_o r_o + w' r \quad (1.4.12)$$

dove, la somma di tutte le quote deve sommare a 1: $w_o + i' \cdot w = 1$. Mediante alcuni passaggi semplici possiamo ottenere la relazione esistente tra il rendimento del portafoglio e il risk-free, inoltre riconosciamo la parte dei rendimenti in eccesso rispetto al titolo non rischioso:

$$z_p = r_p - r_o = \begin{pmatrix} r_1 - r_o \\ \dots \\ r_I - r_o \end{pmatrix} = w' \cdot z \quad (1.4.13)$$

Sottraendo il risk-free a entrambe le parti si ottengono gli extrarendimenti, li indicheremo con z , del portafoglio in funzione degli extrarendimenti dei titoli rischiosi. Il valore atteso dell'extrarendimento di un titolo rischioso è chiamato premio al rischio in quanto rappresenta il compenso atteso per il rischio incorso

detenendolo al posto di un titolo non rischioso. Dato che *il risk-free*, r_0 , non è stocastico, il suo valore atteso e la varianza degli extrarendimenti sono dati da:

$$\mu = E(z) = \mu_r - r_0 \mathbf{i}' \quad (1.4.14)$$

e

$$\Sigma = Var(z) = \Sigma_r \quad (1.4.15)$$

Si nota che: a causa delle fluttuazioni del risk-free nel corso del tempo, le varianze e covarianze campionarie degli extrarendimenti non coincidono con quelle dei rendimenti, quindi è facile verificare che la differenza è molto piccola se la variabilità nel tempo del titolo privo di rischio è piccola rispetto al rendimento dei titoli rischiosi. Segue che, la differenza fra i momenti secondi empirici di r e z non è significativa. I rendimenti e gli extrarendimenti devono *verificare alcune ipotesi* per poter avere validità sulle analisi che si vogliamo svolgere, alcune di queste sono:

- I) entrambi devono essere realizzazioni indipendenti provenienti dalla stessa distribuzione normale multivariata;
- II) il valore atteso dei dati varia nel tempo e questo ci porta ad avere un'immagine distorta della redditività attesa dei titoli, una soluzione a questo problema è interpretare i rendimenti attesi come somma del rendimento risk-free e del premio al rischio, quindi la variabilità è dovuta interamente alle oscillazioni del risk-free;
- III) i rendimenti calcolati con la capitalizzazione continua corrispondono, a meno di una costante additiva, agli errori. Ipotesi diverse sulla distribuzione e sul grado della dipendenza dei rendimenti consecutivi equivalgono a considerare versioni alternative del random walk. Bisogna conservare però, l'ipotesi di rendimenti IID e asintoticamente normali.
- IV) alla fine, ma non meno importante, è l'analisi di correlazione tra i rendimenti contemporanei di titoli diversi. Infatti la correlazione fra titoli può essere più o meno forte, i rendimenti delle azioni di due società che operano nello stesso settore, ad esempio, tendono a essere più correlate di quelle di società che operano in settori diversi, con correlazioni che evolvono durante il ciclo macroeconomico.

CAPITOLO 2

CML, CAPM E INDICI DI PERFORMANCE

In passato, cioè circa fino agli anni '50 l'aspetto del rischio di un investimento finanziario non veniva esplicitamente considerato, e anche se questo veniva fatto, lo si affrontava solo dal punto di vista qualitativo. Soltanto dopo alcuni anni, si ha la nascita della *Moderna Teoria del Portafoglio*, che fa riferimento essenzialmente al contributo di *H. M. Markowitz*. Infatti, è proprio Markowitz a dare inizio all'*analisi quantitativa* a partire dai modelli media-varianza. I modelli successivi sono stati sviluppati per misurare e dare un prezzo al rischio e quello più noto è il *Capital Asset Pricing Model (CAPM) di Sharpe*, di cui approfondiremo le conoscenze in questo capitolo. Vedremo inoltre, quanto sia importante analizzare un fondo in un ottica bidimensionale dato che, tra rendimento e rischio esiste una relazione diretta. Questo significa che, il gestore può ottenere rendimenti sempre più elevati incrementando in modo progressivo il livello di rischiosità del portafoglio gestito, ma d'altra parte aumenta anche la possibilità di andare incontro a perdite sempre più consistenti. Ultimamente invece, si discute molto sulla relazione di trade-off esistente tra rendimento e rischio, cioè se il rendimento identifica quella misura che il risparmiatore cerca di massimizzare, il rischio all'opposto è una misura che i gestori cercano di minimizzare. Il confronto tra i vari strumenti finanziari e quindi tra diversi fondi comuni, vedremo che viene semplificato con la nascita delle misure *Risk-Adjusted Performance Measures (RAPM)* oggetto, anche queste misure, di approfondimento in questo capitolo.

2.1 CML, MODELLO DI MARKOWITZ

Sono gli anni cinquanta, quando *Harry Markowitz* diede origine alla teoria del portafoglio, che diventerà nota, in seguito, come la teoria moderna del portafoglio. L'obiettivo dell'analisi era la valutazione delle attività finanziarie, la quale si doveva basare su due parametri: *rendimento e rischio*. Da allora, anche se sono passati molti anni, più di mezzo secolo, l'obiettivo dell'analisi non è cambiato però, dall'altra parte le analisi di valutazione delle attività finanziarie, come ad esempio: per i fondi, si sono modificate, migliorandosi (come vedremo nei paragrafi che seguono), anche se i concetti base sono sempre quelli di Markowitz.

Il ***modello proposto da Markowitz*** è caratterizzato da un orizzonte temporale monoperiodale e fa riferimento ai *modelli statistici media-varianza*, assumendo come variabili il *rendimento atteso* del portafoglio $E(r)$ o media dei rendimenti dei singoli sottoperiodi e la *rischiosità* legata al portafoglio, assunta pari alla varianza del rendimento del portafoglio $Var(r)$. Il *processo di selezione del portafoglio ottimale* P^* proposto da Markowitz può essere riassunto in tre fasi:

- i) Separazione dei portafogli efficienti dai portafogli inefficienti in base al criterio del rendimento medio e della varianza, o meglio sono definiti ***portafogli efficienti*** l'insieme di tutti i portafogli che per ogni livello di rendimento atteso minimizzano il rischio, misurato dalla varianza del rendimento e definendo una "*frontiera efficiente*". La curva rappresentata da tali punti è identificata da tutte le scelte ottimali di investimento.
- ii) L'individuazione delle *curve di "isoutilità"*, o meglio della specifica funzione di utilità dell'investitore rispetto ai portafogli, tenendo presente il suo grado di avversione al rischio.
- iii) Infine, si ha la determinazione del portafoglio più adeguato, che deve soddisfare la combinazione prescelta dall'investitore ed è definita graficamente dal punto di tangenza tra frontiera efficiente e sistema delle curve di isoutilità.

La presenza del portafoglio r_f , il così detto *Risk-free*, definito da ***James Tobin*** nel 1958. La novità che accompagna l'introduzione del portafoglio risk-free, sta nell'aggiungere al modello media-varianza un'attività r_f , dal rendimento certo r_f e rischio nullo, $Var(r_f) = 0$. Di conseguenza, si ha che il *portafoglio ottimale* viene individuato dal punto di tangenza tra la curva della frontiera efficiente e la

"*capital market line*", cioè la retta che interseca l'asse delle ordinate in corrispondenza del tasso r_f e la cui inclinazione positiva rappresenta il "premio per il rischio". Approfondiremo questo argomento nei paragrafi successivi tenendo presente che, l'esistenza di un titolo risk-free è una delle affermazioni fondamentali alla base del *CAPM* e basandosi anche sull'ipotesi di aspettative omogenee da parte degli investitori, elimina l'aspetto soggettivo nella determinazione del portafoglio ottimo.

La ricerca ha fatto veramente tanto negli ultimi anni, introducendo metodi e strumenti nuovi per la valutazione delle attività finanziarie, anche se i principi base da cui hanno origine sono sempre quelli, e tutto questo lo si deve proprio alla teoria di Markowitz. Oramai, è possibile avere un ordinamento totale nella classificazione dei portafogli rischiosi, evidenziando l'esistenza di un rapporto di correlazione tra un determinato titolo e le variazioni dell'indice di mercato, nel senso che: *mentre una parte del rendimento è da attribuire esclusivamente allo specifico titolo, la parte restante del rendimento complessivo dello stesso è spiegabile attraverso la relazione con il mercato.* La possibilità di classificare i portafogli ci permetterà anche di valutare la performance delle attività finanziarie.

2.1.1 COSTRUZIONE DELLA CML E FE

L'analisi della valutazione del portafoglio è diventata recentemente uno degli argomenti più studiati nell'econometria dei mercati finanziari. In questo paragrafo analizzeremo alcuni dei fattori rilevanti legati alla costruzione della frontiera efficiente.

Supponiamo che, un risparmiatore decida in data 0 come investire la propria ricchezza $W(0)$ fino a data 1, scegliendo fra *un* titolo non rischioso e *I* titoli rischiosi. La sua ricchezza alla data 1 sarà pari a:

$$W_1 = W_0(1 + w_0 r_0 + w' r) = W_0(1 + r_0 + w' z) \quad (2.1.1)$$

$r_0 =$ rendimento titolo non rischioso

$w_0 =$ quota titolo non rischioso nel portafoglio

dove: $w =$ vettore quote ($I \times 1$) titoli rischiosi nel portafoglio

$r =$ vettore rendimenti ($I \times 1$) titoli rischiosi

$z = (r - r_0) =$ vettore extrarendimenti ($I \times 1$) dei titoli rischiosi

$w_0 = (1 - i'w) =$ quota investita nel titolo non rischioso

Alla data 0, si decide ovviamente la quota da investire (w) invece la ricchezza che avremo a data 1, è incerta ($W(1)$). Assumiamo, lo si ha per definizione, che i mercati siano perfetti, nel senso che:

- non esistono imposte;
- non esistono costi di transazione;
- non esistono vincoli alla vendita allo scoperto o al frazionamento dei titoli.

La preferenza dell'investitore sulla ricchezza futura sono descritte dalla sua funzione utilità $U(W(1))$. Il problema dell'investitore consiste nel determinare w in modo da rendere massima l'utilità attesa $E[U(W(1))]$. Markowitz ha fornito una soluzione esplicita, basata sull'ipotesi che l'utilità attesa sia una funzione dei primi due momenti della distribuzione di $W(1)$:

$$E[U(W_1)] = E(W_1) - \eta \cdot \text{Var}(W_1) \quad (2.1.2)$$

dove η rappresenta un parametro positivo che riflette l'avversione al rischio. Questa espressione non è altro che **l'utilità attesa**, funzione crescente del valore atteso della ricchezza futura e decrescente della varianza futura. *L'ipotesi di utilità attesa **media-varianza** vale se la funzione di utilità è quadratica:*

$$U(W_1) = W_1 - \eta \cdot W_1^2 \quad (2.1.3)$$

La formulazione, sopra scritta, viene considerata implausibile in quanto implica avversione assoluta al rischio crescente. Tale problema però, trova un rimedio da *Chamberlain nel 1983*. Infatti, lui mostrò che l'analisi media-varianza è valida per qualsiasi funzione di utilità se la sua *distribuzione dei rendimenti è di tipo ellittico*, o meglio dirsi: *si assume la normalità per i rendimenti e si può dimostrare che la distribuzione normale multivariata è un caso particolare della distribuzione ellittica*.

Un'altra affermazione si riferisce all'utilità attesa media-varianza che rappresenta un'approssimazione della vera utilità attesa, qualunque essa sia e lo si

dimostra utilizzando lo sviluppo in *serie di Taylor* al secondo ordine rispetto a W_1 in un intorno di W_0 .

In base alla ricchezza futura e alla funzione di utilità attesa media-varianza si ha:

$$E(W_1) - \eta \text{Var}(W_1) = W_0(1 - r_0) + W_0[E(w'z) - \eta W_0 \text{Var}(w'z)] \quad (2.1.4)$$

La soluzione di questo problema dipende dalla variazione di $\eta W_0 = \kappa$ e anche di σ^2 e μ , livelli obiettivi di varianza ed extrarendimento atteso:

$$\begin{aligned} w_* = \arg \max_w [E(w'z)] & \quad e \quad w_* = \arg \min_w [\text{Var}(w'z)] \\ \text{s.v. } \text{Var}(w'z) = \sigma^2 & \quad \text{s.v. } E(w'z) = \mu \end{aligned} \quad (2.1.5)$$

matricialmente:

$$\begin{aligned} w_* = \arg \min_w [w' \Sigma w] \\ \text{s.v. } w' \mu = \mu \end{aligned} \quad (2.1.6)$$

La soluzione dell'ottimizzazione vincolata deriva dalla costruzione della funzione lagrangiana:

$$L(w, \lambda) = \frac{1}{2} w' \Sigma w - \lambda (w' \mu - \mu) \quad (2.1.7)$$

Si ottiene che:

$$\begin{aligned} w_* &= \lambda \Sigma^{-1} \mu \\ w_{0*} &= 1 - \iota' w_* \end{aligned} \quad (2.1.8)$$

da cui si ha:

$$\lambda_* = \frac{\mu}{\mu' \Sigma^{-1} \mu} \quad (2.1.9)$$

Qualunque sia l'extrarendimento atteso obiettivo, μ , le quote dei titoli rischiosi nei portafogli ottimali sono proporzionali a $\Sigma^{-1} \mu$. Al variare di μ e quindi di λ^* , viene definita la frontiera dei portafogli efficienti, cioè l'insieme di tutti i portafogli che per ogni livello di rendimento atteso minimizzano il rischio, misurato questo dalla varianza del rendimento. Si ha:

$$\sigma = \pm \frac{\mu}{(\mu \Sigma^{-1} \mu)^{1/2}} \quad (2.1.10)$$

L'asse orizzontale misura lo scarto quadratico medio σ dell'extrarendimento e quello verticale misura il valore atteso μ , cioè si ha una coppia di rette che si intersecano nell'origine. Ovviamente solo quella individuata dal segno positivo rappresenta portafogli scelti da un investitore razionale, (σ, μ) è una semiretta, chiamata Capital Market Line (**CML**):

$$\mu = p s_* \cdot \sigma \quad (2.1.11)$$

Tale risultato rappresenta la relazione tra rischio e extrarendimento atteso, valido per i portafogli e i titoli efficienti in senso media-varianza. La pendenza è definita come performance di Sharpe (1966) dei portafogli efficienti data da:

$$ps_* = (\mu' \Sigma^{-1} \mu)^{1/2} = \frac{\mu}{\sigma} \quad (2.1.12)$$

Tutti i portafogli appartenenti alla frontiera efficiente hanno performance di Sharpe ps_* e posso ottenerli combinando fra loro due portafogli efficienti:

- 1.) il portafoglio che contiene solo il titolo non rischioso (**To**);
- 2.) e il portafoglio che contiene solo titoli rischiosi (**M**).

Risultato questo noto come, **teorema di separazione di due fondi** (Tobin, '58). La semiretta che parte da To e passa per M è la frontiera dei portafogli efficienti in presenza di un titolo non rischioso, To nullo (vedi Figura 2.2.1).

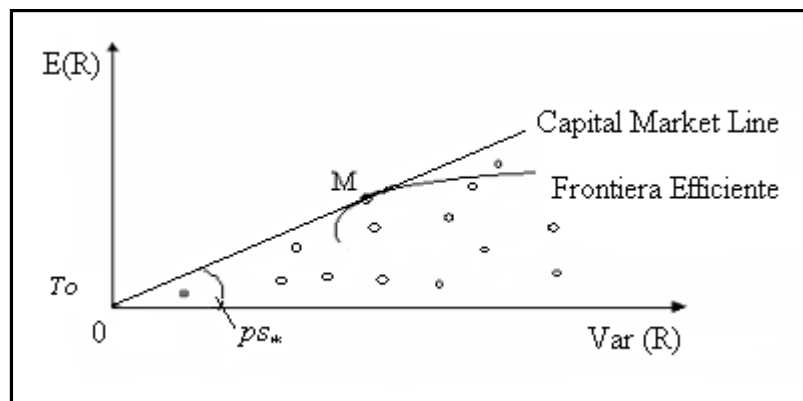


Figura 2.2.1: Capital Market Line senza risk-free, To nullo
(Fonte: Pastorello. S., *Rischio e Rendimento*, cap.2, pag.55)

I portafogli appartenenti alla frontiera efficiente compresi fra To e M investono una quota positiva di W_0 in entrambi, mentre quelli a destra investono una quota negativa di W_0 nel titolo non rischioso, o meglio si indebitano al tasso r_0 e investono il ricavato e la ricchezza iniziale in M . I due portafogli sono composti:

$$T_o : w_{0T_o} = 1, w_{T_o} = 0, \mu_{T_o} = 0, \sigma_{T_o} = 0 \quad (2.1.13)$$

$$M : w_{0M} = 0, w_M = \frac{w_*}{t'w_*}, \mu_M = \frac{\mu' \Sigma^{-1} \mu}{t' \Sigma^{-1} \mu}, \sigma_M = \frac{(\mu' \Sigma^{-1} \mu)^{1/2}}{t' \Sigma^{-1} \mu} \quad (2.1.14)$$

Titoli e portafogli non efficienti sono rappresentati da punti situati al di sotto della CML. La posizione della frontiera efficiente o la composizione teorica dei portafogli efficienti sono funzioni non lineari dei primi due momenti della

distribuzione degli extrarendimenti e sono perciò ignoti. Il modo più conveniente e naturale è quello di sostituire i momenti empirici stimati sulla base di una serie storica con quelli teorici. Dato che i momenti campionari sono stimatori consistenti dei momenti teorici e che sono trasformazioni continue dei momenti teorici di z , **il teorema di Mann e Wald** garantisce che le stime ottenute sostituendo i momenti empirici a quelli teorici sono anche esse consistenti.

In questa analisi, *non tutte le combinazioni di attività rischiose* sono possibili, ma esistono sotto alcuni vincoli:

i.) *vincolo di ammissibilità*:
$$\sum_{i=1}^I w_i = 1 \quad (2.1.15)$$

ii.) alcune combinazioni sono dominate rispetto ad altre, hanno lo stesso rendimento ma varianze inferiori;

iii.) infine, l'insieme dei portafogli efficienti (non dominati) si ottiene come soluzione del problema di ottimo:

$$\begin{aligned} \min_w \quad & w' \Sigma w \\ \text{s.v.} \quad & w' \mathbf{1} = 1 \\ & w' \boldsymbol{\mu} = \mu_p \end{aligned} \quad (2.1.16)$$

Si minimizza il rischio (che si accetta di sopportare) e in modo equivalente si massimizza il rendimento atteso, per un fissato livello di rischio.

Un risultato importante è il seguente: al variare del rendimento atteso del portafoglio, la varianza è figurata da una parabola nel piano (μ, σ^2) e da un'iperbole nel piano (μ, σ) ; anche se, in generale la frontiera efficiente è rappresentata nel piano (σ, μ) ed è quella maggiormente utilizzata.

In presenza di titolo non rischio, si suppone che il rendimento $r_f = \text{risk_free}$ (To è diverso da zero), abbia le proprietà elencate sopra. La presenza del titolo risk-free risulta particolarmente importante per il portafoglio di tangenza (P^*) e il problema di ottimo diventa (non esiste il vincolo di ammissibilità, *Figura 2.2.1*):

$$\begin{aligned} \min_w \quad & w' \Sigma w \\ \text{s.v.} \quad & w' \boldsymbol{\mu} + (1 - w' \mathbf{1}) r_f = \mu_p \\ & w' (\boldsymbol{\mu} - \mathbf{1} r_f) + r_f = \mu_p \end{aligned} \quad (2.1.17)$$

All'aumentare del numero I di titoli rischiosi la frontiera efficiente si allarga rendendo ammissibili nuove combinazioni di rischio e rendimento, ma queste dipendono dalle varianze e dalle correlazioni esistenti tra i vari asset.

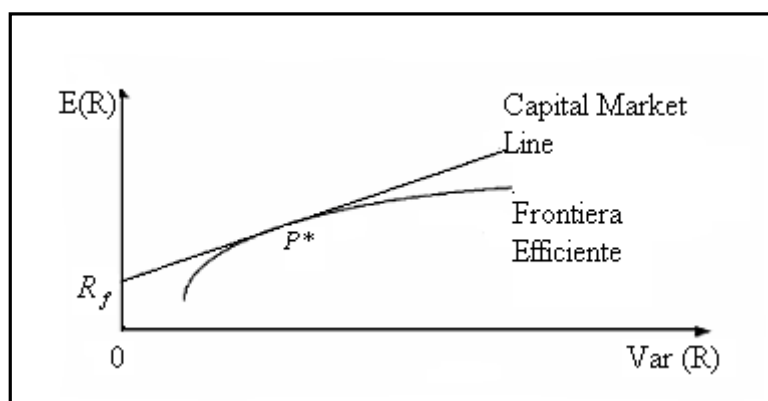


Figura 2.2.2⁹: Capital Market Line con titolo non rischioso

(Fonte: Pastorello. S., *Rischio e Rendimento*, cap.2)

All'aumentare del numero degli asset si hanno maggiori possibilità di diversificazione che permettono di poter raggiungere combinazioni di rischio e rendimento in precedenza non efficienti.

2.2 CAPM: REGRESSIONI IN SERIE STORICA

Quanto visto nel paragrafo precedente è considerato come il punto di partenza per la costruzione del CAPM, infatti l'analisi media-varianza di un titolo non rischioso insieme ai seguenti presupposti ci permettono di definire il CAPM:

- a) a *data 0* i risparmiatori decidono come investire la loro ricchezza fino a *data 1*;
- b) si ha a disposizione un titolo non rischioso e I titoli rischiosi, con vettore di extrarendimenti z di valore atteso μ e matrice di varianza Σ ;
- c) i mercati sono perfetti;
- d) ogni individuo può investire o indebitarsi senza vincoli nel r_f ;
- e) ogni agente ha come utilità attesa media-varianza (con avversioni al rischio diverse), e esiste un numero elevato di investitori *price-taker*, cioè le loro azioni non hanno effetto sulle quotazioni;

⁹ creato in Immagine-Paint;

f) infine, gli investitori hanno le stesse aspettative per μ e Σ .

Tutte queste *ipotesi sono sufficienti* per assicurare l'esistenza di una frontiera dei portafogli efficienti in senso media-varianza, uguale per tutti i risparmiatori. Supponiamo che e , sia un portafoglio appartenente alla frontiera efficiente teorica, la sua *composizione rispetto al titolo non rischio* è data da:

$$\begin{aligned} w_e &= k \Sigma^{-1} \mu \\ \text{con} \quad k &= \frac{\sigma_e^2}{\mu_e} \end{aligned} \quad (2.2.1)$$

dove k è un coefficiente di proporzionalità. Con alcuni passaggi si ha che:

$$\mu = \frac{1}{k} \Sigma w_e \quad (2.2.2)$$

quindi, gli extrarendimenti attesi dei titoli rischiosi sono proporzionali alla covarianza fra z e l'extrarendimento $z(e)$ del portafoglio e :

$$\text{Cov}(z, z_e) = \text{Cov}(z, w_e' z) = \text{Var}(z) w_e = \Sigma \cdot w_e \quad (2.2.3)$$

Gli extrarendimenti attesi di tutti i titoli sono proporzionali alla rispettiva covarianza con qualsiasi portafoglio sulla frontiera efficiente teorica.

Se p è un portafoglio qualsiasi composto da un titolo non rischioso e titoli rischiosi allora $w(p)$ indica la sua composizione rispetto ai titoli che lo compongono. Con alcuni calcoli e passaggi si ha che il *portafoglio e* è caratterizzato da:

$$\begin{aligned} \mu_e &= \frac{1}{k} \sigma_e^2 \quad \text{dove} \quad \mu = \beta^e \mu_e \\ \text{con} \quad \beta^e &= \frac{1}{\sigma_e^2} \Sigma w_e \end{aligned} \quad (2.2.4)$$

Valutando il *titolo i -esimo* e per un qualsiasi portafoglio a cui si è interessati si ha:

$$\mu_i = \beta_i^e \mu_e \quad \text{dove} \quad \beta_i^e = \frac{\sigma_{ie}}{\sigma_e^2} \quad (2.2.5)$$

Questa è una *relazione lineare*, senza intercetta, fra l'extrarendimento atteso di qualsiasi titolo o portafoglio e il rispettivo β rispetto a un portafoglio efficiente qualsiasi; o meglio dicasi, misura il rischio di un titolo con la sua covarianza rispetto a un portafoglio sulla frontiera efficiente teorica. Il tutto si può *verificare empiricamente*, anche se ci si trova molto spesso a dover affrontare due principali ostacoli: la relazione tra momenti teorici e il portafoglio e deve appartenere alla

frontiera efficiente teorica. Quindi, per verificare la relazione (2.2.5) usiamo i momenti empirici e un portafoglio sulla frontiera efficiente empirica, l'ipotesi di una relazione di proporzionalità fra extrarendimenti attesi e beta viene automaticamente accettata. La validità empirica si ha soltanto mediante una teoria economica, basata su ipotesi supplementari, in modo che venga individuato a priori un portafoglio sulla frontiera efficiente teorica a partire dalle condizioni di equilibrio tra domanda e offerta sul mercato dei titoli. Punto di partenza è la determinazione delle domande di titoli da parte dei singoli investitori, rispetto alle loro preferenze descritte da una funzione utilità attesa quadratica (media-varianza) nel piano (σ, μ) , dove le curve di indifferenza associate a un livello di utilità data, sono rappresentate da un fascio di parabole. Al crescere dell'utilità attesa, le parabole si spostano verso l'alto evidenziando combinazioni di extrarendimento atteso e rischio che garantiscono un livello di utilità attesa superiore. Essendo l'obiettivo dell'investitore nel massimizzare la propria utilità attesa, la soluzione è data geometricamente dal punto di tangenza tra la frontiera efficiente in presenza di un titolo non rischioso e la curva di indifferenza più alta possibile.

L'offerta aggregata di titoli rischiosi è costituita dalle quantità di titoli fisicamente disponibili sul mercato, mentre il titolo non rischioso ha offerta netta nulla, cioè il suo rendimento si aggiusta per uguagliare domanda e offerta di prestiti. La condizione di equilibrio tra domanda e offerta di titoli rischiosi implica che il portafoglio di mercato (con le sue quote) coincida con il portafoglio di tangenza M, situato sulla frontiera efficiente teorica. Con una semplice sostituzione delle notazioni si ha:

$$\mu_i = \beta_i \mu_M \quad \text{dove} \quad \beta_i = \frac{\sigma_{iM}}{\sigma_M^2} \quad (2.2.6)$$

Quanto ricavato va sotto il nome di **Security Market Line (SML)** ed è la principale formula del **CAPM**. $\beta(i)$ misura il *rischio sistematico di un titolo*, non diversificabile perché legato alle fluttuazioni del mercato ed è remunerato con un extrarendimento atteso positivo. La formula (2.2.5) si verificata empiricamente solo se si identifica a priori un portafoglio appartenente alla FE. L'unica implicazione verificata del CAPM è che *il portafoglio di mercato sia efficiente in senso media-varianza*, quindi la verifica empirica richiede necessariamente la conoscenza della composizione del portafoglio di mercato, o di una serie storica

dei suoi extrarendimenti. Infatti, **Roll nel 1977** osservò che: *il portafoglio di mercato contiene tutti i titoli rischiosi accessibili agli investitori (azioni, obbligazioni, opere d'arte, immobili, ecc) e per questo motivo al suo posto viene di solito utilizzato un portafoglio m, più limitato, costituito normalmente da un indice azionario*. Il portafoglio m , non è altro che una proxy del vero portafoglio di mercato e non consente di trarre alcuna conclusione sulla sua efficienza in senso media-varianza. Secondo Roll, qualsiasi proxy ben costruita è con ogni probabilità altamente correlata sia con il vero portafoglio di mercato M e anche con un portafoglio sulla FE empirica. *L'efficienza di m è dunque confermata:*

- a) dalla validità del CAPM;
- b) oppure da una elevata correlazione con il portafoglio sulla FE.

Invece, un motivo di *rifiuto dell'ipotesi di efficienza di m* potrebbe essere data da:

- i) una bassa correlazione con M ;
- ii) oppure la non validità del CAPM.

Di conseguenza si ha che, la *critica di Roll* rappresenta un grande ostacolo alla verifica empirica del CAPM ed è uno strumento estremamente interessante nei casi in cui sia necessaria una misura della relazione tra rischio e rendimento atteso di un titolo o un investimento.

La verifica del CAPM empiricamente è stato originariamente suggerito da **Black, Jensen e Scholes (1972)**. Si considera un titolo i qualsiasi e per questo lo si valuta un modello di regressione lineare semplice:

$$z_{it} = \alpha_i + \beta_i z_{mt} + u_{it} \quad (2.2.7)$$

$$E[u_{it} | z_{mt}] = 0 \quad e \quad Var[u_{it} | z_{mt}] = w_i^2$$

dove:
$$\beta_i = \frac{\sigma_{im}}{\sigma_m^2} \quad (2.2.8)$$

$$\alpha_i = \mu_i - \beta_i \mu_m$$

con $t = 1, \dots, T$ (*indice temporale*) e in forma matriciale si ha:

$$\underset{(1 \times 1)}{z_t} = \underset{(1 \times 1)}{\alpha} + \underset{(1 \times 1)}{\beta} \underset{(1 \times 1)}{z_{mt}} + \underset{(1 \times 1)}{u_t} \quad (2.2.9)$$

Mentre, β è dato dal suo vero valore, α per il CAPM scritto sopra implica che sia nullo. Gli stimatori *MQO* di questi parametri sono dati da:

$$\begin{aligned}\hat{\alpha}_i &= \bar{z}_i - \hat{\beta}_i \cdot \bar{z}_m \\ \hat{\beta}_i &= \frac{\hat{\sigma}_{im}}{\hat{\sigma}_m^2}\end{aligned}\tag{2.2.10}$$

I valori trovati non sono altro che i corrispondenti empirici dei valori teorici dei parametri. In particolare il primo è l'*alpha di Jensen* (di cui parleremo in seguito) e rappresenta una stima della remunerazione attesa del titolo *i* non giustificata dalla sua esposizione al rischio di mercato, questa è anche definita come una misura della performance di portafogli gestiti, oggetto di studio nei paragrafi successivi.

Le proprietà dei due parametri ricavati vengono di solito derivate assumendo extrarendimenti $(z_{it}, z_{mt})'$ realizzazioni indipendenti e identicamente distribuite (IID) tratte da una distribuzione *normale bivariata*. Su queste basi la distribuzione degli stimatori è:

$$\begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix} | z_m \sim N \left[\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}; \frac{w_i^2}{T} \cdot \begin{pmatrix} 1 + \frac{\bar{z}_m^2}{\hat{\sigma}_m^2} & -\frac{\bar{z}_m^2}{\hat{\sigma}_m^2} \\ -\frac{\bar{z}_m^2}{\hat{\sigma}_m^2} & \frac{1}{\hat{\sigma}_m^2} \end{pmatrix} \right]\tag{2.2.11}$$

dove la matrice di varianza si riferisce al caso con errori omoschedastici. Per verificare l'ipotesi nulla $H_0: \alpha_i = 0$ associata alla validità del CAPM, o altre ipotesi lineari su questi parametri, possiamo utilizzare la statistica test t:

$$t = \frac{\hat{\alpha}_i}{\hat{w}_i [(1 + \bar{z}_m^2 / \hat{\sigma}_m^2) / (T - 2)]^{1/2}} \sim t_{T-2} \approx N(0,1)\tag{2.2.12}$$

e per T grande, la statistica test è approssimata ad una normale standard. Se gli extrarendimenti $(z_{it}, z_{mt})'$ sono IID ma non normali, la distribuzione dei due parametri e la statistica test precedentemente scritta valgono soltanto come approssimazioni per T sufficientemente grande. In riferimento a quanto visto è da notare che le proprietà di extrarendimenti IID normali può essere valida per osservazioni di frequenza mensile, ma con osservazioni a frequenza più elevata è difficilmente accettabile. Le violazioni di queste ipotesi possono essere accertati utilizzando statistiche test diagnostiche come:

- i) in caso di eteroschedasticità, può essere utile calcolare la varianza degli stimatori dei MQO con la formula di *White*, oppure ricorrere a uno stimatore robusto rispetto a violazioni degli assunti sugli errori come GMM di *Hansen*;

- ii) in caso di autocorrelazione, si valuta il test di *Ljung-Box* per valutare l'ipotesi nulla di assenza di autocorrelazione seriale dei rendimenti e degli extrarendimenti;
- iii) in caso di medie nulle, ipotesi interessante per lo studio effettuato, si fa riferimento alla statistica test *t* ;
- iv) infine, in caso di normalità dei rendimenti e extrarendimenti, l'ipotesi nulla è verificata mediante il test di *Jarque-Bera*.

Per limitare la variabilità dei parametri nel tempo è consuetudine stimare i parametri su campioni da 5 a 7 anni.

Infine, sottolineiamo l'importanza che assume il coefficiente di determinazione R^2 , che misura il grado al quale la performance di un fondo sia correlato ad un benchmark esterno. Per la proprietà dei MQO la *varianza campionaria totale degli extrarendimenti* è scomposta da:

- I) *rischio sistematico*, legato alle oscillazioni di $z(m)$ del titolo i ;
- II) *rischio specifico*, che può essere limitato per diversificazione.

In base a tali componenti l'*indice di determinazione* R^2 può essere scritto come:

$$R^2 = 1 - \frac{\hat{w}_i^2}{\hat{\sigma}_i^2} = \frac{\hat{\beta}_i^2 \hat{\sigma}_m^2}{\hat{\sigma}_i^2} \quad (2.2.13)$$

Questo indice è interpretabile come, *la stima della percentuale di rischio sistematico sul rischio complessivo del titolo i -esimo*. Ad esempio: se un fondo di azioni ordinarie presenta un $R^2 = 0.79$ in relazione ad un indice S&P1500, questo significa che il 79% del comportamento storico del fondo è attribuibile ai movimenti dell'indice S&P1500.

2.3 STRATEGIE DI GESTIONE DEL PORTAFOGLIO

La crescita esponenziale del risparmio gestito è stata accompagnata da un notevole interesse anche verso tutte le problematiche ad esso connesse, sia da parte della letteratura finanziaria e sia dal mondo dell'informazione. Attualmente dati e

ricerche, con frequenza quotidiana vengono resi noti da stampe specializzate e da soggetti istituzionali qualificati per lo studio della gestione collettiva delle attività finanziarie. Negli ultimi anni, una quota rilevante della ricchezza dei paesi economicamente sviluppati (e non) non è investita solo in titoli altamente rischiosi, ma affidata a investitori istituzionali che amministrano Sicav, fondi pensione, fondi comuni d'investimento (protagonisti indiscussi della mutata composizione del patrimonio mobiliare delle famiglie americane e non solo) hedge fund, ecc.

Mediante *la valutazione della performance* si cerca di cogliere, insieme agli studi precedenti sul tema della gestione dei portafogli, il contributo delle innumerevoli analisi empiriche condotte da società di rating, organismi di vigilanza e media. Le analisi si svolgono sempre con l'ottica della **tutela dell'investitore**, con particolare riguardo al *profilo della trasparenza*. Inoltre, risulta fondamentale la scelta del benchmark di riferimento per la valutazione della performance, in quanto per capire l'andamento del fondo e lo stile di gestione in un determinato arco di tempo questo deve essere confrontato con un determinato benchmark. *Il benchmark*, non è altro che un indice o un insieme di indici che sintetizza l'andamento e le opportunità dei mercati in cui investe il fondo, in altre parole costituisce un parametro di confronto dell'andamento del fondo comune. Sempre il gestore è obbligato ad indicare¹⁰ il benchmark che è stato usato per ciascun fondo gestito. *La strategia di gestione di un portafoglio* può essere classificata come:

- a) **Gestione passiva:** in tal caso la composizione del portafoglio viene scelta sulla base di un obiettivo già prestabilito, come la replicazione di un indice, limitando al minimo il numero e il volume delle transazioni. *L'obiettivo del gestore* è quello di *replicare esattamente il benchmark* e l'extrarendimento di tale portafoglio è determinato unicamente dalla sua esposizione al rischio sistematico.
- b) **Gestione attiva:** in questo secondo caso, la composizione del portafoglio ha come obiettivo, quello di avere un rendimento medio superiore a quello di un indice definito dal gestore. Quindi, lo scopo è quello di *“battere” il benchmark* mediante un'accurata selezione dei titoli oppure anticipando le oscillazioni del loro valore. I maggiori costi (di analisi e di transizione) di

¹⁰ all'interno del prospetto informativo;

questa gestione sono giustificabili solo se l'intermediario è in possesso di informazioni non completamente riflesse nei prezzi e che li consentono di realizzare un extrarendimento atteso superiore a quello associato all'esposizione al rischio sistematico. L'extrarendimento atteso addizionale è la performance del portafoglio attivo, ed è vista a volte come la differenza fra il suo premio al rischio e quello del portafoglio passivo a lui più simile. Tale politica di gestione mira a una concentrazione del gestore nell'attività di stock selection e nell'attività tempistica degli interventi sul mercato, *market timing*.

L'investitore si preoccupa prevalentemente delle *reali capacità del money manager di creare extraperformance*: solo se il gestore è in grado di trarre dalle informazioni in suo possesso, come elementi di previsione sull'andamento dei mercati tali da garantire al portafoglio da lui gestito una performance superiore a quella del benchmark (al netto dei maggiori costi che la gestione attiva comporta), appare sensato delegare l'amministrazione del proprio patrimonio mobiliare.

Gli studi di **Eugene Fama** hanno evidenziato come il tentativo di sfruttare le informazioni alla ricerca di "*miss-pricing*" sia essenzialmente un tentativo vano. In quanto, i mercati finanziari, almeno quelli più evoluti, sono sostanzialmente efficienti nel senso che "*prezzano*" i vari titoli presenti sul mercato in modo da riflettere tutte le informazioni disponibili. Una versione meno "integralista" della teoria, riconducibile a Jensen sostiene che, anche qualora vi fossero "nicchie" (aree di inefficienza) nel mercato, la gestione attiva sarebbe comunque incapace di superare, in termini di performance, il mercato a causa dei costi di transazione. Diverse ricerche empiriche sembrano confermare tali affermazioni, le cui principali implicazioni sono: la netta prevalenza dell'asset allocation strategica quale fattore determinante della performance di lungo periodo dei portafogli e la superiorità della gestione passiva su quella attiva, legata ai maggiori costi della seconda. L'individuazione del portafoglio passivo dipende essenzialmente dal fatto che sia nota la composizione del portafoglio attivo, altrimenti l'analisi viene condotta sulla base degli extrarendimenti secondo l'approccio di Jensen.

L'approccio di Jensen fornisce evidenze empiriche a sostegno della tesi secondo cui: i gestori non dispongono di informazioni riservate che consentano loro di ottenere extraperformance. Infatti, **Ippolito** nel 1988 riesce a dimostrare la

capacità dei fondi di battere il benchmark, ma non al netto dei costi di transazione e delle management fees (onorario). Risultati del tutto analoghi a questi ultimi sono ottenuti da **Cesari e Panetta**, anche se loro tendono a sottolineare come ulteriormente contro le presunte "*superior skills*" (coloro che hanno abilità superiori) dei money manager vi siano: la possibile scelta di benchmark inefficienti con cui confrontare i portafogli ed il fenomeno noto come "*survivorship bias*" (selezione derivata dalla sopravvivenza a certi fenomeni). *I fondi con le peggiori performance*, vengono abbandonati dagli investitori, scompaiono (vengono liquidati o incorporati) e inevitabilmente dei loro risultati negativi non si può tenere conto ai fini statistici.

Inoltre, Cesari e Panetta considerano un ulteriore aspetto, cioè *la scarsa persistenza dei risultati*. Non solo gli studi precedentemente citati e altri ancora più empirici dimostrano come nella maggior parte dei casi i *fondi gestiti attivamente* abbiano rendimenti inferiori rispetto al benchmark ed agli *index funds*. Inoltre, anche se vengono individuati i gestori che "*battono*" il mercato, tale evidenza sarebbe di scarsa utilità, dal momento che "*i rendimenti passati non sono indicativi di quelli futuri*" e alla luce del fatto che, con sempre maggiore frequenza, i "*money manager*" si trasferiscono da una S.g.r. all'altra.

Di conseguenza possiamo affermare che, *un'approfondita analisi della performance non deve limitarsi a quantificare i livelli di rischio e rendimento*, ma deve spingersi fino all'individuazione delle componenti dei risultati conseguiti e dell'approccio all'investimento del money manager.

In particolare, focalizziamo l'attenzione sulla *performance attribution* (o *return attribution*), la quale è definita come una metodologia empirica che consente di scomporre il rendimento complessivo di un fondo in singole componenti autonomamente identificabili nel processo di gestione di un portafoglio. Tale processo si articola principalmente in tre attività principali:

- i) **Asset Allocation Strategica**: tale attività permette al gestore di definire il grado di specializzazione/diversificazione del portafoglio. Il livello strutturale di rischio/liquidità e le aree in cui il fondo dovrà investire, suddividendo il patrimonio fra tipologie di titoli (azioni, obbligazioni, liquidità), aree geografiche e settori merceologici sulla base dell'orizzonte temporale prescelto, dei rendimenti attesi, del livello di rischio e delle correlazioni tra le

diverse tipologie di strumenti finanziari. In pratica, *dalla combinazione di asset class elementari* (viste nel primo capitolo) *è possibile costruire infinite asset classes composite* (vedi Tabella 2.3.1). Dalla tabella riportata, si nota che l'asset class "Azionaria", pone come obiettivo finale l'incremento del capitale, cioè si attendono rendimenti alti a scapito di un rischio alto, calcolato secondo DS (deviazione standard o scarto quadratico medio) pari a 5.04, inoltre il tempo dell'investimento rientra tra quelli a lungo termine. Invece, l'asset class "Monetaria" è definita la più elementare in modo assoluto, in quanto fa riferimento ai cash flow velocemente liquidabili con un rischio bassissimo, intorno al 0.15 ed è soggetto ad un orizzonte temporale d'investimento breve, infatti non supera l'anno.

Tabella 2.3.1: *Caratteristiche della composizione delle asset class elementari*

<i>ASSET CLASS</i>	<i>ORIZZONTE TEMPORALE IN MESI</i>	<i>LIVELLO DI RISCHIO IN DS</i>	<i>ESIGENZE D'INVESTIMENTO</i>
<i>Monetaria</i>	Fino a 12	0,15	fondo di scorta rapidamente liquidabile, a breve termine
<i>Obbligazionaria</i>	12-36	0,94	fonte periodica di reddito integrativo
<i>Bilanciata</i>	36-60	3,11	incremento del capitale con prelievi occasionali
<i>Azionaria</i>	Oltre 60	5,04	incremento del capitale come unico obiettivo

Fonte: AreaBanca, www.areaspa.it

- ii) **Asset allocation tattica (market timing):** in questo caso le classi di attività in analisi vengono sovrappesate o sottopesate rispetto ai pesi normali, definiti nella fase precedente, al fine di migliorare il profilo di rischio-rendimento del portafoglio gestito. In altre parole, si cerca di anticipare i cambiamenti nelle variabili macroeconomiche.
- iii) **Stock picking o security selection:** tale attività anche se l'ultima non è meno importante rispetto alle precedenti. Infatti, in questa attività sono favoriti alcuni titoli a discapito di altri, nell'ambito del peso complessivo del portafoglio destinato ad un certo mercato, attraverso vari modelli di analisi tecnica ed un'analisi fondamentale sui singoli titoli.

Gli indirizzi strategici visti sopra, cioè gli *assets allocation strategici* sono descritti in base alle norme del Regolamento del Fondo, che ne definiscono sia la tipologia e sia gli obiettivi generali del fondo stesso. Seguono inoltre, una serie di divieti e di limiti massimi che devono essere rispettati negli investimenti.

Se invece, consideriamo le scelte che vanno ad influenzare *la componente “attiva”*, allora si possono valutare due classi di strategie di gestione dei portafogli e sono:

- I.) **Approccio Top-Down:** analisi che ci permette di definire un portafoglio in base alle analisi di tipo “macro”, e hanno come obiettivo quello di fornire indicazioni legate all’evoluzione dei mercati e dei settori o di analizzare i rischi che si possono incontrare. In tal caso le varie aree in analisi sono tra loro correlate e contribuiscono alla definizione delle scelte strategiche d’investimento. Ovvero, la selezione delle proporzioni d’investimento nelle diverse aree geografiche viene effettuata partendo dall’analisi delle variabili macroeconomiche che potrebbero influenzare gli andamenti dei mercati finanziari internazionali nel medio e lungo periodo.
- II.) **Approccio Bottom-Up:** è invece, un’analisi basata su singole imprese con l’obiettivo di evidenziare le opportunità di profitto. In generale, gli analisti che seguono tali strategie sono degli specialisti in un singolo settore, area o paese. Gli strumenti di analisi partono molto spesso dallo studio del bilancio (quindi da tutti gli indici economici), del posizionamento strategico delle imprese e anche del mercato nel quale opera (*ad esempio: l’impresa in esame*). Quindi, vengono selezionati i titoli azionari che danno buone opportunità di profitto sulla base dell’analisi sia delle singole aziende e sia dei settori produttivi a cui appartengono.

La valutazione della performance, nella sua complessità, da sola non basta perché si basa su informazioni passate e sappiamo che non esiste nessuna garanzia che questo si possa ripetere nel futuro, anche se le previsioni aiutano. Risulta molto importante confrontare la performance con i costi che si sostengono e soprattutto con il tipo di rischio affrontato nello stile di gestione. Vedremo che, il modo più semplice per la valutazione del rischio è misurare la *volatilità*, cioè l’oscillazione della variazione percentuale del valore della quota del fondo nel

tempo. Il fondo migliore non è quello che ha conseguito elevati rendimenti medi, ma quello che ha conseguito elevati rendimenti medi in relazione al rischio.

Mediante l'uso delle misure RAPM, il confronto tra i vari assets classes risulta semplificato dato che l'informazione necessaria a valutare il fondo si basa sul trade-off rendimento/rischio. Nei paragrafi successivi vedremo che il fondo con la misura RAPM più elevata risulta anche essere quello che ha ottenuto la performance migliore nell'ambito della relazione esistente tra rendimento e rischio.

Inoltre, anche se gli Stati Uniti hanno cominciato a studiare il problema della valutazione della performance verso la fine del 1960, sono però tutt'ora aperte le polemiche su *quale sia la misura di performance in grado di rappresentare più correttamente la distribuzione storica o prospettica dei rendimenti di un investimento*. In effetti, nell'impossibilità di individuare gli indicatori più validi in assoluto, è almeno importante che l'investitore selezioni gli strumenti che risultino il più conformi possibile all'obiettivo che si desidera raggiungere. Vedremo in seguito, cosa significa questo per i fondi comuni in analisi e quali indici sono maggiormente informativi.

2.4 DEFINIZIONE DELLA PERFORMANCE

La performance di Sharpe per un portafoglio p lo abbiamo definito come:

$$ps_p = \frac{\mu_p}{\sigma_p} \quad (2.4.1)$$

e tale misurazione per la sua semplicità è definito come *uno degli strumenti di valutazione della performance più utilizzati nella stampa finanziaria*. Il difetto principale è quello di correggere l'extrarendimento atteso usando il *rischio totale*, questo lo rende una misura di performance appropriata per un risparmiatore che investe tutta la propria ricchezza in un unico portafoglio.

Grinblatt e Titman (1989) considerano le origini della performance di un portafoglio attivo e le sue componenti principali. I risultati ottenibili fanno

valuteremo una cluster analysis per i 50 fondi comuni d'investimento statunitensi, ma prima definiremo il benchmark e le critiche che da sempre accompagnano la scelta di tale indice, il quale è essenziale nella valutazione della performance del portafoglio.

2.4.1 SCELTA DEL BENCHMARK

Quanto visto nei paragrafi precedenti, *formule 2.2.5 e 2.2.6* sottolineano che, il rischio sistematico di un portafoglio è misurato dai coefficienti della regressione lineare sugli extrarendimenti con variabili che approssimano le fonti di rischio non idiosincratiche. Per la valutazione della performance nei portafogli passivi vedremo che viene utilizzato come indice del rischio sistematico *il benchmark*.

Il benchmark è un indice di mercato, riferito ad un determinato paniere-campione e di solito utile a definire la *politica di gestione di un fondo comune* per quanto riguarda: il tipo di titoli acquistabili, la quota del patrimonio investita in ogni classe, il grado di rischio, il tipo di gestione (attiva o passiva), ecc. In generale, affinché un indice (o una combinazione di indici) possa essere validamente impiegato come benchmark, è auspicabile che possieda le seguenti caratteristiche:

- *Trasparenza*: in quanto, gli indici devono essere calcolati con regole replicabili autonomamente dall'investitore.
- *Rappresentatività*: è quindi importante che le classi, nonché i titoli inclusi negli indici debbano riflettere le opportunità di investimento disponibili.
- *Replicabilità*: in teoria e così come anche in pratica, gli indici dovrebbero essere completamente replicabili con attività acquistabili direttamente sul mercato.

In seguito a un ragionamento simile, si che il confronto tra un portafoglio costruito teoricamente e un portafoglio in cui si possa effettivamente investire, comporta una serie di problematiche legate alla ponderazione dei costi di gestione e alla tassazione. Ovviamente che, se il benchmark è costruito con attività realmente disponibili allora è ben rappresentativo di una realistica misura di performance.

Negli Stati Uniti, il benchmark è nato come uno strumento di confronto tra il risultato della gestione ed un parametro di riferimento del mercato (scelto ad hoc, cioè assets classes scelti in modo scrupoloso), il tutto è svolto con la massima trasparenza attorno all'informazione finanziaria.

In Italia, non diversamente dal resto dei mercati finanziari mondiali, il benchmark a cui si fa riferimento detiene le stesse funzioni ed è definito in modo più accurato dal Testo Unico.

Molto spesso è necessario descrivere sinteticamente il benchmark adottato per ciascun fondo (o asset class) che si analizza e soprattutto si deve tenere presente che il parametro deve essere conforme ai principi previsti dal Regolamento Consob. Le funzioni svolte dal benchmark sono diverse, ma tutte dipendono dall'obiettivo finale del gestore. Nel caso in cui non è possibile per uno o più fondi individuare un benchmark (*ad esempio*, per i fondi flessibili), occorre indicare le ragioni di tale impossibilità. Inoltre, nel descrivere lo stile gestionale, va indicata la relazione esistente tra parametro di riferimento prescelto come benchmark e l'obiettivi del fondo. Infine, il benchmark può essere utilizzato anche a fini di valutazione della qualità dei prodotti di investimento ed è corretto se e solo se si tiene presente il Regolamento della Consob. In particolare, in questo lavoro si farà riferimento alla definizione secondo la quale: *il benchmark è un indice rispetto al quale esiste una relazione esatta fra rischio ed extrarendimento atteso per tutti i portafogli passivi*, o meglio dicasi attribuisce performance nulla a tutti i portafogli passivi, quindi ogni portafoglio collocato sulla FE teorica può essere un benchmark.

Se supponiamo che, \mathbf{b} sia un portafoglio appartenente alla FE, allora lo si considera come benchmark e per le *proprietà dei portafogli efficienti* si ha:

$$\begin{aligned} z_i &= \beta_i \cdot z_b + u_i \\ \beta_i &= \frac{\text{Cov}(z_i, z_b)}{\text{Var}(z_b)} \end{aligned} \quad (2.4.3)$$

dove, i titoli sono sempre $i=1, \dots, I$ e la media dell'errore è nulla. In generale, l'extrarendimento del portafoglio \mathbf{p} è dato da:

$$\begin{aligned} z_p &= \beta_p z_b + u_p \\ \beta_p &= \sum_{i=1}^I w_i \beta_i \quad e \quad u_p = \sum_{i=1}^I w_i u_i \end{aligned} \quad (2.4.4)$$

si nota che per un portafoglio passivo β è costante e la media dell'errore è nulla e in base al valore atteso degli extrarendimenti del portafoglio si ha:

$$\mu_p = E(\beta_p z_b) + E(u_p) = \underset{1^\circ}{E(\beta_p)} \mu_b + \underset{2^\circ}{Cov(\beta_p, z_b)} + \underset{3^\circ}{E(u_p)} \quad (2.4.5)$$

il tutto valutato senza la necessità di conoscere la composizione di w . Il primo addendo della *formula 2.4.4* è l'extrarendimento atteso del portafoglio p in corrispondenza della sua esposizione media al rischio sistematico, mentre il secondo e il terzo rappresentano l'extrarendimento atteso supplementare generato rispettivamente dalla relazione fra β_p e z_b , e pure dal fatto che per certi titoli l'extrarendimento atteso può non coincidere con quello individuato dal rispettivo rischio sistematico. In un portafoglio passivo (per ipotesi) il primo addendo è diverso da 0 e tutto il resto è nullo, mentre in quello attivo la performance generata dalla disponibilità di informazione privata è generata dalla somma degli altri due termini, che sono:

- 1.) **timing**: che rappresenta la capacità da parte del gestore di anticipare le oscillazioni del benchmark e di approfittarne modificando $\beta(p)$;
- 2.) **selectivity**: invece, rappresenta la capacità di individuare i titoli collocati al di sopra della relazione rischio e rendimento atteso definita da b .

La performance del portafoglio attivo in complessivo può essere interpretata come la differenza fra il suo extrarendimento atteso e quello di un portafoglio passivo che ha in media la stessa esposizione al rischio sistematico. Come enunciato prima il portafoglio b è collocato sulla FE dei portafogli passivi, per questo motivo è considerato come *un benchmark*. Sotto ipotesi di equilibrio sul mercato delle attività rischiose il portafoglio di mercato M è un benchmark valido, ma come evidenziato da Roll (1977), M non è direttamente osservabile, quindi usiamo come proxy un portafoglio m che può non essere efficiente.

Roll nel 1978 riuscì a dimostrare che, la graduatoria di un gruppo di fondi sulla base della loro performance può essere completamente sovvertita passando da un benchmark efficiente a uno inefficiente (e viceversa). Dalla evidenza empirica si è verificato che il benchmark rappresenta il passaggio decisivo nella valutazione della performance, come vedremo anche per i fondi comuni in analisi. Molto spesso si fa riferimento a un *benchmark multivariato* (composto di extrarendimenti), il quale lascia sostanzialmente inalterata la validità della

performance del portafoglio a condizione di modificare leggermente le definizioni di timing e selectivity. In generale, si definisce con **timing** la capacità di prevedere le componenti sistematiche del rendimento (portafoglio appartenente alla FE) e invece con **selectivity** la capacità di prevedere la componente idiosincratca degli extrarendimenti.

Nelle analisi successive si farà riferimento a misure di performance che richiedono la conoscenza di una *serie storica di rendimenti, quindi anche degli extrarendimenti del portafoglio attivo p e del benchmark b* . In base alle informazioni ottenibili dai rendimenti/extrarendimenti (a priori) e dal **benchmark univariato**, svolgeremo un'analisi sulla performance dei fondi comuni d'investimento statunitensi, ricordando che le abilità del gestore influenzano solo la selectivity e non il timing, per questa ragione valuteremo cosa avviene in assenza del timing.

2.5 MISURE DI PERFORMANCE SUI RENDIMENTI

Il calcolo dei rendimenti è solo una delle fasi della procedura di valutazione della performance degli investimenti. Nella prassi gestionale, il processo di valutazione è profondamente cambiato negli ultimi decenni, rifondato sulla moderna teoria del portafoglio, da un semplice calcolo di rendimento si è passati a una complessa "*misurazione congiunta*" di rischio e di rendimento. Alla luce di quanto visto nei paragrafi precedenti assumiamo di esaminare un portafoglio attivo con **timing nullo**, quindi come punto di partenza naturale per valutare la performance è la regressione:

$$z_{pt} = \alpha_p + \beta_{pt} z_{bt} + u_{pt} \quad (2.5.1)$$

per $t = 1, \dots, T$ ed errori per i quali valgono tutte le proprietà che rendono efficienti i MQO, e questa osservazione è verificata se le osservazioni di $(z_{pt}, z_{bt})'$ sono IID, cioè, realizzazioni indipendenti tratte dalla stessa distribuzione normale. Nella

formula 2.5.1, la *selectivity coincide con la performance* ed è misurata da α_p , il cui vero valore è: $[\mu_p - \beta_p \mu_b]$, o meglio l'extrarendimento atteso non giustificato dall'esposizione al rischio sistematico. Come abbiamo visto nella *formula 2.4.3* se il portafoglio b è efficiente allora l'alpha di tutti i portafogli passivi deve essere nullo, a condizione di misurare i rendimenti al lordo delle imposte, costi di transazione e commissioni varie. Invece, nel portafoglio attivo α_p , l'alpha di Jensen, è diverso da zero se si hanno informazioni su come selezionare i titoli aventi un premio per il rischio superiore a quello associato all'esposizione al rischio non diversificabile.

Nei paragrafi successivi valuteremo misure di performance ricavabili *dai modelli econometrici* (regressioni lineari in base alla *formula 2.5.1*), con l'ipotesi di assenza del timing, e misure definite *Rap (risk-adjusted performance)*.

2.5.1 MISURE DI PERFORMANCE BASATE SU MODELLI ECONOMETRICI

Gli indici principali, informativi della performance dei fondi, che si basano su modelli econometrici, cioè ricavabili mediante le regressioni lineari degli extrarendimenti costruiti in riferimento a un determinato risk-free e benchmark, sono due: l' α di Jensen e il β di un fondo o portafoglio di fondi. Vediamo in particolare come sono definite queste due misure di performance e come possiamo verificare la loro significatività, alla luce dei risultati del CAPM.

2.5.1.1 MISURA ALPHA DI JENSEN

Nel 1968 l'economista *Michael Jensen* sviluppò una *misura di performance basata su modelli econometrici* al fine di analizzare l'abilità del gestore di un fondo d'investimento. L'idea di fondo è di prevedere i prezzi futuri delle attività finanziarie in modo da determinare la capacità del gestore di selezionare i titoli

sottovalutati. Tale misura, comunemente è denominata come alfa di Jensen (α) ed è definita come: il rendimento incrementale o extrarendimento che un fondo di investimento ha prodotto rispetto alla redditività che avrebbe dovuto offrire sulla base del suo livello di rischio sistematico, misurato dal β . Analiticamente l'indice alpha di Jensen è ottenibile mediante le regressioni lineari sui fondi:

$$\begin{aligned}\alpha_p &= \bar{r}_p - \bar{r}_{CAPM} \\ \bar{r}_{CAPM} &= \bar{r}_f + \beta_p (\bar{r}_b - \bar{r}_f)\end{aligned}\quad (2.5.2)$$

Dove, \bar{r}_{CAPM} rappresenta il rendimento che il fondo avrebbe dovuto offrire in base al proprio livello di rischio di mercato, nell'ambito del CAPM. Ricordiamo che secondo il CAPM, $u(i)$ è indipendente da $z(m)$ e la media degli errori è nulla. Quindi, un gestore qualificato dovrebbe scegliere fondi (o titoli) da inserire in portafoglio in modo tale che:

$$z_p = \alpha_p + \beta_p z_m + u_i \quad (2.5.3)$$

ovviamente, α_p deve essere significativamente diverso da 0. *I prodotti gestiti* che presentano valori *significativamente positivi* nel tempo dell'alfa di Jensen (che può essere espresso in termini percentuali, risultando quindi di facile comprensione) significa che sono riusciti "a battere il mercato", cioè hanno prodotto un rendimento superiore a quello atteso in base al rischio sistematico assunto. Segue in questo modo un certo **extrarendimento** determinato dalla capacità del money manager di posizionarsi con maggior peso sui titoli sottovalutati che compongono il benchmark da un lato, e dall'altro di ridurre l'esposizione verso le attività sopravvalutate. Ma si ha un discorso opposto per i fondi di investimento caratterizzati da valori significativamente negativi nel tempo dell'alfa, i cui gestori non si sono dimostrati particolarmente abili nelle scelte relative alla composizione del portafoglio. Si può valutare la significatività di tale indice in base alla statistica test T confrontata con la t-Student a (I-2) gradi di libertà. Il test va a verificare che l'intercetta della regressione sia significativamente diversa da 0, cioè:

$$T_\alpha = \frac{\hat{\alpha}}{\sqrt{\frac{\sum_{i=1}^I u_i^2 \sum_{i=1}^I x_i^2}{(I-2) \sum_{i=1}^I (x_i - \bar{x})^2}}} \cong t_Student_{(I-2)} \approx N(0,1) \quad \text{per } I > 30 \quad (2.5.4)$$

Lo stesso test, lo possiamo applicare anche per la verifica della significatività dell'indice β , cioè possiamo verificare se significativamente β è diverso da zero e poi se significativamente diverso da uno.

2.5.1.2 MISURA BETA DEL PORTAFOGLIO

La moderna teoria del portafoglio sostiene che il rischio maggiormente rilevante da considerare e per il quale l'investitore viene compensato non è quello complessivo, misurato tramite la *standard deviation*, ma quello sistematico o di mercato, rappresentato dal coefficiente denominato β (*evidenzia la sensibilità del fondo rispetto ai movimenti di mercato*).

Il rischio sistematico, β , costituisce la parte di rischio non diversificabile di un portafoglio e rappresenta la componente di rischiosità che un potenziale sottoscrittore di un fondo comune dovrebbe maggiormente considerare. Uno dei fattori che rendono preferibile un fondo rispetto ad altre attività finanziarie è la riduzione del rischio ottenibile attraverso la diversificazione e il β viene utilizzato per misurare la sensibilità dei rendimenti di un asset (titolo o fondo) in relazione al mercato (indicato con m oppure da un benchmark, b) e viene espresso dalla seguente formula oppure ricavato dalla regressione lineare (*vedi Formula 2.5.3*):

$$\beta_i = \frac{\mu_i - r_f}{\mu_b - r_f} \quad (2.5.5)$$

In funzione degli extrarendimenti possiamo calcolare il β , come in funzione anche del *portafoglio di benchmark*, b :

$$\beta_i = \frac{Cov(z_i, z_b)}{Var(z_b)} \quad (2.5.6)$$

β è un indice non diversificabile, perché legato alle fluttuazioni del mercato, ma ci consente di poter classificare i fondi in base al suo valore soglia.

- I) $\beta \sim 1$: tipico dei titoli neutrali;
- II) $\beta < 1$: definisce i titoli difensivi;
- III) $\beta > 1$: caratteristico dei titoli aggressivi.

I titoli che presentano un valore del beta superiore a 1 tenderanno a salire, in condizioni favorevoli, e a scendere, in fasi avverse, più rapidamente rispetto alle

variazioni dell'indice di mercato; un comportamento opposto caratterizzerà i titoli con beta inferiore a 1. Un portafoglio di titoli con un beta elevato è dunque più rischioso di un portafoglio che contiene titoli con beta basso.

2.5.2 MISURE DI RAP - RISK ADJUSTED PERFORMANCE

Il rischio degli investimenti, come visto anche sopra, è distinto in *rischio globale* e in *rischio sistematico* non diversificabile, il noto β . Il rischio globale viene misurato con la deviazione standard dei rendimenti, $Var(r_i)$, invece il rischio non diversificabile dal coefficiente "beta" della regressione dei rendimenti del portafoglio sui rendimenti di un benchmark o del mercato. Analiticamente si ha:

$$\begin{aligned} r_i &= r_f + \beta_{i,b}(r_b - r_f) + e_i \\ Var(r_i) &= \beta_{i,b}^2 Var(r_b) + \sigma_i^2(e_i) \end{aligned} \quad (2.5.7)$$

La *misura di performance*, in generale, è definita come rapporto tra una misura di rendimento e una misura di rischio ed è quindi impossibile dal punto di vista logico scegliere una *misura di performance come standard unico*, poiché ciascuna misura è adeguata a uno scopo particolare e orientata da un particolare punto di vista. Un aiuto rilevante arriva da una serie di indici denominati *misure di risk-adjusted performance (RAP)*, indici questi che consentono l'analisi dei risultati realizzati dai fondi (o titoli) considerando nello stesso tempo sia il rendimento ottenuto e sia il livello di rischio assunto per raggiungere tale rendimento.

La particolarità di queste misure è quella di dare una misura di rendimento rettificata per il livello di rischio che è stato assunto per ottenere tale risultato. La rettifica avviene rapportando il rendimento del fondo alla sua rischiosità. Mediante queste misure il confronto tra i vari prodotti risulta semplificato, in quanto tutta l'informazione necessaria a valutare (in base al trade-off rendimento-rischio) il fondo di investimento è basato sul valore assunto dall'indice. Il fondo con la misura RAP più elevata risulta il migliore, spiegabile questo dal fatto che si è collocato sulla migliore posizione nell'ambito della relazione esistente tra rendimento e rischio.

In particolare, l'indice Rap elaborato dall'economista Modigliani, va a valutare lo stesso fenomeno considerato dall'indice di Sharpe rappresentandolo però, in modo diverso: nell'indice di Sharpe l'inclinazione della retta è identificata dalla relazione rendimento/rischio del fondo, mentre l'indice Rap rappresenta la misura del rendimento che ogni fondo dovrebbe garantire per un livello di rischio analogo a quello del mercato, cioè è una trasformazione lineare dell'indice di Sharpe (S). Analiticamente si ha:

$$\begin{aligned}
 M_p^2 &= \frac{\sigma_b}{\sigma_p} \bar{r}_p + \left(1 - \frac{\sigma_b}{\sigma_p}\right) \bar{r}_f \\
 &= \frac{\sigma_b}{\sigma_p} (\bar{r}_p - \bar{r}_f) + \bar{r}_f = \sigma_b S + \bar{r}_f
 \end{aligned}
 \tag{2.5.8}$$

Dalla *Formula 2.5.8*, notiamo che se il livello di rischio del fondo è maggiore rispetto a quello del mercato, allora occorrerà vendere una parte delle risorse investite e dirottarle verso l'attività priva di rischio e viceversa, se il rischio del fondo è inferiore a quello del mercato.

Oltre all'indice di Modigliani ci sono altri indici che fanno parte nelle misure RAP e questi saranno oggetto di studio nei paragrafi successivi. In generale, il principio guida delle RAPM è basato sull'assunzione che, la valutazione di fondi non può avvenire su basi assolute, ma deve essere effettuata su basi relative, in quanto si considerano in modo congiunto tutti quei elementi che hanno influenzato la performance ottenuta.

2.5.2.1 INDICE DI SHARPE

Il famoso *indice di Sharpe* fu proposto originariamente nel 1966 ed è attualmente una delle misure RAP più importanti, anche se nel tempo è stato oggetto di studi e di alcune modifiche, in questa tesi si farà riferimento alla versione originale.

Attraverso l'indice di Sharpe, è possibile confrontare e classificare fondi omogenei per benchmark. Infatti, essendo la misura basata sulla relazione esistente tra rischio e rendimento, risulta che il fondo con il più alto indice di Sharpe è anche quello che ha prodotto il rendimento più alto in base al proprio livello di

volatilità. E' inoltre possibile determinare l'indice di Sharpe del benchmark, con cui confrontare quello dei fondi che hanno scelto quel parametro di riferimento:

$$S_b = \frac{\bar{r}_b - \bar{r}_f}{\sigma_b} \quad (2.5.9)$$

\bar{r}_b : rendimento medio del benchmark

dove: \bar{r}_f : rendimento medio del risk - free

σ_b : deviazione standard del benchmark

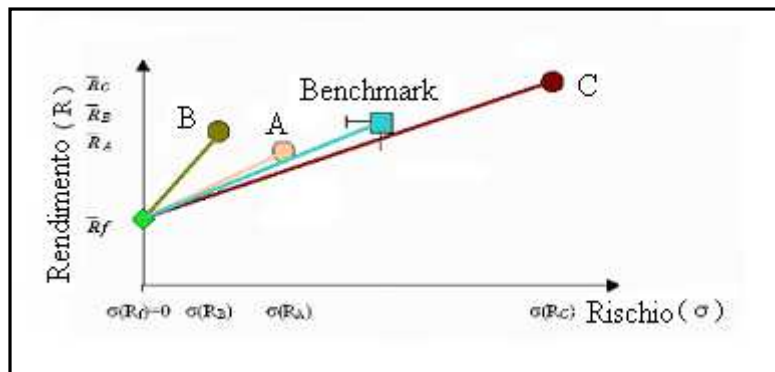


Figura 2.5.1: Interpretazione grafica dell'indice di Sharpe

(Fonte: La valutazione finanziaria di un portafoglio gestito)

Graficamente (vedi Figura 2.5.1), se nello spazio rischio-rendimento collochiamo i fondi e l'attività priva di rischio, ed in seguito uniamo tramite una semiretta ogni punto indicante il fondo con l'attività risk-free, il fondo migliore risulterà quello sulla linea con maggiore pendenza, o meglio più alto valore del indice di Sharpe. Ad esempio, si ha che il fondo contrassegnato dalla lettera B, pur avendo fatto registrare risultati inferiori in termini di rendimento assoluto essendo $r_B < r_C$, risulta essere superiore nell'ottica bidimensionale rischio/rendimento, la cui misura è fornita dalla pendenza della semiretta \bar{R}_f di B. Proprio per le informazioni preziose che l'indice (vedi Formula 2.5.4) ci fornisce e per la sua semplicità nel calcolarlo, lo si valuterà anche per i 50 fondi comuni di investimento statunitensi, argomento questo del capitolo 4.

2.5.2.2 INDICE DI TREYNOR

L'indice di Treynor, detto anche "reward to volatility ratio", o meglio premio per unità di rischio. Graficamente è definito (vedi Figura 2.5) dalla pendenza o inclinazione della security market line passante per P. Analiticamente si ha:

$$tr_p = \frac{\bar{r}_p - \bar{r}_f}{\beta_p} = \frac{\mu_p}{\beta_p} \quad (2.5.10)$$

Dove: μ_p : extra rendimento atteso del portafoglio(o del fondo)
 β_p : rappresenta il β del portafoglio(o del fondo)

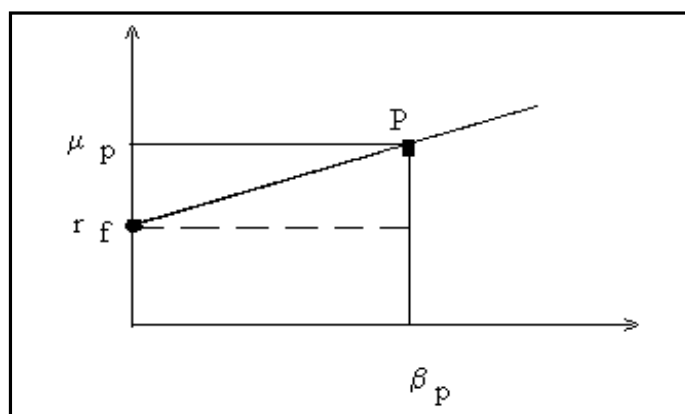


Figura 2.5.2: Interpretazione grafica dell'indice di Treynor¹¹

Mediante l'indice di Treynor si misura la performance dei portafogli attivi con, rendimenti in eccesso per unità di rischio non diversificabile dove, il β di riferimento è stimato (per i fondi) mediante regressioni lineari (pendenza della SML, metodo MQO). Tale indice è un'altra misura RAP e se questo valore risulta grande allora significa che l'inclinazione della SML è grande, o meglio il rendimento in eccesso oppure gli extrarendimenti attesi per unità di rischio sono elevati. Se gli investitori diversificano il rischio, l'indice di Treynor è definito come l'indicatore RAP più corretto, in quanto il mercato paga, in tale caso, solo i rischi non diversificabili. Invece, in un mercato in equilibrio tutti i fondi devono avere un valore dell'indice identico (o uguale) tra di loro. I confronti che possiamo fare mediante l'indice di Treynor sono prevalentemente due: confronti tra fondi diversi oppure tra fondi e un benchmark.

¹¹ creato in Immagine-Paint;

L'indice di Treynor si differenzia dall'indice di Sharpe in quanto, viene calcolata la rischiosità del portafoglio considerando l'andamento di mercato. Al denominatore del rapporto troviamo β , assumendo implicitamente che l'investitore detenga quote del portafoglio di mercato, analiticamente:

$$z_p = \beta_p z_m + u_p \quad \text{con} \quad \beta_p = \sum w_i \beta_i \quad \text{e} \quad u_p = \sum w_i u_i \quad (2.5.11)$$

Dove, il fondo p è formato da i_titoli e la varianza totale è definita come:

$$\sigma_p^2 = \beta_p^2 \sigma_i^2 + \sigma_p^2(u_p) \quad (2.5.12)$$

Se si ha un fondo ben diversificato allora, $\sigma_p^2(u_p) \rightarrow 0$ e $\sigma_p = \beta_p \sigma_b$ quindi:

$$S_p = \frac{\bar{z}_p}{\sigma_p} = \frac{\bar{z}_p}{\beta_p \sigma_b} = \frac{tr_p}{\sigma_b} \quad (2.5.13)$$

In generale, questo però non vale, cioè il CAPM non è sempre valido in quanto i gestori dei fondi non diversificano completamente. Oltre a questa relazione, tra indice di Treynor e quello di Sharpe, ne esiste un'altra tra tr_p e α_p anche questa è piuttosto stretta ed importante. Molto spesso $tr_b = \mu_b$ dato che per costruzione $\beta_b = 1$, quindi:

$$tr_p = \frac{\alpha_p}{\beta_p} + tr_b \quad (2.5.14)$$

Quindi, questo indice misura l'opportunità di usare, come leva finanziaria, l'indebitamento al tasso risk-free per sovra investire nel portafoglio che ha il maggior premio per il rischio a parità di rischiosità (definita dal β del portafoglio).

2.5.2.3 INFORMATION RATIO

L'*Information Ratio* è un indice derivato dalla generalizzazione dell'indice di Sharpe, dove il tasso risk-free è sostituito dal benchmark. Analiticamente si ha:

$$S_p = \frac{\bar{z}_p}{\sigma_p} = \frac{E(r_p - r_f)}{\sigma(r_p - r_f)} \quad \text{sostituisco il tasso risk-free con} \quad (2.5.15)$$

$$\text{il benchmark e otteniamo} \quad IR_p = \frac{E(r_p - r_b)}{\sigma(r_p - r_b)} = \frac{E(TE)}{\sigma_{TE}} \quad (2.5.16)$$

Dalla *Formula 2.5.15* notiamo che, la differenza tra il valore atteso del fondo e rendimento del benchmark è detta *Tracking Error*. Il benchmark ha per definizione IR nullo, inoltre nel caso in cui TE medio è negativo diventa molto difficile interpretare IR.

Molto spesso è di interesse trovare le quote di un portafoglio rischioso e di solito questo studio è effettuato sulla base di *un benchmark e una funzione obiettivo* che massimizza lo scostamento dal benchmark fissando una soglia massima di rischio. Nel condurre questa analisi si valuta l'indice di *tracking error*, il quale assume diverse definizioni a seconda degli obiettivi fissati degli analisti. Analiticamente si hanno due metodi di calcolo del TE:

Il primo: serve a confrontare un insieme di fondi caratterizzati da obiettivi d'investimento differenti, quindi da benchmark e livelli di rischio diversi, è possibile calcolare la differenza tra il rendimento del portafoglio e quella del benchmark:

$$TE^1 = r_p - r_b \quad (2.5.17)$$

Il secondo: è definito come la radice del rischio massimo che l'investitore può assumere ed è ottenuta in base all'allocazione ottima rispetto al benchmark, ricavata mediante:

$$\begin{aligned} \max_w w' \mu - w'_b \mu &= OP \\ \text{s.v. } w' \mathbf{1} &= 1 \\ (w - w_b)' \Sigma (w - w_b) &= \sigma_D^2 = TE^2 \end{aligned} \quad (2.5.18)$$

Dove:

- **OP:** è la “*over-performance*” (*extrarendimento atteso*) rispetto al benchmark, con livello massimo di rischio ammissibile, definito sugli scostamenti tra i pesi del portafoglio che si vuole costruire ed i pesi degli asset rischiosi nel portafoglio benchmark;
- **TE:** rappresenta il rischio massimo che può essere sopportato, il tracking error. I pesi che vengono determinati da tale funzione obiettivo sono anche definiti come pesi ottimi e definiscono una nuova frontiera efficiente. In presenza del risk-free si può anche determinare il nuovo portafoglio di tangenza rispetto alla frontiera efficiente basata sul TE.

Nel caso in cui $OP = 0$, allora l'obiettivo è quello di replicare il benchmark, cioè si vuole fare gestione passiva. Nell'analisi dei 50 fondi comuni d'investimento statunitensi si utilizzerà l'indice IR calcolato come nella *formula 2.5.15*.

Ad esempio: nel caso in cui vogliamo identificare *un portafoglio efficiente*¹², cioè quando voglio avere la minima varianza per gli extrarendimenti (per ogni asset) rispetto a un dato portafoglio target (quello inserito nel programma per il calcolo della FE ha un benchmark dato dallo *S&P 1500*¹³), con lo scopo di ottenere gli extrarendimenti. La media e la deviazione standard degli extrarendimenti sono spesso chiamati *rendimenti attivi (Active Return)* e *rischio attivo (Active Risk)*. Il rischio attivo è a volte definito come il *Tracking Error* e quando l'obiettivo è quello di determinare un portafoglio target ristretto all'analisi allora il portafoglio a cui si fa riferimento è proprio quello determinato della frontiera efficiente (FE). In specifico, assumo per l'analisi che il portafoglio target sia espresso da un *vettore di pesi indicizzato (index)*, quindi i rendimenti di questa serie sono espressi come una combinazione lineare degli 50 assets disponibili.

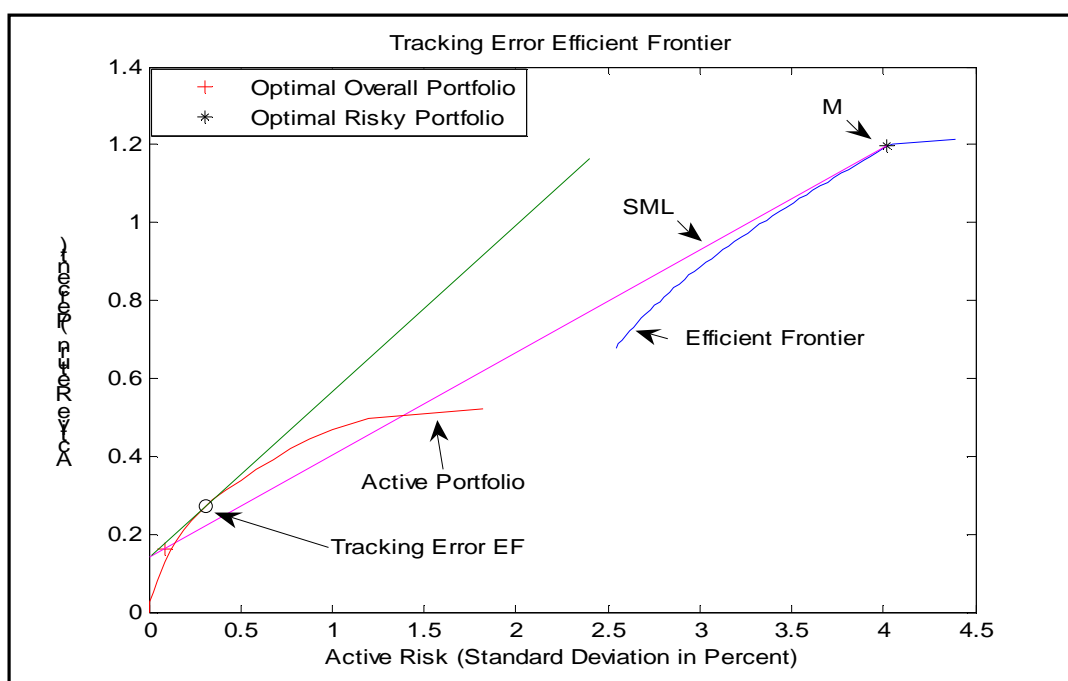


Figura 2.5.3: Costruzione della frontiera che minimizza il Tracking Error

¹² esempio costruito in base al portafoglio in analisi composto dai 50 fondi comuni d'investimento statunitensi;

¹³ osservato dal gennaio 2002 al dicembre 2006;

I risultati ottenuti dalla costruzione della FE che *minimizza l'Active Risk*, quindi il tracking error, sono rappresentati graficamente mediante la *Figura 2.5.3*, il tutto avviene per un certo valore degli extrarendimenti. Segue inoltre che, lo “zero-risk/zero-return” portafoglio ha un pratico significato economico, cioè rappresenta tutti i fondi che pienamente sono riferiti al portafoglio INDEX stesso. Inoltre, ogni tracking error del portafoglio efficiente (cioè, ogni riga di *ActiveWeights*) soddisfano il *budget attivo* e rappresenta l’allocazione del portafoglio dei fondi rispetto a quello INDEX. Di conseguenza l’obiettivo di identificare opportunità di investimento va a buon fine, in quanto la gestione attiva ci porta a un rapporto tra rendimento/rischio che è più efficiente rispetto al “*benchmark*”, o meglio dicasi, si hanno opportunità di ottenere extrarendimenti. Dalla frontiera del Tracking Error, la parte iniziale e la parte finale, si ha la conferma sui valori di $\alpha(i)$ calcolati mediante le regressioni lineari (vedi *Tabella 4.2.1*). Infatti, i primi 10-11 assets non hanno $\alpha(i)$ significativo e tanto meno gli ultimi 10, queste sono anche le parti che si trovano al di sotto della SML messa a confronto con il Tracking Error della FE (vedi *Figura 2.5.3*). Tutta la parte centrale degli assets, riferiti agli *Small Value - S.Blend - Large Growth - L.Value*, ci portano ad avere degli extrarendimenti ($\hat{\alpha}_i$), leggermente significativi, quindi sono le due categorie (Small e Large) sopra citate, che di nuovo si distinguono per il loro comportamento efficiente.

2.5.2.4 INDICE DI SORTINO

L'indice di Sortino è una delle misure RAP più recenti, infatti risale al 1991 ed è definita come una misura di rendimento corretto per il rischio, basato però sul concetto di *rendimento minimo accettabile (minimum acceptable return, MAP)* per un determinato investitore. Così come, l’indice di Sharpe e anche quello di Sortino seguono la stessa logica, si differenziano soltanto per il fatto che quest’ultimo fa riferimento a un rischio asimmetrico, cioè al “*downside risk*”, che valuta solo le deviazioni negative rispetto a un rendimento obiettivo minimo che può essere

benissimo rappresentato dal tasso privo di rischio (r_f). Analiticamente l'indice è definito come:

$$SO = \frac{r_p - r_f}{DD} \quad (2.5.19)$$

Dove:

r_p : rendimento mensile del portafoglio (o fondo);

r_f : rendimento obiettivo minimo dell'investitore, oppure il risk - free;

DD : downside deviation, deviazione standard degli scostamenti negativi

$$DD^2 = \sigma^2(D) = \text{Var} [r_{P,t} I(r_{P,t} < 0)] \quad (2.5.20)$$

Quindi, è definito come differenza tra il rendimento del portafoglio (o del fondo) e il rendimento minimo accettabile dal risparmiatore, molto spesso coincidente con il tasso risk-free, il tutto diviso dalla deviazione standard degli scostamenti negativi dei rendimenti del portafoglio, nonché misura gli extrarendimenti per unità di rischio di perdita (*downside deviation*).

La somiglianza dell'indice di Sortino con quello di Sharpe permette di poter approfondire l'analisi sui fondi, selezionando come i più opportuni e i migliori. O meglio, tra tutti quelli che hanno indice di Sharpe uguale (a parità di tale indice) soltanto quelli che presentano una minore volatilità verso il basso (*downside deviation*), sono i più appetibili, cioè tra diversi portafogli alternativi si sceglierebbe di certo il portafoglio che garantisce indice SO più elevato.

In conclusione, si può dire che gli investitori non devono temere tutto il rischio a cui il portafoglio è soggetto, ma in particolare devono cercare di evitare proprio il "downside risk". Inoltre, i vari indici non sono necessariamente concordi in quanto, l'aggiustamento per il rischio è differente, ed è ovvio perché, si considerano diverse fonti di rischio per ogni indice.

Nell'analisi strumentale vedremo in particolare come questi indici sono definiti per i fondi comuni d'investimento statunitensi e come ci aiutano a riconoscere i fondi migliori sia ad ex-ante e sia a ex-post, cioè dopo avere effettuato la cluster analysis, basata sia sulle serie storiche e anche sugli indici di performance. Tale argomento sarà oggetto di studio e di approfondimento nel capitolo quattro.

CAPITOLO 3

CLUSTER ANALYSIS

I metodi di classificazione, noti come metodi di *cluster analysis* o *clustering*, hanno lo scopo di classificare le unità statistiche attraverso l'uso di procedure che, di solito, sono applicabili quando su ogni unità statistica sono state rilevate le modalità dei parametri oppure si utilizzano direttamente le informazioni derivate dalle serie dei rendimenti. Fu **K. Pearson** che affrontò per primo lo studio della classificazione dal punto di vista statistico verso la fine del XIX secolo. Da allora ai giorni nostri gli *algoritmi di clustering* si sono moltiplicati e differenziati nei diversi ambiti applicativi.

In particolare dalla seconda metà degli anni '50 alcune delle tecniche di raggruppamento hanno ricevuto una più ampia *trattazione teorico-metodologica* grazie alla corrispondenza con la *teoria dei grafi*¹⁴. Le tecniche di analisi dei gruppi, si sono largamente usate nei più svariati campi di ricerca come: fisica, scienze sociali, economia, medicina e di recente anche in finanza, dove la classificazione dei dati disponibili è un momento essenziale nella ricerca dei modelli interpretativi della realtà.

Successivamente, di pari passo agli sviluppi delle *tecnologie di calcolo*, si è posta maggiore attenzione agli aspetti algoritmici delle tecniche di raggruppamento, argomento questo che sarà approfondito nei paragrafi successivi. *L'evoluzione degli strumenti da calcolo automatico* ha consentito di affrontare senza difficoltà la complessità computazionale che è insita in molti dei metodi di classificazione e che in precedenza aveva spinto i ricercatori ad orientarsi verso tecniche di analisi dei gruppi che erano più facilmente applicabili. Si è resa così

¹⁴ Mignani e Montanari, 1994;

possibile la *produzione di diversi algoritmi di classificazione*, sempre più complessi dal punto di vista computazionale, ma anche sempre più efficienti nel trarre informazioni dai dati attraverso una loro opportuna classificazione.

Prima di valutare le tecniche sopra citate cominciamo a introdurre i concetti di prossimità e distanza fra le unità statistiche, in quanto sono la base di quasi tutte le metodologie.

3.1 PROSSIMITA' E DISTANZA

Lo scopo principale dei vari metodi che analizzeremo sono utili a *classificare e raggruppare le unità statistiche in gruppi omogenei*. In concreto sarà di interesse valutare la relazione fra le righe della matrice dei dati, corrispondenti alle unità statistiche. Misurare la 'vicinanza' tra le unità statistiche, o equivalentemente la loro 'lontananza' non è tanto facile, anzi il modo in cui questo viene effettuato è fondamentale. Le possibili scelte per come quantificare, poter confrontarle, questa 'vicinanza' sono tantissime, e variano sia con la natura delle variabili in gioco, sia con gli obiettivi da raggiungere. Risulta di conseguenza necessario introdurre la nozione di **prossimità** per la 'vicinanza' e si usa il termine generico **dissimilarità** per riferirci alle misure di 'lontananza'.

Gli indici di prossimità tra coppie di unità statistiche forniscono le informazioni preliminari indispensabili per poter individuare gruppi di unità omogenee. Un indice di prossimità tra due generiche unità statistiche $x(i)$ e $x(j)$ è definito come una funzione dei rispettivi vettori riga nella matrice dei dati:

$$IP_{ij} = f(x'_i, x'_j) \quad (3.1.1)$$

Altrimenti, considerando che per l' unità statistica i -esima si hanno a disposizione p variabili, x_{i1}, \dots, x_{ip} di cui parte possono essere quantitative e altre qualitative. L'obiettivo è quello di allocare ad un stesso gruppo quelle unità che sono tra loro più simili. Per quanto riguarda invece, la *dissimilarità $d(i,j)$* tra le unità statistiche (righe diverse) i e j , questa si basa sulla composizione delle

dissimilarità valutata per ciascuna delle p variabili osservate, diciamo $d_s(x_{is}, x_{js})$ per $s = 1, \dots, p$ oppure valutata per fondi diversi.

Molto spesso ci si trova ad affrontare problemi abbastanza complessi, cioè si ha un dataset di grandi dimensioni per questo motivo è molto spesso conveniente sintetizzare prima l'informazione contenuta nei dati e poi analizzarlo con i metodi più appropriati. Sia per il dataset di questa tesi anche per altri lavori conviene utilizzare gli indici di prossimità, in quanto aggregano i rendimenti più simili mediante la distanza più opportuna, ovviamente parliamo di dati quantitativi.

In generale, in riferimento a *come si presentano le variabili* si utilizzano anche vari criteri specifici per la loro analisi, in particolare:

- i) per caratteri (o variabili) di tipo quantitativo gli indici di prossimità vengono definiti come **distanze**¹⁵;
- ii) se invece, le variabili sono di tipo qualitativo, allora si usano gli **indici di similarità**;
- iii) per i dati contenuti in una tabella di contingenza, può essere impiegata anche la **distanza chi-quadrato**;
- iv) infine, per variabili miste, o meglio alcune quantitative e altre qualitative, si usano appropriati **indici di prossimità**.

3.1.1 DISTANZA EUCLIDEA

La misura più comunemente usata è la distanza Euclidea e in base ad essa consideriamo una matrice di dati contenete solo **variabili quantitativi** (o *dicotomiche*). Supponiamo che x e y siano le righe della matrice dei dati allora si dice che $d(x,y)$ è una **distanza tra due osservazioni se soddisfa le seguenti proprietà**, $\forall x, x', y, z \in \mathfrak{R}^p$:

– *Identità*: $d_j(x, x) = 0$ (3.1.2)

– *Non negatività*: $d_j(x, x') \geq 0$ (3.1.3)

¹⁵ di largo uso è quella Euclidea;

$$\text{— Simmetria:} \quad d_j(x, x') = d_j(x', x) \quad (3.1.4)$$

$$\text{— Disuguaglianza triangolare:} \quad d_j(x, y) + d_j(y, z) \geq d_j(x, z) \quad (3.1.5)$$

Per poter raggruppare le unità statistiche, solitamente si considera la distanza tra tutte le unità presenti nella matrice dei dati. L'insieme di tali distanze viene rappresentata in una matrice di distanze. Una generica matrice delle distanze è strutturata nel seguente modo:

$$\Delta = \begin{bmatrix} 0 & \dots & d_{1i} & \dots & d_{1n} \\ \dots & 0 & \dots & \dots & \dots \\ d_{i1} & \dots & 0 & \dots & \dots \\ \dots & \dots & \dots & 0 & \dots \\ d_{j1} & \dots & d_{ni} & \dots & 0 \end{bmatrix} \quad (3.1.6)$$

dove il $d(ij)$ misura la distanza tra i vettori riga $x(i)$ e $x(j)$. La *distanza Euclidea* è la misura di prossimità più utilizzata ed è definita per ogni coppia di indici i e j come, la radice quadrata della differenza tra i rispettivi vettori, nello spazio euclideo p -dimensionale:

$$d_{ij} = d(x_i, x_j) = \left[\sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2} \quad (3.1.7)$$

Questa distanza si distingue dalle altre per il fatto di essere fortemente influenzata da una singola *differenza elevata tra i valori in una dimensione*, dovuta al fatto che il quadrato amplifica molto tale differenza. Le dimensioni caratterizzate da una scala di misura differente sono spesso la fonte di differenze eccessive. Per superare tale limitazione, la distanza Euclidea viene calcolata, non sulle variabili originarie, bensì su opportune trasformazioni di queste.

La scelta più comune consiste nella *standardizzazione delle variabili*, e a seguito di questo, ogni variabile statistica contribuisce al calcolo della distanza con uguale peso e invariante rispetto a trasformazioni lineari. Se le *variabili sono standardizzate*¹⁶ risulta:

$$\begin{aligned} d_{ij}^2 &= 2(1 - r_{ij}) \quad \text{con} \quad i, j = 1, \dots, p \\ r_{ij} &= 1 - d_{ij}^2 / 2 \end{aligned} \quad (3.1.8)$$

Dove: r_{ij} indica il *coefficiente di correlazione* fra le osservazioni $x(i)$ e $x(j)$, che rappresentano due fondi diversi, di conseguenza si ottiene la distanza euclidea tra

¹⁶ in senso statistico: media nulla e varianza unitaria;

due osservazioni e una funzione del coefficiente di correlazione tra le stesse. La distanza più semplice e la più accurata quando i dati sono multi-dimensionali è proprio la distanza Euclidea. Se siamo in \mathfrak{R}^2 oppure in \mathfrak{R}^3 questa misura non è altro che la distanza geometrica tra le unità (o le variabili).

Molto spesso quando si valutano degli algoritmi (*ad esempio*: in MatLab) bisogna fare attenzione al tipo di distanza che si propone per le unità in analisi, in quanto si potrebbe valutare una distanza non corretta e l'algoritmo non lo tiene presente, portandoci a delle distanze lontane da quelle reali. Altri algoritmi propongono in automatico la misura di distanza più opportuna a seconda del tipo di analisi che si vuole svolgere in base all'analisi selezionata. *Ad esempio*: per ricerca tecnico-scientifica, si usa la distanza di Minkowsky.

In base alla tipologia delle variabili in analisi, si hanno altre *misure di distanza*, non meno importanti, le quali sono:

a) *Per variabili quantitative, di solito trovano uso:*

- *Distanza di Mahalanobis*: è una misura basata sulla matrice di covarianza (Σ) semplice delle unità statistiche in analisi ed è definita come:

$$d_{ij} = \left\{ (x_i - x_j)^T \Sigma^{-1} (x_i - x_j) \right\}^{1/2} \quad (3.1.9)$$

- *Distanza di Minkowsky*: è una misura di distanza calcolata come:

$${}_{\lambda} d_{ij} = \left[\sum_{s=1}^p w_s (x_{is} - x_{js})^{\lambda} \right]^{1/\lambda} \quad \text{con } \lambda \geq 1 \quad (3.1.10)$$

Alcuni casi particolari sono conseguenza del valore assunto dal parametro λ :

- se $\lambda = 1$, ${}_{\lambda} d_{ij}$ definisce la distanza media assoluta, cioè la misura di city-block;
- se $\lambda = 2$, ${}_{\lambda} d_{ij}$ caso particolare in cui definisce la distanza Euclidea;
- se $\lambda > 3$, raramente assunto (*ad esempio*: in lavori tecnico-scientifici¹⁷).
- *Distanza di Manhattan (city-block)*: tale distanza è definita come semplice differenza media tra i dati. In molti casi questa misura si equivale a quella Euclidea semplice, però bisogna fare attenzione agli outliers perché basta che una singola differenza sia elevata e tutto salta. La distanza city-block è calcolata come:

¹⁷ analisi non empirica;

$${}_M d_{ij} = \sum_{s=1}^p w_s |x_{is} - x_{js}| \quad (3.1.11)$$

- *Metrica di Canberra*: questa è una misura basata sulla differenza in valore assoluto delle due osservazioni, normalizzata per la somma in valore assoluto delle singole osservazioni ed è calcolata come:

$${}_c d_{ij} = \sum_{s=1}^p \frac{|x_{is} - x_{js}|}{|x_{is}| + |x_{js}|} \quad (3.1.12)$$

- *Distanza di Chebychev*: tale misura è appropriata nei casi in cui si definiscono due unità come ‘differenti‘ tra loro. Se la distanza massima per riga delle unità in analisi è diversa ed è definita come:

$$d(x, y) = \text{Max } |x_i - y_i| \quad (3.1.13)$$

b) *Per variabili qualitative e tabelle di contingenza, si usano le seguenti distanze:*

- *Distanza “Percent disagreement”*: questa misura è particolarmente utilizzata se il dataset comprende variabili di natura qualitative ed è calcolata come:

$$d(x, y) = (\text{Number of } x_i \neq y_i) / i \quad (3.1.14)$$

- *Distanza del chi-quadrato*: la seguente distanza è molto utilizzata per valutare il grado di diversità fra le righe (o colonne) di una tabella di contingenza ed esprime una reciproca distanza tra gruppi di unità statistiche che hanno in comune la stessa modalità di riga (o colonna). Bisogna notare che tale misura non è di prossimità fra le unità in analisi, ma per due modalità h e k di una variabile di una tabella a doppia entrata, significa che è possibile calcolare la distribuzione di frequenza subordinata tra la variabile h e k . In questo modo ho le due distribuzioni subordinate, o meglio i profili riga di h e k (analisi delle corrispondenze). La distanza del chi-quadrato tra le due modalità note, corrisponde alla distanza Euclidea ponderata tra le relative distribuzioni subordinate, dove $h=1, \dots, H$ e $k=1, \dots, K$ con $h \neq k$, calcolata come indicato nella formula (3.1.15). Il tutto è moltiplicato per (n/n_j) , che misura l’importanza della modalità j -esima della variabile in colonna rispetto al complesso delle modalità.

$$d_{h,k} = \left[\left\{ \sum_{j=1}^K [(n_{hj} / n_h) - (n_{kj} / n_k)]^2 \right\} * [n / n_j] \right]^{1/2} \quad (3.1.15)$$

Dopo aver visto la pluralità delle distanze esistenti, ci si accorge che è importante non sbagliare nel definire in modo corretto il tipo di distanza tra le unità statistiche in analisi. Le scelte iniziali molto spesso possono portarci a dei risultati errati e per questo motivo che la maggior parte delle distanze sopra citate non vengono utilizzate, si fa quasi sempre riferimento alle distanze tradizionali come quella Euclidea. A volte le variabili qualitative vengono codificate in modo numerico proprio per il motivo sopra citato. Ovvio che, le distanze sono tante proprio per fare fronte alla vasta gamma dei studi scientifici e non, tipiche del XXI _esimo secolo.

3.1.2 INDICE DI SIMILARITA'

Valutiamo un insieme finito di osservazioni $u(i) \in U$, si dice **indice di similarità** un'applicazione $S(u_i, u_j) = S_{ij} \in \mathbf{R}$ che soddisfa le seguenti proprietà:

$$\text{— Non negatività:} \quad S_{ij} \geq 0 \quad \forall u_i, u_j \in U \quad (3.1.16)$$

$$\text{— Normalizzazione:} \quad S_{ii} = 1 \quad \forall u_i \in U \quad (3.1.17)$$

$$\text{— Simmetria:} \quad S_{ij} = S_{ji} \quad \forall u_i, u_j \in U \quad (3.1.18)$$

Gli indici di similarità a differenza delle distanze, possono essere applicati a **variabili qualitative** qualunque, in quanto non sono definiti dai vettori riga, ma definiti con riferimento agli elementi di un insieme. Tali indici assumono valori nell'intervallo $[0,1]$, anziché un qualunque valore non negativo. Il complemento a un indice di similarità è detto indice di dissimilarità e rappresenta una classe di indici di prossimità più ampia delle distanze. Così come un una distanza, un indice di dissimilarità soddisfa le proprietà (3.1.3) e (3.1.4), tuttavia la proprietà di normalizzazione corrisponde solo a una delle due implicazioni della proprietà di identità delle distanze e inoltre devono soddisfare la (3.1.5). In linea teorica questo indice può essere calcolato anche per le *variabili quantitative*, anche se sarebbe di *scarsa utilità*, in quanto distinguerebbe qualcosa soltanto se le due unità statistiche

avessero valori osservati uguali o diversi, ma senza riferimento alla misura di tale diversità.

Dal punto di vista operativo *i principali indici di similarità* fanno riferimento a matrici dei dati contenenti **variabili dicotomiche**, *ad esempio* con variabili a più modalità qualitative. Le frequenze a cui di solito si fa riferimento per il calcolo degli indici sono:

- i) **CP**: ‘Co-Presenze’, definita come frequenza assoluta di fenomeni contemporaneamente presenti nelle due unità (*positive matches*);
- ii) **PA**: ‘Presenza-Assenza’;
- iii) **AP**: ‘Assenza-Presenza’;
- iv) **CA**: ‘Co-Assenza’, fenomeno trascurabile per l’assenza di importanza nella determinazione della rassomiglianza tra due unità.

Gli indici di similarità sviluppati differiscono con riferimento alla considerazione della CA, in particolare si ha:

I) *Indice di similarità di Russel e Rao*:

$$S_{ij} = \frac{CP}{p} \quad (3.1.19)$$

l’indice è in funzione delle CP ed è dato dal rapporto del numero delle co-presenze e del numero totale di variabili binarie **p**;

II) *Indice di similarità di Jaccard*:

$$S_{ij} = \frac{CP}{CP + PA + AP} \quad (3.1.20)$$

valuta il rapporto del numero delle co-presenze e il numero dei caratteri a esclusione di quelli co-assenze.

III) *Indice di similarità di Sokal e Michener*:

$$S_{ij} = \frac{CP + CA}{p} \quad (3.1.21)$$

in questo caso l’indice è detto anche ‘*simple matching coefficient*’ e il suo complemento a uno corrisponde alla media del quadrato della distanza euclidea fra i due vettori associati alle unità statistiche:

$$1 - S_{ij} = \frac{1}{p} d_{ij}^2 \quad (3.1.22)$$

Tale indice è uno degli indici di similarità più usati, definito anche come *'binary distance'*.

In generale, per le **variabili dicotomiche**, diversamente da quanto visto sopra, spesso è molto più conveniente fare riferimento a *indici di dissomiglianza* anche se il concetto di fondo è molto simile al caso precedente, basato sulla similarità. Supponiamo di avere variabili X_{ih} che possono assumere valori 0 o 1, per $i=1,\dots,n$ e $h=1,\dots,k$. Parlando in termini di dissomiglianza tra le due osservazioni X_i e X_j allora si può rappresentare la seguente tabella:

X_i	X_j	
	1	0
1	a	b
0	c	d

Dove: - a : numero di variabili che valgono 1 per ambedue le osservazioni;

- b : numero di variabili che valgono 1 per la i -esima e 0 per la j -esima osservazione;

- c : numero di variabili che valgono 0 per la i -esima e 1 per la j -esima osservazione;

- d : numero di variabili che valgono 0 per ambedue le osservazioni.

In totale: $a+b+c+d = k$ e da cui gli indici di dissomiglianza calcolabili sono:

- *Il coefficiente di dissomiglianza semplice* è basato sul rapporto delle variabili che risultano discordanti:

$$d_{ij} = \frac{b + c}{k} \quad (3.1.23)$$

- *Il coefficiente di Jaccard* è utilizzato per le variabili dicotomiche asimmetriche, che indicano la presenza di una certa caratteristica. L'assenza della caratteristica da ambedue le osservazioni non dovrebbe contribuire ad aumentare il grado di somiglianza:

$$d_{ij} = \frac{b + c}{a + b + c} \quad (3.1.24)$$

Entrambi i metodi sopra citati, sulle variabili dicotomiche, hanno la loro importanza, in quanto sempre prima di applicare un clustering è necessario valutare le tipologie delle variabili in analisi e in un secondo momento si va a

scegliere lo strumento più consono per la loro analisi. Inoltre, bisogna anche sapere l'obiettivo del clustering, in modo da selezionare quel indice che è maggiormente in linea con lo scopo dell'analisi.

3.1.3 METODO DEL MULTIDIMENSIONAL SCALING

Se fin adesso si sono valutati i casi in cui si mostrava come calcolare le distanze tra le osservazioni in base a una data matrice dei dati o tabella, adesso studiamo il caso in cui si hanno a disposizione solo le distanze tra le osservazioni (in termini di distanze) e si desidera ricostruire i valori delle osservazioni che si trovano dietro di esse. Molto spesso le distanze sono calcolate secondo una misura di dissimilarità e si desidera riprodurle in termini di distanza Euclidea, così da ottenere una rappresentazione delle osservazioni in un piano bidimensionale.

I metodi di multidimensional scaling hanno come obiettivo quello di rappresentare le osservazioni, i cui valori osservati sono ignoti (o non espressi in valori numerici) in uno spazio euclideo di piccola dimensione (\mathbb{R}^2). La rappresentazione viene effettuata preservando il più possibile le distanze originarie. Questo metodo è una rappresentazione euclidea per osservazioni di piccole dimensioni, in modo tale che minimizzi un'appropriata distanza tra le distanze originarie e le nuove distanze euclidee.

I metodi multidimensional scaling differiscono per le modalità in cui l'ultima distanza è stata definita. La scelta più comune è la *funzione stress* definita da:

$$fs = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - d_{ij})^2} \quad (3.1.25)$$

dove: δ_{ij} sono le distanze originarie (o dissimilarità) tra ogni coppia di osservazioni, e d_{ij} sono le corrispondenti distanze tra le coordinate riprodotte.

I metodi multidimensional scaling metrici cercano k vettori n -dimensionali a valori \mathbb{R} , ognuno dei quali rappresenta una coordinata delle n osservazioni, espressa da d_{ij} , minimizzi la funzione stress al quadrato. Il caso più comune è $k=2$, perciò il risultato della procedura può essere convenientemente rappresentato

in un grafico di dispersione e la soluzione ottenibile è definita come “*least square scaling*”. Una variante di questa distanza è il *Sammon mapping* che:

$$\min_d \sqrt{\sum_{i=1}^n \sum_{j=1}^n \frac{(\delta_{ij} - d_{ij})^2}{\delta_{ij}}} \quad (3.1.26)$$

possiamo osservare che: quando le distanze tra gli oggetti sono espresse da una distanza euclidea si può mostrare che la soluzione del problema precedente corrisponde al punteggio delle componenti principali che si otterrebbe se la matrice dei dati fosse disponibile.

E' possibile definire *metodi multidimensional scaling non metrici*, per cui la relazione da preservare tra le distanze originarie e quelle riprodotte non è necessariamente di tipo Euclideo.

3.2 ANALISI DI RAGGRUPPAMENTO

L' *analisi dei gruppi*, o *cluster analysis*, un insieme di tecniche atte a ridurre il numero dei dati, unendo vari dati in un solo gruppo (*cluster*¹⁸) in base a una qualche “*somiglianza*” o “*vicinanza*”. Si cerca cioè, di ridurre il numero delle righe della matrice dei dati, sostituendo a tutte le righe che contengono i dati confluiti in un singolo *cluster*, un dato (eventualmente fittizio) rappresentativo di tutto il *cluster* stesso. Questa procedura permette di formare dei *gruppi omogenei*, secondo un certo criterio, a cui poi attribuire un certo numero di caratteristiche proprie di tutti i componenti del gruppo, e solo di questi (almeno una caratteristica deve differire da gruppo a gruppo).

Le tecniche della *cluster analysis* consentono di individuare dei raggruppamenti statisticamente significativi, sia all'interno di un gruppo di oggetti sia tra varie unità statistiche. La struttura dei gruppi omogenei di unità statistiche può essere interpretata come una riduzione delle dimensioni dello spazio \mathfrak{R}^2 , tuttavia completamente differente da quanto accade, ad esempio: con il metodo delle

¹⁸ Gordon, 1988;

componenti principali. Infatti, nei metodi di raggruppamento le n unità vengono riunite in g sottoinsiemi ($g < n$), mentre nell'approccio delle componenti principali le p variabili statistiche vengono trasformate in k nuove variabili ($k < p$).

Per quanto riguarda la **costruzione dei cluster** si può effettuare in molti modi, sia in funzione della scelta del criterio di "**misura della somiglianza**" (o della "**dissomiglianza**") tra i dati, sia delle **diverse strategie di raggruppamento**. Prima di approfondire qualsiasi percorso, bisogna effettuare delle scelte che riguardano i seguenti punti:

- 1.) *Scelta delle variabili da utilizzare*: per poter effettuare questa scelta bisogna tener conto di tutti gli aspetti rilevanti per il conseguimento degli obiettivi prefissati e di tutte le variabili necessarie a tale fine, tenendo presente che l'uso di variabili poco significative porta inevitabilmente a un peggioramento dei risultati, influenzando così il risultato finale. Nell'analisi che vedremo nel capitolo 4, le unità statistiche sono definite dai rendimenti, i quali sono normali e per il clustering utilizzeremo le serie storiche dei rendimenti di 50 fondi comuni d'investimento e alcuni indici che valutano la performance dei fondi comuni di investimento. Di conseguenza passiamo dai rendimenti a degli indici, dove ogni riga contiene valori di indici diversi e ogni colonna rappresenta il valore assunto dall'indice per ogni fondo. In generale una classificazione può considerarsi soddisfacente quando non mostra un'eccessiva sensibilità a piccoli cambiamenti dell'insieme di variabili utilizzate, quindi è necessario prima di effettuare una cluster analysis effettuare indagini esplorative accurate (che possano suggerire possibili configurazioni finali per la classificazione). Molto spesso, al fine di poter visualizzare e interpretare il risultato della cluster analysis è opportuno ridurre la dimensionalità della matrice dei dati, ad esempio mediante il metodo delle componenti principali (non analizzato).
- 2.) *Il metodo di formazione dei gruppi*: si distinguono metodi gerarchici e metodi non gerarchici. I primi consentono di ottenere una successione di raggruppamenti (partizioni) con numero di gruppi da n a 1 , partendo dalla più semplice in cui tutte le unità sono distinte, fino a quella in cui tutti gli elementi appartengono a un unico gruppo. I secondi invece, permettono di

raggruppare le n unità statistiche in un numero di gruppi fissato (soggettivamente) a priori.

-Metodi di raggruppamento $\left\{ \begin{array}{l} \text{Metodi non gerarchici} \\ \text{Metodi gerarchici} \Rightarrow \text{M. divisivi agglomerativi} \end{array} \right.$

3.) *L'indice di prossimità da utilizzare:* viene definito in base alle variabili a disposizione, cioè fra le unità statistiche per calcolare la matrice delle distanze fra di esse:

- se le variabili sono prevalentemente **quantitative** si ricorre alla *distanza Euclidea*;
- se sono prevalentemente **qualitative** si utilizza l'*indice di similarità*;
- se i **dati sono aggregati** in tabelle di contingenza si ricorre alla *distanza del chi-quadrato* fra le modalità.

Sottolineiamo di nuovo l'importanza di standardizzare le variabili per evitare che alcune pesino più di altre nella determinazione dei risultati finali, poi è necessario stabilire come verrà calcolata la prossimità fra i gruppi ottenuti nelle diverse fasi della procedura e stabilire quali unità statistiche sono quelle maggiormente significative e rappresentative per il gruppo. In base a quest'ultima scelta vedremo come si differenzia la classificazione gerarchica.

4.) *La determinazione dei criteri di valutazione dei gruppi ottenuti:* si basa sulla valutazione dei risultati di raggruppamento ottenuti significa, quindi bisogna verificare che i gruppi siano coerenti con l'obiettivo primario della cluster analysis e che, ovviamente soddisfano le condizioni di coesione interna e separazione esterna. Vi è proprio un **trade-off** fra l'ottenimento di gruppi omogenei, caratteristica che è tipicamente funzione crescente del numero dei gruppi scelti, e la necessità di ottenere una rappresentazione parsimoniosa che richiede, al contrario, un numero ridotto di gruppi.

Ogni scelta tra questi criteri porta, in genere, a classificazioni differenti, ciò significa che, in una classificazione due dati apparterranno allo stesso gruppo mentre apparterranno a gruppi diversi per un'altra classificazione.

3.2.1 METODI GERARCHICI DI CLASSIFICAZIONE

I metodi di classificazione gerarchica permettono, con numerosi passaggi, di *ripartire gli individui in classi* mediante un processo che viene ripetuto a diversi livelli e che è rappresentato graficamente per mezzo un *diagramma ad albero*, o *dendrogramma*. In base alla strategia che viene impiegata nell'effettuare l'analisi, i metodi di classificazione gerarchica possono essere di tipo:

- i) **"aggregativo"** (o **"agglomerativo"**): attraverso una serie di successive unioni (o fusioni) dei singoli individui in gruppi sempre più ampi, finché tutti gli individui risultano inseriti in un unico gruppo (*costruzione ascendente delle gerarchie*).
- ii) **"divisivo"** (o **"scissorio"**): suddividono il complesso delle unità di rilevazione da classificare in sottoinsiemi sempre più ristretti, finché si giunge a ottenere i singoli individui (*costruzione discendente delle gerarchie*).

Consideriamo *un esempio*¹⁹, graficamente ha come albero quello rappresentato dalla *Figura 3.2.1*, dove (per semplicità) si considerano solo 7 unità statistiche. I rami descrivono classificazioni successive delle unità, o meglio individuano divisioni successive delle unità in cluster e alla radice tutte le unità statistiche sono contenute in una sola classe. *In verde* si indica il metodo gerarchico agglomerativo e *in blu* il metodo scissorio, in particolare il procedimento a cui si fa riferimento è il seguente:

- *In verde* (direzione del metodo agglomerativo), si parte dai rami con $n=7$, e si misura la distanza Euclidea per 7 unità. Le prime due unità più vicine sono (1,2), ripetiamo la misura della distanza e di nuovo troviamo le due unità più vicine, cioè (3,4). Si continua con tale procedimento fino a quando non si riesce ad avere un unico gruppo rappresentato dalla radice (1,2,3,4,5,6,7).
- *In blu* (direzione del metodo scissorio), si parte dalla radice comprendente tutte le unità statistiche e si valuta la distanza più breve che mi permette di dividere la radice in due rami (1,2) e (3,4,5,6,7). Procedo nello stesso modo fino a quando, ogni unità non rappresenti un ramo.

¹⁹ Data Mining', McGraw-Hill , esempio valutato per 7 unità;

Le successive partizioni individuate da un dendrogramma sono “nidificate”, ciò significa che nei metodi gerarchici gli elementi che vengono uniti (o divisi) a un certo passo resteranno uniti (o divisi) fino alla fine del processo di classificazione.

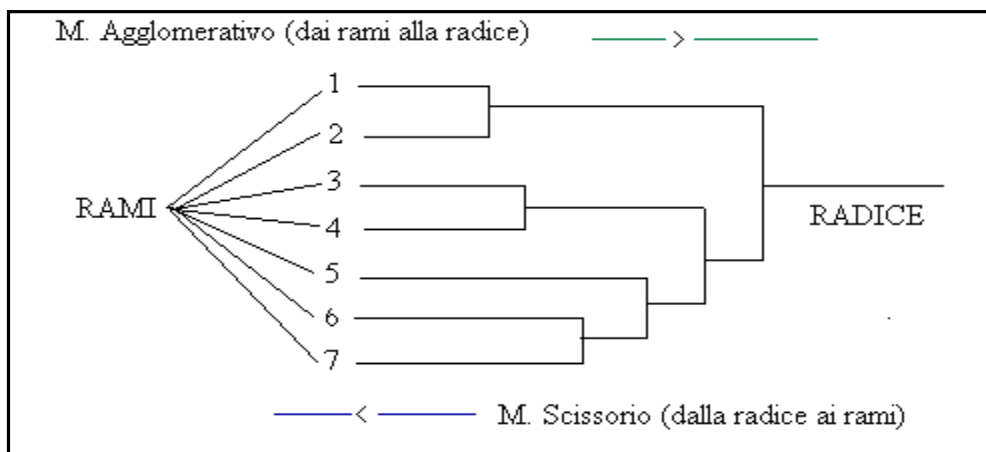


Figura 3.2.1: La struttura del dendrogramma per 7(sette) unità statistiche

Tabella 3.2.1: Partizioni corrispondenti al dendrogramma in Figura 3.2.1

Numero Cluster	Cluster
7	(1) (2) (3) (4) (5) (6) (7)
6	(1,2) (3) (4) (5) (6) (7)
5	(1,2) (3,4) (5) (6) (7)
4	(1,2) (3,4) (5) (6,7)
3	(1,2) (3,4) (5,6,7)
2	(1,2) (3,4,5,6,7)
1	(1,2,3,4,5,6,7)

Fonte: 'Data Mining', McGraw-Hill

Secondo il primo metodo si ha il vantaggio di ridurre il numero di partizioni da confrontare, rendendo le procedure computazionale più efficiente, ma anche lo svantaggio di non “correggere” errori di classificazione commessi nei passi precedenti (vedi Tabella 3.2.1).

In generale, la procedura statistica utilizzata è la seguente (*procedimento valido anche per la Figura 3.2.1*):

- i. *Inizializzazione*: si parte dalle n unità statistiche da classificare, ogni elemento rappresenta un cluster ;

- ii. *Selezione*: si selezionano i due cluster *più simili* (o vicini) rispetto alla misura di prossimità prefissata (*ad esempio*: la distanza Euclidea dato che i dati sono quantitativi);
- iii. *Aggiornamento*: → del numero dei clusters (n-1),
→ e della matrice delle distanze;
- iv. *Ripetizione*: si ripetono i passi definiti in (ii) e (iii) per n-1 di volte;
- v. *Arresto*: tutti gli elementi vengono definiti da un unico cluster (per il dendrogramma riportato nell'esempio ho (1,2,3,4,5,6,7)).

Esistono vari tipi di procedure, fra i **metodi di tipo aggregativo**, derivate da una diversa stima delle misure di similarità o di distanza fra gli individui o i gruppi di individui e bisogna distinguere tra, i metodi che richiedono come input la matrice delle distanze e quelli che richiedono i dati. La distanza tra due *cluster* dipende dalla distanza scelta tra i singoli dati, ma anche da una definizione "*globale*" che coinvolge tutti i dati dei due diversi *cluster*, cioè dal tipo di misura adottata della distanza tra *cluster*.

Le possibili misure basate sulla tipologia della distanza sono **sei** e si dividono in "**misure puntuali**" e "**misure globali**".

a) **Misure puntuali**

1. *Criterio del legame singolo* o del "vicino più vicino" (*Minimum or Nearest-Neighbour Method*): la misura di distanza tra due gruppi è la *minima distanza* tra tutte le coppie di punti di cui il primo elemento è nel primo gruppo e il secondo nel secondo. Si ha che:

$$d(C_1, C_2) = \min(d_{rs}) \quad r \in C_1, s \in C_2 \quad (3.2.1)$$

2. *Criterio del legame completo* o del "vicino più lontano" (*Maximum or Furthest-Neighbour Method*): la misura di distanza tra due gruppi è la massima distanza tra tutte le coppie di punti di cui il primo elemento è nel primo gruppo e il secondo nel secondo:

$$d(C_1, C_2) = \max(d_{rs}) \quad r \in C_1, s \in C_2 \quad (3.2.2)$$

b) **Misure globali**

1. *Criterio del legame medio* o legame medio fra gruppi (*Within Groups Clustering Method*): si considerano tutte le distanze ($n_1 n_2$) tra le coppie di punti di cui il primo elemento è nel primo gruppo e

il secondo nel secondo, e si prende la media, non ponderata, di tutte queste distanze. Si uniscono i gruppi più vicini con questa distanza:

$$d(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} d_{rs} \quad (3.2.3)$$

2. Criterio della media o legame medio nei gruppi (*Unweighted Pair-Groups Method Average*): si definisce la distanza tra due gruppi come la media delle distanze tra le coppie di punti che appartengono all'unione dei due gruppi. Si uniscono i gruppi in modo che la distanza media fra tutti le coppie di punti del gruppo risultante sia la minore possibile;
3. Criterio del centroide o *baricentro*: per ogni gruppo si calcola il baricentro dei dati, o meglio *il punto con coordinate pari alla media delle coordinate dei punti* e si usa come distanza tra due gruppi la distanza tra i due baricentri:

$$d(C_1, C_2) = d(\bar{x}_1, \bar{x}_2) \quad (3.2.4)$$

4. Il criterio di Ward: si aggregano i gruppi in modo che l'incremento di varianza nei nuovi gruppi sia il minimo possibile, tutte le volte che si aggiunge un unità statistica a un gruppo la varianza aumenta e lo scopo è quello di minimizzare tale aumento. Precisamente la **devianza totale (T)** delle p variabili, corrisponde a n volte la traccia della matrice dei dati, viene scomposta in due parti: la *Devianza nei gruppi (W: Within groups)* e la *Devianza fra i gruppi (B: Between groups)*. In termini formali, data una partizione in g gruppi:

- La *devianza totale delle p variabili (T)* corrisponde alla somma delle devianze delle singole variabili rispetto alla corrispondente media generale \bar{x}_s :

$$T = \sum_{s=1}^p \sum_{i=1}^n (x_{is} - \bar{x}_s)^2 \quad (3.2.5)$$

- La *devianza nei gruppi (W)* è data dalla somma delle devianze di gruppo:

$$W = \sum_{k=1}^g W_k \quad (3.2.6)$$

dove: W_k rappresenta la devianza delle p variabili nel gruppo k -esimo (di numerosità n_k) descritta da:

$$W_k = \sum_{s=1}^p \sum_{i=1}^{n_k} (x_{is} - \bar{x}_{sk})^2 \quad (3.3.7)$$

- E poi, la devianza fra i gruppi, B , è data dalla somma (calcolata su tutte le variabili) delle devianze ponderate delle medie di gruppo rispetto alla corrispondente media generale:

$$B = \sum_{s=1}^p \sum_{k=1}^g (\bar{x}_{sk} - \bar{x}_s)^2 \quad (3.2.8)$$

Nel metodo di Ward a ogni passo della procedura si aggregano tra loro i gruppi che comportano il minor incremento della devianza nei gruppi, cioè W (significa maggiore incremento in B) questo significa che si ha la maggiore coesione possibile intera.

Tra tutti gli altri metodi precedentemente visti si nota che, il metodo alla Ward non richiede il calcolo delle distanze, il suo input può essere la matrice dei dati originari e nel nostro caso saranno utilizzati i rendimenti mensili dei fondi comuni oppure indici di performance. Inoltre, lo possiamo vedere anche come una variante del metodo del centroide che invece richiede le distanze.

3.2.1.1 VALUTAZIONE DELLA CLASSIFICAZIONE DEL METODO GERARCHICO

Un algoritmo gerarchico è utilizzato per ottenere una famiglia di partizioni della n unità statistiche di partenza, o meglio un successione di n classificazioni delle suddette unità, con gruppi sempre in decrescenza (*da n a 1*). I gruppi sono caratterizzati da, minimo W , coesione interna e separazione esterna, massimo B . A ogni passo che si va avanti si calcola la bontà della corrispondente partizione ottenuta, in modo da scegliere quale sia più consona al raggiungimento degli obiettivi dell'analisi.

Un primo criterio di classificazione è quello derivato dalla *misurazione della prossimità* tra i gruppi uniti a ogni passo, che può suggerire di arrestare il processo quando tale misura sale bruscamente.

Invece, un criterio di giudizio più frequentemente utilizzato è quello basato sulla *scomposizione della devianza totale* delle p variabili, con riferimento al metodo Ward ($T=W+B$): per tale criterio si definisce valida una classificazione caratterizzata da una bassa devianza entro i gruppi (W) e da un elevato valore della devianza fra i gruppi (B). Alcuni indici che si riferiscono alla devianza sono:

- **Indice R^2** : assume valori nell'intervallo $[0,1]$ ed è espresso come:

$$R^2 = 1 - \frac{W}{T} = \frac{B}{T} \quad (3.2.9)$$

Se circa 1, allora la partizione corrispondente è ottimale, in quanto unità appartenenti allo stesso gruppo sono simili tra loro e i gruppi sono separati tra loro. Invece, se circa 0 allora la bontà dell'analisi di classificazione diminuisce. Al crescere dei numeri dei gruppi cresce l'omogeneità e allo stesso modo cresce R^2 .

- **Indice pseudo-F**: è un indice che si basa sulla versione inferenziale del criterio di Ward e misura il rapporto tra la varianza tra i gruppi e quella nei gruppi. Sia c un certo livello della procedure, corrispondente a un numero di gruppi pari a c , e sia n il numero di osservazioni in esame, allora l'indice è definito come:

$$F_c = \frac{B/(c-1)}{W/(n-c)} \quad (3.2.10)$$

Al diminuire di c diminuisce anche l'indice, in quanto la varianza tra gruppi dovrebbe diminuire e quella all'interno dei gruppi dovrebbe aumentare. Se c'è una caduta brusca significa che si sono uniti gruppi molto diversi tra loro.

- **Indice di RMSSTD**: (*Root Mean Square Standard Deviation*) è una variante del indice R^2 e considera solamente la parte della devianza nei gruppi 'aggiuntiva', che si forma al corrispondente passo della procedura di classificazione gerarchica:

$$RMSSTD = \sqrt{\frac{W}{p(n_h - 1)}} \quad (3.2.11)$$

Dove, se consideriamo il passo h -esimo ($h = 2, \dots, n-1$) e mostriamo che se si ha un forte incremento di tale indice allora i due gruppi che si sono uniti sono fortemente eterogenei ed è opportuno arrestare la procedura al passo precedente.

- **Indice di SPRSQ:** è simile al precedente soltanto che misura il contributo “aggiuntivo” del passo h -esimo della procedura, il cosiddetto R^2 semiparziale definito come:

$$SPRSQ = \frac{(W_h - W_r - W_s)}{T} \quad (3.2.12)$$

dove, h è il nuovo gruppo derivato dalla fusione dei gruppi r e s . Tale indice misura l’incremento della devianza all’interno del gruppo ottenuto unendo i gruppi r e s . Un bruscolo innalzamento indica che si stanno unendo gruppi eterogenei ed è opportuno arrestarsi al passo precedente.

Nell’analisi che si svolgerà sul dataset dei fondi comuni d’investimento USA, per valutare la bontà della cluster analysis considereremo due criteri²⁰ simili in parte a quelli precedentemente visti però nuovi, per quanto riguarda la tipologia dei dati a cui fanno riferimento, cioè a degli indici finanziari:

I) **Variance Ratio Criterion:**
$$VRC = \frac{\text{trace}(B)/(g-1)}{\text{trace}(W)/(n-g)} \quad (3.2.12)$$

La *partizione migliore* è quella che ha *VRC massimo*, cioè minima traccia di W . Dove, $(n-g)$ sono i gradi di libertà della matrice W e $(g-1)$ sono i gradi di libertà della matrice B , con n indicatore dei numeri delle osservazioni in esame. Tale indice è simile a quello pseudo-F.

II) **Marriott’s Criterion:**
$$MC = g^2 \frac{\det(W)}{\det(T)} \quad (3.2.13)$$

La *partizione ottimale*, in questo caso, si ha se l’indice *MC* è *minimo* ed è un indice invariante rispetto alle trasformazioni lineari e delle matrici non singolari.

Entrambi gli indici dipendono dal numero del gruppo g e sono ottimizzati rispetto ad esso. Inoltre, sono identificati in base al numero dei gruppi e il valore dell’indice *VRC* (*MC*) aumenta (*diminuisce*) monoticamente al variare di g , numero dei gruppi.

²⁰ in riferimento alla matrice delle devianze T , B e W ;

3.2.2 METODI NON GERARCHICI DI CLASSIFICAZIONE

I metodi di classificazione non gerarchico (*Raghanvan e Birchand, 1979*) ci permettono di ottenere n unità statistiche in g gruppi ($g < n$) il cui numero viene definito a priori da colui che svolge la classificazione. Tali algoritmi sono di solito aggregativi e producono una sola partizione dove ad ogni passo dell'algoritmo si rimette in discussione la partizione ottenuta. Le classi ottenute ad ogni iterazione intermedia vengono infatti cancellate e il processo di aggregazione ricomincia a partire dai nuovi centri. L'inizializzazione del processo di classificazione è necessariamente data da una qualche scelta di un insieme di g centri iniziali.

Alcune differenze cruciali tra i due procedimenti per ottenere dei clusters sono: gli algoritmi gerarchici risultano spesso eccessivamente onerosi in termini di calcoli che richiedono e un modo per ridurre la quantità dei calcoli, è di scegliere a priori il numero dei cluster, l'algoritmo non è più gerarchico e genera un'unica partizione delle n unità in g clusters. Nel caso gerarchico l'algoritmo cerca ad ogni passo la migliore scissione o aggregazione tra cluster, invece nel caso dei metodi non gerarchici l'algoritmo partiziona le unità in un numero predefinito di gruppi basandosi sulla ottimizzazione di un qualche criterio (predefinito). Il punto cruciale per iniziare il tutto è indicare il numero dei g centri di partenza intorno a cui aggregare le unità statistiche. Inoltre, a differenza dei metodi gerarchici l'assegnazione di un oggetto ad un cluster non è irrevocabile, o meglio le unità vengono riassegnate ad un diverso cluster se l'allocazione iniziale risulta inappropriata.

In algoritmi di classificazione non gerarchica *la procedura d'analisi* si può schematizzare come:

- i. *Scelta del numero dei gruppi* (g) e conseguente scelta di una classificazione iniziale delle n unità statistiche in tali gruppi;
- ii. *Valutazione del 'trasferimento'* di ciascuna unità dal gruppo di appartenenza a un altro gruppo. Si massimizza la coesione interna dei gruppi e viene calcolata la funzione obiettivo causata dallo spostamento, solo se rilevante il trasferimento è permanente;
- iii. *Ripetizione* del punto precedente fino a quando non viene soddisfatta la regola di arresto.

Come si può intuire tale metodo:

- è molto *più veloce*, in quanto *non richiede nemmeno il calcolo della matrice delle distanze*;
- è costruito a partire da un algoritmo molto *più stabile* rispetto a quello gerarchico;
- gli algoritmi non gerarchici si rivelano (per le diverse ragioni elencate sopra) molto più adatti per dataset di grandi dimensioni, per i quali i metodi gerarchici sarebbero poi, anche troppo lenti.

Tuttavia, il numero dei modi in cui è possibile *suddividere n elementi in g gruppi* non sovrapposti è molto ampio, specie per i dati reali, ed è impossibile ottenere e confrontare tutte queste combinazioni. Di conseguenza, risulta difficile massimizzare globalmente la funzione obiettivo e quindi gli algoritmi di classificazione non gerarchica ci portano a soluzioni vincolate, spesso corrispondenti a massimi locali. Alcuni *aspetti critici* di questo metodo consistono soprattutto nella necessità di definire a priori il numero dei gruppi, ma il criterio maggiormente utilizzato per prendere tale decisione consiste nella ripetuta conduzione dell'analisi con differenti valori di g e nella determinazione della soluzione migliore confrontando appropriati indici della bontà della classificazione (come R^2 e pseudo-F).

Fra i metodi di classificazione non-gerarchica (definiti anche come tecniche di ottimizzazione, *Optimization Techniques*) più noti abbiamo la ***K-means Cluster Analysis***, nota come *c-means*. I metodi "*c-means*" operano direttamente sulla *matrice dei dati* e suddividono gli elementi in gruppi i cui centri, inizialmente valori medi di ciascun parametro, vengono progressivamente modificati attraverso *un processo iterativo che minimizza una funzione della distanza cartesiana* (MQG, minimi quadrati generalizzati) quindi gli elementi dei vari gruppi hanno dei rispettivi centri. Questi metodi non sono gerarchici e occupano piccoli spazi nel calcolatore.

Rispetto ai *metodi "c-means"*, il *fuzzy c-means*²¹ (FCM) di **J. C. Bezdek**, ci consente in più di definire l'appartenenza ai gruppi in modo non definitivo ma "*elastico*" mediante una *funzione continua i cui valori vanno da 0 a 1*, indice del grado di appartenenza di ciascun elemento (*o sito*) ai vari gruppi. Ogni elemento

²¹ altro metodo di clustering non gerarchico;

ha quindi un "grado di appartenenza" in ciascun gruppo, i valori più prossimi ad 1 indicano un grado maggiore di appartenenza, cioè una maggiore "similarità" o somiglianza dei parametri di un partizione con quelli degli altri elementi del gruppo. Il numero dei gruppi può essere variante a piacere, di conseguenza variano anche i numeri dei gruppi, di conseguenza è possibile che varino pure i valori della funzione di appartenenza dei singoli elementi ai vari gruppi.

Per *ottimizzare il numero dei gruppi* esistono test di validità:

- I) alcuni test consentano di scegliere il numero dei gruppi sulla base del grado della loro "*definizione*" (partition), cioè della differenziazione l'uno dall'altro;
- II) altri sono metodi di ottimizzazione matematica, ad esempio l'FCM offre la possibilità di valutare il livello di adeguatezza o le scelte alternative di ripartizione in un certo numero di gruppi e del trasferimento di uno o più elementi da un gruppo all'altro sulla base dei riferimenti passati. Si ha in questo modo la possibilità di interazione tra le scelte della macchina e quelle dell'esperto che inserisce i dati (all'occorrenza) in base ai nuovi criteri .

3.4 L'APPROCCIO DI SHARPE, BROWN E GOETZMAN

Le analisi concentrate sugli strumenti finanziari, richiedono molta attenzione e accuratezza, in quanto le analisi che possiamo svolgere sono tantissime, si può partire a volte da semplici analisi matematico-econometriche e spaziare poi in analisi di tipo statistico-economico, cioè valutare i rendimenti, calcolare i vari indici di performance, ecc. A volte l'obiettivo non è soltanto il modo statistico mediante il quale vengono selezionati i fondi (*tutto quello che vale per gli individui è applicabile a dei fondi comuni d'investimento*), ma soprattutto come questi vengono in seguito analizzati.

La ricerca più attuale sui fondi comuni di investimento trova sostegno da parte di tre grandi scienziati, *Sharpe (1992)*, *Brown e Goetzman (1997)*. Il primo propone di valutare le regressioni lineari alla Sharpe, ma il problema è quello di trovarsi con rendimenti molto correlati serialmente e non si riesce ad arrivare a un buon modello, cioè rappresentativo per il mercato. Mentre *i secondi*, considerano il *metodo del 'k-means'* per i rendimenti delle serie storiche mensili, in quanto secondo gli studi converge in modo efficace a un valore ottimo, e si arriva a risultati rappresentativi solo dopo avere svolto molte prove.

In riferimento sia ai risulti ottenuti da Sharpe e anche da Brown e Goetzman, si arriva a un concepire un nuovo metodo di analisi dei fondi comuni, considerando *i rendimenti passati*. Questo metodo è basato su un'analisi esplorativa più accurata dei rendimenti e in base alle informazioni globali che si hanno a priori, si valuta un'analisi di raggruppamento (mediante clustering) che consente di poter confrontare le informazioni a priori con i risultati ottenuti a posteriori. Solo quindi in un secondo momento si introduce la parte di analisi finanziaria e tutte le proposte o controproposte, sostenute dagli analista.

In questa tesi si fa riferimento proprio all'approccio sopra citato, nonché noto come *l'approccio di Sharpe, Brown e Goetzman*, si valuta una *classificazione basata su metodi statistici* a partire da un'analisi esplorativa accurata e poi si prosegue con l'analisi finanziaria dei fondi. Il *procedimento logico* per l'analisi del dataset, cioè delle serie storiche dei prezzi dei 50 fondi comuni statunitensi, è il seguente:

- 1.) *si esegue un'analisi esplorativa dei prezzi e dei rendimenti, calcolando gli indici più opportuni per la valutazione della performance;*
- 2.) *in base sia ai rendimenti e sia agli indici di performance risultanti per i fondi comuni, svolgo una cluster analysis;*
- 3.) *in seguito si valutano le regressioni alla Sharpe dei gruppi ottenuti a posteriori e li confrontiamo con i gruppi a priori.*

Questa analisi ci permette di identificare, non solo le classi dei fondi più simili tra loro a priori e a posteriori, ma anche a valutare le caratteristiche dei gruppi ottenuti a posteriori mediante la cluster analysis, legate al comportamento dei fondi rispetto al mercato. L'analisi è presentata nel capitolo successivo.

CAPITOLO 4

CLUSTERING DEI FONDI COMUNI D'INVESTIMENTO USA

4.1 I FONDI COMUNI STATUNITENSIS

Fondata nel 1984 negli Stati Uniti, *Morningstar* è una società indipendente, leader nell'analisi e nella valutazione dei fondi e dei prodotti di risparmio gestito (non solo fondi, ma anche hedge fund, ETF, fondi chiusi, ecc.). È presente in 20 paesi e nei suoi database sono archiviati dati su oltre 145.000 strumenti di investimento, con l'obiettivo di fornire un'informazione sempre chiara, oggettiva e aggiornata ai risparmiatori. Nei mercati Italiani è presente dal 18 aprile del 2001 ed offre dati ed analisi su circa 6.000 fondi e servizi informativi a tutte le primarie istituzioni finanziarie. *Morningstar* ha sviluppato un sistema di rating basato su una misura denominata "*risk-adjusted*" rating (*RAR*). Per valutare un fondo non basta guardare il rendimento passato, ma bisogna tenere in considerazione il *rischio assunto dal gestore e i costi sostenuti dall'investitore*. Ogni fondo di investimento viene *collocato e valutato all'interno di una categoria* e svolge la funzione di parametro di riferimento costruita non sulla base dei benchmark generici come quelli dichiarati nei prospetti informativi, ma della specifica composizione dei portafogli gestiti.

Il rating Morningstar è una *misura quantitativo-statistica* di sintesi dei tre fattori e viene definita mediante le stelle: si parte da 5 per i migliori dei fondi e 1 per i peggiori (*vedi Figura 4.1.1*). Gli elementi chiave o i fattori di sintesi della metodologia di valutazione sono: *la categoria, il rendimento corretto per il rischio*

(*risk-adjusted return*) e i costi. Ma vediamo in particolare come queste quantità influenzano tale misura:

– **LE CATEGORIE:** il sistema di classificazione di Morningstar raggruppa i fondi in modo omogeneo, tenendo conto delle politiche di investimento dichiarate, ma verificandone l’attendibilità e la coerenza pratica attraverso l’analisi dei portafogli e dei singoli titoli che li compongono. Morningstar chiede alle società di inviare i dati completi sul portafoglio e fa un’analisi dettagliata prima di procedere all’inserimento di un fondo in una determinata categoria, perché spesso le politiche d’investimento non sono sufficienti per una corretta assegnazione. I parametri rivelanti per la classificazione non sono solo le aree geografiche o i settori, ma

▪ per i fondi azionari²²:

- la capitalizzazione delle società è: **large, small, mid-cap**

- e le prospettive di crescita sono: **value, blend, growth.**

▪ per i fondi obbligazionari:

- la durata finanziaria (*duration*) è: **long, medium, short**

- e l’affidabilità dell’emittente.

– **IL RENDIMENTO CORRETTO PER IL RISCHIO:** per il calcolo del rating vengono presi in considerazione i rendimenti mensili di un fondo negli ultimi 5 *anni* e la performance è “corretta” per il rischio, ossia tiene conto della volatilità dei risultati nell’orizzonte temporale considerato. Il presupposto, in generale, è che l’investitore preferisca ottenere rendimenti regolari nel tempo piuttosto che grandi guadagni e perdite clamorose, cioè che sia *avverso al rischio*.

– **I COSTI:** anche se spesso non viene preso in considerazione dall’investitore, è una variabile importante nella determinazione del rendimento. Il rating di Morningstar considera i costi di commissione che gravano sul fondo, in particolare quelle di gestione, di sottoscrizione e di riscatto. Molti prodotti, soprattutto comparti di SICAV estere, hanno più classi, che condividono lo stesso portafoglio, ma profili commissionali differenti; per questo motivo la valutazione è data separatamente.

²² In base ad essi, si è scelto anche il dataset in analisi;

– **LE STELLE:** (vedi Figura 4.1.1) di rating sono assegnate, su base mensile, all'interno di ciascuna categoria, tenendo presente quanto enunciato sopra si fa riferimento al seguente schema: al primo 10% dei fondi vengono assegnate 5 stelle, al successivo 22.5% quattro stelle e all'ulteriore 35% tre stelle. All'ultimo 10% è assegnata una stella e al rimanente 22.5% due stelle.

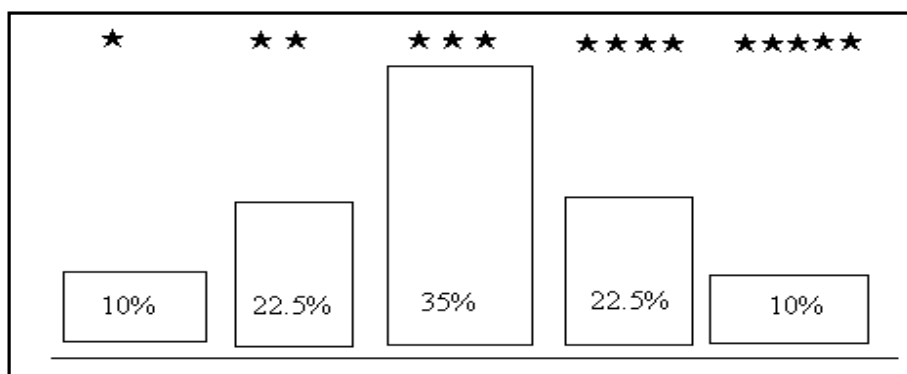


Figura 4.1.1: I criteri di assegnazione delle stelle di Morningstar rating

(Fonte: [www. MORNINGSTAR..com](http://www.MORNINGSTAR.com))

– **FONDI SENZA RATING:** non sempre è possibile attribuire il rating²³ a un fondo. I casi più comuni di mancata assegnazione sono:

- il fondo ha meno di 3 anni di vita;
- le informazioni sono troppo scarse per cui il fondo non può essere assegnato a nessuna categoria;
- il fondo ha effettuato continui cambi di categoria a seguito da variazioni nell'asset allocation, così che i rendimenti storici non sono più rilevanti;
- la categoria ha un numero troppo esiguo di fondi perché l'attribuzione del rating sia significativa.

L'utilizzo del rating risale al 1985 (USA) ed è un punto di partenza nella valutazione di un fondo a non va considerato come un'indicazione di acquisto o vendita. Nelle decisioni di investimento, il risparmiatore deve sempre tenere in considerazione la propria asset allocation, valutare la propria tolleranza al rischio e le esigenze di reddito e liquidità. Non esiste un fondo migliore in assoluto, ma fondi che meglio si adattano ai propri obiettivi e alle proprie necessità.

Morningstar classifica i fondi venduti facendo riferimento a delle analisi, fondamentali queste per valutare ogni singolo titolo che costituisce il portafoglio.

²³ come ad esempio avviene per un fondo incluso nel portafoglio in analisi, lo SSRGX;

Attraverso l'esatta identificazione delle caratteristiche delle singole posizioni del fondo è possibile capire qual è l'universo d'investimento del fondo, con quale mercato e quali concorrenti va misurato. O meglio, identificare dove investire effettivamente il fondo significa creare dei gruppi omogenei, all'interno dei quali è poi corretto e ha senso confrontare le performance.

Sulle caratteristiche dei titoli acquistati dal gestore, ogni fondo viene assegnato una categoria Morningstar e posizionato all'interno dello style box, la matrice a **nove** quadranti che identificano la tipologia del portafoglio, e i diversi gradi di rischio a cui è esposto (*vedi Figura 4.1.2*).

Lo style box azionario fornisce le due caratteristiche essenziali di un fondo che investe in azioni:

1. *la capitalizzazione*
2. *e lo stile delle società presenti nel portafoglio.*

Si tratta di due elementi importanti, in quanto spiegano non solamente la performance, ma anche quali sono i rischi a cui il fondo può andare incontro. Ogni società è soggetta a una valutazione da parte del mercato, valutazione che può essere superiore o inferiore a quella contenuta nei dati di bilancio. Sono le *società Growth* quelle che presentano una storia in crescita e un buon potenziale di incremento di valore del capitale nel futuro. La valutazione è effettuata prendendo in considerazione i cosiddetti multipli, ovvero i rapporti: *Price/Earning* o *Price/Cash Flow*. Al contrario, un titolo *Value* risulta sottovalutato dal mercato: il suo prezzo è basso rispetto al livello dei dividendi, degli utili e del valore contabile dell'azienda. A differenza dei primi, hanno un rischio di deprezzamento inferiore. La dimensione della società è un fattore che influenza la performance del fondo, quindi quelle con bassa capitalizzazione di mercato sono anche più volatili e spesso i fondi che vi investono non possono costruire posizioni importanti senza rischiare di influenzare il prezzo Borsa.

Lo style box obbligazionario è caratterizzato da *due fattori di rischio-opportunità* presenti in ogni fondo obbligazionario e questi sono:

1. *l'esposizione ai movimenti dei tassi di interesse*
2. *e la qualità del credito degli emittenti,*

per il *totale di nove combinazioni*, divise tra livelli *basso, medio e alto*. Gli strumenti di lunga durata sono i più rischiosi rispetto a quelli con scadenza breve,

perché più sensibili ai movimenti dei tassi di mercato. Per stimare la sensibilità dell'obbligazione delle variazioni del costo del denaro, si utilizza la *duration*, indice sintetico funzione della durata di un titolo obbligazionario e della ripartizione delle cedole. Più la duration è elevata, maggiori sono le oscillazioni del prezzo del titolo in seguito a politiche monetarie restrittive o espansive delle Banche Centrali. Il secondo fattore analizzato è la *qualità del credito* e l'esposizione a titoli di debito come obbligazioni, comporta un rischio di inadempienza da parte degli emittenti elevato se si tratta di organismi non governativi, paesi emergenti o società con basso merito. Il primo passo per determinare l'esposizione al credito consiste nel considerare il rating medio assegnato dalle agenzie specializzate a ciascuna obbligazione in portafoglio e a ogni rating viene dato un punteggio numerico, in seguito si calcola la media del portafoglio che indica il livello medio di rischio di credito contenuto nel fondo.

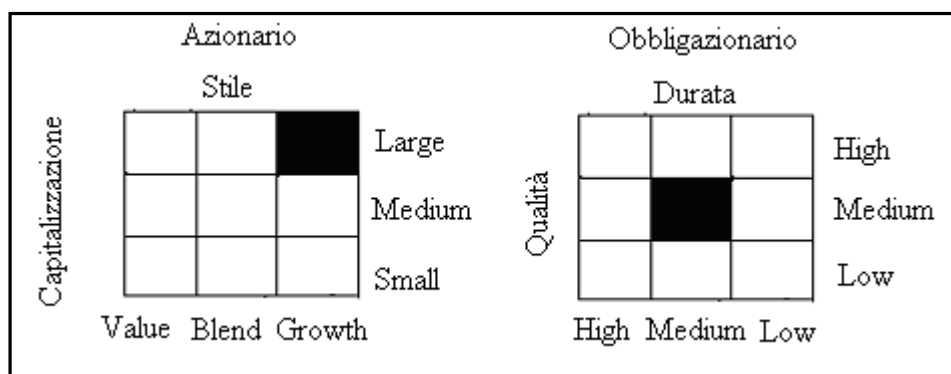


Figura 4.1.2: Gli style box azionario e obbligazionario statunitense

(Fonte: www.MORNINGSTAR.com)

4.2 CARATTERISTICHE DEL DATASET

Il dataset è un portafoglio abbastanza diversificato di 50 fondi comuni d'investimento statunitensi appartenenti a diverse categorie Morningstar, in quanto si distinguono sia per la capitalizzazione, sia per lo stile di gestione e quindi anche

per il Morningstar Rating (vedi Appendice pag. 139, Tabella A.1); in particolare derivano da 7 categorie:

- 11 fondi comuni “Small Growth”;
- 9 fondi “Small Value”;
- 10 fondi “Small Blend”;
- 5 fondi “Large Growth”;
- 5 fondi “Large Value”;
- 5 fondi “Large Blend”
- e 5 fondi “Mid-Cap-Growth”.

La ricerca dei fondi è casuale, o meglio è effettuata senza tener conto della composizione del benchmark e questo rende variabile lo stile di investimento del portafoglio, a cavallo tra i Large e i Mid-Cap, tra i Value e i Blend. Per ogni categoria è stato selezionato il fondo con maggiore permanenza nel mercato²⁴ mentre, il numero dei fondi per ogni categoria è diverso in quanto, non è difficile reperire società “Small” con lunga permanenza nel mercato e viceversa, non è facile reperire società “Large” con tale caratteristica. L’ordine in cui si presentano nell’elenco sopra citato viene mantenuto anche per la loro analisi in MatLab.

MatLab è un programma statistico-matematico molto efficiente e negli ultimi anni si è trasformato, anzi si è arricchito nel contenuto e nell’analisi matematico-statistico dei dati. Per l’analisi che segue si utilizzeranno i pacchetti finanziari e statistici, mentre il dataset è caratterizzato dalle seguenti componenti:

- **I dati:** sono serie storiche dei prezzi di *chiusura aggiustati con frequenza mensile* e valutati in Dollaro (\$);
- **Risk-free:** è rappresentato dal “1-Month Treasury Constant Maturity Rate”, scaricato dalle risorse del *Board of Governors of the Federal Reserve System*;
- **Benchmark:** invece, è dato dallo “S&P 1500²⁵” *indice composito* formato dallo S&P500, S&P 400 Mid-Cap e dallo S&P 600;
- **Periodo di valutazione:** Gennaio 2002-Dicembre 2006, 5 anni, 60 mesi;
- **Rendimenti:** per l’efficienza che garantiscono nell’analisi della performance si usano i rendimenti percentuali, cioè la *capitalizzazione semplice*.

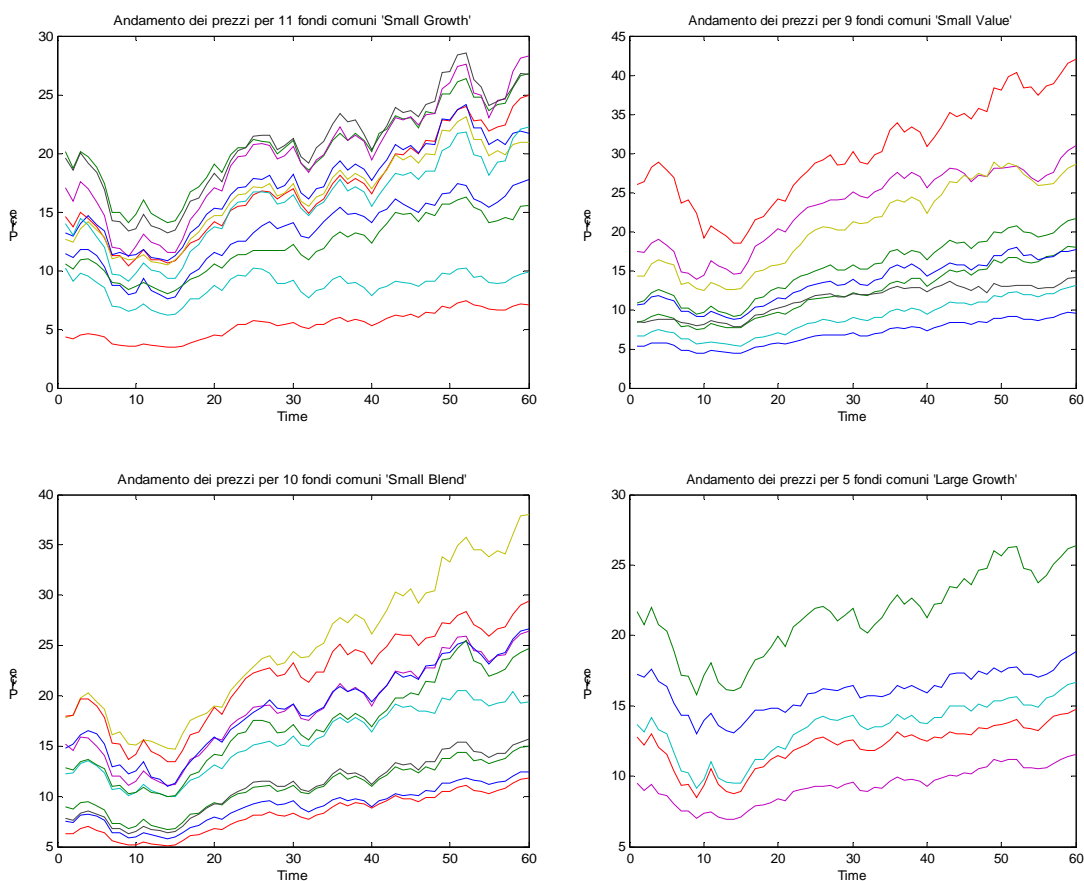
²⁴ l’unico criterio per la selezione dei fondi che devono costituire il portafoglio;

²⁵ scelto proprio per la sua composizione, in questo modo è meglio rappresentativo per i fondi scelti;

Iniziamo ad analizzare il dataset partendo da un'analisi dei prezzi, così verifichiamo tutti i presupposti teorici e le proprietà tipiche dei rendimenti. In seguito si prosegue con l'analisi del portafoglio e degli indici di performance, in un secondo momento si valuterà *una cluster analysis sui rendimenti mensili* dei 50-fondi comuni d'investimento statunitensi e in questo modo possiamo valutare gli effetti che si hanno dal clustering sulla performance del portafoglio, che è poi anche lo scopo di questa tesi.

4.2.1 ANALISI DEI PREZZI E DEI RENDIMENTI

I prezzi dei fondi comuni d'investimento statunitensi in analisi non sono stazionari nel tempo e risultano quindi non adeguati per poter valutare l'analisi sulla performance dei fondi comuni e una conferma di quanto enunciato si ha dalla *Figura 4.2.1* che descrive l'andamento dei prezzi dei 50 fondi comuni.



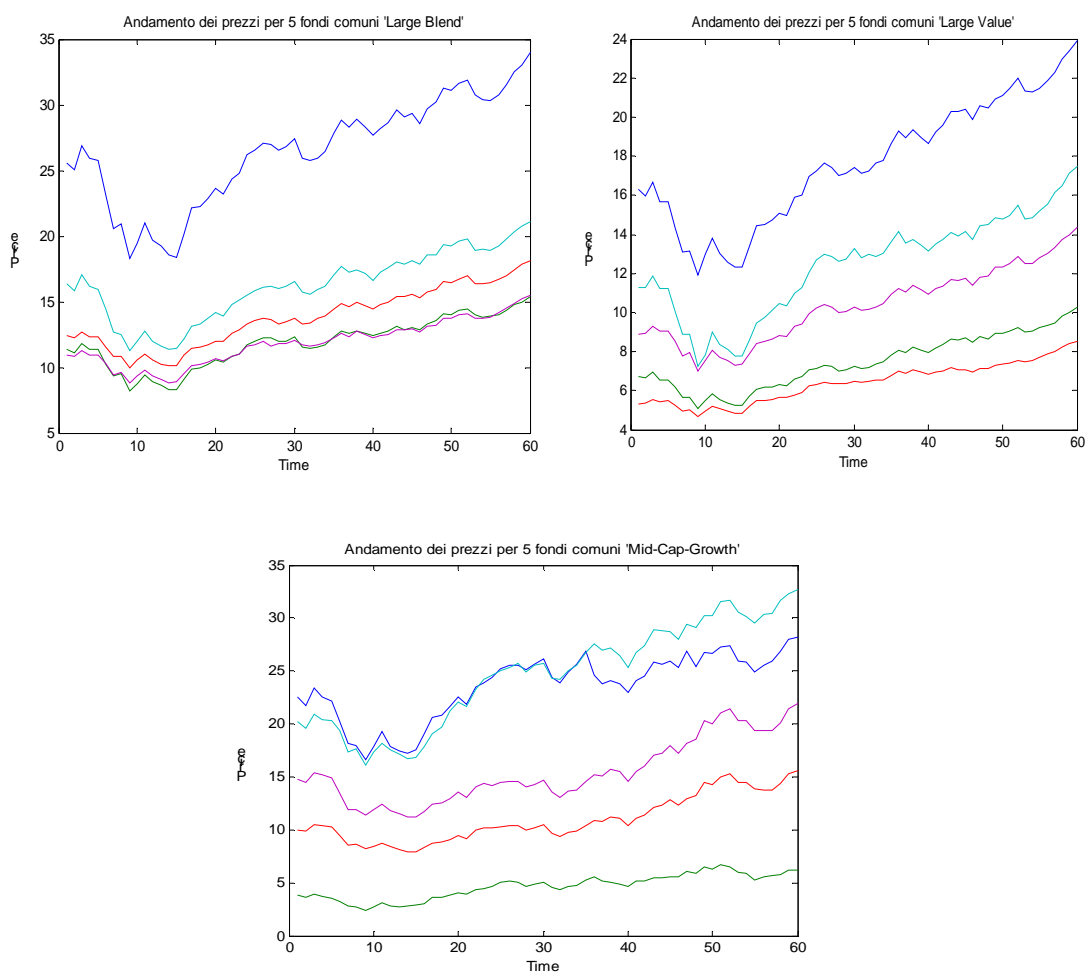


Figura 4.2.1: *Andamento dei prezzi per i 50-fondi comuni d'investimento statunitensi suddivisi per categorie, nel periodo Gennaio 2002–Dicembre 2006.*

I grafici riportati sono sette, tanto quanti sono i gruppi delle categorie di provenienza (a priori) dei 50 fondi e si basano sulle rispettive serie storiche dei prezzi per i 60 mesi in osservazione, cioè 5 anni.

Le categorie di provenienza sono 7^{26} su 9^{27} fra quelle definite nei paragrafi precedenti e sono diverse per capitalizzazione, stile di gestione e Morningstar rating, di conseguenza ci si aspetta che i fondi inclusi all'interno di un gruppo siano tra loro molto simili. Infatti, dai grafici si osserva chiaramente un andamento dei prezzi dei fondi, lungo i cinque anni, molto simile all'interno dei sette gruppi e questo è indice di correlazione interna tra i fondi. Osserviamo inoltre che, i prezzi base dei fondi appartenenti ad uno stesso gruppo, sono leggermente diversi anche se l'andamento nel tempo è molto simile, o meglio: il primo grafico (vedi Figura

²⁶ elenco al paragrafo 4.2;

²⁷ vedi paragrafo 4.1, pag. 91;

4.2.1) rappresenta l'andamento dei fondi *Small Growth* con prezzi base tra (5:20) e un solo fondo si colloca nella parte inferiore, cioè parte da un prezzo base di circa 5 anche se l'andamento nel tempo è simile ai fondi inclusi in quel stesso gruppo. Il secondo grafico include fondi *Small Value* con prezzi base tra (5:25), indice di una forte oscillazione dei prezzi base dei fondi, però le serie storiche hanno di nuovo lo stesso andamento nel tempo (o almeno molto simile). Il terzo grafico si riferisce ai fondi *Small Blend* con prezzi base tra (5:18) e andamento simile nel tempo delle serie storiche dei prezzi. Possiamo notare che in via generale l'andamento nel tempo delle serie storiche dei prezzi dei fondi, interni ai sette gruppi, è praticamente identico per questo motivo si suppone una forte correlazione interna tra di essi, questa verifica lo rinviamo nei paragrafi successivi. Inoltre, le serie storiche dei prezzi dei fondi in analisi hanno medie, per definizione, non nulle e per questo motivo si ha non stabilità e di conseguenza le serie storiche dei prezzi non sono adeguate per nessun tipo di analisi, ma in aiuto ci vengono i rendimenti.

I rendimenti sono caratterizzati da un andamento approssimativamente stazionario nel tempo e hanno medie, da un punto di vista esplorativo, circa nulle. Nelle *Figure 4.2.2a e 4.2.2b*, continua l'analisi esplorativa sui fondi in analisi e questa volta sono riportati i grafici dei *rendimenti percentuali* per i 50 fondi che a priori sono raggruppati in sette classi (ricordiamo diverse per capitalizzazione, stile di gestione e Morningstar rating), per un arco temporale di 60 mesi. L'andamento dei rendimenti per ogni gruppo (o classe) è molto simile, infatti all'interno dei gruppi i rendimenti quasi si sovrappongono, con differenze nell'ampiezza delle bande molto piccola, distinguibili dai vari colori che vengono attribuiti ai diversi fondi. *Ad esempio*: il primo grafico rappresenta i fondi *Small Growth* e se prima in questo gruppo una serie storica²⁸ si distingueva per il prezzo base basso, adesso questa differenza non è più osservabile, tranne per un picco negativo che si osserva nei primi 12 mesi e poi nel tempo si assorbe. In particolare (*vedi Appendice, pag. 138, Tabella A.2*) per tutti e sette i grafici si ha un comportamento dei rendimenti normale, o meglio *Accetto sempre* l'ipotesi nulla di distribuzione normale dei rendimenti, valutato mediante il *test JBS (Jarque-Bera Statistic)*. La stazionarietà in media dei rendimenti è stata verificata all'inizio con il calcolo dell'indice di Kurtosi e quello di Skewness, in seguito mediante il *t-test* che ci porta ad *Accettare*

²⁸ rappresentato da un fondo;

quasi sempre l'ipotesi nulla (H_0) di media nulla dei rendimenti per ogni fondo. L'indice di Kurtosi per la media dei rendimenti normali è circa 3 e nel nostro caso alcuni rendimenti sono minori di 3 e altri maggiori di 3, ma in generale abbiamo medie dei rendimenti leptocurtiche, $k > 3$.

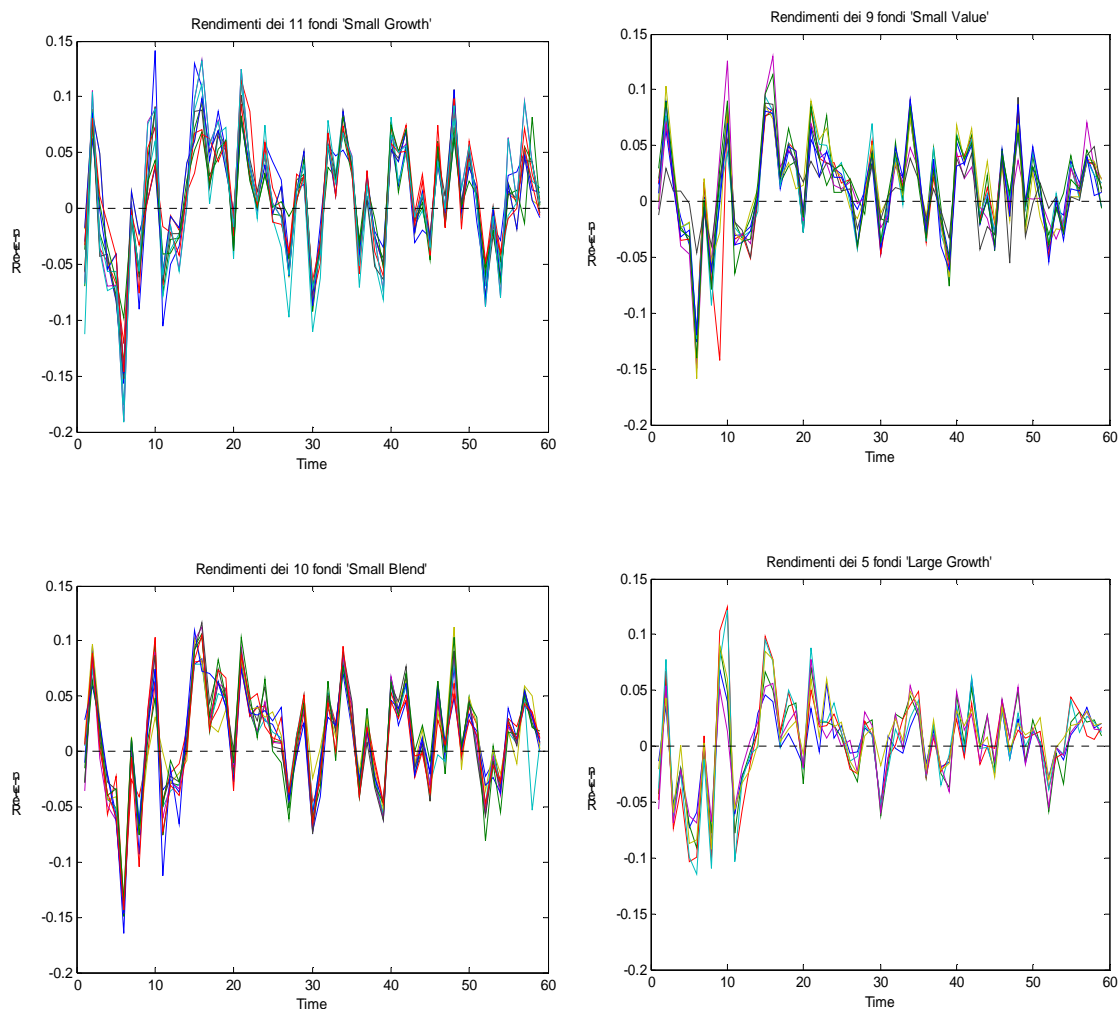


Figura 4.2.2a: Andamento dei rendimenti mensili dei 50 fondi comuni statunitensi in base alle 7 categorie d'origine da Gennaio 2002 a Dicembre 2006.

Questo fenomeno è tipico degli strumenti finanziari, dato che rendimenti lontani dal valore atteso corrispondono a risultati economici (positivi o negativi) rilevanti. L'indice Skewness invece, misura l'asimmetria dei dati attorno al valore medio e per la simmetria della distribuzione normale questo è 0. Nel nostro caso si ha una leggera asimmetria a sinistra, anche questa tipica degli strumenti finanziari, come anche dei fondi comuni d'investimento. Per quanto riguarda il t-test, questo è confrontato con un $\chi^2_{0.95,25}$ pari circa a 37.65 e ci porta ad Accettare l'ipotesi nulla

quasi sempre (vedi *Appendice, pag. 140, Tabella A.2, colonna 7*), perché in realtà ci sono tre casi in cui il test rifiuta l'ipotesi nulla e sono in particolare i fondi: *RCBSX* (Small Value), *DFSCX* (Small Blend) e *BOGIX* (Small Blend). I tre casi non ci limitano nel definire, in via generale, che: i rendimenti hanno media nulla, sono incorrelati²⁹ e indipendenti³⁰, o meglio i rendimenti dei fondi per i sette gruppi sono asintoticamente normali.

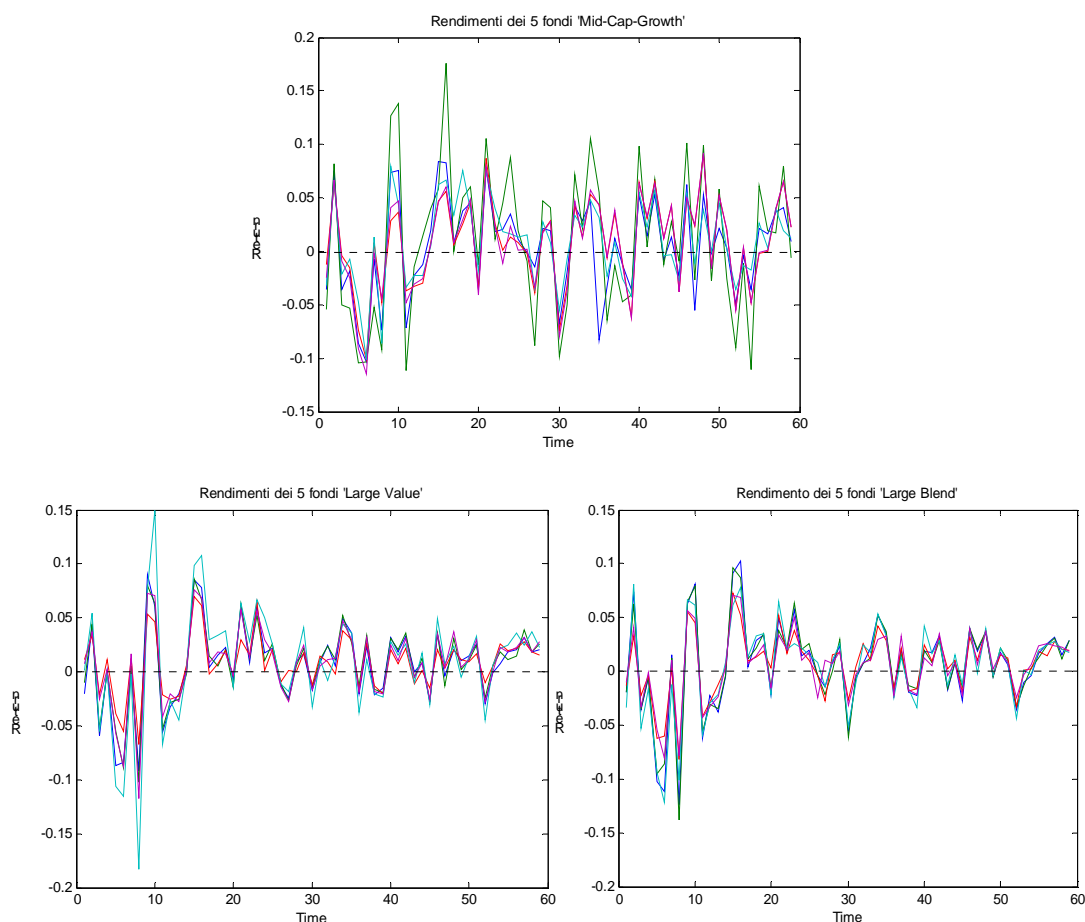


Figura 4.2.2b: *Andamento dei rendimenti mensili dei 50 fondi comuni statunitensi in base alle 7 categorie d'origine da Gennaio 2002 a Dicembre 2006.*

Stesse analisi si applicano sia al *benchmark (S&P1500)* e anche all'*asset risk-free*, e i risultati ottenuti confermano quanto precedentemente visto (vedi *Appendice, pag. 140, Tabella A.2, righe 52-53*). Graficamente (vedi *Figure 4.2.3 e Figura 4.2.4*) l'andamento di questi due assets, sia per quanto riguarda i prezzi e

²⁹ test sulle autocorrelazioni;

³⁰ test di Ljung-Box;

anche i loro rendimenti per il periodo in analisi³¹, è quello tipico delle serie storiche finanziarie: *prezzi volatili e rendimenti asintoticamente normali*.

I rendimenti, per le varie proprietà caratterizzanti insieme al benchmark e all'asset risk-free, sono i punti di riferimento per l'analisi degli extrarendimenti e quindi anche della performance del portafoglio in analisi. L'ipotesi di assenza di autocorrelazione è molto importante, sia per i rendimenti e sia per gli extrarendimenti in quanto è la *verifica dell'ipotesi di normalità*, che consente la buona finalità delle varie analisi svolgibili sui fondi e altri strumenti finanziari.

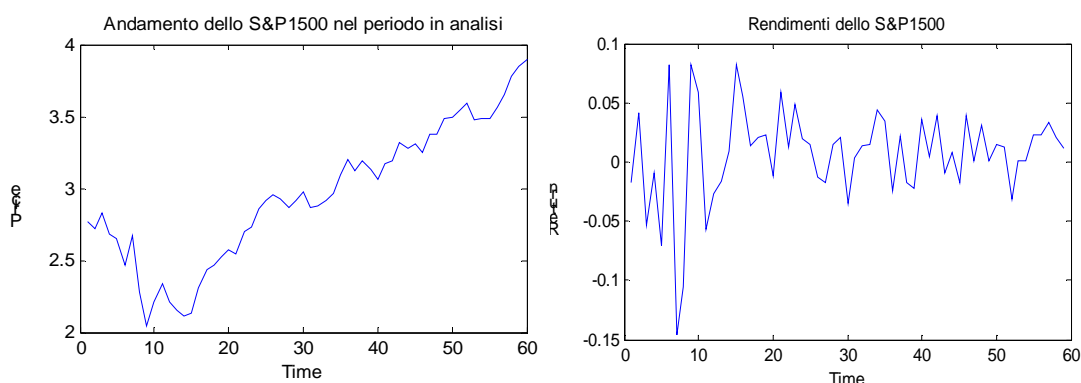


Figura 4.2.3: Rappresentazione dell'andamento del prezzo e del rendimento dello S&P1500 mensili, per il periodo Gennaio 2002 - Dicembre 2006.

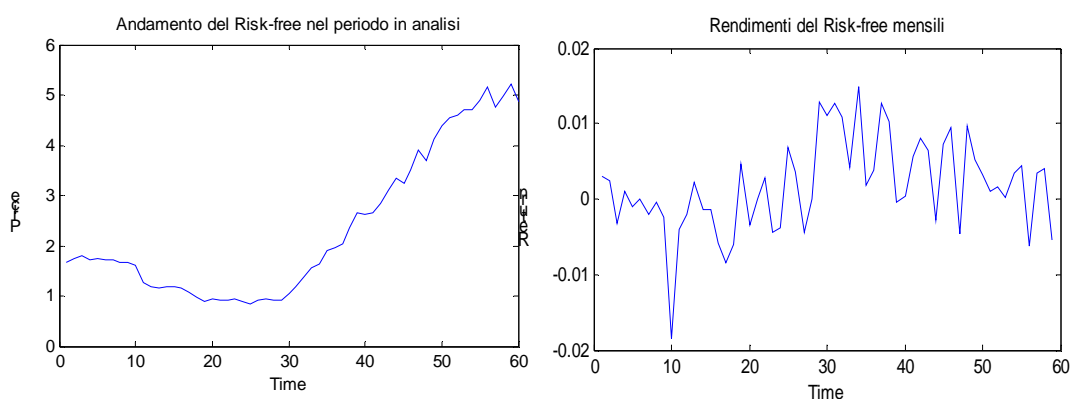


Figura 4.2.4: Rappresentazione dell'andamento del prezzo e del rendimento del Risk-free mensili, per il periodo Gennaio 2002 - Dicembre 2006.

La statistica *test bilaterale* LS^{32} (Lilliefors) considera il vettore degli extrarendimenti e in base a tale input verifica l'ipotesi nulla di distribuzione

³¹ Gennaio 2002 – Dicembre 2006;

³² Vedi Appendice, pag. 140, Tabella A.2, colonna 9;

normale degli extrarendimenti ed è confrontato con il valore soglia 0.1153, valore impostato da MatLab. Anche in questo caso si accetta sempre l'ipotesi nulla, cioè gli extrarendimenti sono distribuiti in modo asintoticamente normale. In conclusione si deduce che, i vari test confermano *la normalità asintotica sia dei rendimenti medi e anche degli extrarendimenti*, questo ci consente di proseguire con l'analisi della performance del portafoglio.

4.2.2 PERFORMANCE DEL PORTAFOGLIO

Mediante i 50 fondi comuni statunitensi costruiamo la frontiera efficiente dove l'asset risk-free è rappresentato dal primo valore assunto dal "1-Month Treasury Constant Maturity Rate" in data 01/01/2002. L'inclinazione della SML, cioè la sua pendenza, non è altro che la performance di Sharpe dei portafogli efficienti:

$$ps^* = \frac{\mu}{\sigma} = 0.1105 \quad (4.2.1)$$

Tutti i portafogli sulla frontiera efficiente hanno performance di Sharpe ps^* e sono ottenibili dalla combinazione dei due portafogli (*vedi Appendice pag. 139, Figura A.1*) efficienti, quello Risk-free (con solo il titolo non rischioso) e quello contenente il portafoglio con solo titoli rischiosi. MatLab ci consente di poter definire la composizione del portafoglio ottimale (investitore avverso al rischio) ripartito come descritto nell' *Appendice pag. 142, Tabella A.3*. Il valore di 'RiskyFraction' non eccede 1 (100%) questo implica che, la tolleranza specifica al rischio non consente di investire nel titolo rischioso, ma la parte di ricchezza in possesso è meglio investirla in asset risk-free.

Vediamo adesso altri indici di performance rispetto alla SML e in riferimento ai rendimenti dei fondi comuni d'investimento. Dato che non conosciamo i pesi delle quote (w) allora usiamo come proxy lo *S&P1500*, per poter valutare un modello di regressione lineare semplice per ogni asset. Questo metodo ci consente di avere una verifica empirica del CAPM, considerando i vari coefficienti in base al loro significato economico. Mediante le stime delle regressioni univariate per gli extrarendimenti mensili dei 50 fondi comuni d'investimento statunitensi osservati

nel periodo Gennaio 2002-Dicembre 2006 e con l'indice di mercato lo *Standard & Poor's 1500 Composite* si ottengono i risultati riportati nella *Tabella 4.2.1*. I parametri alpha e beta sono stimati mediante i MQO, invece gli standard error degli stimatori vengono valutati secondo White: in questo modo si ha la consistenza dei risultati anche per residui con media nulla e assenza di autocorrelazione, ma eteroschedastici (*tipico del nostro caso: residui non autocorrelati, però eteroschedastici*).

Le stime di beta (*vedi Tabella 4.2.1, colonna 5*) sono superiori al valore di 0.40 e quando si considera l'ipotesi nulla per $\hat{\beta}_i = 0$, la statistica test è confrontata con i quantili di una normale standard e ci porta a Rifiutare sempre l'ipotesi nulla, perché tutti gli assets presentano una stima di β significativamente diverso da zero per ogni livello di significatività fissato (sia per $\alpha=5\%$ e anche per $\alpha=1\%$).

Tabella 4.2.1: *Stima dei MQO $\hat{\alpha}_i$ e $\hat{\beta}_i$ sugli extrarendimenti mensili dei 50 fondi d'investimento statunitensi per il periodo Gennaio 2002 – Dicembre 2006.*

<i>Categoria</i>	<i>Codice³³ numerico</i>	<i>Fondi</i>	$\hat{\alpha}_i$	$\hat{\beta}_i$	β -test ¹	R ²
<i>Small Growth</i>	1	AFCAX	0,003	0,80	1,37''	0,33
	2	GTSBX	0,005	0,71	2,21	0,34
	3	ALSRX	0,005	0,74	2,03	0,36
	4	ASMGX	-0,004	0,77	1,35''	0,27
	5	POPAX	0,005	0,83	1,03''	0,29
	6	PMGAX	0,006	0,47	3,59	0,16
	7	QUASX	0,001	0,73	1,85''	0,31
	8	PMCIX	0,006	0,48	3,58	0,15
	9	MSASX	0,003	0,60	3,40	0,30
	10	ANOIX	0,005	0,50	3,50	0,18
	11	POPCX	0,004	0,82	1,03''	0,29
<i>Small Value</i>	12	ASVIX	0,007	0,52	4,38	0,28
	13	ARTVX	0,010*	0,48	4,48	0,23
	14	CSSBX	0,006	0,40	4,16	0,12
	15	TSVUX	0,008**	0,55	3,78	0,27
	16	RCBSX	0,006	0,59	3,26	0,27
	17	BERWX	0,009**	0,41	4,42	0,14
	18	AVALX	0,005	0,46	6,03	0,31
	19	AMRRX	0,005	0,55	3,85	0,28
	20	AASVX	0,008**	0,60	3,14	0,28
<i>Small Blend</i>	21	SMEAX	0,005	0,59	3,13	0,26
	22	ASCIX	0,005	0,70	2,24	0,31
	23	PNSEX	0,007	0,52	3,98	0,25
	24	SMCIX	0,004	0,54	3,54	0,24

³³ questo codice quando faccio il clustering mi identifica il fondo ad esso associato;

	25	<i>DFTSX</i>	0,005	0,73	1,96''	0,33
	26	<i>KDSAX</i>	0,009**	0,47	4,18	0,20
	27	<i>DFSCX</i>	0,008	0,68	2,30	0,29
	28	<i>CRSIX</i>	0,006	0,60	2,98	0,26
	29	<i>BOGIX</i>	0,008	0,60	2,85	0,25
	30	<i>SSRCX</i>	0,004	0,73	2,07	0,35
<i>Large Growth</i>	31	<i>ENGRX</i>	-0,002	0,60	5,98	0,50
	32	<i>CSTGX</i>	-0,0009	0,68	3,34	0,47
	33	<i>AFEAX</i>	-0,0002	0,80	1,75''	0,47
	34	<i>AFUAX</i>	-0,0012	0,81	1,68''	0,48
	35	<i>LCGAX</i>	-0,0006	0,57	4,98	0,44
<i>Large Value</i>	36	<i>AVAIX</i>	0,002	0,66	3,99	0,51
	37	<i>AMWYX</i>	0,003	0,61	4,61	0,47
	38	<i>TWEAX</i>	0,005	0,43	8,84	0,42
	39	<i>PDLAX</i>	0,003	0,90	0,82''	0,48
	40	<i>ABVAX</i>	0,004	0,58	4,80	0,44
<i>Large Blend</i>	41	<i>GTVCX</i>	0,0006	0,71	2,74	0,43
	42	<i>LCBAX</i>	0,0008	0,72	2,85	0,48
	43	<i>AGALX</i>	0,003	0,52	7,16	0,52
	44	<i>AGWFX</i>	0,00008	0,67	3,27	0,44
	45	<i>CHTRX</i>	0,002	0,51	6,37	0,43
<i>Mid-Cap-Growth</i>	46	<i>ADEGX</i>	-0,0002	0,69	2,69	0,38
	47	<i>CHCYX</i>	0,003	1,16	-1,01''	0,47
	48	<i>ATHAX</i>	0,004	0,53	4,13	0,27
	49	<i>ROCAx</i>	0,004	0,59	4,07	0,37
	50	<i>TWGTx</i>	0,003	0,57	3,55	0,28
<p>Gli simboli (**) indicano stime diverse da zero per un livello di significatività del 0.05 (0.01), con standard error calcolati secondo White.</p> <p>¹ Statistica test bilaterale per β con ipotesi $H_0: \beta = 1$, accetto H_0 per un livello di significatività del 0.05, valori segnati con (").</p>						

Invece, se si considera l'ipotesi nulla per $\hat{\beta}_i = 1$, si notano delle anomalie osservabili nella *sesta colonna della Tabella 4.2.1*, la quale riporta i valori della statistica test applicata. Si osservano solo dieci fondi, quindi un quinto del portafoglio, appartenenti a categorie diverse, che assumono un valore di β non significativamente diverso da 1, comportamento questo tipico dei fondi che hanno un andamento neutrale rispetto ai cambiamenti del mercato. La metà di questi fondi (5 su 10) appartengono al primo gruppo, cioè ai fondi *Small Growth*, i quali comprendono le piccole società caratterizzate da una forte crescita. L'unico asset con stima di β maggiore di 1 è il '*CHCYX*', o meglio il fondo tenderà a salire in condizioni di mercato favorevoli e scendere in condizioni opposte, ma in tal caso le perdite avvengono molto più rapidamente rispetto all'indice di mercato. Si può affermare che un fondo con β alto è più rischioso rispetto a quello con un β inferiore a 1, in quanto un valore di β piccolo è indice di una posizione difensiva

assunta dal fondo. Si nota inoltre che, *circa l'80%* degli assets hanno una stima di β *significativamente inferiore a 1*, quindi in generale gli assets assumono una posizione difensiva. I coefficienti di determinazione sono bassi e compresi tra 0.14 e 0.50, tipico per quanto avviene nelle analisi finanziarie, in quanto non esiste una correlazione perfetta nel campione, vale a dire, sussistono delle differenze tra il valore previsto degli extrarendimenti e il loro valore effettivo. Di conseguenza tale indice di bontà indica una scarsa correlazione tra gli extrarendimenti dei fondi effettivi con quelli stimati, o meglio il mercato non riesce a catturare tutta l'informazione per poter stimare accuratamente l'andamento dei fondi.

Da valutare con attenzione sono anche gli alpha di Jensen (*vedi Tabella 4.2.1, colonna 3*), i quali nella maggior parte non risultano significativamente diversi da zero. Alcuni assets hanno una stima di α negativa, ma non essendo significativamente diverse da zero, non possiamo dire che sono legate a una cattiva gestione dei fondi. Come previsto dal CAPM le stime di α , in generale, non sono significative, tranne per cinque assets appartenenti alle categorie "Small Value" e "Small Blend". I valori di alpha significativi sono lo stesso piccoli, mediamente sono intorno al 1%, indice questo del rendimento ottenuto dal portafoglio gestito rispetto a quello dato dal CAPM³⁴.

Invece, per quanto riguarda il comportamento degli indici legati alla rischiosità degli assets rispetto al mercato si osservi la *Tabella 4.2.2*, dove si riportano i risultati ottenuti per i 50 fondi comuni USA in analisi. Osservando i valori degli indici ottenuti si possono fare le seguenti osservazioni:

- *L'indice di Sharpe* del mercato è circa 0.121, valore che viene raggiunto e in alcuni casi superato, prevalentemente dai fondi appartenenti ai gruppi *Small Growth, Small Value e Small Blend*. Di conseguenza possiamo dire che sono le società piccole (in base alla capitalizzazione) quelle con indice di Sharpe più alto e in certi casi addirittura superiore rispetto a quello del mercato, e questo ci porta a pensare che i fondi delle piccole società, indipendentemente dalla tipologia di crescita, garantiscono rendimenti alti e sono i migliori nel mercato, quindi i fondi appartenenti a queste categorie sono quelli che hanno creato più valore per unità di rischio. Si può affermare che sono queste le categorie ad aver prodotto maggiormente extrarendimento, sempre in base al

³⁴ ci si riferisce a un portafoglio non gestito con beta costante e alpha nullo;

proprio livello di volatilità, per il periodo in analisi. In particolare, sono i fondi appena inseriti nel mercato che assumono valori di Sharpe pari o superiori al valore dell'indice di mercato (*vedi Tabella 4.2.2, colonna 4*). Invece, i fondi delle *Large Growth* e *Large Blend* mostrano valori per l'indice di Sharpe bassi, in generale inferiori al valore 0.07, quindi i fondi appartenenti a questo gruppo non sono di certo tra i migliori e non creano in modo significativo valore aggiunto. Infine, i fondi *Mid-Cap-Growth* sono abbastanza misti, in quanto alcuni si avvicinano all'indice di Sharpe di mercato, ma solo un fondo lo supera il "CHCYX", creando così valore aggiunto. Per quanto riguarda gli altri assets, si osserva che sono abbastanza bassi e si allontanano di parecchio dal valore soglia del mercato, quindi non sono significativi, in quanto non creano nessun valore aggiunto. Tale esito non è proprio inaspettato, in quanto sappiamo che all'interno del gruppo si hanno una mistura di fondi con caratteristiche di crescita e di capitalizzazione diverse.

- *L'indice di Treynor* degli assets in analisi non supera il valore di 0.010 ed è in generale molto vicino al valore di Treynor trovato per il mercato. Questo ci porta a pensare che, gli extrarendimenti per gli assets per unità di rischio sono relativamente in linea con l'indice del mercato e sono solo i fondi *Small Value* e *Small Blend* ad avere valore di tale indice superiore a quello determinato per il mercato, ovvero gli extrarendimenti attesi per l'unità di rischio sono elevati, quindi i fondi appartenenti a questi due gruppi sono i migliori anche per quanto riguarda l'indice di Treynor.
- *L'indice di Sortino* è in molti casi non calcolabile (NC³⁵), dovuto questo alla presenza di rendimenti mensili, per molti degli assets, non negativi. Infatti, sono stati registrati molti rendimenti mensili non negativi quindi, non viene soddisfatta la condizione necessaria per il calcolo dell'indici di Sortino per la maggior parte degli fondi, indice questo della loro bontà:

$$DD^2 = \sigma^2(D) = \begin{cases} \text{Var}[r_{P,t}] & \text{se } r_{P,t} < 0 \\ 0 & \text{se } r_{P,t} > 0 \end{cases} \quad (4.2.2)$$

³⁵ NC, non calcolabile: $SO = \frac{r_p - r_f}{DD}$ se il rendimento mensile risulta positivo;

- *Il Range* è definito mediante il rendimento massimo e quello minimo raggiunti nel periodo di osservazione, per ogni asset del portafoglio. Tale valore oscilla tra [0.15;0.34], indice questo della variabilità dei rendimenti garantiti nel tempo dal portafoglio e sono soprattutto gli assets appartenenti alla classe *Small* ad avere valori di range abbastanza alti, indice questo della crescita in valore (rendimenti crescenti nel tempo) per gli assets nel tempo. Mentre, gli assets *Large* hanno una variazione del range più contenuta, tranne per un asset "PDLAX" della *Large Value* che registra 0.332 spiegabile questo, dalla categoria a cui è legato, in quanto deve creare valore nel tempo.

Tabella 4.2.2: *Misure di performance per i 50 fondi comuni d'investimento statunitensi legate al rischio e con Risk-free variabile nel tempo*

Categoria	Codice numerico	Fondi	ps_i^{*36}	tr_i	Sortino	Range³⁷
<i>Small Growth</i>	1	AFCAX	0,106	0,006	NC	0,297
	2	GTSBX	0,063	0,004	NC	0,231
	3	ALSRX	0,126	0,007	NC	0,214
	4	ASMGX	0,008	0,0004	-0,058	0,289
	5	POPAX	0,128	0,008	NC	0,323
	6	PMGAX	0,118	0,007	NC	0,254
	7	QUASX	0,072	0,004	NC	0,254
	8	PMCIX	0,117	0,007	NC	0,254
	9	MSASX	0,085	0,005	NC	0,181
	10	ANOIX	0,114	0,007	NC	0,260
	11	POPCX	0,117	0,007	NC	0,323
<i>Small Value</i>	12	ASVIX	0,135	0,008	NC	0,211
	13	ARTVX	0,176	0,010	NC	0,197
	14	CSSBX	0,111	0,007	NC	0,232
	15	TSVUX	0,157	0,009	NC	0,218
	16	RCBSX	0,131	0,008	NC	0,279
	17	BERWX	0,162	0,010	NC	0,262
	18	AVALX	0,108	0,006	NC	0,149
	19	AMRRX	0,114	0,007	NC	0,211
	20	AASVX	0,161	0,010	NC	0,254
	<i>Small Blend</i>	21	SMEAX	0,115	0,007	NC
22		ASCIX	0,121	0,007	NC	0,256
23		PNSEX	0,143	0,008	NC	0,222
24		SMCIX	0,104	0,006	NC	0,221
25		DFTSX	0,132	0,008	NC	0,259
26		KDSAX	0,177	0,010	NC	0,246
27		DFSCX	0,165	0,010	NC	0,262
28		CRSIX	0,138	0,008	NC	0,255
29		BOGIX	0,157	0,009	NC	0,238

³⁶ indice di Sharpe per i-esimo fondo;

³⁷ determinato come differenza tra i rendimenti: [Max-Min] per un dato fondo;

	30	SSRCX	0,115	0,007	NC	0,249
<i>Large Growth</i>	31	ENGRX	0,004	0,0002	-0,029	0,157
	32	CSTGX	0,035	0,002	-0,013	0,180
	33	AFEAX	0,026	0,002	-0,003	0,228
	34	AFUAX	0,040	0,002	-0,017	0,237
	35	LCGAX	0,032	0,002	-0,009	0,147
<i>Large Value</i>	36	AVAIX	0,080	0,005	NC	0,182
	37	AMWYX	0,090	0,005	NC	0,188
	38	TWEAX	0,100	0,006	NC	0,137
	39	PDLAX	0,100	0,006	NC	0,332
	40	ABVAX	0,100	0,006	NC	0,194
<i>Large Blend</i>	41	GTVCX	0,059	0,003	NC	0,227
	42	LCBAX	0,064	0,004	NC	0,235
	43	AGALX	0,075	0,004	NC	0,155
	44	AGWFX	0,049	0,003	NC	0,203
	45	CHTRX	0,068	0,004	NC	0,152
<i>Mid-Cap-Growth.</i>	46	ADEGX	0,045	0,003	-0,003	0,188
	47	CHCYX	0,125	0,007	NC	0,287
	48	ATHAX	0,099	0,006	NC	0,192
	49	ROCAX	0,104	0,006	NC	0,181
	50	TWGTX	0,086	0,005	NC	0,207
<i>Indice Treynor di mercato</i>			0,005			
<i>Indice Sharpe di mercato</i>			0,121			

4.3 CLUSTER ANALYSIS DEI FONDI COMUNI USA

Lo scopo del clustering è quello di dare una visione completa e informativa sulla performance del portafoglio rispetto al mercato. La tesi ha l'obiettivo di applicare le tecniche di cluster analysis ad un caso reale. Dato un portafoglio di 50 fondi comuni d'investimento statunitensi, appartenenti a diverse categorie Morningstar, si dovrà stabilire a posteriori del clustering il miglior raggruppamento possibile e valutare la loro performance. I risultati ottenuti saranno confrontati con quelli veri (le 7 categorie da cui provengono i fondi) e si riprenderà l'analisi della performance alla luce dei nuovi gruppi, verranno poi discusse discordanze o le nuove informazioni aggiuntive derivabili dal clustering.

Le analisi sono state svolte con l'ausilio del pacchetto software *MatLab 7.1* e a tale fine è stato predisposto uno script (*cluster_analysis.m*) che genera gli output

grafici in modo efficiente, mettendo in evidenza le caratteristiche di quello che viene individuato come “miglior clustering possibile”.

Il dataset “*rendimenti dei 50 fondi comuni d’investimento U.S.A.*”, estrapolato in riferimento ai dati originari, è una matrice di ordine 50x59, ogni riga corrisponde a un asset, mentre le colonne rappresentano le osservazioni dei rendimenti passati. Da un punto di vista tecnico, ricevuta in input la matrice dei rendimenti, lo script ha generato la matrice delle distanze (o matrice di dissimilarità), basandosi sul metodo della distanza euclidea. Punto d’inizio per l’elaborazione degli algoritmi è quello di minimizzare la varianza interna e massimizzare quella esterna ai cluster. Successivamente, viene costruito l’albero gerarchico dei cluster quindi viene tracciato il dendrogramma, *Figura 4.3.1*. Mediante una linea tratteggiata di distanza Euclidea 0.4, mettiamo in evidenza il probabile numero ottimale di clusters a cui il dataset può essere suddiviso. I clusters individuati sono tre e i fondi comuni inclusi in questi gruppi sono identificati soltanto mediante il codice numerico³⁸ riportato in ascissa, questo raggruppamento vedremo che si ritrova nelle analisi successive. Nel rappresentare graficamente i dati insorge un problema di visualizzazione, in quanto si hanno 50 unità statistiche³⁹, mentre in ‘default’ si prevede una lettura e analisi più facilità di 30 variabili. Di conseguenza vengono selezionate e rappresentate soltanto 30 delle unità statistiche, le 20 unità non presenti nel dendrogramma *non vengono perse*. In default si unificano le unità statistiche con distanze simili in un’unica foglia mediante dei “links” rappresentando al massimo 30 delle unità in analisi. In seguito a un ciclo iterativo il dataset viene segmentato con un numero di clusters compreso tra 2 e 4. L’indice ‘*cophenetic correlation coefficient*’ misura **la bontà del clustering** e se vicino a 1 significa che la classificazione effettuata dal cluster analysis riflette bene il dataset. Nel nostro caso l’indice è **0.65**, di conseguenza possiamo dire che il clustering riflette in modo discreto il dataset cioè, al 65%. Osservando la *Figura 4.3.2* si può notare che in ascissa sono collocati i rendimenti dei 50 fondi comuni d’investimento in analisi in modo compattato, in quanto i più simili sono in automatico unificati e i numeri riportati in ascissa rappresentano i fondi maggiormente significativi per un certo valore di rendimento.

³⁸ codice numerico per ogni fondo, *vedi Appendice pag. 139, Tabella A.1, colonna 2;*

³⁹ rappresentano i 50 fondi comuni d’investimento;

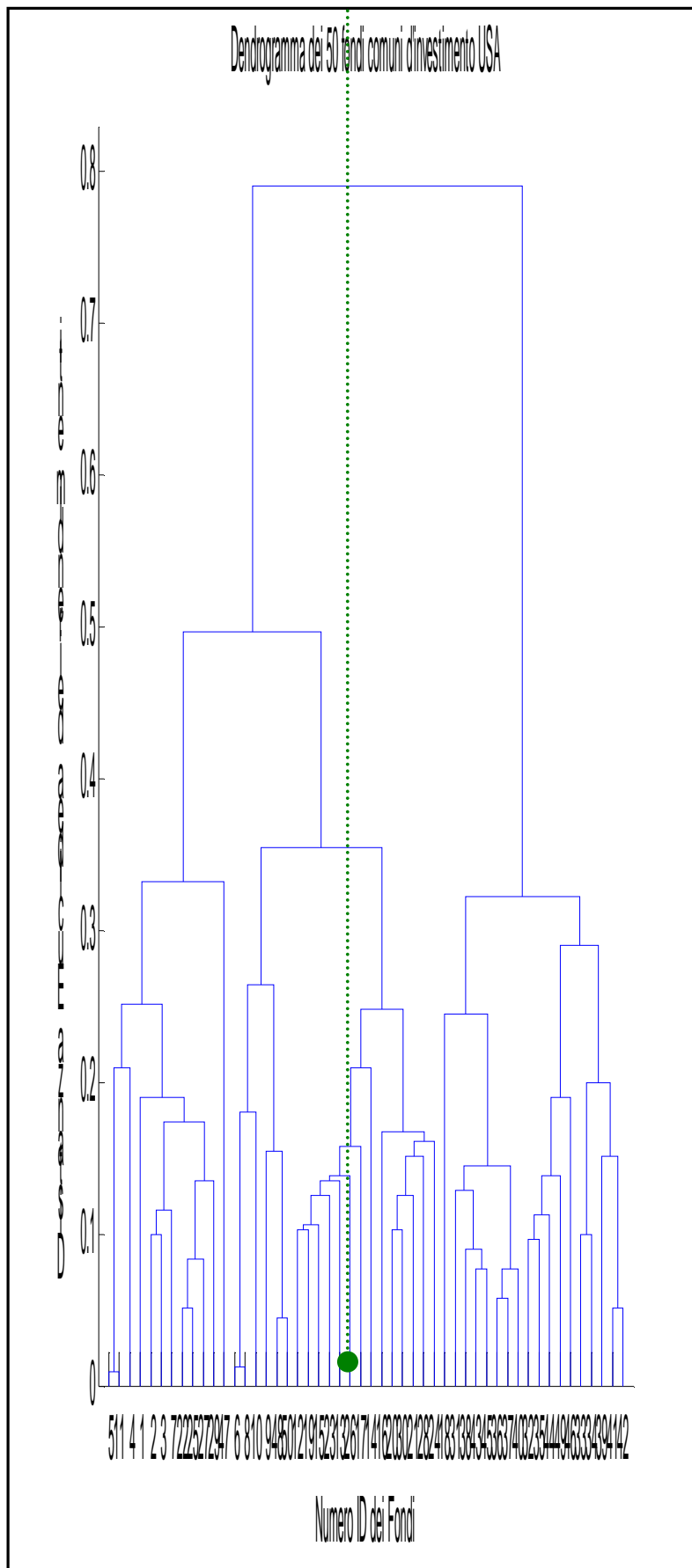


Figura 4.3.1: Dendrogramma per i rendimenti mensili dei 50 fondi comuni d'investimento statunitensi per il periodo

Gennaio 2002 - Dicembre 2006

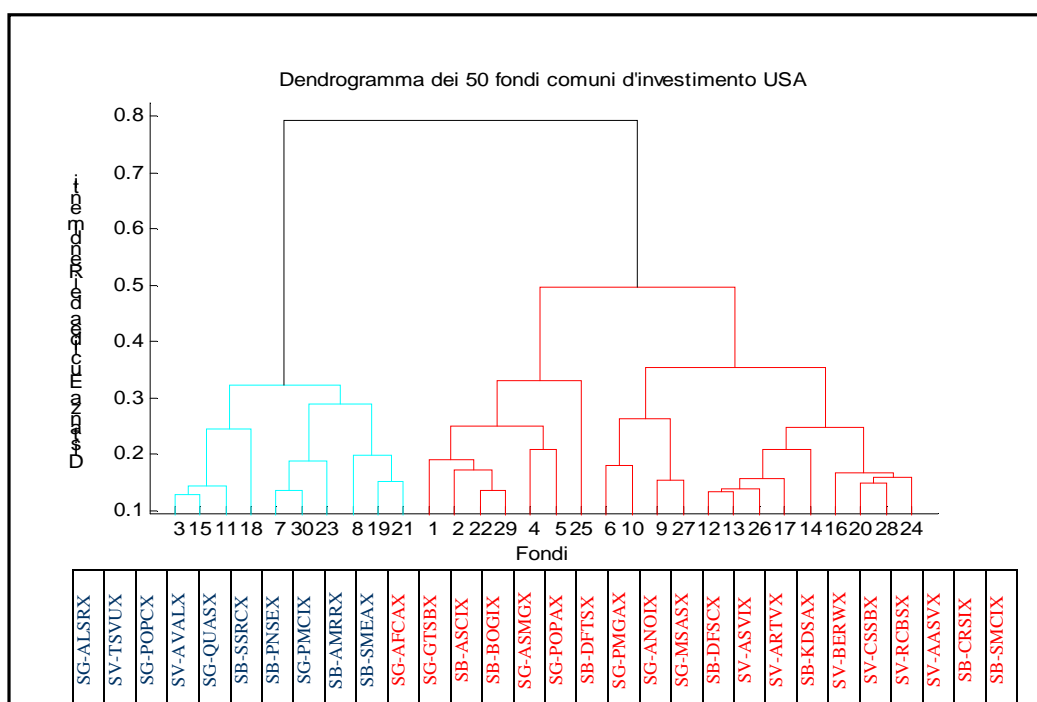


Figura 4.3.2: Dendrogramma in 'default' per i rendimenti mensili dei 50 fondi comuni d'investimento statunitensi

O meglio dicasi, ogni numero in ascissa, definisce il fondo maggiormente rappresentativo tra quelli aventi valori di rendimento simili, il programma⁴⁰ definisce solo uno dei fondi ed è per questo motivo che troviamo soltanto 30 numeri⁴¹, corrispondenti ai fondi che tra i più simili hanno una varianza di rendimento leggermente più elevata. I nomi e le categorie di appartenenza dei fondi sono riportati nella parte inferiore del grafico, mentre in ordinata viene riportata la loro distanza euclidea⁴². A sinistra del dendrogramma sono collocati quasi esclusivamente i fondi appartenenti ai gruppi Small Value e Small Blend che al loro interno possono però contenere fondi Large e Mid-Cap aventi simili caratteristiche dal punto di vista della varianza dei rendimenti; mentre, a destra del albero troviamo prevalentemente fondi Small Growth che benissimo possono rappresentare altri fondi delle classi Large e Mid-Cap sempre raggruppati in base al criterio della varianza dei rendimenti registrati per ogni fondo. Si nota che, sia il ramo superiore e sia il ramo a destra del grafico hanno distanze alte, indice questo della forte variabilità tra i gruppi formati, di conseguenza si ha la necessità di

⁴⁰ vantaggio offerto del programma, in quanto ci porta a una facile lettura dei dati;

⁴¹ numero massimo in default;

⁴² criterio della minima varianza all'interno dei gruppi e massima varianza all'esterno di esso;

calcolare l'indice di inconsistenza, che tiene in considerazione la media delle distanze. Il *coefficiente di inconsistenza* è calcolato come:

$$Y(k,4) = \frac{(CT(k,4) - Y(k,1))}{Y(k,2)} \quad (4.3.1)$$

Dove, k è il numero delle osservazioni all'interno dei gruppi iniziali che si formano. I nodi, che non portano più a nessun'altra ramificazione hanno un coefficiente di inconsistenza pari a 0 (vedi Appendice pag. 142, Tabella A.4).

Per ogni cluster formato sono state generate le matrici W (Within), B (Between) da cui deriva la matrice delle distanze totali (Totale). Le matrici delle distanze sono state elaborate mediante il metodo del centroide⁴³ e ward, con lo scopo di calcolare l'indice MC (Marriott's Criterion) e VRC (Variance Ratio Criterion), necessari per stabilire il numero ottimale di clusters in cui suddividere il dataset.

Utilizzando *l'indice VRC*, risulta che la miglior segmentazione è quella che divide il dataset in tre clusters e graficamente (vedi Figura 4.3.3 e Tabella 4.3.1) si ha:

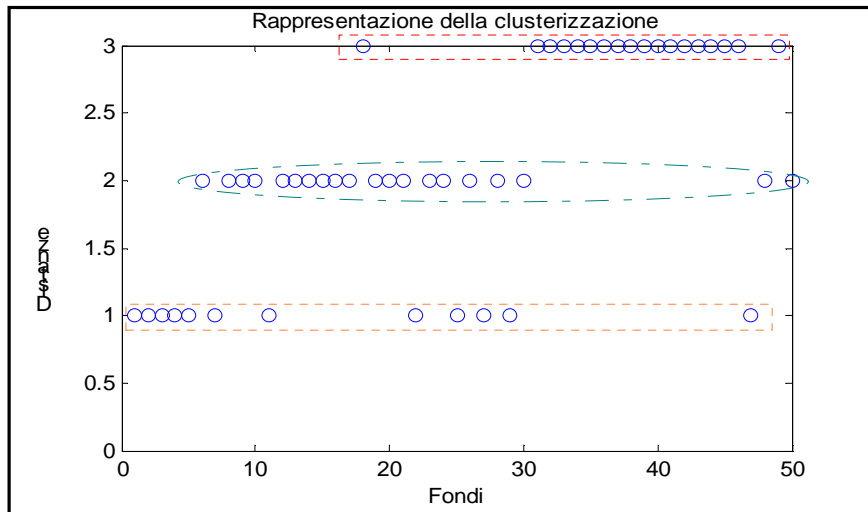


Figura 4.3.3: Rappresentazione grafica per i 3-clusters, secondo l'indice VCR

Grazie alla rappresentazione grafica possiamo capire meglio la composizione dei clusters generati, in particolare questi sono strutturati nel seguente modo:

- 1.) *il primo cluster:* include fondi della classe Small Growth e Small Blend, soltanto un fondo appartiene al gruppo Mid-Cap-Growth, in particolare i numeri identificativi sono: {[1:5],7,11,22,25,27,29,47};

⁴³ Metodo Centroid;

- 2.) *il secondo cluster*: comprende fondi derivati dalla classe Small Growth, Small Value, Small Blend e soltanto due fondi appartenenti al gruppo Mid-Cap-Growth, i numeri identificativi sono: {6,8,9,10,[12:17],[19:21],23,24,26,28,30,48,50};
- 3.) *il terzo cluster*: include prevalentemente fondi della classe Large Growth, Large Value, Large Blend, un soltanto fondo Small Value e un altro Mid-Cap-Growth, i numeri identificativi sono: {18,[31:46],49}.

Tabella 4.3.1: *Composizione dei tre clusters per i 50 fondi comuni d'investimento statunitensi, ottenuti mediante l'indice VCR.*

<i>Numero fondi per Cluster</i>	<i>1°Cluster ID-Nome-Classe-MR</i>	<i>2°Cluster ID-Nome-Classe-MR</i>	<i>3°Cluster ID-Nome-Classe-MR</i>
1	1-AFCAX-SG-3	6-PMGAX-SG-3	18-AVALX-SV-3
2	2-GTSBX-SG-3	8-PMCIX-SG-3	31-ENGRX LG-2
3	3-ALSRX-SG-4	9-MSASX-SG-2	32- CSTGX-LG-2
4	4-ASMGX-SG-1	10-ANOIX-SG-4	33- AFEAX-LG-3
5	5-POPAX-SG-3	12-ASVIX-SV-3	34- AFUAX-LG-4
6	7-QUASX-SG-3	13-ARTVX-SV-4	35-LCGAX-LG-2
7	11-POPCX-SG-3	14-CSSBX-SV-3	36-AVAIX-LV-3
8	22-ASCIX-SB-3	15-TSVUX-SV-4	37-AMWYX-LV-3
9	25-DFTSX-SB-4	16-RCBSX-SV-2	38-TWEAX-LV-3
10	27-DFSCX-SB-3	17-BERWX-SV-3	39-PDLAX-LV-4
11	29-BOGIX-SB-4	19-AMRRX-SV-2	40-ABVAX-LV-3
12	47-CHCYX-MCG-3	20-AASVX-SV-4	41-GTVCX-LB-2
13	-	21-SMEAX-SB-2	42-LCBAX-LB-2
14	-	23-PNSEX-SB-3	43-AGALX-LB-3
15	-	24-SMCIX-SB-3	44-AGWFX-LB-1
16	-	26-KDSAX-SB-4	45-CHTRX-LB-2
17	-	28-CRSIX-SB-3	46-ADEGX-MCG-4
18	-	30-SSRCX-SB-NA	49-ROCAX-MCG-3
19	-	48-ATHAX-MCG-4	-
20	-	50-TWGTX-MCG-3	-
<i>Legenda:</i> - SG: Small Growth; - SV: Small Value; - SB: Small Blend; - MCG: Mid-Cap-Growth. - LG: Large Growth; - LV: Large Value; - LB: Large Blend; -1-2-3-4-5: indicano il numero di stelle secondo il Morningstar Rating.			

Diversamente, l'utilizzo *dell'indice MC* ha individuato quattro gruppi di clusters, come classi rappresentative ottimali di sottoinsiemi da utilizzare per la segmentazione del dataset.

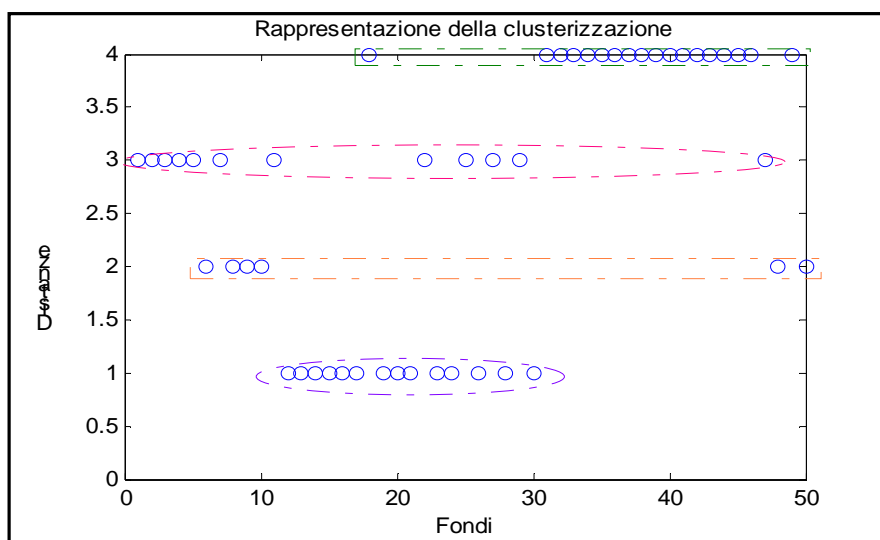


Figura 4.3.4: Rappresentazione grafica per i 4-clusters, secondo l'indice MC

I quattro clusters sono strutturati nel seguente modo (vedi Fig. 4.3.4 e Tab. 4.3.2):

- 1) *il primo cluster:* è definito da quasi soltanto fondi Small Value e Small Blend, ossia: {[12:17],[19:21],23,24,26,28,30};
- 2) *il secondo cluster:* è composto da fondi Small Growth e due fondi Mid-Cap-Growth, i quali sono: {6,[8:10],49,50};
- 3) *il terzo cluster:* è un misto di fondi Small Growth, Small Blend e uno soltanto fondo Mid-Cap-Growth: {[1:5],7,11,22,25,27,29,47};
- 4) *il quarto cluster:* è prevalentemente composto da fondi Large Growth, uno soltanto è Small Value e un altro Mid-Cap-Growth: {18,[31:46],49}.

Tabella 4.3.2: Composizione dei quattro clusters per i 50 fondi comuni d'investimento statunitensi, ottenuti mediante l'indice MC.

<i>N. f. per Cluster</i>	<i>1°Cluster</i> ID-Nome-Classe-MR	<i>2°Cluster</i> ID-Nome-Classe-MR	<i>3°Cluster</i> ID-Nome-Classe-MR	<i>4°Cluster</i> ID-Nome-Classe-MR
1	12-ASVIX-SV-3	6-PMGAX-SG-3	1-AFCAX-SG-3	18-AVALX-SV-3
2	13-ARTVX-SV-4	8-PMCIX-SG-3	2-GTSBX-SG-3	31-ENGRX LG-2
3	14-CSSBX-SV-3	9-MSASX-SG-2	3-ALSRX-SG-4	32- CSTGX-LG-2
4	15-TSVUX-SV-4	10-ANOIX-SG-4	4-ASMGX-SG-1	33- AFEAX-LG-3
5	16-RCBSX-SV-2	48-ATHAX-MCG-4	5-POPAX-SG-3	34- AFUAX-LG-4
6	17-BERWX-SV-3	50-TWGTX-MCG-3	7-QUASX-SG-3	35-LCGAX-LG-2
7	19-AMRRX-SV-2		11-POPCX-SG-3	36-AVAIX-LV-3
8	20-AASVX-SV-4		22-ASCIX-SB-3	37-AMWYX-LV-3
9	21-SMEAX-SB-2		25-DFTSX-SB-4	38-TWEAX-LV-3

10	23-PNSEX-SB-3		27-DFSCX-SB-3	39-PDLAX-LV-4
11	24-SMCIX-SB-3		29-BOGIX-SB-4	40-ABVAX-LV-3
12	26-KDSAX-SB-4		47-CHCYX-MCG-3	41-GTVCX-LB-2
13	28-CRSIX-SB-3			42-LCBAX-LB-2
14	30-SSRCX-SB.NA			43-AGALX-LB-3
15	-			44-AGWFX-LB-1
16	-			45-CHTRX-LB-2
17	-			46-ADEGX-MCG-4
18	-			49-ROCAX-MCG-3
<p>Legenda - SG: Small Growth; - SV: Small Value; - SB: Small Blend; - MCG: Mid-Cap-Growth. - LG: Large Growth; - LV: Large Value; - LB: Large Blend; -1-2-3-4-5: indicano il numero di stelle secondo il Morningstar Rating.</p>				

Le analisi effettuate hanno portato a due conclusioni diverse tra loro, dovuto questo ai differenti criteri utilizzati per poter stabilire il numero ottimale dei clusters in cui suddividere l'insieme dei dati.

La formazione dei gruppi ottenuti mediante il criterio VCR si basa prevalentemente sulla capitalizzazione e sull'indice di Morningstar rating. Mentre, l'indice MC tende a raggrupparli in un modo che assomiglia maggiormente alla classificazione reale, o meglio fa riferimento sia alla capitalizzazione, allo stile di gestione e anche alla categoria di Morningstar rating dei fondi, ma questo non significa che è la migliore scelta, in quanto per l'analisi della performance vedremo che è il criterio VCR a portarci verso il raggruppamento migliore. Gli assets si raggruppano prevalentemente in funzione della *capitalizzazione*; i *Small* definiscono il primo⁴⁴ e il secondo cluster, i *Large* nella loro vasta gamma compongono il terzo cluster (*vedi Tabella 4.3.1*). Infatti, il criterio VCR di clustering ci porta ad avere soltanto tre gruppi (con numero di fondi al loro interno variabili dai 12 ai 20), quindi l'informazione a priori viene elaborata in modo differente⁴⁵ ottenendo dei gruppi maggiormente informativi attorno alla performance del portafoglio. Tale metodo, per effettuare il raggruppamento, fa riferimento al criterio della minima varianza all'interno dei clusters e massima varianza tra loro e anche alla misura di Morningstar rating. Il risultato derivato dal metodo VCR trova sostegno dai confronti grafici in riferimento ai vari indici di performance e tutta questa analisi sarà presentata nel paragrafo successivo.

⁴⁴ definiamo come errore grossolano la presenza di un fondo Mid-Cap-Growth nel primo cluster;

⁴⁵ oggetto di approfondimento nei paragrafi che seguono;

Una conferma della bontà della scelta del “miglior raggruppamento” mediante il metodo gerarchico, si ha mediante il metodo non gerarchico delle ***k-means***. La convalida di quanto trovato dall’analisi svolta in precedenza, dopo alcuni tentativi mediante il procedimento delle *k-means*, ci porta ad avere la suddivisione del dataset in tre clusters (vedi *Figura 4.3.5*). I clusters sono al loro interno il più omogenei possibili rispetto alla varianza ⁴⁶(si nota che il secondo cluster è più omogeneo rispetto agli altri) ed eterogenei tra loro.

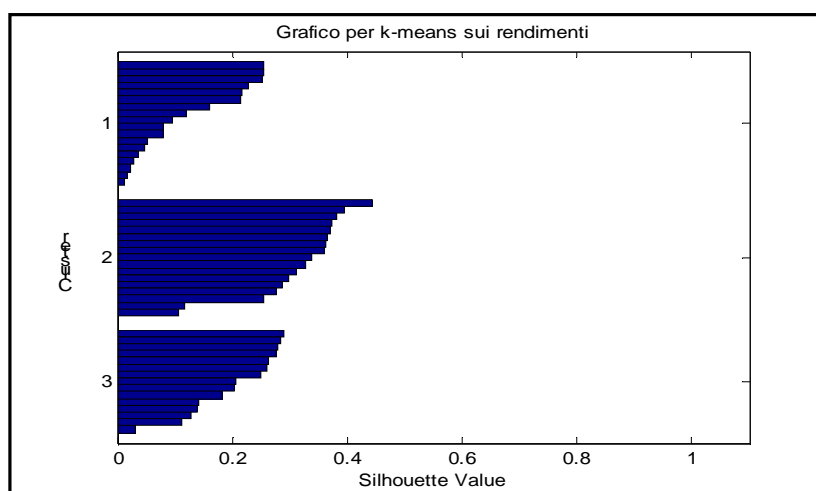


Figura 4.3.5: Raggruppamento dei 50 fondi comuni d’investimento statunitensi in base ai rendimenti di 60 mesi con il metodo delle *k-means*, caso non gerarchico.

4.4 EFFETTI SULLA PERFORMANCE DOPO IL CLUSTERING

La prima cosa, che viene naturale considerare dopo aver effettuato il clustering, è quella di considerare come sono formati i gruppi rispetto al criterio della minima varianza interna e massima varianza tra clusters. Infatti, in base ai tre nuovi clusters la suddivisione, diversamente dall’a priori, avviene per la capitalizzazione e il Morningstar rating e questo lo possiamo vedere da un primo confronto tra le medie dei fondi per i rispettivi tre clusters (*vedi Figura 4.4.1*).

⁴⁶ per costruzione;

Il primo gruppo con categorie ‘S. Growth, S.Blend e Mid.C.G.’, mediamente è collocato nella fascia superiore del grafico, cioè i fondi appartenenti a questo cluster hanno medie uguali o superiori al **0.6%** (linea rossa nella Figura 4.4.2),

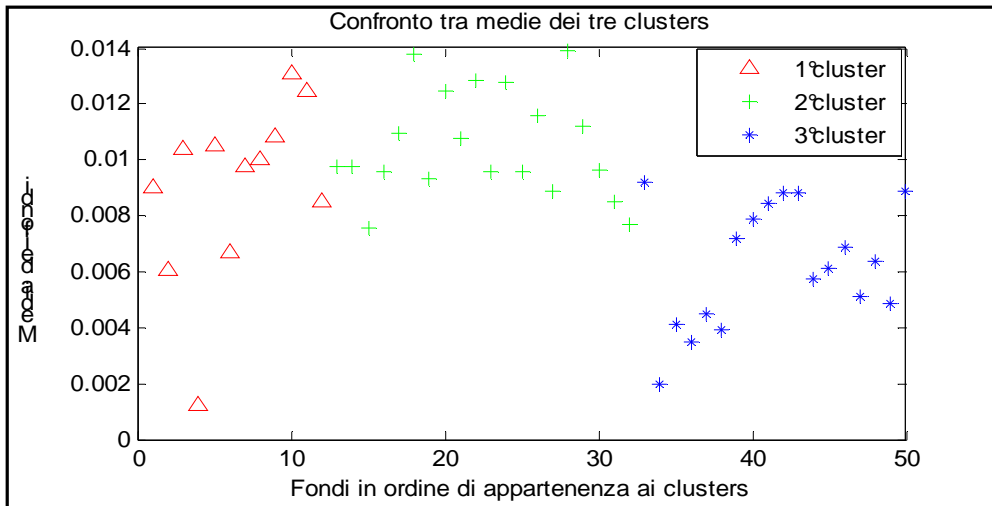


Figura 4.4.1: Confronto tra i clusters in riferimento ai rendimenti medi mensili dei fondi comuni d’investimento statunitensi.

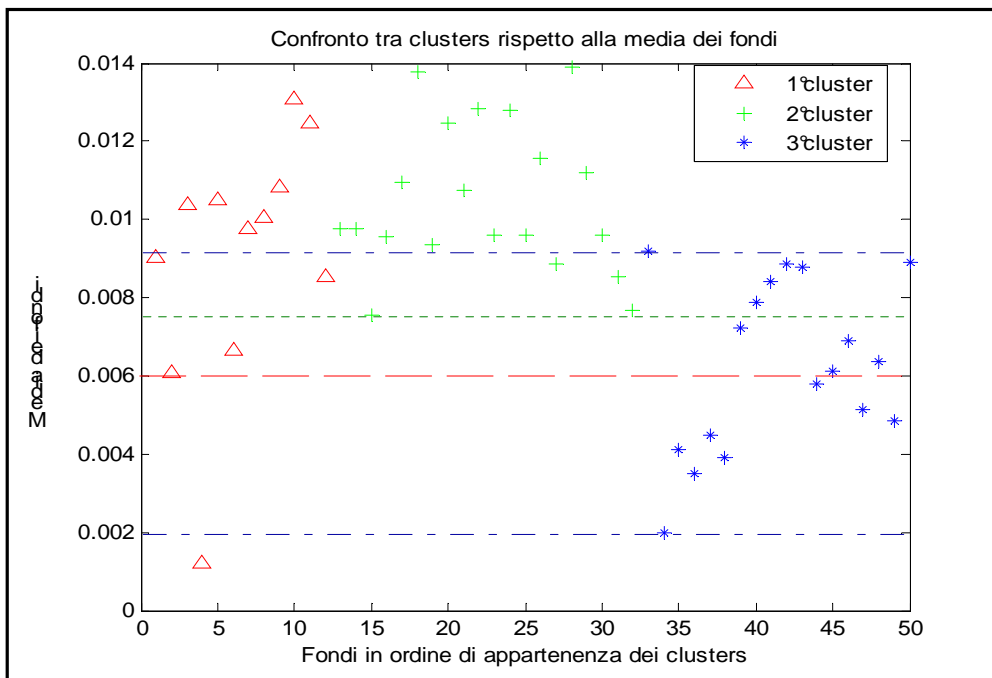


Figura 4.4.2: Confronto tra i clusters per bande dei rendimenti medi mensili dei fondi comuni d’investimento statunitensi.

tranne se non ci fosse per un errore dovuto alla presenza del fondo ‘ASMGX’, presente nella parte inferiore del grafico, si distingue anche per un indice di

Morningstar Rating basso, cioè 1. (vedi Figura 4.4.2 per le bande delle medie e la Tabella 4.3.1). Errori di questo genere non sono prevedibili, in quanto i gruppi ottenuti dagli algoritmi di clustering non sono sempre efficienti e la presenza degli errori grossolani non deve stupire molto. La media (*dei rendimenti*) interna del primo gruppo è **0.90%** ed è la media più alta registrata, in quanto comprende fondi avventi un indice di Morningstar rating alto (da 3 a 4 stelle), tranne per il dato anomalo che ha un Morningstar rating pari ad 1 stella.

Nel secondo gruppo si includono prevalentemente fondi della categoria ‘*Small*’ indipendentemente dallo stile di gestione che hanno però, l’indice di Morningstar rating è medio-alto (vedi Tabella 4.3.1), sia ha un misto di 2, 3 e 4 stelle. Le loro medie sono superiori al 0.70%⁴⁷. La media interna del secondo gruppo è **0.11%** ed è il gruppo con la media interna più bassa tra i cluster ottenuti.

Invece, per quanto riguarda il terzo gruppo, questo include solo fondi della categoria ‘*Growth*’ e fondi avventi un indice di Morningstar rating variabile tra 1, 2, 3 e 4 stelle. Si nota un raggruppamento dei fondi con medie comprese tra [0.3;0.9]% (*bande color blu nella Figura 4.4.2*), tranne per un dato⁴⁸ anomalo, e una media interna al cluster pari al **0.62%** punti percentuali. Quindi, il clustering ci ha portato ad avere informazioni importanti per quanto riguarda le medie dei nuovi gruppi e a distinguere tre clusters diversi tra loro fondamentalmente per la capitalizzazione e il Morningstar rating.

In un secondo momento, dopo aver avuto informazioni aggiuntive sulla media dei fondi che compongono i clusters⁴⁹, osserviamo cosa avviene quando valutiamo la varianza (vedi Figura 4.4.3). Graficamente si nota una suddivisione ben definita dei fondi comuni d’investimento statunitensi in base alla loro varianza per gruppo di cluster di appartenenza.

Il 1° cluster (*forma a triangolo e in rosso*) include fondi con varianze elevate, cioè le più alte tra i tre gruppi che abbiamo ottenuto. Per definizione i clusters hanno variabilità interna minima ed esterna massima. Inoltre, graficamente si ha la verifica di questa proprietà fondamentale per la costruzione “del miglior raggruppamento”, valida per tutti e tre i clusters.

⁴⁷ linea color verde nella Figura 4.4.2;

⁴⁸ fondo anomalo, in quanto ha la media più bassa all’interno del terzo cluster;

⁴⁹ si distinguono per medie diverse tra loro;

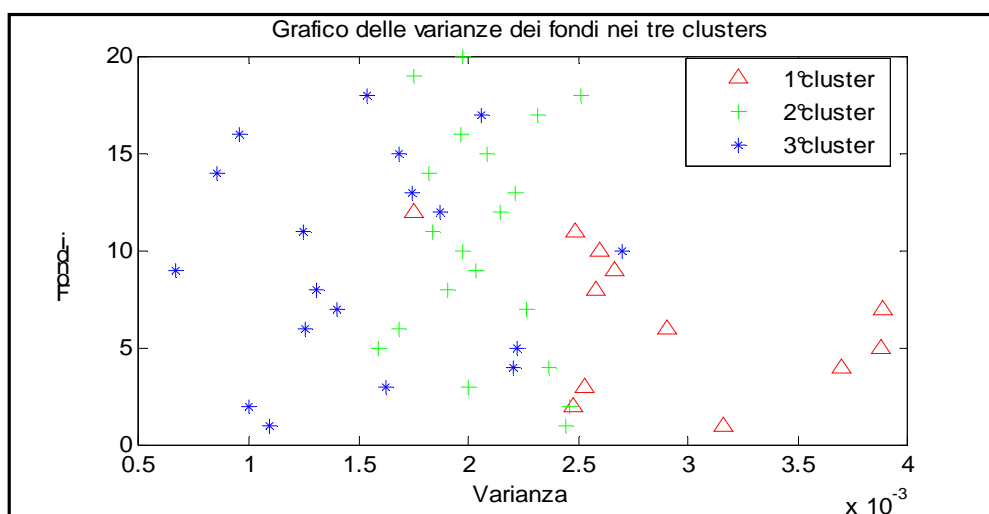


Figura 4.4.3: *Confronto grafico tra i tre clusters rispetto alle varianze interne dei rendimenti mensili dei fondi comuni d'investimento statunitensi.*

Mentre, il primo gruppo è completamente localizzato a destra (*forma a triangolo ed in rosso*) ed include fondi con varianza alta, tranne per un fondo anomalo che si distingue anche per la sua media bassa. Il secondo cluster si trova in centro (*fascia centrale forma a '+' ed in verde*) ed è assoggettato a una variabilità più limitata e anche contenuta rispetto agli altri due clusters. Inoltre, nella parte bassa del grafico notiamo la sovrapposizione di due fondi appartenenti però a due clusters diversi, al primo (GTSBX – SG) e al secondo (PMCIX-SG), (*vedi Tabella 4.3.1, pag. 112, terza riga*). Di conseguenza, essendo posizionato in centro, i fondi hanno varianza limitata ma dall'altra parte, una media più alta quindi il gestore può usare questa informazione per poter guadagnare in extrarendimento perché, tutti mirano a trovare i fondi che, a scapito di una varianza discreta portino ad avere rendimenti alti. Tutto questo ci porta a collegare le differenze in media (sopra descritte) con la variabilità tra i gruppi, da cui possiamo ricavare informazione preziosa sulla possibilità di creare un portafoglio che garantisca guadagno.

Infine, il terzo cluster (*forma a stella ed in blu*) è presente nella parte sinistra del grafico ed include fondi che hanno una variabilità più contenuta, anche se notiamo un valore anomalo alla sua destra, il quale precedentemente si distingue anche per la media bassa. Il comportamento dei dati anomali è spiegabile dal fatto che (*vedi Figura 4.4.4*), il clustering sui rendimenti dei fondi comuni d'investimento ci porta un'altra informazione aggiuntiva, o meglio, i gruppi che prima avevano una

media bassa adesso si trovano ad avere anche una varianza bassa (a sinistra del grafico), quello con media intermedia come il 1° cluster (a triangolo ed in rosso) è posizionato a destra e il secondo cluster che include fondi con medie alte è posizionato in centro ed in alto.

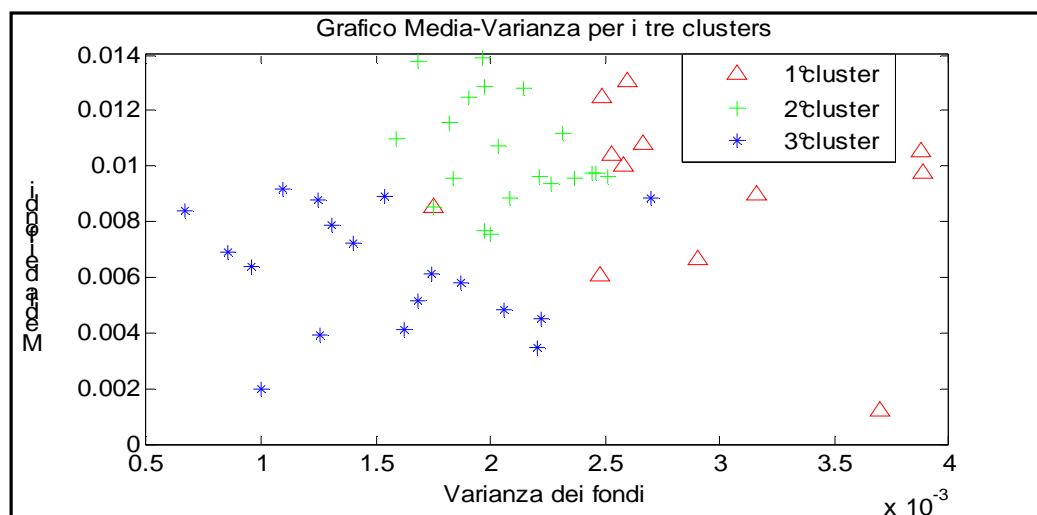


Figura 4.4.4: Rappresentazione del confronto media-varianza interna per i tre clusters sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi.

La prima osservazione da questo confronto è, di nuovo, legata alla predisposizione che hanno i tre clusters nel grafico. Il terzo cluster (forma '*' e in blu), avente media interna al gruppo 0.11%, è situato nella parte più a sinistra e in basso del grafico media-varianza, o meglio, si intuisce che i fondi appartenenti a questo cluster offrono la possibilità di avere rendimenti contenuti a scapito di un rischio relativamente basso. Questa informazione per un gestore è molto importante, in quanto localizza i fondi meno remunerativi e aventi un rischio basso; infatti, in esso rientrano tutti i fondi 'Large'. Per quanto riguarda il secondo cluster⁵⁰, la media interna è del 0.90% e lo porta a localizzarsi nella parte centrale del grafico e in alto, a scapito di una varianza compresa tra $(1.5:2.5) \times 10^{-3}$: *indice questo dell'aumento del rendimento all'aumentare della rischiosità, così come è previsto anche da Markowitz*. Questa osservazione è valida anche per il primo cluster⁵¹ localizzato nella parte più a destra e in alto (tranne per un fondo, che come abbiamo spiegato prima è risultato di un errore grossolano dalla parte degli

⁵⁰ forma a '+' e in verde;

⁵¹ forma a triangolo e in rosso;

algoritmi di clustering). Si osserva di nuovo la coincidenza tra due fondi che appartengono a clusters diversi e questi sono gli medesimi fondi dell'osservazione precedente: due Small Growth - GTSBX e PMCIX – appartenenti rispettivamente al primo e al secondo cluster. Tale sovrapposizione ci informa che il clustering non è del tutto corretto, in quanto due fondi appartenenti alla stessa categoria e avventi stesso indice di Morningstar rating sono definiti inutilmente in due gruppi diversi, però vediamo in seguito se ritroviamo questo risultato, magari gli indici di performance che propongono sono diversi?

Viene naturale pensare, dopo tali verifiche, come si comportano le frontiere efficienti? (vedi Figura 4.4.5) Ci sono possibilità di combinarli?

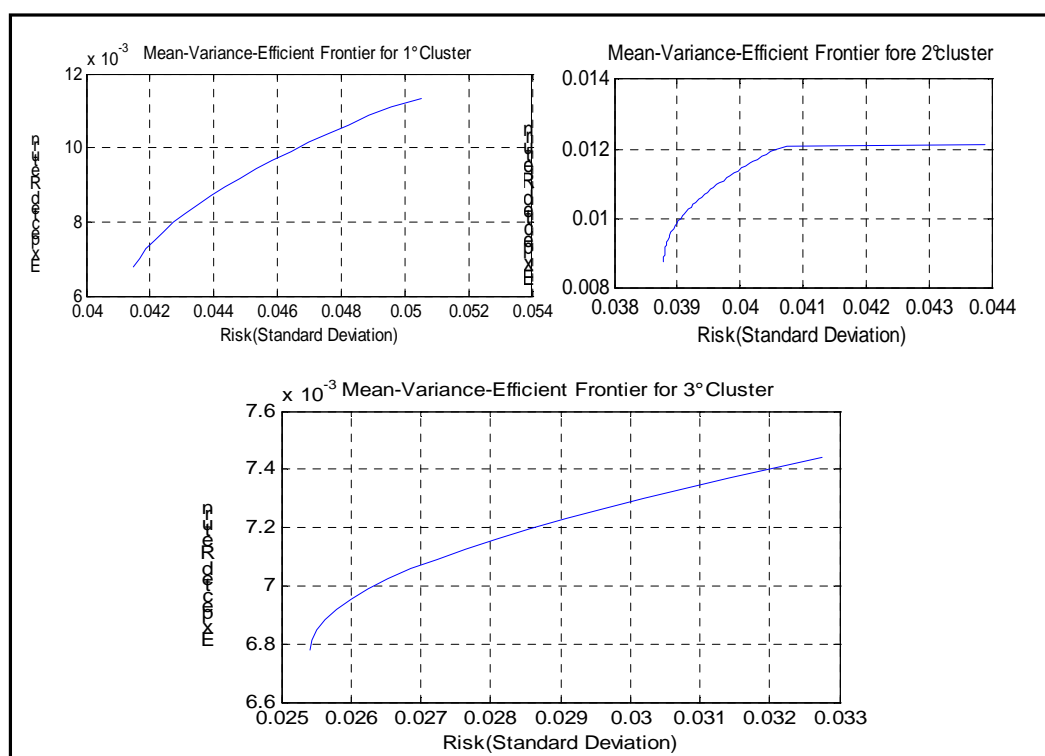


Figura 4.4.5: Confronto delle frontiere efficienti per i tre clusters sui fondi comuni

Le frontiere efficienti (vedi Figura 4.4.5), costruite a partire da gruppi ottenuti mediante il clustering, sono informative della differenza in media e varianza tra i tre clusters. Il secondo cluster si distingue per l'alto rendimento e per una varianza contenuta, quindi è anche ciò a cui ogni gestore mira. Il grafico della EF è informativo del fatto che, la maggior parte dei fondi inclusi in questo cluster ha medie che tendono a stabilizzarsi verso un certo valor medio (0.012), però si distinguono per la loro varianza; sarebbe dunque opportuno filtrare parte dei fondi

che offrono rendimenti con la stessa media ma che dall'altra parte hanno una varianza molto più elevata. Le altre due EF sono informative del fatto che il primo e il terzo cluster sono formati in modo tale che ci sia all'aumentare della media dei rendimenti anche un aumento della varianza. Il programma non ci consente di avere un confronto grafico contemporaneamente per tutte e tre le EF, però l'intuizione ci porta a pensare che tra la EF del 1° e del 2° cluster ci sia un'intersezione, proprio in riferimento al punto definito dai due fondi sopraccitati (GTSBX e PMCIX), indice questo di somiglianza tra i due e della possibilità di combinare tali portafogli per ottenerne uno più efficiente.

In seguito a questi indici di variabilità, consideriamo adesso gli effetti sulla performance dei 50-fondi comuni d'investimento statunitensi in riferimento agli indici: *Alpha* che misura la variazione (maggiore o minore) della performance dei vari fondi rispetto al rispettivo valore atteso, calcolati in funzione del rischio assunto, cioè *Beta* (vedi Figura 4.4.6 e Figura 4.4.7).

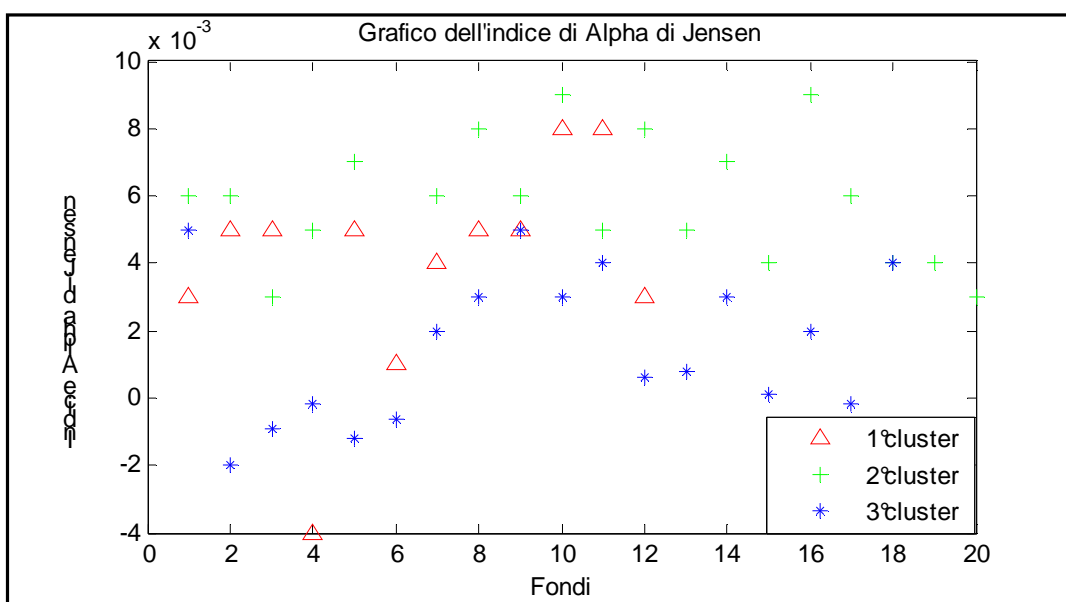


Figura 4.4.6: Confronto tra i tre clusters sui rendimenti mensili dei fondi comuni d'investimento statunitensi rispetto all'indice Alpha di Jensen.

Osservando i valori concentrati nella parte superiore del grafico, appartenenti al secondo cluster (forma '+' e in verde); si può affermare che i diversi fondi comuni di capitalizzazione *Small* hanno valori di alpha decisamente elevati e fra i tre clusters sono quelli che garantiscono maggior rendimento, quindi di nuovo tale cluster si distingue per la migliore performance. Tale indice è indicatore

dell'abilità del gestore nello *stock picking* e quindi della sua capacità di creare valore aggiunto. Mentre, il primo cluster composto da fondi Small G. & Small B. (*forma a triangolo e in rosso*) è concentrato nella parte centrale del grafico, *tranne per un valore anomalo individuato fin dall'inizio dell'analisi di performance dei clusters, si tratta del quarto fondo del 1° cluster 'ASMGX'* il quale ha un valore di alpha fortemente negativo indice della scarsa abilità del fondo di creare valore aggiunto nel tempo e questo è confermato dal suo indice di Morningstar rating che è pari a uno. Dall'altra parte notiamo un'altra anomalia, definita dalla coincidenza tra il nono fondo del primo (*DFTSX-SB-4*) e del secondo (*RCBSX-SV-2*) cluster, tali fondi si ritrovano ad avere stesso indice alpha pari a 5×10^{-3} e sono propensi a creare valore aggiunto nel tempo anche se hanno una misura di Morningstar rating diversa, indice questo del loro diverso posizionamento nel mercato. La coincidenza tra questi due fondi è informativa del fatto che il fondo *RCBSX* appartenente al secondo cluster molto probabilmente seguirà il comportamento del fondo *DFTSX*, che per il mercato è ben definito e premiato con 4 stelle di rating, sia per il rendimento corretto per il rischio assunto dal gestore, sia per i costi sostenuti dall'investitore e sia per la sua capitalizzazione. Ricordiamo che, a scapito di un determinato rischio si ha la possibilità di creare extrarendimenti, o meglio valore aggiunto, se si detiene un portafoglio costruito come il secondo cluster. Il terzo e il primo cluster si trovano nella parte più bassa e quella intermedia del grafico e questo posizionamento rispetto al valore assunto dell'indice alpha non è sempre positivo, in quanto conferma la scarsa abilità per tutti i fondi inclusi in questi due clusters di creare extrarendimento nel tempo. Il gestore in base a questo grafico di certo sceglierebbe di costruire un portafoglio che assomigli il più possibile al secondo cluster, alpha raggiunge valori elevati, quindi si includono i fondi che assicurano un maggior rendimento per il portafoglio gestito rispetto a quello di equilibrio (come costruito dal Capital Asset Pricing Model, CAPM) e di un portafoglio non gestito (quello di mercato) avente lo stesso Beta. Da un'analisi dei valori di Beta per ogni cluster (*vedi Figura 4.4.7*), possiamo osservare la sensibilità dei fondi inclusi in un determinato cluster rispetto alle oscillazioni del mercato, o meglio possiamo stabilire quanto variano i fondi congiuntamente al mercato. Notiamo, inoltre, che in seguito ai movimenti del mercato nascono oscillazioni incontrollabili da parte dei gestori attorno al

portafoglio ed è proprio questo fenomeno che ci interessa valutare. Il rischio sistematico⁵² è attribuibile prevalentemente a fattori macroeconomici, politici ecc., e rispetto clustering si ha che: il primo cluster (*forma triangolare e in rosso*) assume i valori alti di tale indice ed è poi seguito dal terzo e infine dal secondo cluster. La predisposizione dei clusters è in ordine decrescente per i valori di Beta assunti e questo è un ottimo indicatore delle oscillazioni a cui sono soggetti i fondi comuni per clusters rispetto al mercato.

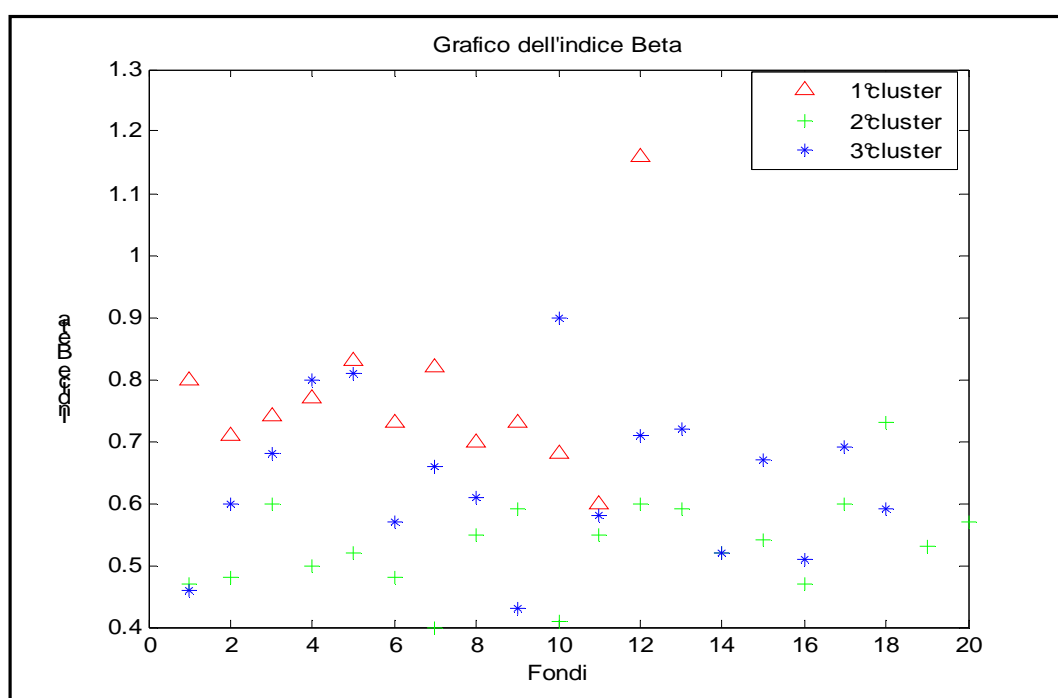


Figura 4.4.7: Confronto dei tre clusters di fondi comuni rispetto all'indice Beta

Il primo cluster⁵³ si distingue per un comportamento *leggermente più aggressivo* rispetto agli altri clusters, in quanto assume valori dell'indice Beta variabili tra [0.6;1.2], il minimo e il massimo sono oltre tutto assunti dagli ultimi due fondi appartenenti al cluster i quali sono rispettivamente BOGIX (Small Blend) e CHCYX (Mid-Cap-Growth), distinguibili l'uno dall'altro sia per la capitalizzazione, sia per stile di gestione e anche per la misura di Morningstar rating. Di conseguenza, rispetto alle caratteristiche degli altri fondi interi al cluster per valori di Beta assunti, possiamo definire l'ultimo fondo localizzato in alto

⁵² espresso mediante Beta;

⁵³ forma a triangolo e in rosso;

come un dato anomalo non soltanto in riferimento al primo cluster ma, rispetto all'intero dataset.

Il secondo cluster⁵⁴ invece, è dominato da valori dell'indice Beta limitati tra [0.4;0.7], indice questo di un comportamento *difensivo* rispetto al mercato. Tale informazione è estremamente chiarificatrice per il portafoglio definito dal secondo cluster, in quanto ogni gestore tranquillamente può dare valutazioni più o meno precise sulla performance di questo portafoglio, in quanto definito da fondi comuni con capitalizzazione *Small*⁵⁵ (Growth, Value, Blend) e 2 *fondi Mid-Cap-Growth*, e inoltre, caratterizzati da una misura del Morningstar rating compresa tra 2, 3 e 4; indice questa di una buona valutazione del rendimento corretto per il rischio assunto dal gestore e i costi sostenuti dall'investitore per tali fondi. Mentre, il terzo cluster⁵⁶ è caratterizzato (precedentemente abbiamo visto che sono inclusi fondi con media bassa e variabilità bassa) da valori dell'indice Beta compresi tra [0.4; 0.9], indice questo di un comportamento molto variabile e di conseguenza, *misto tra il difensivo e il neutrale*, quindi da valutare con molta attenzione da parte del gestore e dell'investitore. In poche parole, l'informazione ottenuta dal clustering sia per l'indice Alpha e sia per l'indice Beta è piuttosto informativa intorno alla performance del portafoglio, in quanto ci porta a concludere, che ogni cluster è distinto prevalentemente sia per la capitalizzazione e sia per l'indice di Morningstar rating.

Le analisi svolte precedentemente ci hanno aiutato a raccogliere informazioni rilevanti a posteriori⁵⁷ del clustering, sia intorno alle caratteristiche dei gruppi formati e sia intorno alla performance del portafoglio costruito dai fondi comuni appartenenti al secondo cluster. In particolare, valutiamo adesso il trade-off tra Alpha e Beta, confronto molto utile, però vediamo se nel nostro caso anche informativo, per avere una visione più completa attorno ai clusters ottenuti e alla performance del portafoglio.

Graficamente⁵⁸ si osserva, che il mix migliore (*significa generare extrarendimenti a basso rischio sistematico*) tra gli indici Alpha e Beta, è definito dal secondo

⁵⁴ forma '+' e in verde;

⁵⁵ indipendentemente dallo stile di gestione;

⁵⁶ forma '*' e in blu;

⁵⁷ fase successiva al clustering;

⁵⁸ vedi Figura 4.4.8;

cluster. Infatti, tale cluster è posizionato in alto, quindi distinto per valori alti di Alpha, e a sinistra del grafico, cioè caratterizzato da valori bassi di Beta. Tutto questo significa, che il cluster lo si può definire come quello ottimale per quanto riguarda la performance raggiungibile, in quanto gli effetti sulla performance sono tra i migliori, perché si riesce ad ottenere maggior valore aggiunto a scapito di un rischio sistematico relativamente basso.

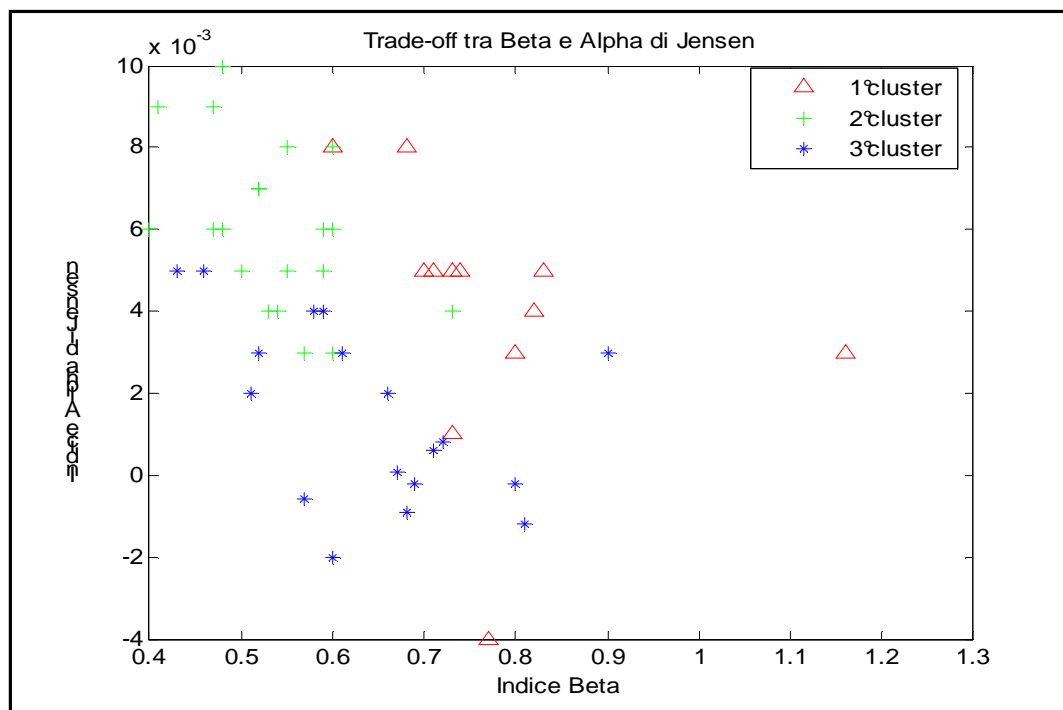


Figura 4.4.8: Confronto dei tre clusters sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi, in riferimento al trade-off tra Alpha e Beta.

Per quanto riguarda il primo (*forma a triangolo e in rosso*) e il terzo (*forma a stella e in blu*) cluster, possiamo annotare che anche essi non si comportano male, anzi il clustering ci permette di ottenere raggruppamenti più selettivi sui fondi comuni. Infatti, nel primo cluster vengono raccolti tutti quei fondi che hanno un extrarendimento positivo e rischio sistematico discreto, tranne per due dati anomali, uno caratterizzato da un rischio sistematico alto⁵⁹ e extrarendimento discreto pari a 3×10^{-3} , mentre l'altro⁶⁰ raggiunge un extrarendimento fortemente negativo -4×10^{-3} e rischio sistematico decisamente alto, entrambi da considerarsi come due fondi a cui il gestore deve fare molta attenzione in quanto

⁵⁹ fondo CHCYX della categoria Mid-Cap-Growth;

⁶⁰ fondo ASMGX della categoria Small Growth;

possono comportare delle perdite ed essere troppo rischiosi. Comunque, un bravo gestore se interpreta bene questa informazione può decisamente decidere di includere anche il primo cluster nel proprio portafoglio, in quanto diventa molto simile al secondo cluster, questo è verificabile graficamente dalla sovrapposizione di due fondi (1°C: *DFTSX-SB-4* e 2°C: *RCBSX-SV-2*). Si può affermare che sia nel primo e anche nel secondo cluster vengono inclusi prevalentemente fondi con misura di Morningstar rating elevata, compresa tra 3-4 (tralasciando i valori anomali) e il gestore può usare al meglio tale informazione per poter costruire un portafoglio più efficiente composto da questi fondi. Il terzo cluster è invece, caratterizzato da fondi abbastanza eterogenei sia per lo stile di gestione, sia per la capitalizzazione e anche per la misura di Morningstar rating, visibile questo graficamente dalla dispersione dei dati. Di conseguenza, possiamo definire tale cluster come un portafoglio composto da fondi che non garantiscono una buona performance, in quanto i possibili profitti sono compensati da probabili perdite dovute alla presenza di fondi altamente rischiosi e altri che non riescono a raggiungere nessun tipo di extrarendimento, informazione questa non deducibile a priori. A posteriori sembra che la valutazione della performance sia condizionata dalla misura di Morningstar rating e dalla capitalizzazione, mentre lo stile di gestione è irrilevante, in quanto rappresenta una caratteristica non significativa per il clustering.

Oltre a valutare le misure basate sulla media - varianza dei clusters, possiamo considerare anche gli effetti sugli altri indici di performance e valutare se eventualmente si ha informazione aggiuntiva derivato dal clustering.

Partiamo dall'indice di Sharpe (*vedi Figura 4.4.9*), graficamente osserviamo che i tre clusters rispetto a tale indice sono abbastanza eterogenei, in quanto assumono decisamente valori diversi per cluster e quello che si distingue per i valori più elevati di Sharpe è il secondo cluster. Si deduce che il secondo cluster ha un comportamento migliore rispetto agli altri, cioè il trade-off tra rendimento e rischio ottenuto⁶¹ è tra i più alti, indice questo di opportunità di guadagno. Fin adesso, tutte le analisi svolte hanno portato a definire il secondo cluster come il migliore sul mercato, anche se sia il primo e il terzo cluster non sono del tutto pessimi basta soltanto manipolarli con la giusta cautela. Ad esempio: se eliminiamo i fondi che ci

⁶¹ analisi confermata dai precedenti risultati, Figura 4.4.8;

risultano anomali, come il quarto fondo del primo cluster (ASMGX), allora possiamo osservare che tutti i cluster vengono definiti in modo da trovarsi sopra lo zero (vedi Figura 4.4.9) ed avere valori dell'indice di Sharpe distinti per clusters, che fondamentalmente si differenziano in base al tipo di capitalizzazione e alla misura di Morningstar rating dei fondi d'investimento inclusi in ogni cluster. L'effetto principale che ottengo dal raggruppamento è una netta divisione dei cluster per diverso indice di Sharpe confermando, per il primo cluster valori intermedi e per il terzo valori bassi. Il gestore, a posteriori del clustering, può dedurre che i fondi di capitalizzazione *Small* e misura MR⁶² compresa tra 2-4, sono tra i migliori⁶³ nel mercato perché garantiscono un trade-off tra rendimento e rischio elevati. Informazione questa rilevate nella valutazione della performance del nostro portafoglio in analisi, in quanto ci permette di effettuare una selezione più efficiente sui fondi da includere e quelli che necessitano di maggior attenzione nella gestione.

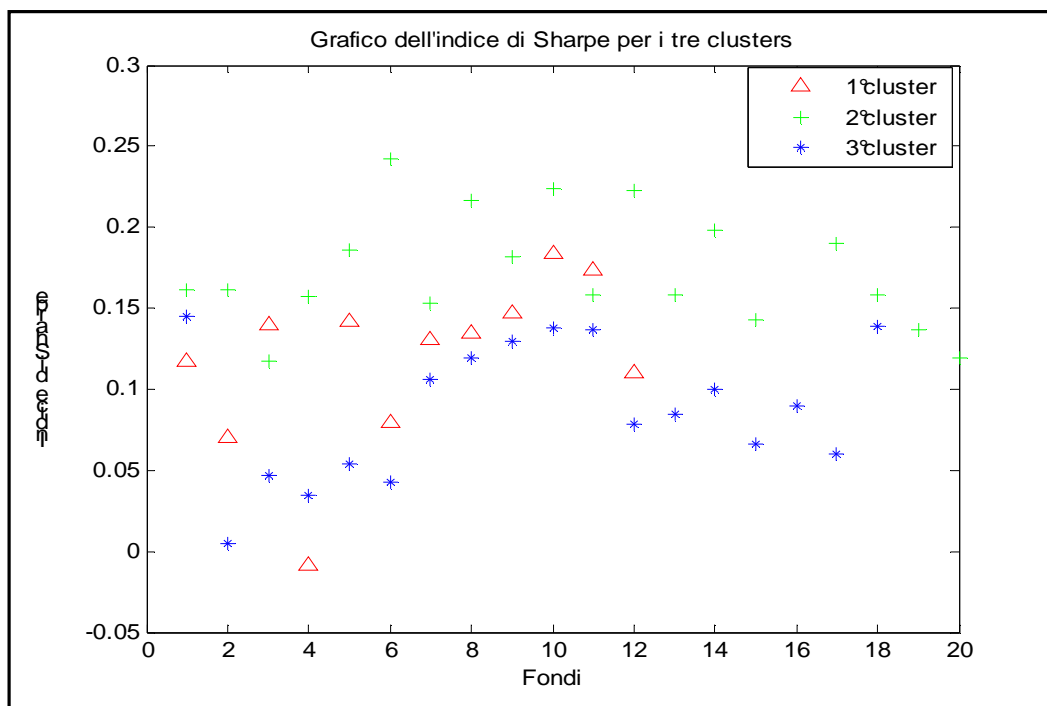


Figura 4.4.9: Rappresentazione grafica degli indici di Sharpe per i tre clusters, basati sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi.

⁶² Morningstar rating;

⁶³ definiscono il primo e il secondo cluster;

Un altro indice di performance utile da valutare è l'indice di Treynor (vedi Figura 4.4.10). Il ragionamento che si potrebbe fare sui tre clusters è molto simile a quello considerato per l'indice di Sharpe. Il cluster che garantisce rendimenti in eccesso è dato dal secondo, in quanto è posizionato nella fascia più alta del grafico.

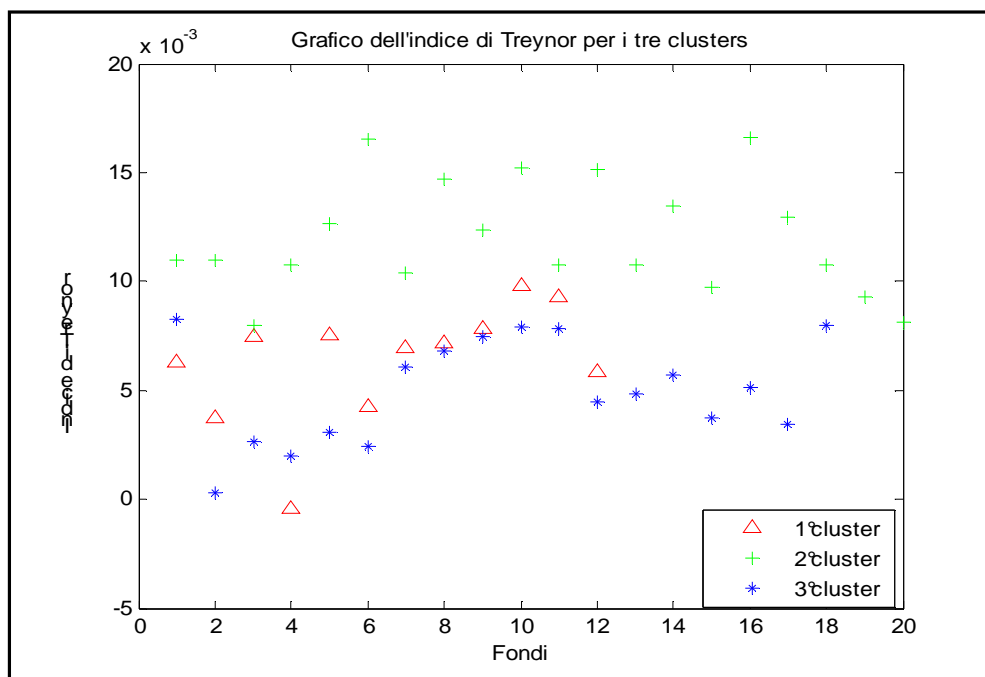


Figura 4.4.10: Rappresentazione grafica dell'indice di Treynor per i tre clusters, basati sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi

I portafogli ottenibili a posteriori sono formati da fondi comuni d'investimento distinti per la capitalizzazione e la misura di MR, si riesce così a generare valore aggiunto, soprattutto se l'informazione derivata dal clustering viene correttamente interpretata. Gli effetti del clustering sono rilevanti e informativi in riferimento alla performance dei fondi in quanto, se si ha la possibilità di gestire solo fondi raggruppati dal primo e dal secondo cluster questi di certo saranno i più vincenti, in quanto garantiscono valore aggiunto. Mentre, il terzo cluster ci informa della possibilità di non riuscire a raggiungere nel tempo valore aggiunto significativo, in quanto l'Alpha dei fondi comuni in esso inclusi è basso (significativamente non diversi da 0) però, i Beta sono alti, indice questo di un comportamento dei fondi molto sensibile rispetto alle oscillazioni del mercato.

Un confronto (vedi Figura 4.4.11), altrettanto informativo, deriva dalla rappresentazione grafica dell'indice di Sharpe e quello di Treynor. Si osserva

immediatamente una particolarità legata all'andamento dei fondi appartenenti al secondo cluster, o meglio tali fondi comuni sono predisposti più in alto rispetto agli altri fondi inclusi rispettivamente nel primo e nel terzo cluster. Decisamente il secondo cluster è quello più efficiente, in quanto garantisce il miglior trade-off e crea valore aggiunto nel tempo a scapito di un rischio moderato.

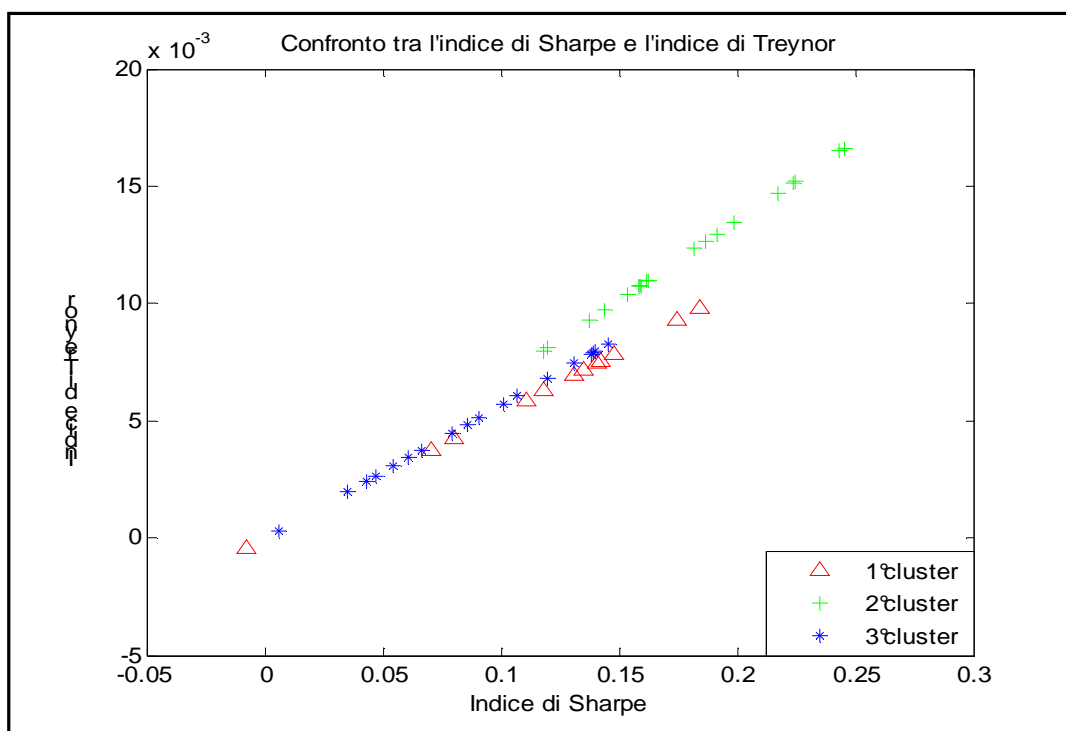


Figura 4.4.11: Confronto tra l'indice di Treynor e di Sharpe per i tre clusters sui rendimenti mensili dei 50 fondi comuni d'investimento statunitensi.

Di conseguenza si ha che, sono prevalentemente il primo e il secondo cluster i principali gruppi contenenti fondi comuni in grado di generare effetti positivi sulla performance, quindi a posteriori otteniamo informazioni rilevanti e utili per poter gestire e valutare al meglio la performance del portafoglio, anche se a posteriori i fondi sono classificati in modo diverso dalla a priori, però di certo questo è un vantaggio sia per il gestore e sia per l'investitore, in quanto permette di valutare in modo più efficiente la performance del portafoglio.

4.4.1 VALUTAZIONI DAL CLUSTERING SUGLI INDICI

In riferimento agli indici di performance calcolati nei paragrafi precedenti, abbiamo valutato un clustering anche su di essi, il quale ci porta alle stesse conclusioni sopra discusse, il dendrogramma è rappresentato nella *Figura 4.4.12*.

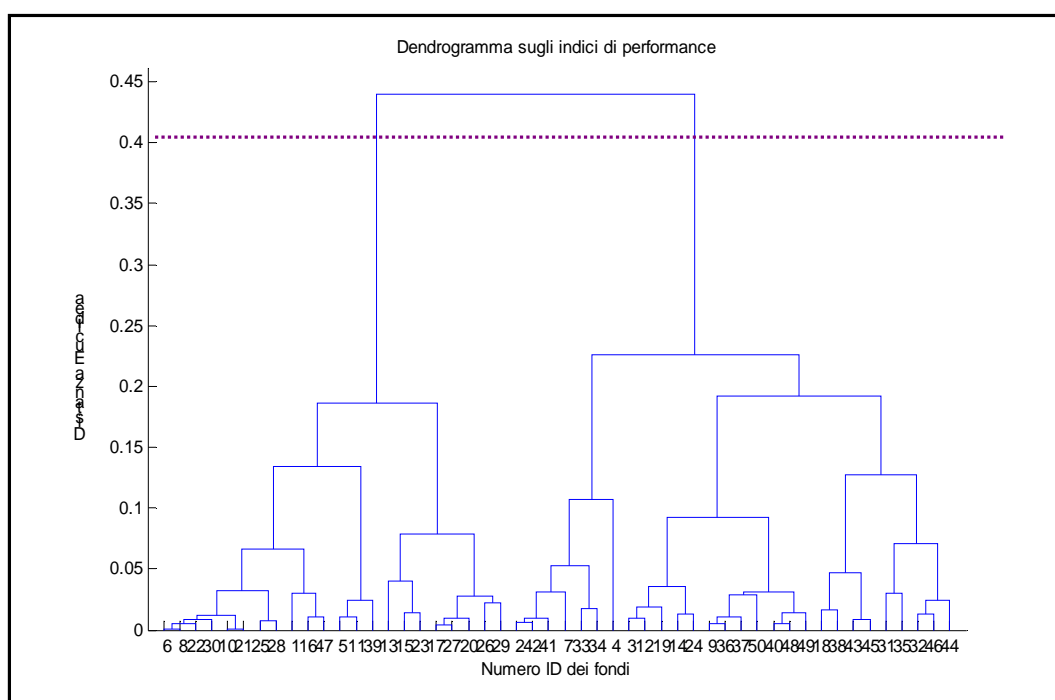


Figura 4.4.12: Dendrogramma sui quattro principali indici di performance, per un valore della distanza euclidea pari a 0.4 si individuano due clusters.

In seguito alle varie analisi e controlli di bontà sui clusters, sia il metodo VCR e MC confermano $g = 2$, i clusters sono due e sono composti nel seguente modo:

- 1° cluster: {1, 5, 6, 8, 10, 11, 13, 15-17, 20-23, 25-30, 39, 47} si ha praticamente l'unificazione dei primi due clusters determinati precedentemente in base al clustering sui rendimenti mensili dei 50 fondi comuni d'investimento. In esso sono inclusi quasi tutti i fondi della categoria Small, tranne per gli ultimi due fondi che presentano soltanto una forte assomiglianza rispetto ad essi anche se derivano da Large Value e Mid-Cap-Growth.
- 2° cluster: {2-4, 7, 9, 12, 14, 18, 19, 24, 31-38, 40-46, 48-50} è definito principalmente dai fondi Large, questo coincide con il terzo cluster analizzato nei paragrafi precedenti. Di nuovo l'informazione a priori viene elaborata nel modo più efficiente e si ottiene in questo modo una valutazione più accurata della performance del portafoglio.

CONCLUSIONI

La valutazione della performance dei fondi comuni d'investimento ha richiesto l'analisi congiunta di due elementi: *il rendimento* ottenuto e *il rischio* che l'investimento comporta. L'analisi è stata condotta in un'ottica bidimensionale poiché tra rendimento e rischio esiste una relazione diretta, vale a dire, il gestore può ottenere rendimenti via via più elevati incrementando progressivamente il livello di rischio del portafoglio gestito, ma aumentando dall'altra parte anche la possibilità di incorrere in perdite sempre più consistenti. Di conseguenza emerge un trade-off tra queste due misure, dato che il rendimento rappresenta una componente che il risparmiatore cerca di massimizzare mentre il rischio è invece, un elemento che gli agenti cercano di minimizzare.

Gli strumenti utilizzati in questa tesi poggiano ad un portafoglio composto da 50 fondi comuni d'investimento statunitensi osservati per un periodo di 5 anni e per la valutazione della performance sono stati calcolati gli indici di RAP, non conoscendo i pesi da applicare agli assets. Il portafoglio è ben diversificato in base alle sette categorie da cui i fondi hanno origine, cioè per capitalizzazione, per stile di gestione e anche per Morningstar rating. A priori in base all'informazione a disposizione sono state calcolate le regressioni lineari definite dal CAPM e in seguito gli indici di performance, i quali mostrano una certa variabilità tra le varie categorie; gli alpha⁶⁴ sono piccoli, prossimi a zero, se nella maggior parte dei casi risultano significativi per una piccola parte sono non significativi, indice questo di una scarsa abilità da parte del gestore di poter creare valore aggiunto. Si distinguono per alpha alti soltanto le categorie Small Value e Small Blend, i loro Alpha sono compresi tra [0.4;0.9]%, mentre i meno significativi sono quelli della

⁶⁴ Per il CAPM sono non significativamente diversi da zero;

categoria Large Growth e Large Blend. Mentre, il calcolo dell'indice Beta porta a dei valori abbastanza elevati per le categorie Small Blend, Large Value e Mid-Cap-Growth, indicatore del fatto che i fondi appartenenti a queste categorie sono abbastanza sensibili alle oscillazioni del mercato, in quanto hanno un comportamento neutrale o aggressivo. Notiamo che nella maggior parte dei casi il comportamento è neutrale, tranne per il fondo 'CHCYX' (Mid-Cap-Growth), che risulta aggressivo rispetto al mercato: in condizioni favorevoli seguirà un rialzo mentre in condizioni opposte un ribasso veloce. Infine, è stato calcolato l'indice di Sharpe per il portafoglio, soltanto però dopo aver valutato la FE, pari a 0.1206. Tale valore è indice della performance storica corretta per il rischio, cioè indica che l'efficienza della gestione è scarsa, in quanto questo valore rappresenta il rendimento registrato dal portafoglio rispetto all'asset Risk-free. Per quanto riguarda i singoli assets, questi mostrano indici di Sharpe in linea con quello del portafoglio⁶⁵. Tutti gli R^2 , indice che riflette la percentuale di oscillazioni del fondo riconducibili alle oscillazioni nel benchmark del fondo (o asset), nel nostro caso per tutti gli assets risultano bassi, o meglio i valori sono compresi tra [0.14; 0.52]. Questo porta ad affermare che, solo in parte la variabilità degli assets in analisi è dovuta alle oscillazioni dei propri benchmark, il resto è dovuto ai fattori macroeconomici e politici.

Alla luce di quanto ottenuto dall'analisi basata sul CAPM è stato valutato un clustering sui rendimenti dei 50 fondi comuni che costituiscono il portafoglio, in modo da analizzare attentamente come sono formati i clusters e quali effetti hanno generato attorno alla valutazione della performance. La cluster analysis sia rispetto ad un'analisi esplorativa e in seguito ad un'analisi più approfondita mediante il metodo VCR e MC, ci porta ad individuare tre clusters, i quali vengono confermati anche mediante il metodo delle k-means:

- *il primo cluster*: include fondi della classe Small Growth e Small Blend, soltanto un fondo appartiene al gruppo Mid-Cap-Growth, in particolare i numeri identificativi sono: {[1:5],7,11,22,25,27,29,47};
- *il secondo cluster*: comprende fondi derivati dalla classe Small Growth, Small Value, Small Blend e soltanto due fondi appartenenti al gruppo Mid-Cap-Growth, i numeri identificativi sono: {6,8,9,10,[12:17],[19:21],

⁶⁵ trovato dal CAPM;

23,24,26,28,30,48,50};

- *il terzo cluster*: include prevalentemente fondi della classe Large Growth, Large Value, Large Blend, un soltanto fondo Small Value e un altro Mid-Cap-Growth, i numeri identificativi sono: {18,[31:46],49}.

In riferimento ad essi è stata svolta un'analisi accurata sul comportamento degli indici di performance. Infatti, i tre clusters non rispecchiano più la suddivisione iniziale, ma seguono una classificazione basata soprattutto sulla *capitalizzazione e la misura di Morningstar rating*. Il primo e il secondo cluster si distinguono per avere medie e varianze alte, mentre il terzo cluster ha una media e varianza bassa; quindi si va a confermare il teorema di Markowitz, o meglio dicasi, i cluster con media interna alta hanno anche varianza elevata, quindi a scapito di rendimenti alti si va incontro anche a un rischio elevato e questo è ben visibile dai confronti grafici (massimizzare i rendimenti significa andare incontro a rischi maggiori). I confronti grafici tra le frontiere efficienti indicano che si ha un'intersezione (non mostrata graficamente perché MatLab 7.1, non prevede più di un input per le FE), tra il 1° e 2° cluster, indice questo del fatto che se combinati potrebbero dare origine a un portafoglio più efficiente. Infatti, entrambi questi clusters hanno indici di Alpha elevati, maggiori di zero (tranne per un asset del primo cluster, che potrebbe essere definito un dato anomalo), mentre il terzo raggruppa tutti i fondi con Alpha bassi. Tale informazione aggiuntiva deriva dal clustering, il quale ci permette di poter classificare i 50 fondi comuni provenienti da 7 categorie diverse, in soli tre gruppi distinti rispetto alla capitalizzazione e alla misura di Morningstar rating, ed ottenere valutazioni più accurate sulla performance del portafoglio. Si deduce che i fondi di capitalizzazione Small e misura di Morningstar rating compresa tra il 2-4, risultano avere un comportamento ottimale, in quanto si riesce ad ottenere un portafoglio più efficiente. Sono i Large, raggruppati nel terzo cluster ad essere distinti per la loro scarsa efficienza. Quindi, se a priori avevamo sette gruppi distinti per capitalizzazione, stile di gestione e Morningstar rating, il clustering ci consente di compattare l'informazione e di individuare soltanto tre clusters, formati da fondi che hanno medie e varianze dei rendimenti simili, e lo stesso dicasi per gli indici di Sharpe e Treynor, tutto questo ci porta a ritenere che l'applicazione del metodo di Sharpe, Brown e Goetzman sia veramente efficiente per l'analisi dei fondi comuni. L'informazione che si riesce ad ottenere è rilevante

per la valutazione della performance del portafoglio, in quanto ci consente di capire quali tra gli assets sono quelli meno efficienti, o meglio sia l'investitore e sia il gestore possono evitare di considerare fondi non remunerativi nel tempo (Alpha basso, indice di Sharpe basso...) in quanto, mediante un clustering si riescono ad individuare i gruppi di fondi che mostrano effetti significativi sulla valutazione della performance. Riducendo i gruppi da sette (a priori) e soli tre (a posteriori) tramite il metodo VCR, si riesce ad ottenere indici di performance simili al interno dei clusters e distinti tra loro. Si può affermare che i clusters sono prevalentemente distinti per capitalizzazione e Morningstar rating e i fondi migliori per un investitore sarebbero quelli inclusi nel 1° e soprattutto quelli del 2° cluster. Nel primo rientrano i fondi di categoria Small Growth e Small Blend (tranne per 1 fondo Mid-Cap-Growth, che si può ritenere sia dovuto a un errore dell'algoritmo di clustering) e nel secondo sono inclusi tutti gli Small (Growth, Value, Blend) e 2 Mid-Cap-Growth. Il clustering riesce a valutare in modo più appropriato il comportamento eterogeneo che avevano a priori i fondi comuni Mid-Cap-Growth, suddividendoli tra i tre clusters. Se gli indici di performance a priori mostravano per i Large valori abbastanza significativi, a posteriori questi vengono raggruppati in un unico cluster, il terzo, che però evidenzia indici di performance prevalentemente scarsi.

Si conclude che, applicare l'algoritmo di clustering sui rendimenti dei fondi comuni d'investimento statunitensi, alla base abbiamo serie storiche dei prezzi, oppure sugli indici di performance, si ottengono informazioni aggiuntive circa equivalenti. Se il clustering sui rendimenti mensili ci porta ad avere tre clusters, quelli sugli indici (α , Sharpe, Treynor, Range e SO) sono soltanto due, in quanto l'informazione viene meglio elaborata in modo tale da unificare i primi due clusters più simili. Si hanno, dunque, effetti positivi sulla valutazione della performance dei fondi in seguito al clustering, perché l'investitore o il gestore sono maggiormente informati sui fondi comuni d'investimento da includere nel portafoglio e di scegliere la composizione migliore e anche la più appropriata alle loro esigenze.

APPENDICE

• COMANDI DI MATLAB (VERSION 7.1)

```
plot(Pr)
RetSeries = price2ret(Pr,[],'Periodic')
plot(RetSeries)
boxplot(RetSeries)
autocorr(RetSeries(1:59,1))
parcorr(RetSeries(1:59,1))
[H,pValue,Qstat,CriticalValue] = lbqtest(RetSeries(1:59,1),25,0.05)
plot(sp1500(1:60,1))
autocorr(sp1500(1:60,1))
parcorr(sp1500(1:60,1))
R_M = price2ret(sp1500(1:60,1),[],'Periodic');
[H,pValue,Qstat,CriticalValue] = lbqtest(R_M(1:59,1),25,0.05)
plot(Rf(1:60,1))
autocorr(Rf(1:60,1))
parcorr(Rf(1:60,1))
Rfree= price2ret(Rf(1:60,1),[],'Periodic')
R_fm=Rfree/12
[H,pValue,Qstat,CriticalValue] = lbqtest(R_f(1:59,1),25,0.05)
normplot(RetSeries)
qqplot(RetSeries)
1) [ p,h ] = signtest(...) returns the result of the hypothesis test, performed at the
5% significance level, in h.
[p,h] = signtest(RetSeries(1:59,1))
2) h = tttest(x) performs a t-test of the hypothesis that the data in the vector x
comes from a distribution with mean zero. The data are assumed to come from a
normal distribution with unknown variance.
[h,p,ci,stats] = tttest(RetSeries(1:59,1))
3) H = jbtest(X) performs the Jarque-Bera test on the input data vector X and
returns JB, the result of the hypothesis test.
[H,P,JBSTAT,CV] = jbtest(RetSeries(1:59,1),0.05)
[H,P,JBSTAT,CV] = jbtest(z_i(1:59,1),0.05)
4) H = lillietest(X) performs the Lilliefors test on the input data vector X and
returns H, the result of the hypothesis test.
[H,P,LSTAT,CV] = lillietest(RetSeries(1:59,1))
I=ones(50,1)
z_i = RetSeries - R_fm*I' ---> Extrarendimenti
z_im = R_M - R_fm
autocorr(az(1:59,1))      dove az = abs(z_i)
autocorr(qz(1:59,1))     dove qz = z_i^2
Mz_i = mean(RetSeries) - mean(R_fm)*I' -----> Media extrarendimenti
```

```

Mz_Mi = mean(R_M) - mean(R_fm) -----> Media extrarendimento del benchmark
rispetto all'asset risk-free.
[ExpReturn, ExpCovariance, NumEffObs] = ewstats(RetSeries)
Mu=ExpReturn-mean(R_fm)' = Mz_i
PortValue=1
NumAssets = 50
[A,b] = pcpval(PortValue, NumAssets)
[ExpSigma, ExpCorrC] = cov2corr(ExpCovariance)
RisklessRate = 0.0168/12
NumPorts = 50
RiskAversion = 3
[PortRisk, PortReturn, PortWts]=portopt(Mu, ExpCovariance, NumPorts)
portopt(Mu, ExpCovariance, NumPorts)
portalloc (PortRisk*100, PortReturn*100, PortWts, RisklessRate*100,
RiskAversion)
[RiskyRisk, RiskyReturn, RiskyWts, RiskyFraction, OverallRisk,
OverallReturn]=portalloc(PortRisk*100, PortReturn*100, PortWts,
RisklessRate*100, RiskAversion)
I=ones(50,1)
z_M =R_M - R_fm
z_i = RetSeries - R_fm*I'
• Tracking Error FE
NumAssets=50
Index = ones(NumAssets, 1)/NumAssets
AbsConSet = portcons('PortValue', 1, NumAssets, 'AssetLims',
zeros(NumAssets,1), ones(NumAssets,1))
ActiveConSet = abs2active(AbsConSet, Index)
[AbsConSet(:,end) ActiveConSet(:,end)]
[ActiveRisk, ActiveReturn, ActiveWeights] =portopt(Mu,ExpCovariance, 21, [],
ActiveConSet)
ActiveRisk = real(ActiveRisk)
plot(ActiveRisk*100, ActiveReturn*100, 'red')
grid('on')
xlabel('Active Risk (Standard Deviation in Percent)')
ylabel('Active Return (Percent)'); title('Tracking Error Efficient Frontier')
TM2 = mean(R_M) - mean(R_fm)
Sh=Mz_i'/ExpSigma'
T=Mz_i'/Beta
• Vettore riga di dimensione (50 - 1)*50 / 2 = 49*25 = 1225 colonne, Y vettore
(1x1225) contenente le distanze Euclidee dei rendimenti;
Y=pdist(RetSeries') oppure YI=pdist(indici)
• Matrice dissimilarità
SQ = squareform(Y)
Z = linkage(Y,'centroid') anche Z = linkage(Y,'ward')
c = cophenet(Z,Y)
[c,D] = cophenet(Z,Y)
r = corr(Y',D','type','spearman')
• Dataset completo
[H,T] = dendrogram(Z,'colorthreshold')

```

```

[H,T] = dendrogram(Z,'colorthreshold','default')
T = cluster(Z,'maxclust',3)
find(T==i) ----> dove i = 1,2,3
stem(T,'DisplayName','T','YDataSource','T')
scatter3(Z(:,1),Z(:,2),Z(:,3),59,T,'filled')
Z = linkage(Y,'centroid')
Z = linkage(Y,'ward')
[IDX,C,sumd] = kmeans(RetSeries',g) ----> dove g è il numero dei gruppi
[s,h] = silhouette(RetSeries',IDX,'euclidean')
plot((1:12),mean(firstc),'*r')
plot((1:20),mean(secondc),'+b')
plot((1:18),mean(thirdc),'*g')
plot((1:12),mean(firstc),'*r',(13:32),mean(secondc),'+g',(33:50),mean(thirdc),'*b')
plot(var(firstc),mean(firstc),'*r',var(secondc),mean(secondc),'+g',var(thirdc),mean(thirdc),'*b')
plot(var(firstc),(1:12),'*r',var(secondc),(1:20),'+g',var(thirdc),(1:18),'*b')
[ExpReturn1, ExpCovariance1, NumEffObs1] = ewstats(firstc)
[ExpSigma1, ExpCorrC1] = cov2corr(ExpCovariance1)
Mu1=ExpReturn1-mean(R_fm)'....
portopt(Mu1, ExpCovariance1, NumPorts)...
plot((1:12),Sh1(1:12,7),'^r',(1:20),Sh2(1:20,8),'+g',(1:18),Sh3(1:18,10),'*b')
plot((1:12),T1(1:12,12),'^r',(1:20),T2(1:20,18),'+g',(1:18),T3(1:18,10),'*b')
plot((1:12),alpha1,'^r',(1:20),alpha2,'+g',(1:18),alpha3,'*b')
plot((1:12),Beta1,'^r',(1:20),Beta2,'+g',(1:18),Beta3,'*b')
plot(Beta1,alpha1,'^r',Beta2,alpha2,'+g',Beta3,alpha3,'*b')
plot(SH1(1:12,12),T1(1:12,7),'^r',Sh2(1:20,18),T2(1:20,8),'+g',SH3(1:18,10),T3(1:18,10),'*b')

```

• TABELLE E GRAFICI

Tabella A.1: Composizione e profilo dei 50 fondi comuni d'investimento USA

CATEGORIA	CODICE NUMERICO	CODICE FONDO	FUND FAMILY	MORNINGSTAR RATING
SMALL GROWTH	1	AFCAX	AFBA Five Star Fund	★★★
	2	GTSBX	AIM Investments	★★★
	3	ALSRX	Alger S. Cap G.I	★★★★
	4	ASMGX	Allegiant S. Cap G.A	★
	5	POPAX	Allianz OCC Opp. A	★★★
	6	PMGAX	Allianz CCM Emerging C.A.	★★★
	7	QUASX	Alliance Bernstein G.A	★★★
	8	PMCIX	Allianz CCM Emerging C.	★★★
	9	MSASX	American Century	★★
	10	ANOIX	American C. New Opp.	★★★★
	11	POPCX	Allianz OCC Opportunity C	★★★
SMALL VALUE	12	ASVIX	American Century	★★★
	13	ARTVX	Aristan	★★★★
	14	CSSBX	Columbia Small	★★★

	15	TSVUX	Consulting Group	★★★★
	16	RCBSX	CNI Charter RCB	★★
	17	BERWX	Berwyn	★★★
	18	AVALX	Aegis Value	★★★
	19	AMRRX	Allegiant	★★
	20	AASVX	American Beacon	★★★★
SMALL BLEND	21	SMEAX	AIM Small Cap Equity A	★★
	22	ASCIX	American Beacon S.C. Idx Inst	★★★
	23	PNSEX	BlackRock S.C.V. Instl	★★★
	24	SMCIX	California Inv. S&P S.C.	★★★
	25	DFTSX	DFA Tax-Managed U.S.	★★★★
	26	KDSAX	DWS Dreman S.Cap Value A	★★★★
	27	DFSCX	DFA U.S. Micro Cap	★★★
	28	CRSIX	CRM S.C. Value Instl	★★★
	29	BOGIX	Bogle S. C. Growth Instl	★★★★
	30	SSRCX	BlackRock Aurora Instl	N/A
LARGE GROWTH	31	ENGRX	AXA Enterprise Growth A	★★
	32	CSTGX	AIM Constellation A	★★
	33	AFEAX	AFBA Five Star L. Cap A	★★★
	34	AFUAX	AFBA Five Star USA Global A	★★★★
	35	LCGAX	AIM Large Cap Growth A	★★
LARGE VALUE	36	AVAIX	Accessor Value Adv	★★★
	37	AMWYX	Activa Value A	★★★
	38	TWEAX	American C. Equity Income A	★★★
	39	PDLAX	Allianz OCC Value Admin	★★★★
	40	ABVAX	AllianceBernstein Value A	★★★
LARGE BLEND	41	GTVCX	AIM Basic Value C	★★
	42	LCBAX	AIM Large Cap Basic Value A	★★
	43	AGALX	Accessor G. Allocation Inv	★★★
	44	AGWFX	AIM Select Equity A	★
	45	CHTRX	AIM Charter A	★★
MID-CAP-GROWTH	46	ADEGX	Advance Capital I Equity G.	★★★★
	47	CHCYX	AllianceBernstein Adv	★★★
	48	ATHAX	American Century Heritage A	★★★★
	49	ROCAx	Aquila Rocky Mountain E.A	★★★
	50	TWGTx	American Century Giftrust Inv	★★★

(Fonte:www.finance.yahoo.com)

Tabella A.2: Statistiche descrittive dei rendimenti e extrarendimenti

Fondi	E(r)	E(z_i)	$\sigma(z_i)$	K*	S*	t-test°	JBS*	LS°	JBS*
AFCAX	0,0090	0,0073	0,0557	3,4729	-0,2643	36,5572	1,2370	0,0805	1,1842
GTSBX	0,0061	0,0043	0,0494	2,8403	-0,4082	30,1975	1,7009	0,0795	1,1436
ALSRX	0,0104	0,0086	0,0499	2,3866	-0,3050	31,3042	1,8399	0,0742	1,7742
ASMGX	0,0012	-0,0005	0,0603	2,9446	-0,3677	36,8877	1,3371	0,0581	1,2268
POPAX	0,0105	0,0088	0,0617	3,3676	-0,4204	34,7308	2,0703	0,0840	1,5014
PMGAX	0,0098	0,0080	0,0490	3,5081	-0,5251	29,7452	3,3459	0,0739	3,4708
QUASX	0,0067	0,0049	0,0534	2,8563	-0,3491	32,6702	1,2490	0,0686	0,8879
PMCIX	0,0097	0,0080	0,0491	3,4803	-0,5158	29,5087	3,1828	0,0762	3,3153
MSASX	0,0076	0,0058	0,0443	2,2262	-0,1281	28,2989	1,6330	0,0731	1,7758
ANOIX	0,0096	0,0078	0,0482	3,3549	-0,2394	23,0130	0,8732	0,0788	1,0177
POPCX	0,0098	0,0080	0,0618	3,3724	-0,4239	34,6509	2,1082	0,0820	1,5319
ASVIX	0,0110	0,0092	0,0395	4,1423	-0,5945	28,1803	6,6826	0,0600	5,8588
ARTVX	0,0138	0,0121	0,0406	3,4125	-0,4399	29,0543	2,3208	0,0527	1,8698
CSSBX	0,0093	0,0076	0,0472	3,9610	-0,7438	24,8322	7,7112	0,0928	5,9844
TSVUX	0,0125	0,0107	0,0432	3,4978	-0,4367	31,8217	2,4842	0,0699	2,3396
RCBSX	0,0107	0,0090	0,0447	5,2057	-0,1763	40,9033	12,2658	0,1011	14,8584
BERWX	0,0128	0,0111	0,0441	5,3097	-0,6573	28,8791	17,3628	0,0886	14,3425
AVALX	0,0092	0,0074	0,0328	3,1029	0,3352	18,7300	1,1307	0,0628	1,9530
AMRRX	0,0096	0,0078	0,0425	3,1484	-0,2468	29,9618	0,6530	0,0566	0,5675
AASVX	0,0128	0,0111	0,0459	3,9399	-0,4167	31,3488	3,8794	0,0694	3,1920
SMEAX	0,0096	0,0079	0,0466	4,8653	-0,8846	29,9135	16,2484	0,1004	13,6882

ASCIX	0,0100	0,0083	0,0503	3,3284	-0,3581	34,0378	1,5260	0,0617	1,0903
PNSEX	0,0116	0,0098	0,0423	4,0528	-0,6845	32,8386	7,3313	0,1048	6,1682
SMCIX	0,0089	0,0071	0,0452	3,2369	-0,3855	33,8163	1,5991	0,0610	1,2464
DFTSX	0,0108	0,0091	0,0512	3,1580	-0,3060	36,5423	0,9818	0,0486	0,6380
KDSAX	0,0139	0,0121	0,0439	4,0677	-0,3361	18,2705	3,9133	0,0692	3,8920
DFSCX	0,0131	0,0113	0,0505	3,2257	-0,3133	39,3084	1,0904	0,0754	0,7969
CRSIX	0,0112	0,0095	0,0477	3,9961	-0,8225	35,1645	9,0912	0,1170	6,3118
BOGIX	0,0125	0,0107	0,0494	3,0769	-0,3252	40,9857	1,0546	0,0638	0,9322
SSRCX	0,0096	0,0079	0,0497	3,4554	-0,3901	35,7975	2,0064	0,0784	1,3521
ENGRX	0,0020	0,0003	0,0314	3,5477	-0,5263	22,4906	3,4613	0,0894	1,2063
CSTGX	0,0041	0,0024	0,0400	2,6592	-0,4076	23,0445	1,9189	0,1147	1,1645
AFEAX	0,0035	0,0018	0,0465	3,7508	-0,1681	19,4820	1,6634	0,0963	2,3818
AFUAX	0,0045	0,0028	0,0467	3,8283	-0,2903	27,9643	2,5155	0,0984	2,3252
LCGAX	0,0039	0,0022	0,0351	2,6553	-0,4105	30,9926	1,9490	0,0926	1,7304
AVAIX	0,0072	0,0055	0,0371	3,8382	-0,4434	28,1581	3,6607	0,1137	1,9340
AMWYX	0,0079	0,0061	0,0358	4,0386	-0,5389	23,5076	5,5081	0,0939	3,5862
TWEAX	0,0084	0,0067	0,0256	3,9773	-0,2170	28,5949	2,8109	0,1010	3,0860
PDLAX	0,0088	0,0071	0,0515	5,9242	-0,7765	23,1768	26,9509	0,1032	22,6049
ABVAX	0,0088	0,0071	0,0351	5,1327	-0,6524	26,7678	18,7987	0,1136	12,1229
GTVCX	0,0058	0,0040	0,0429	4,4126	-0,6524	27,7155	9,0900	0,1009	5,8877
LCBAX	0,0061	0,0044	0,0413	4,8616	-0,7537	24,9940	14,1045	0,1050	9,6132
AGALX	0,0069	0,0051	0,0290	3,8567	-0,6243	27,8411	5,3660	0,1100	2,5028
AGWFX	0,0051	0,0034	0,0407	4,0828	-0,8099	28,4341	9,3322	0,1176	5,9836
CHTRX	0,0064	0,0046	0,0307	3,6543	-0,5459	26,0512	3,9828	0,1203	1,5485
ADEGX	0,0048	0,0031	0,0450	2,5604	-0,2904	23,1893	1,3041	0,0722	0,5824
CHCYX	0,0103	0,0086	0,0686	2,3390	0,0679	20,3566	1,1195	0,0584	0,8261
ATHAX	0,0085	0,0068	0,0415	2,7096	-0,3762	23,1893	1,5988	0,0619	2,0360
ROCAx	0,0089	0,0072	0,0389	3,0073	-0,3678	28,4881	1,3303	0,0460	0,6542
TWGTx	0,0077	0,0059	0,0440	2,8379	-0,4955	24,4540	2,4787	0,0818	2,8162
S&P1500	0,0066			5,7309	-1,0646	13,7676	29,4784	0,1180	-----
Risk-free	0,0017			3,7860	-0,2019	22,3547	1,9183	0,0646	-----
$E(r)$				2,7638	-0,45		1,8051		
$Var(sp)$					0,0017				
$Var(Rf)$					3,76e-5				

I test sono stati valutati per $\alpha = 5\%$

t-test^o è confrontato con un $\chi^2_{0,95,25} = 37,6525$ K* = Kurtosi S° = Skewness

JBS* (sui rendimenti) confrontato con $\chi^2_{0,95,2} = 5,9935$ JBS* (sugli extrarendimenti)

LS° è confrontato con un valore soglia 0.1153, generato in default da MatLab 7.1

Figura A.1: Frontiera Efficiente del portafoglio con i 50 - fondi comuni USA

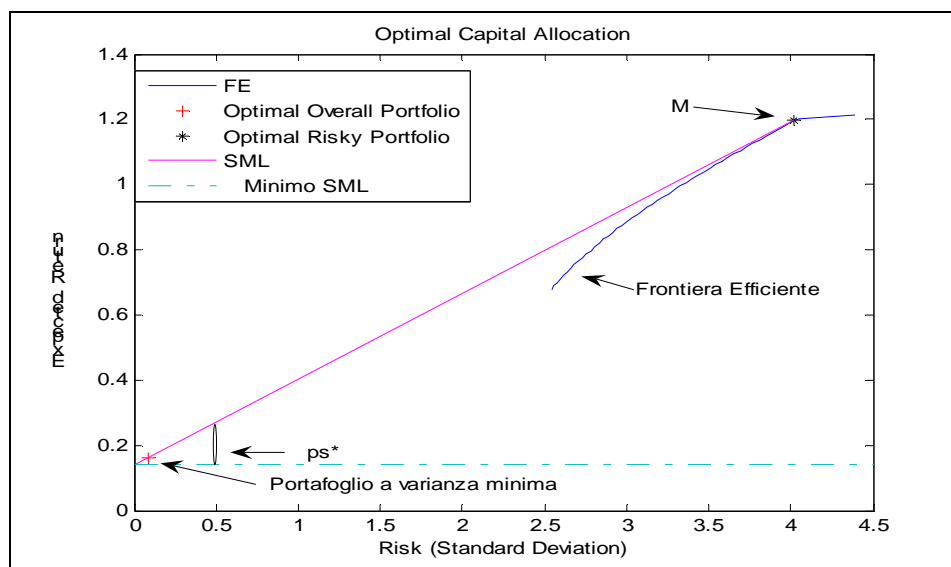


Tabella A.3: Composizione del portafoglio ottimo

Portafoglio con titolo non rischioso	<i>RiskyRisk</i>	<i>RiskyReturn</i>	<i>RiskyWts</i>	<i>RiskyFraction</i>
	4.0210	1.1991	$W_{(1x50)}$	0.0218
Portafoglio con titoli rischiosi	OverallRisk		OverallReturn	
	0,078		0.1631	

Tabella A.4: Valori del coefficiente di inconsistenza tra i 'links' nei clusters

MEDIA TRA LINKS	STD DEI LINKS	LINKS NEL GRUPPO	COEFFICIENTE DI INCONSISTENZA
0,001	0	1	0
0,0110	0	1	0
0,0445	0	1	0
0,0511	0	1	0
0,0514	0	1	0
0,0560	0	1	0
0,0744	0	1	0
0,0657	0,0137	2	0,7071
0,0669	0,0219	2	0,7071
0,0817	0,0103	2	0,7071
0,0962	0	1	0
0,0977	0	1	0
0,0997	0	1	0
0,1002	0	1	0
0,1021	0	1	0
0,1021	0,0027	2	0,7071
0,1037	0,0105	2	0,7071
0,1075	0,0109	2	0,7071
0,1091	0,0121	3	1,1406
0,1126	0,0147	2	0,7071
0,0974	0,0280	3	1,1143
0,1151	0,0155	4	1,1528
0,0899	0,0426	3	1,07598
0,1146	0,0203	3	1,0732
0,1195	0,0167	5	1,0604
0,0943	0,0338	6	1,4180
0,1247	0,0234	3	1,0333
0,1011	0,0707	2	0,7071
0,0989	0,0771	2	0,7071
0,1257	0,0212	6	1,4503
0,1334	0,0259	4	1,0155
0,1402	0,0271	5	0,9986
0,1094	0,0419	6	1,4865
0,0958	0,1199	2	0,7071
0,1330	0,0403	4	1,36739
0,1209	0,0489	7	1,4125
0,1245	0,0638	4	1,1524
0,1078	0,1413	2	0,7071
0,1374	0,0367	7	1,9247
0,1158	0,0649	7	1,9944
0,1470	0,0428	13	2,3509
0,1313	0,0749	10	1,5966
0,1305	0,1028	5	1,2889
0,14647	0,0705	9	2,0061
0,14423	0,0800	17	2,2265
0,1495	0,0932	11	1,9520
0,1536	0,0773	19	2,5913
0,1632	0,1016	31	3,2816
0,16947	0,1300	49	4,78632

BIBLIOGRAFIA

- A. Azzalini, B. Scarpa (2004), “*ANALISI DEI DATI E DATA MINING*”, SPRINGER.
- Bodie, A. Kane and A. Marcus (2001), “*INVESTMENTS*”, Irwin, McGraw-Hill.
- Frank Höppner, Frank Klawonn, Rudolf Kruse, Thomas Runkler (2000), “*FUZZY CLUSTER ANALYSIS*”, LTD, John Wiley & Sons.
- [Fabbris](#) L. (1997) “*STATISTICA MULTIVARIATA*”, McGraw-Hill, Milano.
- Jakša Cvitanić and Fernando Zapatero (2000), “*INTRODUCTION TO THE ECONOMICS AND MATHEMATICS OF FINANCIAL MARKETS*”, London, The MIT Press Cambridge, Massachusetts.
- J. Van Ryzin (1977), “*CLASSIFICATION & CLUSTERING*”, ACADEMIC PRESS.
- K. Jajuga, A. Sokolowski, H. – H. Bock (2002), “*CLASSIFICATION, CLUSTERING AND DATA ANALYSIS*”, SPRINGER.
- Mehmed Kantardizic (2003), “*DATA MINING: Concepts, Models, Methods and Algorithms*”, WILEY-INTERSCIENCE - J. WILEY & SONS, INC.
- Michael J.A.Berry, Gordon S.Linoff (2001), “*DATA MINING*”, APOGEO.
- Paolo Giudici (seconda edizione - 2003) “*DATA MINING*”, McGraw-Hill.
- [Sharpe](#), W.F. (1992), “ASSET ALLOCATION MANAGEMENT STYLE AND PERFORMANCE MEASUREMENT” - J. Portfolio Manage, Winter.
- Sergio Pastorello (2001), “*RISCHIO E RENDIMENTO - Teoria Finanziaria e Applicazioni Econometriche*”, il Mulino.
- [Bandyopadhyay](#), S., Murthy, C.A. and Pal, S.K., (1995), “*Pattern classification with genetic algorithm*” - *Pattern Recognition Lett.* 16, Pages 801–808.
- [Banfield](#), J.D. and Raftery, A.E. (1993), “*Model-based Gaussian and non-Gaussian clustering*” - *Biometrics* 49, Pages 803–821.
- [Bohte](#), Z., Cepar, D., Kosmelj, K. (1988), “*Clustering of Time Series, in COMPSTAT*” - *Physica-Verlag, Wurzburg*, Pages 587–593.
- Brown, S.J. and Goetzman, W.N (1997), “*Mutual fund styles*” – *J. Fin. Econom.* 43, Pages 373-399.

- [Cowgill](#), M.C., Harvey, R.J. and Watson, L.T. (1999), "A genetic algorithm approach to cluster analysis" - *Comput. Math. Appl.* 37, Pages 99–108.
- [Connor](#), G. and Korajczyk, R., (1986), "Performance measurement with the arbitrage pricing theory: a new framework for analysis" - *J. Finan. Econom.* 15, Pages 373–394.
- [Chatterjee](#), S., Laudato, M. and Lynch, L.A. (1996), "Genetic algorithm and their statistical applications: an introduction" - *Comput. Statist. Data Anal.* 22, Pages 633–651.
- [Cucurachi](#), P.A. (2000), "I peer groups nella valutazione della performance dei fondi comuni di investimento azionari" - *Econom. Manage.* 1, Pagine 103–112.
- D.E.A. Giles and R. Draeseke (2001), "Econometric modelling using pattern recognition via the fuzzy C-means algorithm". In: D.E.A. Giles, "Computer-aided econometrics", Marcel Dekker, New York.
- F. Pattarin, S. Paterlini and T. Minerva, "Clustering Financial Time Series: an application to mutual funds style analysis", *ScienceDirect - Computational Statistics & Data Analysis*, Vol.47, Issue 2, Pages 353-372, 1 September 2004
- [Friedman](#), H.P. and Rubin, J. (1967), "On some invariant criterion for grouping data" - *J. Amer. Statist. Assoc.* 63, Pages 1159–1178.
- H. Frigui and R. Krishnapuram (1999), "A robust competitive clustering algorithm with applications in computer vision" - *IEEE Trans Pattern Anal Mach Intell* 21 - 5, Pages 450–465.
- Jessica Lin, Michail Vlachos, E. Keogh and Dimitrios Gunopulos (2002), "Iterative Incremental Clustering of Time Series" - *Computer Science & Engineering Department, California.*
- [Krishna](#), K. and Murthy, M.N. (1999), "Genetic k-means algorithm" - *IEEE Trans. System Man Cybernet.* 29 - 3, Pages 433–439.
- [Lucas](#), L., Riepe, M.W. (1996), "The role of returns-based style analysis: understanding, implementing and interpreting the technique" - Working Paper, Ibbotson Associates.
- [Marriott](#), F.H.C. (1971), "Practical problems in a methods of cluster analysis" - *Biometrics* 27, Pages 501–514.
- [Marriott](#), F.H.C. (1982), "Optimization methods of cluster analysis" - *Biometrics* 69-2, Pages 417–422.
- Matteo Marsili (2006), "Dissecting financial markets: Sectors and States", INFM - *Documento consultivo.*
- M. Gavrilov, D. Anguelov, P. Indyk and R. Motwani (2000), "Financial Time-Series Clustering: Mining the Stock Market: Which measure is Best?" - *Proc. Sixth Int. Conf. Knowledge Discovery and Data Mining (KDD)*, Pages 487-496.

- M. Scionti and J.P.Lanslots, “*Stabilisation diagrams: Pole identification using fuzzy clustering techniques*” – *ScienceDirect: Advances in Engineering Software*, Vol.36, Issues 11-12, Pages 768-779.
- [Paterlini](#), S. and Minerva, T. (2003), “Evolutionary Approaches for Cluster Analysis”. In: Bonarini, A., Masulli, F. and Pasi, G., “*Soft Computing Applications*” - Pages. 167–178, Springer, Berlin.
- [Srinikanth](#), R., George, R., Warsi, N., Prabhu, D., Petri, F.E. and Buckles, B.P. (1995), “A variable-length genetic algorithm for clustering and classification” - *Pattern Recognition Lett.* 16, Pages 789–800.
- S. Miccichè, F.Lillo and R. N. Mantenga (2005), “*Correlation Based Hierarchical Clustering in Financial Time Series*”, INFM.
- U. Desai, P. Debajyoti and R. Kirkpatrick (1985), “A realization approach to stochastic model reduction”, *Int J Control* 24 - 4, Pages 821–838.
- Riferimenti dal quotidiano: “*Il Sole 24 ore*”, in modo continuo dal giugno 2007.

WEBGRAFIA

- <http://www.consob.it>
- <http://www.nber.org/papers/w9222>
- <http://www.borsaitalia.it>
- <http://www.mtsspa.it>
- <http://www.eurotlx.com>
- <http://www.finance.yahoo.com>
- <http://www.MORNINGSTAR.com>
- http://www.performancetrading.it/Documents/PGTeoria/Teo_Mercato.htm
- <http://www.operativetrading.it>
- <http://www.ScienceDirect.com>
- <http://www.elsevier.com>
- <http://www.wikimedia.it>
- <http://www.economiamc.org>
- <http://www.antiqui.it>

- <http://www.webdice.it>
- <http://www.host.uniroma3.it/facoltà/economia>

*Ognuno di noi è un fiume,
Ogni fiume segue una linea,
Ogni linea retta passa per due punti,
Ogni punto è l'inizio di una nuova frase,
Ogni frase è la nascita di un pensiero.*