



UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

**CORSO DI LAUREA SPECIALISTICA
IN SCIENZE STATISTICHE ECONOMICHE,
FINANZIARIE E AZIENDALI**

RELAZIONE FINALE:

**METODI STATISTICI PER LA GESTIONE AUTOMATICA
DEI FONDI COMUNI**

RELATORE: CH.MO PROF. LISI FRANCESCO

LAUREANDO: CANELLA FRANCESCO

MATRICOLA: 513607 - SEFA

ANNO ACCADEMICO: 2005-2006

A Valentina

Indice e Sommario.

0. Introduzione.	p. 3
1. I fondi comuni d'investimento.	p. 7
1.1 Che cosa sono i fondi comuni.	p. 7
1.2 Le categorie di fondi comuni.	p. 8
1.3 A che cosa servono.	p. 10
1.4 Società di gestione, ruolo del gestore e come lavora.	p. 11
1.5 Diaman s.r.l.	p. 15
1.6 La metodologia usata da Diaman.	p. 16
1.7 Definizione del problema.	p. 19
2. La descrizione dei metodi.	p. 25
2.1 Metodo di formazione dei gruppi.	p. 26
2.2 La misura della distanza.	p. 29
2.3 La scelta delle variabili.	p. 33
2.4 Come ottenere i gruppi dal dendogramma.	p. 38
3. L'analisi e i risultati.	p. 41
3.1 L'analisi.	p. 43
3.2 I risultati.	p. 47
4. Conclusioni.	p. 59
5. Riferimenti bibliografici.	p. 61

0. Introduzione

I fondi comuni d'investimento si sono diffusi enormemente negli ultimi vent'anni anche nel nostro paese, al punto che quasi tutti i risparmiatori ne hanno sottoscritto qualcuno almeno una volta o perlomeno sono stati sollecitati a farlo dalla propria banca o da un promotore finanziario.

L'offerta di strumenti di investimento è ormai vastissima e sempre più articolata e all'aumentare delle opportunità diventa difficile scegliere bene come allocare il risparmio sia per le famiglie che per gli enti istituzionali quali banche, Sim o assicurazioni. Alla società di gestione e ai gestori è anzitutto demandato il compito di riuscire a creare e promuovere prodotti competitivi e che rispettino le esigenze dei clienti. Le armi in possesso dei gestori per perseguire questi obiettivi sono i risultati di gestione e di conseguenza gli stili di gestione.

Il modo principale utilizzato per valutare la gestione è il confronto fra il rendimento di un fondo comune e quello di un indice di mercato (*benchmark*). Infatti per le sue caratteristiche di rappresentatività, viene considerato dagli intermediari come riferimento per capire se un titolo o un fondo dalle caratteristiche analoghe ha registrato, in un dato periodo, migliori o peggiori *performance* in termini di rendimento rispetto al *benchmark* stesso.

A seconda dell'obiettivo che il gestore si pone si possono identificare due principali stili distinti di gestione: la gestione passiva o la gestione attiva.

La gestione passiva si basa sull'assunzione che i mercati siano efficienti e che quindi per ottenere le migliori performance sia meglio avere un portafoglio ampiamente diversificato che rappresenti l'intero mercato di riferimento. La gestione attiva si basa sull'ipotesi che i mercati siano inefficienti almeno localmente e che non tutte le informazioni rilevanti siano incorporate nei prezzi dei titoli.

Di conseguenza i gestori meglio informati possono sfruttare le inefficienze e conseguire extraperformance, cioè rendimenti superiori a quelli di mercato. La gestione attiva può essere messa in pratica adottando due principali metodologie d'investimento, spesso utilizzate insieme: la scelta dei titoli sulla base dell'analisi fondamentale (*stock picking*) e la scelta dei tempi di investimento in base all'analisi dell'evoluzione del mercato (*market timing*).

Esistono poi fondi a gestione quantitativa nei quali il processo di selezione dei titoli e di *asset allocation* è totalmente affidato ad elaborazioni matematiche. Il vantaggio dei fondi a gestione quantitativa sta nella ferrea disciplina e quindi nella totale mancanza di influenze psicologiche o emotive, come possono essere le mode per certi settori borsistici e nei costi di gestione molto bassi.

E' in quest'ultima categoria di gestori che può essere collocata Diaman s.r.l., una società di consulenza a enti istituzionali con la quale ho avuto l'opportunità di collaborare. Sfruttando un breve periodo di stage ho potuto sperimentare i problemi concreti che comporta la gestione di un fondo sia in termini di costi che di scelta delle metodologie da utilizzare.

Il caso che ho analizzato in questa tesi riproduce la situazione reale della gestione di un portafoglio di 30 titoli selezionati da un paniere di quasi 400, tutti di area euro. La metodologia utilizzata da Diaman prevede la divisione dei titoli in 5 gruppi, secondo informazioni extra-campionarie sui settori, e una selezione di 6 titoli da ogni gruppo fino a formare il portafoglio. L'aggiornamento del portafoglio avviene con cadenza settimanale sulla base di un indicatore di acquisto e di vendita noto nell'analisi tecnica come forza relativa.

Per la fase che si riferisce al raggruppamento vengono poi proposte in alternativa alcune possibili soluzioni basate principalmente sull'analisi dei dati. Attraverso le tecniche statistiche di classificazione basate sull'analisi dei gruppi vengono formulate tre proposte: la prima è di raggruppare i titoli in base alle traiettorie dei rendimenti, la seconda in base alla struttura di autocorrelazione delle serie dei rendimenti e dei rendimenti al quadrato, la terza in base ad una opportuna distanza tra modelli. Si propone infine una quarta possibilità basata su informazioni extra-campionarie quali la nazionalità dei mercati in cui i titoli vengono quotati.

Scopo della tesi è, in primo luogo, valutare la bontà della metodologia nel suo complesso. Per far questo confronteremo le varie strategie proposte (intese come la combinazione tra procedura di raggruppamento e procedura di selezione dei titoli)

con un portafoglio di riferimento (*benchmark*) secondo determinati criteri. In secondo luogo ci si propone di isolare l'effetto netto apportato dalla fase di selezione dei titoli mediante la forza relativa e vedere qual è l'effetto congiunto delle due procedure. Quest'ultimo risultato è stato ottenuto in due passi: prima depurando ogni strategia di partenza dall'effetto raggruppamento tramite estrazione casuale dei gruppi; poi, confrontando i risultati ottenuti rispetto ad una strategia che prevedesse tecniche casuali per entrambe le procedure.

1. I fondi comuni d'investimento.

1.1 Che cosa sono i fondi comuni.

Il fondo comune d'investimento è una cassa collettiva dove confluisce il risparmio di una pluralità di investitori, denaro che viene investito in valori mobiliari da società di gestione del risparmio (Sgr) iscritte ad un apposito Albo.

Lo scopo principale di un fondo comune è quello di ottenere, attraverso una gestione collettiva, vantaggi in termini di rendimento, minori costi, maggiore potere contrattuale nell'investimento e diversificazione del portafoglio. Di fatto, chi investe personalmente i propri risparmi non diversifica con efficacia il rischio, non dispone di sufficienti informazioni e di analisi accurate, non può accedere ad alcuni mercati o ricorrere a tecniche sofisticate di gestione del rischio e può incorrere in maggiori costi.

Il fondo comune d'investimento in generale non garantisce comunque alcun rendimento a meno che non sia dichiaratamente un fondo a capitale garantito, protetto o con garanzia di rendimento: il ritorno per il sottoscrittore dipende soprattutto dalle attività in cui è investito.

Il patrimonio del fondo è gestito in monte ed è ripartito in quote, ciò vuol dire che il sottoscrittore è proprietario di una fetta del patrimonio e non di questo o quel titolo in cui è investito il patrimonio del fondo stesso.

Il valore di una quota, visibile giornalmente, non è altro che il valore di mercato degli impieghi del fondo – al netto delle spese e delle commissioni di gestione – diviso per il numero delle quote in circolazione in quel momento. La performance o rendimento di un fondo comune rappresenta la variazione di tale quota.

Le risorse versate dai sottoscrittori del fondo costituiscono un patrimonio separato da quello delle società di gestione: queste ultime sono in buona parte controllate da banche e da compagnie di assicurazioni e soddisfano requisiti di solidità.

I titoli che fanno parte del patrimonio del fondo sono custoditi in una banca depositaria, che è un istituto completamente separato dal patrimonio della società che lo gestisce. Le società di gestione sono, inoltre, soggette alla doppia vigilanza di Consob e Banca d'Italia.

I fondi comuni d'investimento sono aperti: l'ingresso dei nuovi sottoscrittori è libero, così come è consentito in qualsiasi momento rientrare in possesso del proprio investimento, che viene liquidato dalle società di gestione in base al valore della quota. Il patrimonio dei fondi comuni, pertanto, non è fisso, perché può aumentare o diminuire in relazione ai flussi di raccolta e a quelli di rimborso: se le sottoscrizioni superano i riscatti, il patrimonio registra un incremento, in caso contrario subisce una diminuzione.

I fondi sono severamente disciplinati anche sotto il profilo organizzativo dei soggetti che li gestiscono e sotto il profilo della diversificazione dei portafogli.

Analoghe ai fondi sono le Sicav (Società di investimento a capitale variabile) sia italiane sia estere autorizzate al collocamento in Italia; l'unica differenza è che prevedono il diritto di voto in capo al sottoscrittore nelle assemblee della Sicav (il cui patrimonio è diviso in azioni).

Fondi e Sicav appartengono, insieme ai fondi mobiliari e immobiliari chiusi, alla categoria degli Organismi di investimento collettivo del risparmio.

Nei fondi i costi sono rappresentati dalle commissioni d'ingresso a carico dei sottoscrittori e da quelle di gestione, prelevate dal fondo. Qualche fondo presenta anche commissioni di rimborso (a carico dei sottoscrittori) e molti commissioni di incentivo (prelevate dal patrimonio del fondo) che vengono applicate quando il rendimento del fondo supera una certa soglia.

1.2 Le categorie di fondi comuni.

Esistono sul mercato vari tipi di fondi comuni con caratteristiche abbastanza definite in termini di profilo rischio/rendimento atteso. Vi sono fondi di liquidità (o monetari) che investono soprattutto in titoli obbligazionari a breve termine; fondi obbligazionari che investono in titoli obbligazionari a medio-lungo termine; fondi

azionari domestici che investono soprattutto in titoli azionari del mercato italiano; fondi azionari internazionali che diversificano all'estero.

Dal 1° luglio 2003, i fondi comuni d'investimento sono divisi in cinque macrocategorie in base al peso dell'investimento azionario: Assogestioni, l'associazione di categoria delle società di gestione di fondi, ha fissato limiti d'investimento per ognuna di esse. Ogni macrocategoria si divide in categorie sulla base di differenti criteri. La classificazione Assogestioni è adottata da tutti i fondi di diritto italiano, da tutti i fondi lussemburghesi storici e da una buona parte dei fondi esteri armonizzati.

Fondi di liquidità. Non possono investire in azioni e destinano tutto il portafoglio a obbligazioni e liquidità. I fondi di liquidità non possono investire in strumenti privi di rating e gli strumenti finanziari detenuti in portafoglio devono avere un rating non inferiore ad A2 (Moody's) o A (S&P). La duration massima del portafoglio è 6 mesi, e non è ammessa la copertura del rischio di cambio. I fondi di liquidità si differenziano in quattro categorie sulla base della valuta di emissione dei titoli in portafoglio: fondi di liquidità area euro, area dollaro, area yen, area altre valute.

Fondi obbligazionari. Non possono investire in azioni (con l'eccezione degli obbligazionari misti, che possono investire fino al 20% nelle Borse) e pertanto destinano tutto il portafoglio a obbligazioni e liquidità. Si prestano a soddisfare l'esigenza di chi vuole far crescere il capitale nel medio periodo (3-5 anni). Su questi archi temporali, il loro rendimento finanziario è solitamente superiore a quello dei fondi di liquidità, ma l'oscillazione della quota (volatilità) è superiore. La volatilità degli obbligazionari è tanto maggiore quanto più è lunga la scadenza media dei titoli in portafoglio e tanto maggiore l'esposizione del rischio valutario. Si tenga presente che l'esposizione al rischio valutario dipende non solamente dalla percentuale di titoli in valuta detenuti in portafoglio, ma anche dal fatto che questi titoli siano coperti dal rischio di cambio oppure no.

Assogestioni prevede 16 categorie di fondi obbligazionari, di cui 14 specializzate e 2 non specializzate. In realtà, sono operative 15 categorie obbligazionarie. La classificazione dei fondi obbligazionari specializzati è basata su due criteri: rischio di mercato (valuta di portafoglio e duration) e rischio di credito (giurisdizione dell'emittente, tipologia dell'emittente, merito creditizio). I loro investimenti consistono in un investimento principale che copre almeno il 70% del portafoglio e

un investimento residuale pari al massimo al 30% rappresentato da strumenti obbligazionari.

Fondi bilanciati. Investono in un mix di titoli di Stato, obbligazioni e azioni, in Italia e all'estero con un unico vincolo: la quota del portafoglio destinata alle azioni deve essere compresa tra il 10% e il 90%. Si adattano a chi vuol far crescere il capitale nel medio-lungo periodo (oltre 5 anni), con un rendimento potenzialmente superiore a quello degli obbligazionari e una volatilità ovviamente maggiore per via della componente azionaria. Sono divisi in 3 categorie: bilanciati obbligazionari bilanciati e bilanciati azionari a seconda del peso delle azioni in portafoglio (rispettivamente: 10-50, 30-70 e 50-90 percento).

Fondi azionari. Investono prevalentemente (almeno il 70% del portafoglio) in azioni e sono adatti a soddisfare l'esigenza di chi vuole far crescere il capitale nel lungo periodo (7-10 anni e oltre). La rischiosità dei fondi azionari, in generale, cresce all'aumentare della specializzazione. I fondi azionari sono divisi in 18 categorie (tutte operative), caratterizzate da un investimento principale pari almeno al 70% del portafoglio con emittenti o specializzazioni definiti dalle categorie, e da un investimento residuale pari al massimo al 30% del portafoglio in titoli obbligazionari di qualunque emittente e in liquidità nella valuta del mercato di riferimento oppure in euro.

Fondi flessibili. Non rientrano in alcuna delle precedenti categorie, e sono i fondi più speculativi, perché investono senza restrizioni – secondo i criteri fissati dal regolamento – in liquidità, obbligazionario, azionario, euro e non euro a seconda delle aspettative del gestore. Su questi fondi, a differenza di tutti gli altri, non c'è obbligo di confronto con il parametro di riferimento (*benchmark*).

1.3 A che cosa servono i fondi comuni.

I fondi comuni servono a soddisfare una lunga serie di bisogni finanziari della famiglia, perché, al di là di ogni classificazione, si differenziano per il rischio/rendimento atteso. La gamma dei bisogni finanziari che può essere soddisfatta dai fondi comuni è ampia: si parte dalla gestione delle eccedenze di liquidità di famiglia: il conto corrente bancario, sul quale si possono accumulare queste eccedenze temporanee, infatti, non è un investimento, ma un mezzo di pagamento: le cifre che restano sul conto tendenzialmente perdono potere d'acquisto.

C'è poi un'altra liquidità importante, che è rappresentata dalla riserva di sicurezza che in famiglia dovrebbe essere sempre tenuta a portata di mano per fronteggiare oneri imprevisti. Ebbene, in tutti questi casi possono funzionare i fondi comuni di liquidità perché sono quelli meno rischiosi: questo significa che, quand'anche il riscatto dovesse essere molto vicino alla sottoscrizione, le probabilità di chiudere in perdita l'investimento sono minime.

I fondi comuni sono poi particolarmente adatti al bisogno finanziario che forse è più comune nell'ambito delle famiglie, la crescita del capitale per vari scopi: ad esempio l'acquisto di una casa, di un'auto, ecc. Fondamentale è l'orizzonte temporale entro il quale si vuole realizzare questa crescita: più è lungo, più rischi si possono prendere e, conseguentemente, più è elevata la remunerazione che è legittimo attendersi alla conclusione dell'investimento.

Infine, i fondi comuni si possono prestare anche all'esigenza di fruire di una rendita periodica: è una necessità che si può manifestare non solamente quando si va in pensione, ma in qualsiasi fase della vita, allorché le entrate correnti siano insufficienti per mantenere il tenore di vita desiderato.

1.4 Società di gestione, ruolo del gestore e come lavora.

Alla società di gestione è anzitutto demandato il compito di promuovere il fondo, cioè di provvedere alla sua costituzione e di raccogliere mezzi finanziari presso i risparmiatori mediante l'emissione e il collocamento delle quote. In secondo luogo la società di gestione deve gestire le disponibilità monetarie ottenute dai partecipanti investendole nella o nelle categorie di strumenti indicati nel regolamento del fondo e nel prospetto informativo sottoposti alla Banca d'Italia che accerta la corrispondenza del loro contenuto alle disposizioni vigenti in tema di completezza e di trasparenza dell'informazione. Infine, dove il tipo di fondo lo prevede, la società di gestione ha la responsabilità di procedere al rimborso delle quote del fondo al partecipante che abbia esercitato la domanda di riscatto, mettendo a disposizione di quest'ultimo, il corrispondente controvalore nei tempi e nei modi indicati dal prospetto informativo. La società di gestione (Sgr promotore) può delegare le scelte di investimento in talune predefinite aree ad un'altra Sgr o ad altri intermediari finanziari abilitati a prestare servizi di gestione del patrimonio.

Come ricordato, la gestione è effettuata in monte e non su base individuale, come avviene di regola nel servizio di gestione patrimoniale propriamente detto. Ciò significa non solo che è effettuata nell'interesse di un insieme di soggetti numeroso, non definito *ex-ante* e mutevole nel tempo, ma soprattutto che ha per oggetto un patrimonio comune o indiviso sulle cui componenti i partecipanti non hanno alcuno specifico titolo.

L'attività di gestione si svolge mediante operazioni di acquisto e di vendita e ogni altro atto di amministrazione che sia ritenuto opportuno o utile per incrementare il valore del fondo ed eventualmente distribuirne i proventi ai partecipanti e che non sia precluso dalle norme legislative, dalle disposizioni emanate dagli organi di vigilanza e dalle clausole del rendimento del fondo.

Nel regolamento e nel prospetto informativo di ciascun fondo sono indicate le classi di strumenti finanziari (azioni, derivati, obbligazioni, fondi) nelle quali il gestore può investire il patrimonio e viene descritta in sintesi la politica di gestione che si intende seguire. Il processo decisionale relativo alla distribuzione dell'investimento tra le diverse categorie di attività finanziarie (azioni, obbligazioni, liquidità) si chiama *asset allocation*. L'*asset allocation* si suddivide in strategica e tattica. Le scelte di *asset allocation* strategica sono finalizzate all'ottimizzazione del rapporto rendimento/rischio in relazione all'orizzonte temporale e alle aspettative dell'investitore. Le scelte di *asset allocation* tattica sono costituite dalla possibilità di modificare temporaneamente le scelte strategiche di lungo periodo per sfruttare le opportunità che il mercato offre nel medio e breve periodo.

La politica di gestione descritta nel regolamento e nel prospetto informativo è sintetizzata nel cosiddetto *benchmark*, cioè in un indice rappresentativo della composizione del portafoglio del fondo ideale in termini di classi di attività finanziarie (*asset class*), intese come azioni, titoli di Stato e liquidità, titoli domestici ed esteri, settori e aree geografiche.

Il confronto fra il rendimento di un fondo comune e quello di un indice di mercato, pur con tutti i limiti del caso, è un modo per valutare la gestione. Infatti per le sue caratteristiche di rappresentatività, viene considerato dagli intermediari come riferimento per capire se un titolo o un fondo dalle caratteristiche analoghe ha registrato, in un dato periodo, migliori o peggiori *performance* in termini di rendimento rispetto al *benchmark* stesso. Non tutti gli esperti sono concordi sul ruolo del *benchmark*. Se per alcuni è un buon parametro con cui confrontare la

performance del fondo e per avere una descrizione del profilo rischio/rendimento atteso, per altri non è corretto utilizzarlo come strumento di valutazione della performance dal momento che rappresenta per sua natura un portafoglio strategico da cui il gestore del fondo si può anche allontanare nel breve periodo per corretti motivi di gestione finanziaria. Secondo questa posizione, il *benchmark* riveste un ruolo fondamentale di comunicazione al pubblico della struttura tipica del portafoglio del fondo comune, utilizzabile per la stima del rischio atteso più che del rendimento atteso.

Di fatto replicare o battere questo indicatore diventa uno dei principali obiettivi del gestore e come raggiungere tale obiettivo diventa il vero problema. A seconda dell'obiettivo che il gestore si pone si possono identificare due principali stili distinti di gestione: la gestione passiva o la gestione attiva.

La gestione passiva si basa sull'assunzione che i mercati siano efficienti e che quindi per ottenere le migliori performance sia meglio avere un portafoglio ampiamente diversificato che rappresenti l'intero mercato di riferimento. Il gestore passivo, quindi, definisce l'*asset allocation* strategica e la mantiene assicurandosi che l'andamento del valore del portafoglio del fondo sia sostanzialmente corrispondente all'andamento degli indici di riferimento. Numerosi fondi di investimento sono a gestione passiva e ottengono quindi rendimenti lordi uguali ai rendimenti di mercato e di conseguenza il rendimento netto per l'investitore è tanto più alto quanto più contenuti sono i costi di gestione.

La gestione attiva si basa sull'ipotesi che i mercati siano inefficienti almeno localmente e che non tutte le informazioni rilevanti siano incorporate nei prezzi dei titoli. Di conseguenza i gestori meglio informati possono sfruttare le inefficienze e conseguire extraperformance, cioè rendimenti superiori a quelli di mercato. La gestione attiva può essere messa in pratica adottando due principali metodologie d'investimento: *stock picking* e *market timing*.

Nel primo caso il gestore dedica il proprio tempo nell'analisi delle singole azioni cercando di individuare, nell'ambito dei titoli potenzialmente investibili, quelli che offrono migliori opportunità in termini di rischio/rendimento. Spesso in questi casi si ricorre all'analisi fondamentale, analisi effettuata sulla base dei dati di bilancio, di informazioni societarie e di mercato. Nel secondo caso, il gestore cerca di prevedere l'evoluzione del mercato azionario individuando il momento più idoneo all'implementazione delle proprie strategie di investimento. Una strategia di questo

tipo può avvalersi dell'analisi tecnica. Il presupposto dell'analisi tecnica è che la conoscenza degli eventi borsistici passati facilita la previsione di quelli futuri, perché le motivazioni e le variabili che influenzano i comportamenti umani tendono a ripetersi e quindi la semplice serie storica dei prezzi di un titolo contiene informazioni utili a predirne l'andamento futuro. I dati storici delle quotazioni possono essere studiati su un grafico (analisi grafica) o elaborati con calcoli di varia natura per ricavare indicatori di acquisto e vendita (analisi quantitativa).

Ovviamente sia per quanto riguarda l'analisi fondamentale sia per quanto riguarda l'analisi tecnica diventa importante per il gestore capire quali variabili scegliere, come valutare gli indicatori e stabilire le azioni da compiere in base all'orizzonte temporale sul quale si vuole operare. Non esiste un modo comune di operare ma ogni gestore in base alla propria esperienza utilizza ora l'uno ora l'altro indicatore.

Esistono poi fondi a gestione quantitativa nei quali il processo di selezione dei titoli e di *asset allocation* è totalmente affidato ad elaborazioni matematiche. Il vantaggio dei fondi a gestione quantitativa sta nella ferrea disciplina e quindi nella totale mancanza di influenze psicologiche o emotive, come possono essere le mode per certi settori borsistici e nei costi di gestione molto bassi.

Va comunque detto che nella pratica i gestori non sempre scelgono rigidamente l'uno o l'altro stile: il rendimento dei fondi è il risultato che si ottiene mescolando gli sforzi volti a scegliere i titoli migliori con il tentativo di prevedere i movimenti del mercato in generale.

Va aggiunto poi che soprattutto per le gestioni attive risulta ancor più difficile il confronto con il *benchmark* dal momento che il fondo comune d'investimento subisce dei costi per la gestione attiva del portafoglio mentre il *benchmark* per definizione è rappresentato da uno o più indici opportunamente pesati e non è soggetto a costi di alcun genere.

I costi per la gestione possono essere individuati nei costi di negoziazione, di *settlement* e di *slippage*. I costi di negoziazione sono attribuibili al *broker* che è un intermediario finanziario che esegue gli ordini di acquisto e vendita ricevuti dal cliente (per conto terzi). In qualità di negoziatore, riceve una commissione per il servizio fornito sia al momento dell'acquisto che al momento della vendita.

Il costo di registrazione (*settlement*), a differenza dei costi di negoziazione, è un costo fisso, indipendente dall'ammontare dell'operazione, e va dato alla banca

depositaria del fondo ed è il costo di registrazione dell'operazione di acquisto o vendita.

Infine con il termine *slippage* ci si riferisce a quella differenza che si registra, praticamente sempre, tra prezzo di mercato (come riportato dai circuiti informatici o dalla stampa) e prezzo effettivo di esecuzione: è la differenza fra il prezzo stabilito a priori per una vendita (acquisto) di un'attività finanziaria ed il prezzo a cui effettivamente viene eseguito sul mercato. Tale differenza, non rappresenta sempre una cattiva esecuzione dell'ordine da parte del *broker*, quanto piuttosto un prezzo che si paga alla liquidità non perfetta dei mercati. Anche sui mercati telematici, dove i circuiti riportano in tempo reale i prezzi a cui vengono concluse le transazioni, non è sempre detto che il prezzo che si ottiene per effettuare un'operazione sia esattamente pari a quello quotato giusto un secondo prima: l'ordine potrebbe essere troppo grosso oppure troppo piccolo; potrebbe darsi che alcune controparti siano uscite (od entrate) dal (nel) mercato proprio in quel momento o più semplicemente che le condizioni di mercato siano mutate.

Peraltro, il fondo comune reinveste i dividendi incassati sulle azioni possedute, i *benchmark* azionari no; inoltre, spesso, vengono utilizzati dei *benchmark* depurati in modo forfettario dalle imposte sui guadagni in conto capitale (in Italia 12,5%). È quindi inevitabile che il rendimento del fondo comune sia inferiore a quello del *benchmark*, a meno che la gestione attiva sia stata eseguita con una capacità tale da generare rendimenti superiori al *benchmark* per l'intero ammontare dei costi necessari per la gestione.

1.5 Diaman s.r.l.

Nel corso di uno stage, durato alcuni mesi presso la società Diaman s.r.l., ho potuto vedere più da vicino come vengono affrontati i problemi inerenti alla gestione dei fondi e quali sono le soluzioni proposte. Dallo stage è nata una collaborazione col fine di utilizzare le metodologie statistiche nella valutazione e selezione dei titoli.

Diaman s.r.l. è una società di consulenza finanziaria creata con l'obiettivo di fornire un servizio di consulenza a clienti istituzionali quali banche, Sim o assicurazioni.

Tale società agisce nel mercato attraverso metodologie proprie ed algoritmi consolidati che mirano a scegliere con continuità i migliori titoli sui mercati e a determinare con margini d'errore quanto più ristretti possibili i momenti d'acquisto e

vendita dei titoli stessi in maniera tale da massimizzare il rendimento dei fondi detenuti in gestione.

La filosofia guida è quella di investire nei mercati finanziari solo quando e dove conviene e che si preferisce rischiare di non guadagnare molto che rischiare di perdere molto. Per cercare di raggiungere questi obiettivi Diaman ha scelto di praticare una gestione attiva dei fondi basata sull'analisi quantitativa dei titoli con l'obiettivo di cogliere le inefficienze del mercato e di ridurre sempre più al minimo la discrezionalità nelle scelte di gestione. Questa scelta è basata sulla convinzione che influenze psicologiche o emotive possano portare il più delle volte a commettere degli errori mentre nel caso di un processo sistematico e automatizzato tali influenze non condizionano le scelte. Ecco perché Diaman si adopera alla creazione di algoritmi che forniscano segnali d'acquisto e di vendita in forma il più automatizzata possibile e alla ricerca di indicatori che abbiano una buona capacità di previsione sulla base di qualche evidenza empirica o esperienza passata.

Capire l'effettiva efficacia di questi indicatori è uno degli obiettivi di questa tesi. Per realizzarlo cercheremo di metterci in un caso il più reale possibile e in particolare il più vicino alle abitudini di gestione di Diaman.

1.6 La metodologia usata da Diaman.

Il primo passo consiste nella definizione dell'*asset allocation* strategica che viene sviluppata con l'intento di soddisfare le aspettative dei clienti in termini di rischio e rendimento. Una volta stabilito il paniere dei titoli investibili Diaman procede alla divisione del paniere in cinque gruppi attraverso un'analisi di dati extra-campionari basati sui settori. Il portafoglio viene parzialmente aggiornato con frequenza settimanale. In pratica a turno i gruppi vengono analizzati e i titoli relativi a quel gruppo presenti nel portafoglio sostituiti con quelli nuovi.

La scelta di dividere i titoli del paniere in gruppi è una scelta non così intuitiva. Diaman giustifica tale scelta soprattutto da un punto di vista pratico. In base alla propria esperienza afferma che una serie storica finanziaria si muova per il 70% del tempo secondo dei trend definiti locali e qualora si riesca a definire un trend locale su una serie storica, il trend si muove nella stessa direzione per un periodo superiore o uguale a cinque settimane. Sulla base di queste analisi ha scelto di aggiornare il proprio portafoglio con una frequenza di 25 giorni (cioè cinque settimane). Concentrare lo stesso giorno l'acquisto e la vendita di un folto numero di titoli

significherebbe un forte esborso di denaro a causa dei costi di negoziazione, di registrazione e di settlement: sempre sulla base della propria esperienza ha stimato che ci possa attendere un abbassamento della performance giornaliera di portafoglio attorno a 0,3%-0,5%. Dal momento che l'obiettivo è creare una continuità di risultati nella gestione del portafoglio evitando, dove possibile, strappi, soprattutto verso il basso, ha pensato di diluire tali costi nel tempo. Perciò ha scelto di suddividere l'insieme dei titoli considerati in cinque gruppi dai quali estrarre ogni settimana 6 titoli che vanno a rinnovare il portafoglio. In tal modo l'incidenza dei costi è meglio assorbita dal portafoglio, incidendo una volta a settimana solamente per un quinto. Il secondo vantaggio è quello di poter tenere sempre parzialmente aggiornato il portafoglio settimana dopo settimana.

Finora per selezionare i gruppi si è utilizzata la classificazione per settori e i cinque gruppi vengono selezionati in base a valutazioni che devono garantire una certa diversificazione in termini economici. Vengono selezionati cinque settori che siano il meno affini possibile e che garantiscano la crescita attesa maggiore. L'analisi di tali gruppi viene rifatta con frequenza semestrale.

Una volta stabilita l'asset allocation l'aspetto che più interessa è con che logica selezionare da ciascun gruppo i titoli sui cui investire. Quello che Diaman si ripropone è investire in quei titoli che si stanno comportando meglio della media e per far questo si serve di un indicatore dell'analisi tecnica conosciuto come forza relativa (Kaufman, 2005). La definizione di forza relativa, cui si farà riferimento in seguito anche con l'acronimo del termine inglese RS (Relative Strength), si basa sul rapporto tra prezzo di un'azione e quotazione dell'indice azionario cui questa fa riferimento, ovvero:

$$RS_{it} = \frac{P_{it}}{I_t}$$

dove con P_{it} e I_t si intendono rispettivamente il prezzo dell'azione i -esima e la quotazione dell'indice I , entrambi all'istante t .

Diaman, fissato un istante t_0 , indicizza a 100 tutte le serie. RS , P e I diventano, quindi, rispettivamente:

$$RS_{it}^* = \frac{P_{it}^*}{I_t^*} \quad P_{it}^* = 100 \cdot \left(\frac{P_{it}}{P_{i0}} \right) \quad I_t^* = 100 \cdot \left(\frac{I_t}{I_0} \right)$$

In questo modo si ottengono due effetti: il primo è di rendere confrontabili le serie tra di loro; il secondo è quello di depurare ciascuna serie dall'effetto di mercato, in

maniera da poter misurare e confrontare le variazioni relative con maggiore immediatezza.

Successivamente, viene effettuato sulle serie della forza relativa il calcolo di due medie mobili semplici asimmetriche.

L'utilizzo delle medie mobili è largamente diffuso tra le aziende operanti sui mercati finanziari nella determinazione dei trend nelle serie dei prezzi. Infatti, l'applicazione di una media mobile permette di smussare l'andamento di una serie e di eliminare o ridimensionare l'erraticità della serie stessa. Nell'analisi tecnica la media mobile è utilizzata abbinata alla serie dei prezzi per la determinazione dei segnali di acquisto e di vendita: se la serie dei prezzi interseca verso l'alto la media mobile questo è un segnale di acquisto; se la interseca verso il basso allora questo è un segnale di vendita. Questo perché la media mobile tiene conto della media di k ritardi. Se gli ultimi prezzi superano la media vuol dire che stanno salendo, viceversa il contrario.

Va anche ricordato, però, che una media mobile con un basso numero di termini tende a dare falsi segnali d'inversione dei trend (e quindi d'acquisto o di vendita) con frequenza tanto maggiore quanto minore è l'ordine della media stessa, comportando pertanto il rischio di pervenire a decisioni fuorvianti; viceversa all'aumentare del numero di termini su cui viene calcolata la media mobile si registra invece un miglioramento della capacità di discernimento dei segnali forniti. Tuttavia, poiché viene richiesto un numero maggiore di osservazioni passate al fine di poter effettuare il calcolo, medie mobili con un alto numero di termini tendono a segnalare la presenza di un trend con un ritardo sistematico che può inficiare in maniera significativa l'efficacia dell'operazione. Si tratta quindi di trovare un giusto compromesso tra la velocità di reazione e la capacità di individuare un trend a cui si risponde applicando simultaneamente due medie mobili di diverso ordine.

Operativamente si definiscono due medie mobili semplici asimmetriche calcolate sulla serie della forza relativa $MM(k_1)$ e $MM(k_2)$ tali che

$$MM^{(i)}(k_j) = \frac{1}{k_j} \sum_{s=0}^{k_j-1} RS_{it}^*$$

dove i si riferisce al titolo e $j=1,2$ si riferisce alla media mobile con minor numero di termini (k_1) o maggior numero di termini (k_2) e si selezionano di volta in volta quei titoli che ha un istante t fissato hanno il rapporto T maggiore.

$$T = \frac{MM^{(i)}(k_1)}{MM^{(i)}(k_2)} - 1$$

La scelta dell'ordine delle medie mobili viene ottenuta tramite massimizzazione della funzione $r(k_1, k_2)$ rispetto k_1 e k_2 in (t_0, T) e si scrive

$$\max_{k_1, k_2} \{r(k_1, k_2)\} \quad , \quad \text{con } r(k_1, k_2) = \frac{Y_T(k_1, k_2)}{Y_{t_0}(k_1, k_2)} - 1$$

dove $Y_t(k_1, k_2)$ è il valore del rendimento capitalizzato di un portafoglio all'istante $t \in [t_0, T]$ che è stato creato in t_0 ed utilizza per la selezione dei titoli il criterio T .

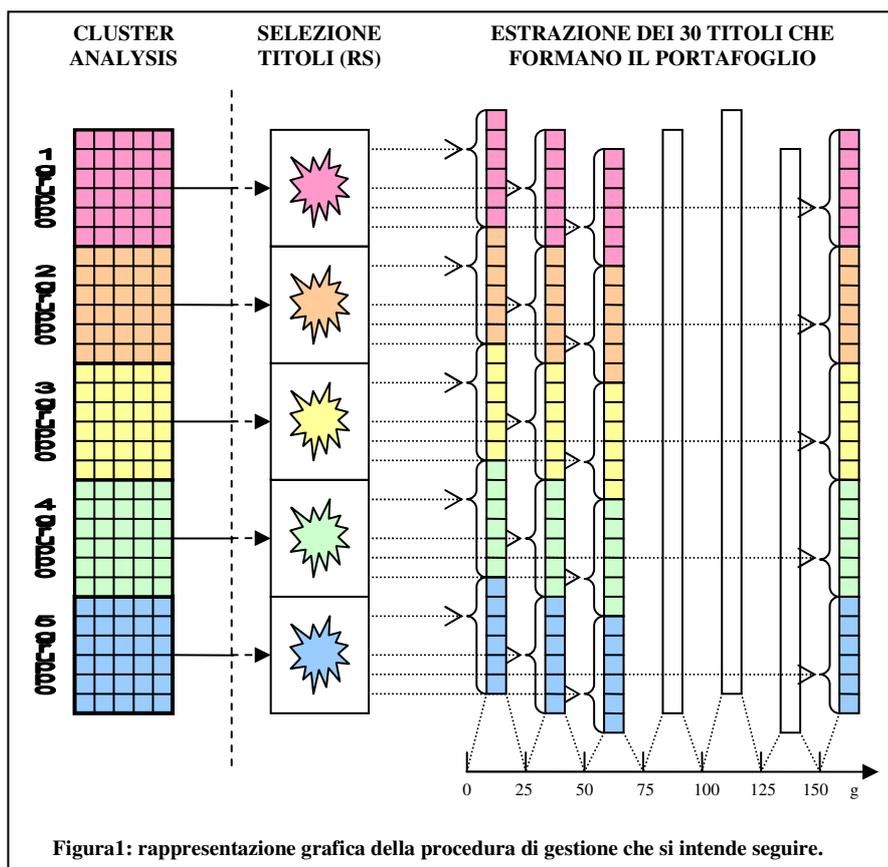
La massimizzazione avviene gruppo per gruppo e non è altro che una serie di prove simulate, al variare degli ordini delle medie mobili, dell'applicazione del trading system sul passato in cui i titoli selezionati sono quelli con il miglior rapporto tra il valore della media mobile di ordine più basso e la media mobile di ordine più alto. Tal scelta si giustifica dal momento che una maggiore distanza tra le due medie implica un trend più evidente e quindi un minore rischio d'incorrere in falsi segnali. La coppia di medie mobili, che possiamo definire ottima, viene utilizzata da quel momento in poi per la selezione dei titoli futuri.

1.7 Definizione del problema.

Le analisi svolte hanno voluto recuperare le metodologie utilizzate da Diaman integrandole con delle proposte alternative di raggruppamento basate sulle tecniche statistiche di *cluster analysis*. Abbiamo simulato di dover gestire un portafoglio che possedesse solamente titoli europei. I titoli del paniere selezionato e sul quale si sono fatte le analisi provengono da mercati di area euro e in particolare: Italia, Francia, Germania, Spagna, Portogallo, Finlandia, Austria, Belgio e Olanda. Fanno parte del paniere tutti i titoli che nei trenta giorni precedenti alla data di formazione del paniere hanno registrato un valore medio di capitale scambiato giornalmente pari o superiore a 1.500.000 €. Questo accorgimento ha l'obiettivo di togliere tutti quei titoli, definiti "sottili" in finanza, cioè che presentano scarsi volumi scambiati, per i quali sarebbe più difficile la vendita o l'acquisto di quantità elevate d'azioni.

Su questo insieme di titoli è stata applicata quindi una metodologia che prevedesse più modi alternativi di formare i titoli in gruppi e una selezione dei titoli da investire basata sulla procedura di Diaman sulla forza relativa (vedi Figura2).

Fine della tesi è, in primo luogo, valutare la metodologia nel suo complesso; in secondo luogo capire quale sia l'effetto netto apportato rispettivamente dalla procedura di selezione dei titoli e quale l'effetto congiunto delle due procedure.



Lo scopo dell'analisi dei gruppi è quello di classificare unità statistiche, per le quali sono state misurate le modalità di differenti variabili in un numero inferiore e relativamente contenuto di gruppi. Alla base degli algoritmi di *clustering* vi è il concetto di distanza. L'obiettivo che ci si prefigge è infatti quello di massimizzare la distanza intra-gruppi e minimizzare, al tempo stesso, quella intra-gruppi.

Quando si affronta il problema della scelta delle variabili è giusto sapere che si tratta di un problema cruciale dell'analisi e che la scelta lascia larghi margini di soggettività. La classificazione dovrebbe fondarsi su tutti gli aspetti che si ritengono importanti per gli scopi prefissati, e questo potrebbe indurre ad ampliare il più possibile le variabili rilevate. Occorre tener presente che l'aggiunta di variabili con scarso potere di discriminazione tra i gruppi, ovvero con bassa qualità dei dati, può peggiorare i risultati rendendo meno nitida la classificazione ottenuta.

Quattro sono i tipi di variabili che si intende proporre.

La prima proposta è di selezionare i gruppi in base ai rendimenti logaritmici dei prezzi di chiusura giornalieri dei titoli.

$$X = \begin{bmatrix} r_{1,1} & \vdots & & & \\ \cdots & r_{i,t} & \cdots & \cdots & \\ & \vdots & \ddots & & \\ & \vdots & & & r_{n,T} \end{bmatrix}, \text{ con } r_{i,t} = \ln(p_t) - \ln(p_{t-1})$$

dove:

$i = 1, \dots, n$ è il numero corrispondente al titolo;

$j = 1, \dots, T$ l'istante temporale;

$r_{i,j}$ è il rendimento logaritmico relativo al titolo i -esimo al tempo j -esimo.

In questo modo ci si propone di catturare le traiettorie simili delle serie dei rendimenti dei titoli.

La seconda proposta invece ha come obiettivo di recuperare la relazione che esiste, ai vari ritardi, tra i rendimenti e i rendimenti al quadrato in modo tale da suddividere i titoli in base alla struttura che manifestano sulla media e sulla volatilità. Per ottenere questo risultato si sono utilizzate le funzioni di autocorrelazione semplici e parziali.

$$X = \begin{bmatrix} \rho_{1,1}^{(1)} & \cdots & \rho_{1,k}^{(1)} & \rho_{1,1}^{(2)} & \cdots & \rho_{1,k}^{(1)} & P_{1,1}^{(1)} & \cdots & P_{1,k}^{(1)} & P_{1,1}^{(2)} & \cdots & P_{1,k}^{(2)} \\ \cdots & \rho_{i,s1}^{(1)} & \cdots & \cdots & \rho_{i,r1}^{(2)} & \cdots & \cdots & P_{i,s2}^{(1)} & \cdots & \cdots & P_{i,r2}^{(2)} & \cdots \\ \cdots & \cdots \\ \rho_{n,1}^{(1)} & \cdots & \rho_{n,k}^{(1)} & \rho_{n,1}^{(2)} & \cdots & \rho_{n,k}^{(1)} & P_{n,1}^{(1)} & \cdots & P_{n,k}^{(1)} & P_{n,1}^{(2)} & \cdots & P_{n,k}^{(2)} \end{bmatrix}$$

dove

$\rho_{i,s1}^{(1)}$ è la funzione di autocorrelazione semplice rispetto alla i -esima serie dei rendimenti logaritmici, al $s1$ -esimo ritardo;

$\rho_{i,r1}^{(2)}$ è la funzione di autocorrelazione semplice rispetto alla i -esima serie dei rendimenti logaritmici al quadrato, al $r1$ -esimo ritardo;

$P_{i,s2}^{(1)}$ è la funzione di autocorrelazione parziale rispetto alla i -esima serie dei rendimenti logaritmici, al $s2$ -esimo ritardo;

$P_{i,r2}^{(2)}$ è la funzione di autocorrelazione parziale rispetto alla i -esima serie dei rendimenti logaritmici al quadrato, al $r2$ -esimo ritardo;

k è il numero di ritardi considerati.

La terza proposta (Otranto, 2004) si basa sulla matrice dei parametri α e β provenienti da un modello AR(1)-GARCH(1,1) per le medie e le varianze condizionate di ciascuna serie Y_t dei rendimenti logaritmici:

$$Y_t - \phi_1 Y_{t-1} = \varepsilon_t \quad \text{con } \varepsilon_t = \eta_t \sigma_t, \quad \varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2), \quad \eta_t \sim N(0,1)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

dove

ϕ è il parametro ar(1) riferito alla i -esima serie;

ω, α, β sono i parametri del modello garch(1,1) stimato per la singola serie.

La matrice dei dati diventa:

$$Z = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \dots & \dots \\ \alpha_n & \beta_n \end{bmatrix}$$

dove α_i, β_i sono i parametri del modello GARCH(1,1) stimato per la i -esima serie con $i = 1, \dots, n$.

Dal momento che le serie da modellare sono numerose si è deciso di selezionare un unico modello per tutte le serie. Si è scelto il GARCH(1,1) perché quello comunemente utilizzato per le serie finanziarie. La presenza di una componente AR(1), invece, ha la funzione di filtrare la serie da eventuali presenze autoregressive sulla media condizionata per una migliore stima della varianza condizionata.

Si aggiunge, infine, una quarta proposta basata non su criteri statistici ma su informazioni extra-campionarie. Ipotizziamo di dividere tutto il paniere dei titoli in cinque gruppi in base alla nazionalità del mercato in cui sono quotati. Questo per vedere se gli sforzi statistici effettuati sono da considerarsi opportuni o se sono facilmente superabili da divisioni più semplici ed immediate.

Una volta effettuate le analisi per i gruppi proseguiremo l'analisi così come è previsto nella procedura Diaman. Utilizzeremo dei dati fuori dal campione utilizzato per la creazione dei gruppi come periodo di *testing* e registreremo come si sarebbero comportati i nostri quattro trading system.

Per la valutazione ci serviremo di grafici, della performance di periodo, di indicatori come il *drawdown* e il tempo medio di recupero e infine della serie dei rendimenti a tre mesi calcolati a una settimana l'uno dall'altro. In cosa consistono e come verranno utilizzati verrà spiegato nel capitolo 3.

2. La descrizione dei metodi.

Una volta definita la strategia utilizzata e prima di mostrare e commentare come si sono svolte le analisi illustriamo con quali metodi è possibile trattare i dati e dividere i titoli in gruppi. In questo capitolo ci occuperemo, quindi, dei metodi statistici per la classificazione delle unità in gruppi omogenei.

Molti studi svolti sinora ammettono l'importanza crescente di riuscire a catturare le caratteristiche delle serie finanziarie (Yee, 2005 o Mcciché et al., 2004), di cercare di raggruppare le serie in gruppi (ad es. Dose e Cincotti, 2005 o Kalpakis, 2001) e, uscendo dall'ambito finanziario, più in generale, tutte le serie temporali proponendo anche approcci e metodi nuovi in alternativa a quelli più comuni (Wang et al., 2004 o Diggs et al., 2003). I metodi più diffusi rimangono attualmente l'analisi discriminante e l'analisi dei gruppi (o cluster analysis).

Nel primo caso è noto a priori che le n unità osservate appartengono a due o più popolazioni differenti. Per ogni unità si conosce il corrispondente vettore dei valori delle p variabili. L'obiettivo dell'analisi discriminante è quello di stabilire un criterio per assegnare correttamente ulteriori unità alla rispettiva popolazione di appartenenza, minimizzando la probabilità degli errori di attribuzione.

L'analisi dei gruppi, o cluster analysis, è invece un metodo tipicamente esplorativo: esso consiste nella ricerca delle n osservazioni p -dimensionali di gruppi di unità tra loro simili, non sapendo a priori se tali gruppi omogenei esistono effettivamente nel *dataset*. La *cluster analysis* ha dunque l'obiettivo di riconoscere gruppi che appaiono con naturalezza nelle osservazioni. Per il tipo di problema che è stato preso in considerazione in questa tesi è quest'ultimo il metodo statistico che utilizzeremo e che illustreremo più da vicino nei prossimi paragrafi, mentre tralascieremo la discussione e l'approfondimento del metodo dell'analisi discriminante.

Presentiamo, quindi, la teoria che sta alla base della cluster analysis attraverso una discussione sulla scelta delle variabili, elencando definendo le misure di distanza i metodi di formazione dei gruppi più comunemente utilizzati.

2.1 Metodo di formazione dei gruppi.

In questo paragrafo ci limitiamo a considerare solamente i metodi che individuano partizioni dell'insieme delle n unità statistiche. L'obiettivo è quello di classificare tali unità in gruppi con le caratteristiche di coesione interna (le unità assegnate ad un medesimo gruppo devono essere tra loro simili) e di separazione esterna (i gruppi devono essere il più possibile distinti).

I metodi di formazione dei gruppi vengono distinti in gerarchici e non gerarchici.

I metodi gerarchici consentono di ottenere una famiglia di partizioni, con un numero di gruppi da n a 1, partendo da quella banale in cui tutte le unità sono distinte per giungere a quella, pur banale, in cui tutti gli elementi sono riuniti in un unico gruppo.

I metodi non gerarchici forniscono un'unica partizione delle n unità in g gruppi, con g fissato a priori.

Da come è stato presentato il problema di cui intendiamo occuparci potrebbe sembrare quest'ultima la scelta migliore. Infatti avendo bisogno di dividere le unità statistiche in cinque gruppi la scelta più naturale potrebbe sembrare quella di porre g uguale a 5. Purtroppo questa modalità nasconde delle insidie. Il vero motivo che ci ha portato ad accantonare tale procedura risiede nel fatto che dipende fortemente dall'unità statistica rispetto alla quale cominciare il raggruppamento.

Un procedimento di questo tipo può fornire più soluzioni possibili e quindi anche molto diverse tra loro introducendo una componente erratica nella scelta dei titoli che può risultare difficile da gestire e soprattutto da motivare. Per questo prenderemo in esame solamente i metodi gerarchici.

Caratteristiche generali dei metodi gerarchici.

Un metodo di classificazione gerarchico presenta le seguenti caratteristiche:

- considera tutti i "livelli di distanza", che indicheremo con γ
- i gruppi che si ottengono ad ogni livello di distanza comprendono i gruppi ottenuti ai livelli inferiori.

Pertanto, quando due o più unità si uniscono tra loro esse non possono venire separate nei passi successivi alla procedura.

Un metodo gerarchico genera una famiglia di partizioni delle n unità partendo da quella in cui tutte le unità sono distinte, ricavando successivamente quelle con $(n-1), (n-2), \dots$ gruppi, per giungere sino a quella in cui tutte le unità sono riunite in un unico gruppo.

Un'ampia classe di metodi gerarchici si fonda sull'impiego iniziale d'una matrice di distanze \mathbf{D} calcolata per le n unità statistiche. In tal caso la procedura seguita per individuare i gruppi d'elementi si articola nelle fasi seguenti:

- i) si individuano nella matrice \mathbf{D} le due unità con la minore distanza e si riuniscono a formare il primo gruppo. Si ottiene una partizione con $(n-1)$ gruppi di cui $(n-2)$ costituiti da singole unità e l'altro formato da due unità.
- ii) si ricalcola la distanza del gruppo ottenuto dagli altri gruppi, ricavando una nuova matrice delle distanze, con dimensioni diminuite di uno.
- iii) Si individua nella nuova matrice delle distanze la coppia di unità con minore distanza, riunendole in un unico gruppo.
- iv) si ripetono le fasi ii) e iii) sino a quando tutte le unità sono riunite in un solo gruppo.

Le differenze tra i vari metodi gerarchici consistono nel criterio utilizzato per calcolare la distanza tra due gruppi di unità.

Prima di vedere da vicino le principali, definiamo alcuni punti di riferimento sui quali basare la descrizione particolare.

Si considerino due gruppi C_1 e C_2 e siano n_1 e n_2 le rispettive numerosità. Sono possibili diverse definizioni di distanza tra i due gruppi, che identificano altrettanti metodi gerarchici. Ecco un possibile elenco: metodo del legame singolo, metodo del legame completo, metodo del legame medio, metodo del legame medio dei gruppi, metodo del centroide e metodo di WARD.

a) metodo del legame singolo.

La distanza tra due gruppi è definita come il minimo delle n_1 e n_2 distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo:

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

b) metodo del legame completo.

La distanza tra due gruppi è definita come il massimo delle n_1n_2 distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo:

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

Adottando questo metodo, tutte le distanze tra le unità del primo gruppo e quelle del secondo sono minori o uguali alla distanza tra i due gruppi così definita.

c) metodo del legame medio.

La distanza tra due gruppi è definita come la media aritmetica delle n_1n_2 distanze tra ciascuna delle unità d'un gruppo e ciascuna delle unità dell'altro gruppo:

$$d(C_1, C_2) = \frac{1}{n_1n_2} \sum_r \sum_s d_{rs}, \quad \text{per } r \in C_1, s \in C_2$$

d) metodo del legame medio dei gruppi.

È una variante del modello precedente. La distanza tra due gruppi è definita come la media aritmetica delle distanze tra tutte le possibili coppie di unità del nuovo gruppo che si ottiene. Indicando con $m = (n_1 + n_2)$ il numero complessivo di unità dei due gruppi considerati, tale distanza è la seguente:

$$d(C_1, C_2) = \frac{1}{m(m-1)/2} \sum_{r=1}^m \sum_{s>r}^m d_{rs}, \quad \text{per } r \in C_1, s \in C_2$$

e) metodo del centroide.

La distanza tra due gruppi è definita come la distanza tra i rispettivi centroidi:

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

Ovviamente, il calcolo del centroide d'un gruppo di unità richiede la conoscenza dei rispettivi valori delle p variabili, che si possono leggere nella matrice dei dati.

Facciamo notare che, per un principio di coerenza, la distanza tra i centroidi deve essere del medesimo tipo utilizzato nella matrice delle distanze.

Il centroide del nuovo gruppo che si forma può essere calcolato in funzione dei centroidi dei due gruppi di partenza:

$$\text{centroide}(C_1, C_2) = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n_1 + n_2}$$

Il metodo del centroide presenta analogie con il metodo del legame medio: in quest'ultimo, infatti, si considera la media delle distanze tra le unità del'uno e

dell'altro gruppo, mentre nel metodo del centroide si individua dapprima un "centro" di ciascun gruppo e poi si misura la distanza tra essi.

f) metodo di WARD.

Lo studio svolto da principio ha visto come protagonisti due metodi: il metodo del legame completo e il metodo di Ward. Poi ci siamo concentrati semplicemente su quest'ultimo dal momento che il primo non forniva dendogrammi chiari e quindi complicava la procedura.

Nel metodo di Ward si definisce esplicitamente una funzione obiettivo. Dato che lo scopo della classificazione è quello di ottenere gruppi con la maggiore coesione interna, si considera la scomposizione della devianza totale (T) delle p variabili in devianza nei gruppi (W) e devianza fra i gruppi (B):

$$T = W + B$$

$$\text{con } T = \sum_{s=1}^p \sum_{i=1}^n (x_{is} - \bar{x}_s)^2, \quad W = \sum_{l=1}^g \sum_{s=1}^p \sum_{i=1}^{n_l} (x_{is} - \bar{x}_{s,l})^2, \quad B = \sum_{s=1}^p \sum_{l=1}^g n_l (\bar{x}_{s,l} - \bar{x}_s)^2$$

dove, data una partizione in g gruppi,

T: è la devianza totale delle p variabili, ottenuta come somma delle devianze delle singole variabili rispetto alla corrispondente media generale \bar{x}_s ;

W: è la devianza nei gruppi, cioè la somma delle devianze di gruppo;

B: è la devianza fra i gruppi, cioè la somma delle devianze delle medie di gruppo rispetto alla corrispondente media generale.

Ad ogni passo della procedura gerarchica si aggregano tra loro i gruppi che comportano il minor incremento della devianza nei gruppi, cioè che assicurano la maggiore coesione interna possibile.

2.2 La misura della distanza.

Se immaginiamo la nostra matrice di dati di dimensioni $n \times p$, dove n sono i titoli e p sono le variabili che tengono i rendimenti giornalieri, possiamo affermare che finora nei paragrafi precedenti abbiamo esaminato le relazioni tra variabili ed i metodi che si propongono di conseguire una riduzione delle dimensioni dello spazio R^p e cioè delle colonne. La formazione di gruppi omogenei di unità, che è lo scopo dei prossimi

paragrafi, può interpretarsi, invece, come una sorta di riduzione delle dimensioni dello spazio R^n , poiché si riuniscono le n unità in g sottoinsiemi, che dovrebbero corrispondere a categorie, o classi, delle stesse.

Per poter affrontare questo tema si deve illustrare il concetto di prossimità tra unità statistiche che per quanto detto sinora individuano i vettori riga nella matrice dei dati. Con il termine generico di prossimità ci si riferisce sia al concetto di rassomiglianza tra le unità, sia a quello di diversità.

Un indice di prossimità tra due generiche unità statistiche è definito come una funzione dei rispettivi vettori riga nella matrice dei dati. Le informazioni fornite dagli indici di prossimità tra coppie d'elementi costituiscono una premessa per l'individuazione di gruppi di unità omogenee.

Gli indici di similarità vengono abitualmente distinti a seconda che si applichino a fenomeni quantitativi oppure qualitativi. Nel nostro caso, avendo noi matrici formate solamente da variabili quantitative, considereremo ed esporremo solamente misure di distanze opportune per le variabili scelte.

Definizione di distanza.

Si dice distanza tra due punti corrispondenti ai vettori $\mathbf{x}, \mathbf{y} \in R^p$ una funzione $d(\mathbf{x}, \mathbf{y})$ che gode delle seguenti proprietà:

1) non negatività:

$$d(x, y) \geq 0 \quad \forall x, y \in R^p$$

2) identità:

$$d(x, y) = 0 \quad \Leftrightarrow x = y$$

3) simmetria:

$$d(x, y) = d(y, x) \quad \forall x, y \in R^p$$

4) disuguaglianza triangolare:

$$d(x, y) \leq d(x, z) + d(y, z) \quad \forall x, y, z \in R^p$$

Uno spazio con riferimento al quale sia definita una distanza è detto spazio metrico.

Con riferimento ad n unità statistiche, calcolando la distanza tra ciascuna delle possibili coppie d'elementi, si ottiene una matrice delle distanze di dimensione $n \times n$. Per la definizione di distanza, tale matrice è simmetrica e presenta valori tutti nulli sulla diagonale principale. Inoltre la matrice è definita positiva.

I tipi di distanza utilizzati.

Distanza euclidea.

È il tipo più noto di distanza. Con riferimento ad una matrice dei dati, con n unità statistiche e p variabili, definiamo la distanza euclidea tra due unità statistiche i e j come la norma (euclidea) della differenza tra i rispettivi vettori:

$${}_2d_{ij} = \|\vec{x}_i - \vec{x}_j\| = \left[\sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2}$$

dove x_{is} e x_{js} sono le coordinate dei punti P_i e P_j nello spazio cartesiano sulla variabile x_s .

Il tipo di distanza considerata è funzione delle differenze in modulo tra i valori che le p variabili presentano in due unità statistiche. La somma di tali differenze ha però significato solo se tutte le variabili sono espresse nella stessa unità di misura. Anche in tale circostanza, tuttavia, le distanze dei tipi precedenti non risultano del tutto appropriate per misurare la dissimilarità tra due generici elementi, poiché esse risultano fortemente influenzate dai caratteri con più elevato ordine di grandezza e con maggiore variabilità. Si possono superare entrambe le limitazioni suddette se le distanze tra le unità vengono calcolate, anziché sulle variabili originarie, su opportune trasformazioni delle stesse. Una scelta molto naturale è quella di considerare gli scostamenti standardizzati di ciascuna delle variabili.

Distanza tra modelli GARCH.

Piccolo (1990) definisce la distanza tra due processi ARMA invertibili e indipendenti $V_{1,t}$ e $V_{2,t}$ come:

$$d(V_{1,t}, V_{2,t}) = \left[\sum_{j=1}^{\infty} (\pi_{j,1} - \pi_{j,2})^2 \right]^{1/2}$$

dove $V_{i,t} = \sum_{j=1}^{\infty} \pi_j V_{i,t-j} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2)$ e $\sum_{j=1}^{\infty} |\pi_j| < \infty$.

Questa misura di distanza ha l'obiettivo di fornire una misura facile da calcolare, applicabile in genere, statisticamente consistente e capace di mettere in relazione molte serie storiche confrontando i modelli ARMA stimati su di esse. La logica della distanza definita dall'autore si basa sul fatto che un qualsiasi processo stocastico appartenente alla classe dei modelli ARMA invertibili è pienamente caratterizzato dalla sequenza $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \dots)$, dove $\boldsymbol{\pi}$ rappresenta il vettore dei coefficienti autoregressivi che caratterizzano la rappresentazione AR(∞) del processo stocastico,

che specifica completamente la distribuzione dei processi ARMA invertibili e dei processi ARMA gaussiani stazionari e invertibili. Dal momento che la sequenza π ci dà tutta l'informazione utile sulla struttura stocastica del processo, la misura della diversità strutturale tra due processi può essere ottenuta confrontando le rispettive sequenze π .

In Otranto (2004) l'autore estende la misura di distanza appena riportata alla famiglia dei modelli GARCH.

Per chiarezza espositiva mostreremo il caso del GARCH(1,1).

Consideriamo due serie storiche che seguano i seguenti modelli:

$$\begin{aligned} Y_{1,t} &= \mu_1 + \varepsilon_{1,t}, & t = 1, \dots, T \\ Y_{2,t} &= \mu_2 + \varepsilon_{2,t}, & t = 1, \dots, T \end{aligned}$$

dove $\varepsilon_{1,t}$ e $\varepsilon_{2,t}$ sono white noise a media zero e varianza rispettivamente $\sigma_{1,t}^2$ e $\sigma_{2,t}^2$.

Supponiamo inoltre che seguano la struttura GARCH(1,1)

$$\begin{aligned} \sigma_{1,t}^2 &= \omega_1 + \alpha_1 \varepsilon_{1,t-1}^2 + \beta_1 \sigma_{1,t-1}^2 \\ \sigma_{2,t}^2 &= \omega_2 + \alpha_2 \varepsilon_{2,t-1}^2 + \beta_2 \sigma_{2,t-1}^2 \end{aligned}$$

nel quale $\omega_i > 0$, $0 < \alpha_i < 1$ e $0 < \beta_i < 1$ e $(\alpha_i + \beta_i) < 1$ ($i = 1, 2$).

Questa scrittura implica che il quadrato dei residui segua un processo ARMA(1,1)

$$\varepsilon_{i,t}^2 = \omega_i + (\alpha_i + \beta_i) \varepsilon_{i,t-1}^2 - \beta_i (\varepsilon_{i,t-1}^2 - \sigma_{i,t-1}^2) + (\varepsilon_{i,t}^2 - \sigma_{i,t}^2), \quad i = 1, 2$$

dove $\varepsilon_{i,t}^2 - \sigma_{i,t}^2$ sono errori a media nulla incorrelati con le informazioni passate.

La rappresentazione AR(∞) è la seguente

$$\varepsilon_{i,t}^2 = \frac{\omega_i}{1 - \beta_i} + \alpha_i \sum_{j=1}^{\infty} \beta_i^{j-1} \varepsilon_{i,t-j}^2 + (\varepsilon_{i,t}^2 - \sigma_{i,t}^2) \quad i = 1, 2$$

In questa forma, i due modelli GARCH(1,1) possono essere confrontati nei termini della misura di distanza proposta da Piccolo (1990) e ottenere la seguente espressione:

$$d(X, Y) = \left[\sum_{j=0}^{\infty} (\alpha_1 \beta_1^j - \alpha_2 \beta_2^j)^2 \right]^{1/2}$$

dove $\alpha_1 \beta_1^j = \pi_{1,j}$ e $\alpha_2 \beta_2^j = \pi_{2,j}$.

che risolvendo le serie diventa:

$$d(X, Y) = \left[\alpha_1^2 \sum_{j=0}^{\infty} \beta_1^{2j} + \alpha_2^2 \sum_{j=0}^{\infty} \beta_2^{2j} - 2\alpha_1 \alpha_2 \sum_{j=0}^{\infty} (\beta_1 \beta_2)^j \right]^{1/2} = \left[\frac{\alpha_1^2}{1 - \beta_1^2} + \frac{\alpha_2^2}{1 - \beta_2^2} - \frac{2\alpha_1 \alpha_2}{1 - \beta_1 \beta_2} \right]^{1/2}$$

2.3 La scelta delle variabili.

Le variabili di partenza che si sono considerate sono i vettori dei rendimenti giornalieri logaritmici calcolati sui prezzi di chiusura di ogni titolo. Questo perché si ritiene che il prezzo di un'azione sintetizzi dentro di sé tutta l'informazione necessaria mentre il suo rendimento ne contenga la dinamicità.

A partire da questo concetto si sono ipotizzate tre delle possibili strade proposte:

- 1) variabili che individuano le traiettorie
- 2) variabili che individuano la struttura autoregressiva
- 3) variabili che permettano una distinzione tra modelli

1) Le traiettorie.

È stato detto che per tener conto delle traiettorie si considera la matrice dei rendimenti logaritmici dei prezzi di chiusura giornalieri delle serie. Dal momento che la matrice così formata contiene oltre 200 osservazioni si è scelto di applicare a tali variabili le componenti principali in modo tale da ridurre considerevolmente il numero. Abbiamo così ottenuto una nuova matrice con un numero di componenti principali tali da spiegare il 90% della variabilità delle variabili iniziali. Tale soglia è stata scelta come compromesso tra il minor numero di variabili e la maggior variabilità spiegata.

Le componenti principali. L'analisi delle componenti principali applicata in via preliminare a tutte le variabili disponibili può essere di aiuto nella scelta delle variabili da utilizzare nella cluster analysis. Se k componenti tengono conto d'una percentuale elevata della varianza totale, si può effettuare la classificazione direttamente sugli scores di tali CP, che costituiscono il segnale degli aspetti rilevanti, mentre le restanti componenti rappresentano i residui, cioè il rumore.

Vediamo nello specifico di cosa si tratta. L'analisi delle componenti principali (ACP) è una metodologia statistica multivariata che, partendo da una matrice dei dati di dimensioni $n \times p$ con variabili tutte quantitative, consente di sostituire alle p variabili (tra loro correlate) un nuovo insieme di variabili – chiamate *componenti principali* (CP) – che godono delle seguenti proprietà:

- 1) sono tra loro incorrelate;
- 2) sono elencate in ordine decrescente della loro varianza.

La prima CP, Y_1 , è la combinazione lineare delle p variabili di partenza avente massima varianza; la seconda CP è la combinazione lineare delle p variabili con

varianza immediatamente inferiore, soggetta al vincolo di essere ortogonale alla componente principale, e così via.

Se le p variabili sono fortemente correlate, un numero $k \ll p$ di CP tiene conto d'una quota elevata della varianza totale, per cui ci si può limitare a considerare solo tali componenti, trascurando le restanti ($p - k$), conseguendo una sensibile parsimonia nella descrizione del data set.

La determinazione della prima CP, y_1 , richiede l'individuazione del vettore p -dimensionale \mathbf{a}_1 dei coefficienti della seguente combinazione lineare delle p variabili, espresse in termini di scostamenti dalle loro medie:

$$y_1 = \mathbf{Z} \mathbf{a}_1, \quad \mathbf{a}_1 = [a_{11}, \dots, a_{1s}, \dots, a_{1p}]$$

La prima CP è per definizione la combinazione lineare di massima varianza. Dato che la varianza totale d'una trasformazione lineare di \mathbf{Z} è esprimibile in funzione della matrice di covarianza \mathbf{S} :

$$\text{VAR}(\mathbf{Z} \mathbf{a}_1) = \mathbf{a}_1' \mathbf{S} \mathbf{a}_1$$

il vettore dei coefficienti \mathbf{a}_1 dovrà essere tale da massimizzare l'espressione precedente. Le soluzioni di tale problema di massimo sono però infinite proporzionali, poiché la combinazione lineare contiene un fattore di scala arbitrario. Si introduce quindi un vincolo, cioè si impone che la norma del vettore dei coefficienti sia uguale ad una costante. Solitamente si assume la condizione di normalizzazione:

$$\mathbf{a}_1' \mathbf{a}_1 = 1$$

sviluppando, con l'impiego dei moltiplicatori di Lagrange, il problema di massimo vincolato si arriva alla seguente equazione:

$$\text{Det}(\mathbf{S} - \lambda_1 \mathbf{I}) = 0$$

La suddetta uguaglianza definisce l'equazione caratteristica della matrice \mathbf{S} , che è un polinomio di ordine p , con p soluzioni chiamate *autovalori*. Essendo la matrice di covarianza semidefinita positiva, gli autovalori sono tutti non negativi. Poiché l'obiettivo è la massimizzazione della varianza della prima CP, si sceglie come λ_1 il massimo di tali autovalori, in quanto vale la seguente uguaglianza:

$$\text{VAR}(\mathbf{Z} \mathbf{a}_1) = \mathbf{a}_1' \mathbf{S} \mathbf{a}_1 = \mathbf{a}_1' \lambda_1 \mathbf{a}_1 = \lambda_1$$

Attraverso questo risultato possiamo definire la *prima componente principale* di p variabili, espresse in termini di scostamenti, come la combinazione lineare:

$$y_1 = \mathbf{Z} \mathbf{a}_1$$

in cui a_1 è l'autovettore corrispondente all'autovettore più grande, λ_1 , della matrice di covarianza S .

Definiamo, più in generale, la v -esima componente principale di p variabili, espresse in termini di scostamenti dalla media, come la combinazione lineare:

$$y_v = Za_v \quad \text{per } v = 1, \dots, k \leq p$$

in cui a_v è l'autovettore associato al v -esimo autovalore λ_v , in ordine decrescente, della matrice di covarianza.

Le CP ottenute partendo dalla matrice di covarianza sono una combinazione lineare degli scostamenti dalla media delle variabili originarie. Il confronto fra tali variabili è però lecito se esse sono espresse nella medesima unità di misura. Inoltre un semplice cambiamento di scala d'una variabile modifica il valore della varianza totale e quindi influenza fortemente i risultati dell'ACP.

Si può superare questa difficoltà considerando le variabili espresse in termini di scostamenti standardizzati dalla media, che equivale ad assumere come punto di partenza dell'ACP la matrice di correlazione invece della matrice di covarianza.

La procedura di calcolo delle CP segue le linee già descritte, tenendo presente che la varianza totale di p variabili standardizzate è uguale a p :

$$VAR_T(W) = tr(R) = p$$

Quindi si può definire v -esima componente principale di p variabili standardizzate come la combinazione lineare:

$$y_v = Wa_v \quad \text{per } v = 1, \dots, k \leq p$$

in cui a_v è l'autovettore associato al v -esimo autovalore λ_v , in ordine decrescente, della matrice di correlazione R .

È bene ricordare che le CP del medesimo data set ottenute dalla matrice di covarianza o dalla matrice di correlazione non sono le stesse e che non è possibile passare da un tipo di analisi all'altro mediante semplici cambiamenti di scala. L'impiego di S oppure di R rappresenta dunque una scelta a priori che condiziona i risultati dell'analisi. Noi abbiamo utilizzato la matrice R .

2) Struttura di autoregressiva.

Il secondo tipo di fenomeno che si vuole considerare come caratteristica discriminante è la struttura autoregressiva. La matrice che è stata considerata tiene conto solamente delle stime con un livello di significatività del 5%. I valori non

significativi sono stati posti uguali a 0. Infine il numero di ritardi considerato è 15, cifra ritenuta sufficientemente elevata ai fini del problema.

Le autocorrelazioni. Come è noto in letteratura (Gallo, 2002) chi intende riconoscere la forma di un processo ARMA sulla serie di osservazioni sui rendimenti di un'attività finanziaria può utilizzare la funzione di autocorrelazione empirica $\hat{\rho}$ come controparte stimata della funzione di autocorrelazione teorica ρ :

$$\hat{\rho}_\tau = \frac{T}{T-\tau} \frac{\sum_{t=\tau+1}^T (r_t - \bar{r}^*)(r_t - \bar{r}^{**})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad \rho_\tau = \frac{E\{(Y_t - \mu)(Y_{t+\tau} - \mu)\}}{E\{[Y_t - \mu]^2\}}$$

dove \bar{r}^* è la media sulle ultime $T-\tau$ osservazioni e \bar{r}^{**} è invece calcolata sulle prime $T-\tau$, Y_t è la variabile casuale indicizzata a t dei rendimenti appartenente al processo stocastico ARMA.

Lo strumento dell'autocorrelazione empirica, mediante il confronto con la funzione di autocorrelazione teorica tipica di determinati processi può infatti suggerire se il processo che ha generato i dati sia autoregressivo o a media mobile e quanti ritardi considerare. Ovviamente tale strumento è più efficace se viene utilizzato in abbinamento alla funzione di autocorrelazione parziale:

$$\hat{P}_\tau = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{\tau-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{\tau-3} & \hat{\rho}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{\tau-1} & \hat{\rho}_{\tau-2} & \cdots & \hat{\rho}_1 & \hat{\rho}_3 \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{\tau-2} & \hat{\rho}_{\tau-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{\tau-3} & \hat{\rho}_{\tau-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{\tau-1} & \hat{\rho}_{\tau-2} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}}$$

Dove $|\cdot|$ rappresenta il determinante della matrice.

Semplificando possiamo dire che quella semplice individua eventuali processi MA mentre quella parziale processi AR.

Le funzioni di autocorrelazioni empirica non si riducono solamente a questo utilizzo. Vengono ampiamente utilizzate anche nella diagnostica di modelli GARCH per la varianza dei rendimenti.

Tenendo presente queste considerazioni e il fatto che non si può e non si vuole stimare un modello ad hoc per ogni serie ma farne risaltare la struttura di correlazione abbiamo deciso di prendere come variabili le autocorrelazioni semplici e parziali sui rendimenti e le autocorrelazioni semplici e parziali dei rendimenti al

quadrato. È possibile sottoporre queste statistiche calcolate sulle osservazioni a verifica d'ipotesi nella quale per ciascun coefficiente l'ipotesi nulla è $H_0: \rho_\tau = 0, \tau = 1, \dots, \tau$. La verifica sulla significatività del singolo coefficiente di correlazione si basa sul fatto che lo standard error dello stimatore sotto l'ipotesi nulla che il corrispondente coefficiente sia uguale a zero è pari a $1/\sqrt{T}$, e che quindi la regione di accettazione al livello di significatività del 5% è approssimativamente uguale a $\pm 2/\sqrt{T}$.

3) Distanze tra modelli.

Il modello ARMA-GARCH. Sia $\{\varepsilon_t\}$ un processo white noise di media nulla e varianza σ_ε^2 . Si dice che $\{Y_t\}$ è un processo autoregressivo a media mobile di ordine (p,q) , e lo si indica con ARMA (p,q) se si può scrivere

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

oppure nella forma più compatta utilizzando gli operatori polinomiali $\theta(B)$ e $\phi(B)$:

$$\phi(B)Y_t = \theta(B)\varepsilon_t \text{ che diventa } \pi(B)Y_t = \varepsilon_t.$$

Se a questo modello aggiungiamo delle ulteriori ipotesi, imponendo che l'andamento della varianza del processo generatore dei dati sia di tipo condizionatamente autoregressivo ci mettiamo nelle condizioni di poter introdurre il modello GARCH per la varianza condizionata.

In termini generali un GARCH (m,n) è un modello in cui la varianza condizionata al tempo t (σ_t^2) è una combinazione lineare di m ritardi dei residui al quadrato e di n ritardi della varianza condizionata, in formule:

$$\sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^n \beta_j \sigma_{t-j}^2, \quad \text{con } \varepsilon_t = \eta_t \sigma_t, \quad \varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2), \quad \eta_t \sim N(0,1)$$

Nel modello GARCH, quindi, le informazioni passate sono sintetizzate dai ritardi della varianza, mentre le novità e la capacità di variazione nel tempo delle stime della varianza condizionata sono racchiuse dal termine ε_{t-j} .

Nelle nostre analisi abbiamo considerato il caso più diffuso in finanza e cioè un GARCH $(1,1)$.

Utilizzeremo i parametri α_1 e β_1 come variabili discriminanti e attraverso una procedura già presente in letteratura calcoleremo una matrice di distanze tra i modelli.

2.4 Come ottenere i gruppi dal dendogramma.

Di norma applicando la regola alla nostra situazione avremmo dovuto fermarci non appena il procedimento di aggregazione fosse arrivato ad un totale di cinque gruppi. Visivamente si può immaginare di avere un dendogramma e di tagliare la figura in corrispondenza di una data distanza in modo tale da ottenere i gruppi stabiliti a priori. Procedere in tal senso, però, può portare alla formazione di gruppi molto eterogenei tra di loro nella numerosità e in alcuni casi costringere il gruppo ad essere non sufficientemente numeroso da garantire il numero minimo di titoli richiesto su cui investire (nel nostro caso sei) o da permettere all'algoritmo basato sulla forza relativa di essere utilizzato con efficacia.

Dal momento che le nostre analisi non possono prescindere da come questa società effettivamente procede e da come concretamente opera sul mercato diamo come dato fisso del problema la costituzione di cinque gruppi di titoli su cui operare ed elaboriamo un metodo alternativo di raggruppamento che funziona come descritto sopra cambiando solamente la tecnica di taglio del dendogramma.

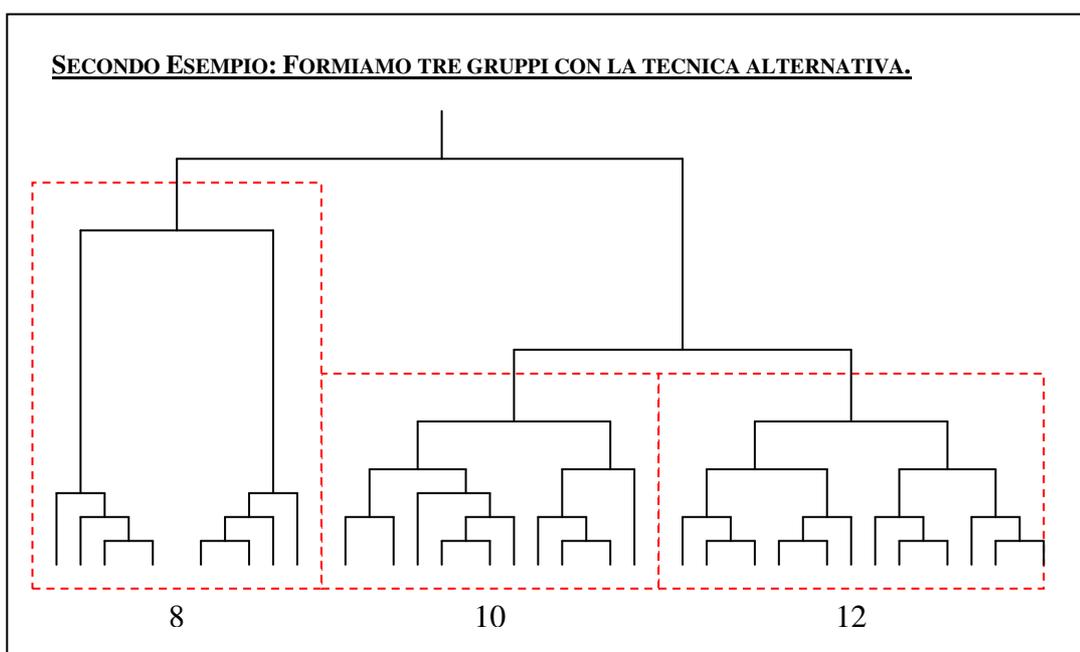
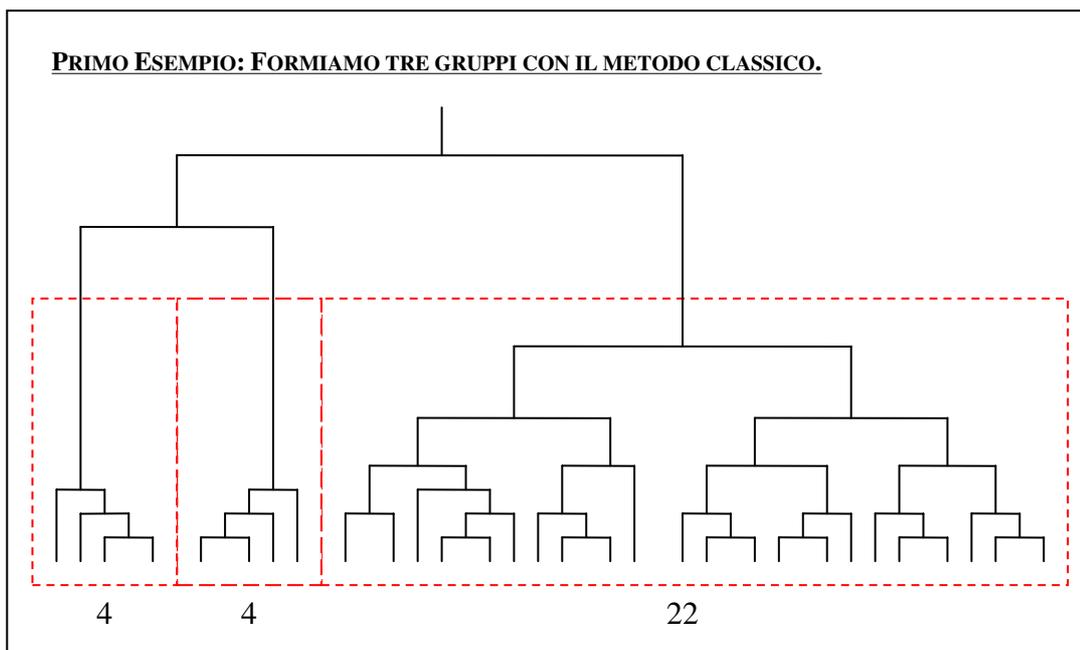
Rifacendoci all'immagine precedente del dendogramma abbiamo creato una routine che invece di tagliare il grafico ad una certa distanza rende variabile questa distanza e crea così dei gruppi discriminati in base alla numerosità.

Più semplicemente ripercorre tutto il dendogramma dall'alto verso il basso fermandosi ad ogni punto di aggregazione. Si ferma su ciascuno dei due gruppi e in base alla numerosità decide se tenerlo o dividerlo ulteriormente.

Ovviamente abbiamo dovuto inserire dei criteri tali per cui la procedura di suddivisione in gruppi fosse il più automatica possibile, per non introdurre elementi soggettivi anche in questa fase dell'analisi, in modo da rendere tale procedimento di assegnazione dei gruppi il meno discrezionale possibile.

I parametri che sono stati decisi a priori si riferiscono alla numerosità dei singoli gruppi. In altre parole sapendo che il nostro obiettivo è arrivare a cinque gruppi abbiamo imposto che ciascun gruppo non possa essere più numeroso del 30% del totale dei titoli da suddividere.

Di seguito abbiamo riportato due esempi, uno in cui si suddivide le unità statistiche in tre gruppi attraverso il metodo classico di slittamento, l'altro con il metodo da noi utilizzato.



Questo modo di procedere, in alcuni casi, ha portato alla formazione di più di cinque gruppi e quindi all'eliminazione dei gruppi meno numerosi. L'eliminazione di molti titoli che questo sistema comporta può sembrare una maniera sconveniente si procedere ma garantisce una uniformità di giudizio, consente cioè, che ogni dendrogramma sia stato trattato con oggettività e che i risultati finali non possano essere inficiati dal tipo di valutazione dell'uno o dell'altro dendrogramma.

3. L'analisi e i risultati.

Nel precedente capitolo sono stati descritti alcuni possibili metodi per raggruppare diversi titoli. In questo capitolo tali metodi verranno applicati a dati reali in un contesto il più possibile realistico. Terremo conto dei costi di negoziazione, di settlement e di slippage che incideranno sul primo rendimento giornaliero del titolo acquistato per un valore che è ragionevole imporre pari a 0,3% del capitale acquistato. Sempre per rimanere all'interno di un contesto il più possibile realistico, sono state considerate le serie dei dati relativi ai giorni in cui sono avvenute le contrattazioni, in caso di festività settimanali sono stati riportati i prezzi della chiusura precedente.

L'analisi è stata effettuata su un paniere di 397 titoli le cui serie dei prezzi di chiusura giornalieri vanno dal 11/04/2001 fino al 30/06/2005 per un totale di 1102 osservazioni per titolo. I titoli del paniere selezionato provengono da mercati di area euro e in particolare: Italia, Francia, Germania, Spagna, Portogallo, Finlandia, Austria, Belgio e Olanda. Fanno parte del paniere tutti i titoli che nei trenta giorni precedenti alla data di formazione del paniere hanno registrato un valore medio di capitale scambiato giornalmente pari o superiore a 1.500.000 €.

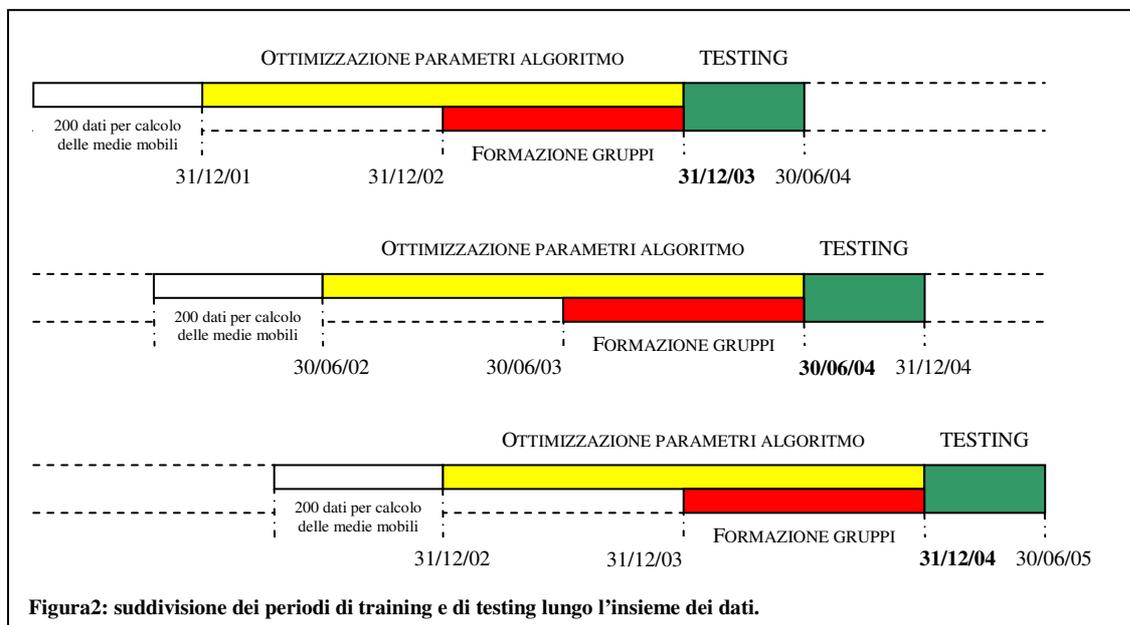
Il portafoglio è composto da un totale di 30 titoli e viene aggiornato ogni settimana attraverso l'acquisto e la vendita di 6 titoli per gruppo.

L'insieme dei dati è suddiviso in tre periodi sovrapposti (come mostrato in Figura2). Ogni periodo di dati è diviso in due parti: una prima parte definita periodo di *training* e una seconda parte definita periodo di *testing*. Il periodo di *training* viene utilizzato per l'ottimizzazione dei parametri (cioè la composizione dei cinque gruppi e l'ordine

delle coppie di medie mobili associate) e consta di due anni di osservazioni; il periodo di *testing* viene utilizzato per registrare il comportamento dell' algoritmo *out of the sample* e consta di sei mesi di osservazioni. I parametri vengono ri-stimati ogni sei mesi per garantire una maggiore flessibilità.

Infine standardizziamo il portafoglio in modo tale abbia valore 100 all'inizio del periodo di testing e che la serie da analizzare sia la serie dei rendimenti capitalizzati giornalmente.

Valuteremo le serie sia nei tre periodi di testing separatamente che congiuntamente. Avremo un periodo di testing complessivo di un anno e mezzo, dal 31/12/2003 al 30/06/2005.



Riprendo in maniera schematica le quattro strategie che intendiamo utilizzare nell'analisi.

La prima strategia prevede di individuare le caratteristiche dei titoli attraverso le traiettorie. Utilizziamo la matrice dei rendimenti logaritmici delle serie e applichiamo le componenti principali per ridurre il numero di variabili. La matrice delle distanze calcolata è quella euclidea.

La seconda strategia prevede di individuare le caratteristiche dei titoli sulla base della struttura di autocorrelazione. Prenderemo come variabili le autocorrelazione semplici e parziali dei rendimenti logaritmici lineari e al quadrato fino al 15-esimo ritardo. Anche in questo caso la matrice delle distanze sarà calcolata attraverso la distanza euclidea.

La terza strategia prevede di individuare le differenze tra le serie in base ai parametri del modello AR(1)-GARCH(1,1). La distanza è calcolata rispetto ai coefficienti α e β come mostrato nel capitolo 2.

La quarta strategia si basa sulla suddivisione per nazionalità: Spagna, Germania, Francia, Italia formano un gruppo ciascuno; Portogallo, Finlandia, Austria, Belgio e Olanda un gruppo insieme. Tale scelta è stata fatta in base alla numerosità dei titoli.

3.1 Le analisi.

Si mostra a titolo esemplificativo l'analisi svolta per la strategia di gestione risultante dall'applicazione della distanza tra modelli GARCH. Tutti gli altri risultati relativi alle altre tre strategie sono ottenuti applicando lo stesso procedimento, l'unica differenza sta nella prima fase dell'analisi che dipenderà dal metodo di selezione dei gruppi applicato.

Primo periodo. I dati utilizzati per l'ottimizzazione dei parametri comprendono una finestra temporale che va dal 11/04/2001 al 31/12/2003.

La prima fase dell'ottimizzazione prevede la divisione dei titoli in gruppi. Si stima un modello AR(1)-GARCH(1,1) per tutte le serie dei rendimenti logaritmici calcolati sull'intervallo temporale compreso tra il 31/12/2002 e il 31/12/2003.

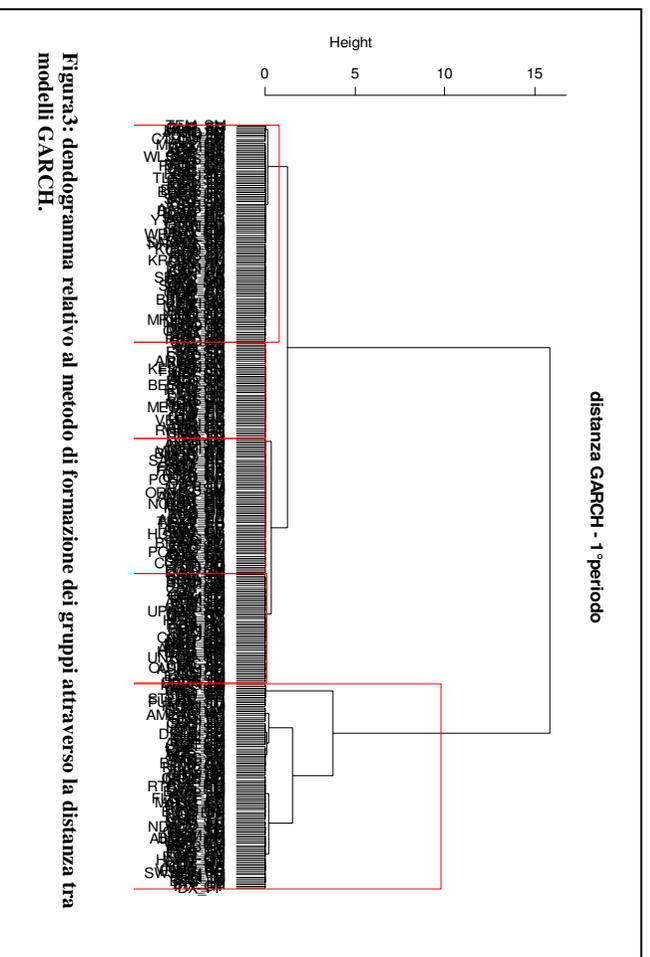


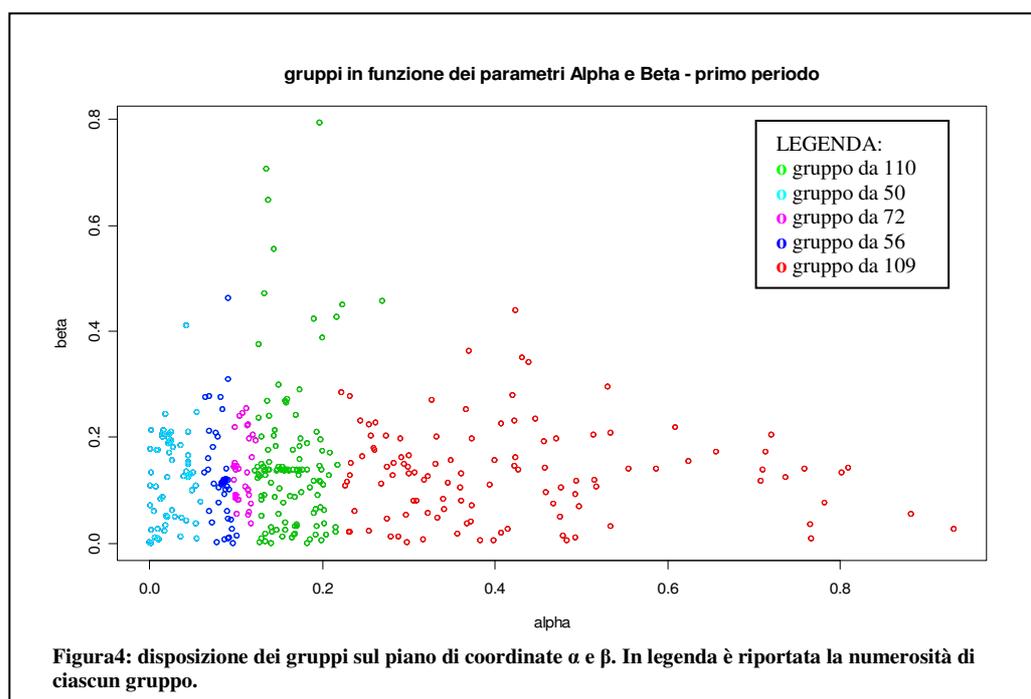
Figura3: dendrogramma relativo al metodo di formazione dei gruppi attraverso la distanza tra modelli GARCH.

I coefficienti α e β stimati vengono utilizzati per la costruzione della matrice delle distanze secondo il criterio illustrato nel capitolo precedente. La matrice, così

ottenuta, viene utilizzata dalla metodologia statistica dell'analisi dei gruppi che attraverso il metodo di Ward, fornisce i cinque gruppi d'interesse. Il dendrogramma è presentato in Figura3.

I titoli sono stati raggruppati con la tecnica di taglio del dendrogramma mostrata nel paragrafo 2.4. ed hanno la seguente numerosità: 110 per il 1° gruppo, 50 per il 2° gruppo, 72 per il 3° gruppo, 56 per il 4° gruppo, 109 per l'ultimo.

A differenza degli altri metodi esaminati utilizzando solamente i due coefficienti α e β come variabili discriminanti su cui calcolare la matrice delle distanze è possibile utilizzare gli stessi come coordinate di un piano su cui sono disposti i gruppi.



La seconda fase dell'ottimizzazione prevede l'ottimizzazione delle medie mobili da utilizzare.

Tale ottimizzazione avviene gruppo per gruppo. Ciò vuol dire che per ogni gruppo ottenuto nella fase precedente si ipotizza un sistema di trading basato sulla forza relativa. Definiremo coppia ottima la coppia di ordini di medie mobili che massimizzano la funzione di rendimento di periodo associata. Il periodo di riferimento, in questo caso è 31/12/2001-31/12/2003.

Le coppie di ordini di medie mobili ottenute sono rispettivamente: (40,65) per il 1° gruppo, (40,105) per il 2° gruppo, (155,185) per il 3° gruppo, (115,130) per il 4° gruppo, (30,180) per l'ultimo. Una volta ottenuti tutti i parametri si passa alla

verifica dell'algoritmo sul periodo di *testing*. La procedura appena mostrata si ripete similmente nei due periodi rimanenti.

Riportiamo, quindi, nelle Tabelle1-4 tutti i parametri ottimizzati metodo per metodo, periodo per periodo, gruppo per gruppo. Sono visualizzate la numerosità del gruppo e la coppia di medie mobili semplici asimmetriche.

Tabella1. Formazione dei gruppi in base alla nazionalità per ciascun periodo di training.					
1°PERIODO	SPAGNA	ALTRI	GERMANIA	FRANCIA	ITALIA
numerosità	57	92	68	95	85
medie mobili	(10,20)	(25,85)	(160,170)	(15,25)	(80,115)
2°PERIODO	SPAGNA	ALTRI	GERMANIA	FRANCIA	ITALIA
numerosità	57	92	68	95	85
medie mobili	(15,25)	(75,95)	(50,75)	(120,140)	(95,115)
3°PERIODO	SPAGNA	ALTRI	GERMANIA	FRANCIA	ITALIA
numerosità	57	92	68	95	85
medie mobili	(15,25)	(10,180)	(20,100)	(120,140)	(15,70)

Tabella2. Formazione dei gruppi in base alle traiettorie per ciascun periodo di training.					
1°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	56	62	57	114	108
medie mobili	(90,100)	(165,175)	(55,125)	(45,55)	(155,175)
2°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	40	100	89	35	100
medie mobili	(170,190)	(110,140)	(50,130)	(115,130)	(45,65)
3°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	58	114	56	100	69
medie mobili	(150,180)	(50,65)	(20,35)	(45,55)	(90,110)

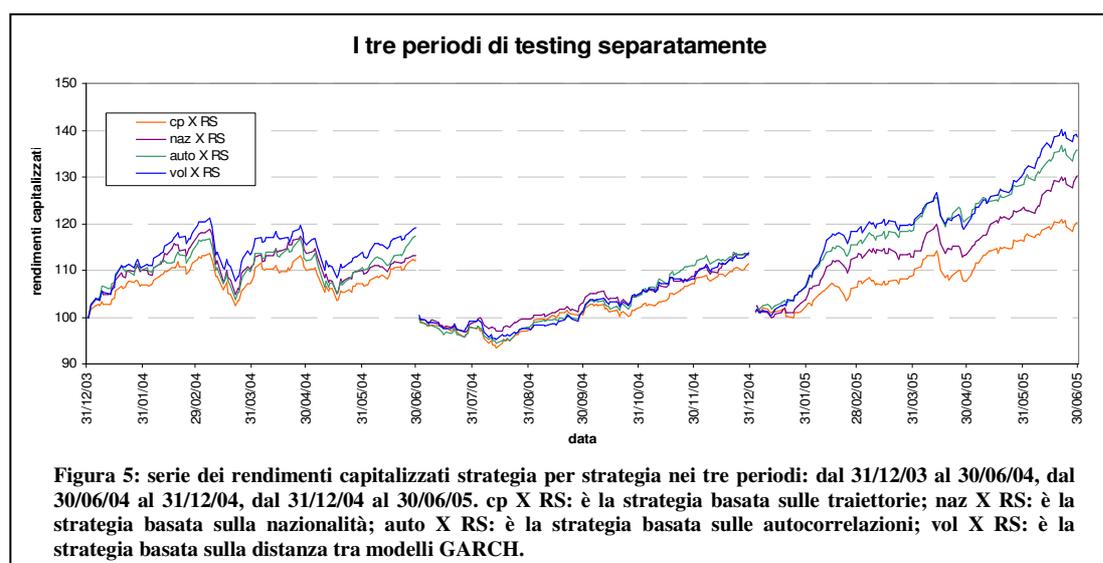
Tabella3. Formazione dei gruppi in base alle autocorrelazioni per ciascun periodo di training.					
1°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	64	52	41	44	81
medie mobili	(110,125)	(20,110)	(170,190)	(100,175)	(50,70)
2°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	46	53	105	80	72
medie mobili	(50,60)	(15,30)	(25,35)	(70,100)	(40,60)
3°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	54	41	40	100	113
medie mobili	(35,65)	(130,140)	(55,155)	(20,90)	(40,70)

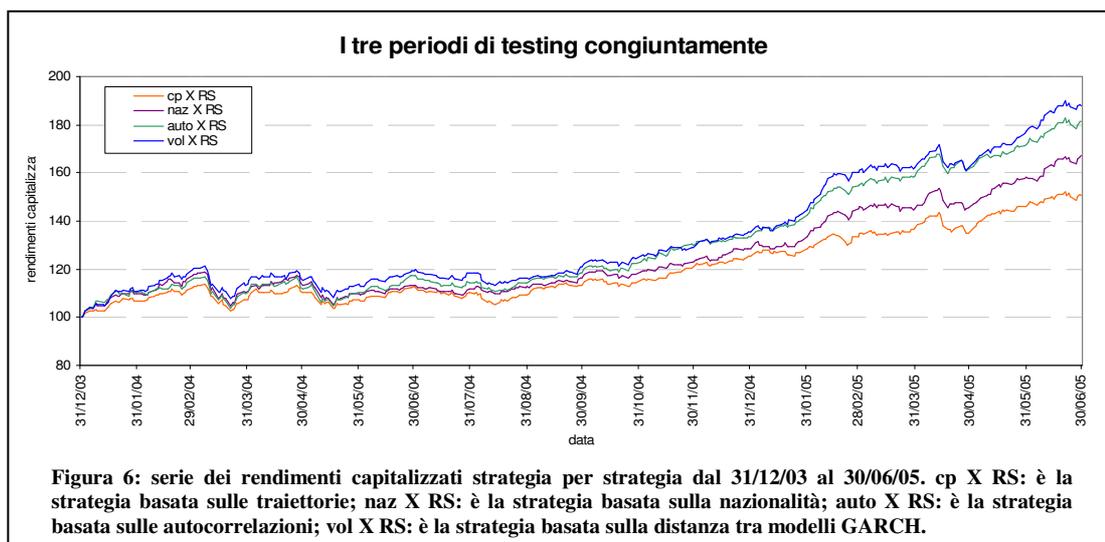
Tabella4. Formazione dei gruppi in base alla distanza tra modelli GARCH per ciascun periodo di training.					
1°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	110	50	72	56	109
medie mobili	(40,65)	(40,105)	(155,185)	(115,130)	(30,180)
2°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	71	57	90	57	85
medie mobili	(10,65)	(45,135)	(115,190)	(35,165)	(50,155)
3°PERIODO	PRIMO	SECONDO	TERZO	QUARTO	QUINTO
numerosità	48	58	117	114	60
medie mobili	(45,60)	(50,60)	(80,90)	(130,165)	(15,185)

Una volta ottenuti i parametri incogniti del problema si procede alla verifica empirica della validità delle strategie ipotizzate attraverso il confronto delle serie dei rendimenti capitalizzati per ciascun portafoglio.

Le Figure5-6 riportano i rendimenti capitalizzati per i portafogli gestiti con le quattro strategie proposte. In Figura5 viene riportato il grafico dei portafogli nei periodi di *testing* procedendo alla standardizzazione dei rendimenti capitalizzati all'inizio di ogni periodo. In Figura6 si riporta invece i risultati standardizzando solo all'inizio del primo periodo mentre i successivi partono con il valore raggiunto alla fine del periodo precedente. Le due figure sono in scale differenti.

La valutazione qualitativa e quantitativa dei risultati verrà approfondito più avanti nel prossimo paragrafo.





3.2 I risultati.

Obiettivo di questo capitolo è valutare le analisi svolte ed in particolare definire alcune evidenze empiriche rispetto alle diverse strategie proposte nella gestione di un fondo comune. Innanzitutto si è interessati a capire se esiste e qual è la strategia migliore da utilizzare sia rispetto alle altre strategie sia in rapporto ad un portafoglio di riferimento (*benchmark*). In secondo luogo si vuole studiare il contributo apportato alla performance della procedura di selezione e successivamente l'effetto congiunto delle due procedure. Questi obiettivi ci impongono di fare due tipi di riflessioni. La prima riguarda i criteri di valutazione di un trading system e quindi gli strumenti di cui intendiamo servirci per la valutazione; la seconda il come isolare i due tipi di effetto di raggruppamento e di selezione dei titoli.

Criteri di valutazione dei trading system.

Come abbiamo esposto nel primo capitolo la valutazione di un fondo e quindi di una strategia va messa a confronto con un indice o *benchmark* rappresentativo del mercato in cui il fondo opera.

Nelle analisi finora effettuate abbiamo creato strategie per fondi investiti in titoli europei e quindi l'indice di riferimento dovrebbe essere l'Eurostoxx 50 (l'indice composto dalle 50 azioni a maggior capitalizzazione quotate nelle principali Borse dell'area Euro).

La scelta che invece effettuiamo va in un'altra direzione. L'indice che utilizziamo è composto da tutti i 397 titoli selezionabili e corrisponde a un portafoglio che abbia come rendimento giornaliero la media equiponderata dei rendimenti giornalieri di tutti i titoli. Dalla Figura 7 si può vedere come questo secondo portafoglio mostri una performance finale di periodo superiore al 20% rispetto all'Eurostoxx 50. Quindi, la scelta non è dettata da una convenienza, quanto dal voler verificare se effettivamente con le strategie proposte si è in grado di catturare quei titoli che hanno rendimenti maggiori degli altri titoli presenti nel paniere iniziale.

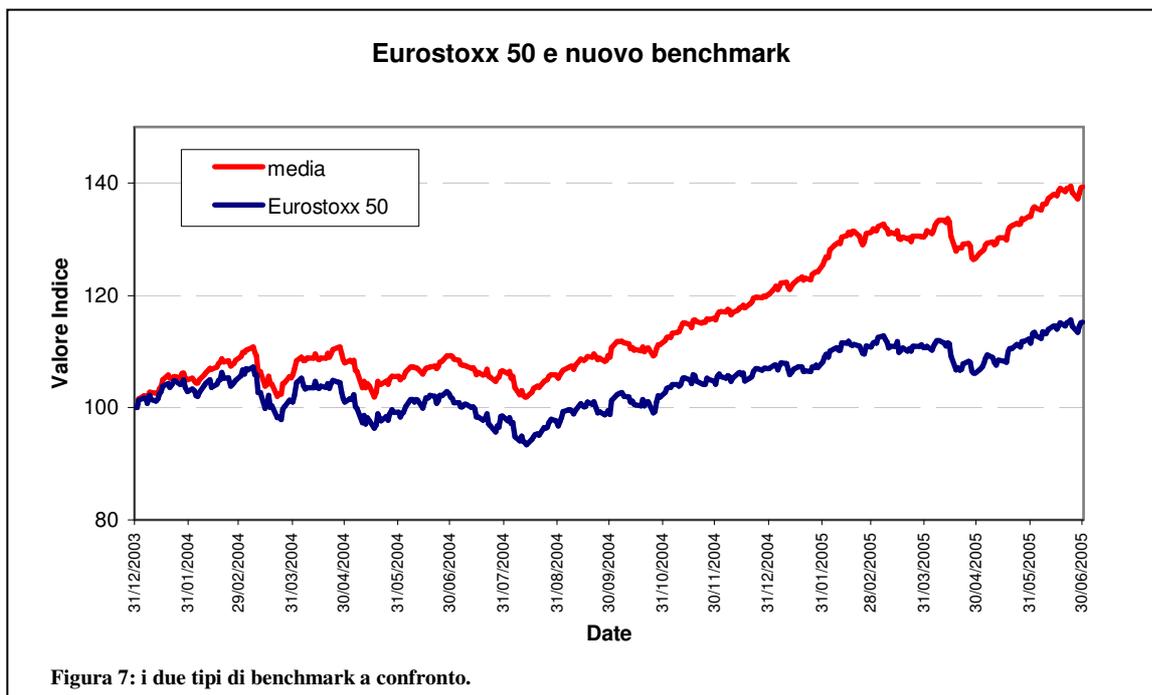


Figura 7: i due tipi di benchmark a confronto.

Assieme al *benchmark* vengono utilizzati anche altri strumenti per la valutazione dei fondi (Kaufman, 2005). Ne vengono presentati quattro e sono: la performance di periodo, la percentuale massima di drawdown, il tempo medio di recupero e la media dei rendimenti a tre mesi:

1. Performance finale di periodo (*Perf.*): si tratta del rendimento totale di periodo al netto delle commissioni.

$$Perf. = (Y_T - Y_0) / Y_0,$$

dove Y_i è il valore del portafoglio o dell'indice al tempo i .

2. Percentuale Massima di drawdown (*MaxDD*): poniamo Y la serie dei rendimenti capitalizzati e $t1$ e $t2$ due istanti di tempo nell'intervallo $(0,T)$, fissiamo

$$M_{t1} = \max_{i \in (0, t1)} Y_i \text{ e } m_{t1} = \min_{i \in (t1, t2)} Y_i, \text{ con } t1 < t2 \text{ e } Y_{t2} > M_{t1}.$$

Definiamo percentuale massima di drawdown:

$$MaxDD = \max_{t1} (m_{t1} / M_{t1} - 1).$$

Essendo il drawdown la differenza in termini assoluti tra un massimo e il successivo minimo, la percentuale massima di drawdown è la perdita massima registrata nell'intervallo di tempo considerato.

3. Tempo medio di recupero (*Recovery*): questo strumento si accompagna e completa quello del drawdown. Si tratta della media dei tempi di recupero o altrimenti detto l'intervallo medio tra l'ultimo massimo e il massimo successivo. Infatti, se è importante sapere che tipo di perdite attendersi è altrettanto importante chiedersi in quanto tempo si è capaci di riprendersi da tali perdite.
4. Media della serie dei rendimenti a 3 mesi (M_3): calcolati lungo tutti i tre periodi di training scalati di una settimana l'uno dall'altro. Questa serie ci dovrebbe permettere di analizzare il trading system nella sua dinamicità e dare una valutazione più chiara sui trading system analizzati.

Effetto procedura selezione dei titoli.

Per valutare l'effetto della selezione dei titoli sono state applicate due procedure che si differenziano solo per il metodo di selezione dei titoli; la selezione dei gruppi, infatti, sarà la stessa e avverrà in maniera del tutto casuale. Per quanto riguarda la procedura di selezione sono stati utilizzati l'indicatore basato sulla forza relativa e una procedura casuale consistente nella estrazione casuale dei titoli da acquistare.

La logica di questa scelta sta nella volontà di eliminare l'effetto congiunto di raggruppamento e selezione. Dal momento che i gruppi sono formati casualmente e in un caso anche la selezione dei titoli è casuale si è scelto di fare 100 replicazioni per ogni strategia analizzata.

In questo caso una valutazione qualitativa basata sul grafico in Figura8 dei rendimenti capitalizzati si pronuncia decisamente a favore di un effetto importante della selezione dei titoli sul risultato finale della strategia: le due serie si allontanano l'una dall'altra sempre di più al passare del tempo. Se confrontiamo, poi, le medie dei rendimenti a tre mesi, quella della strategia basata sulla selezione dei titoli

attraverso la forza relativa è 9,16%, doppia rispetto a quella della strategia completamente casuale che è 3,99% (v. Tabella6). Anche la distribuzione dei rendimenti è più spostata verso valori positivi, più elevati della strategia casuale: tutti i quartini della distribuzione dei rendimenti a tre mesi sono maggiori rispettivamente dei rispettivi quantili della strategia basata su procedure di raggruppamento e selezione casuali.

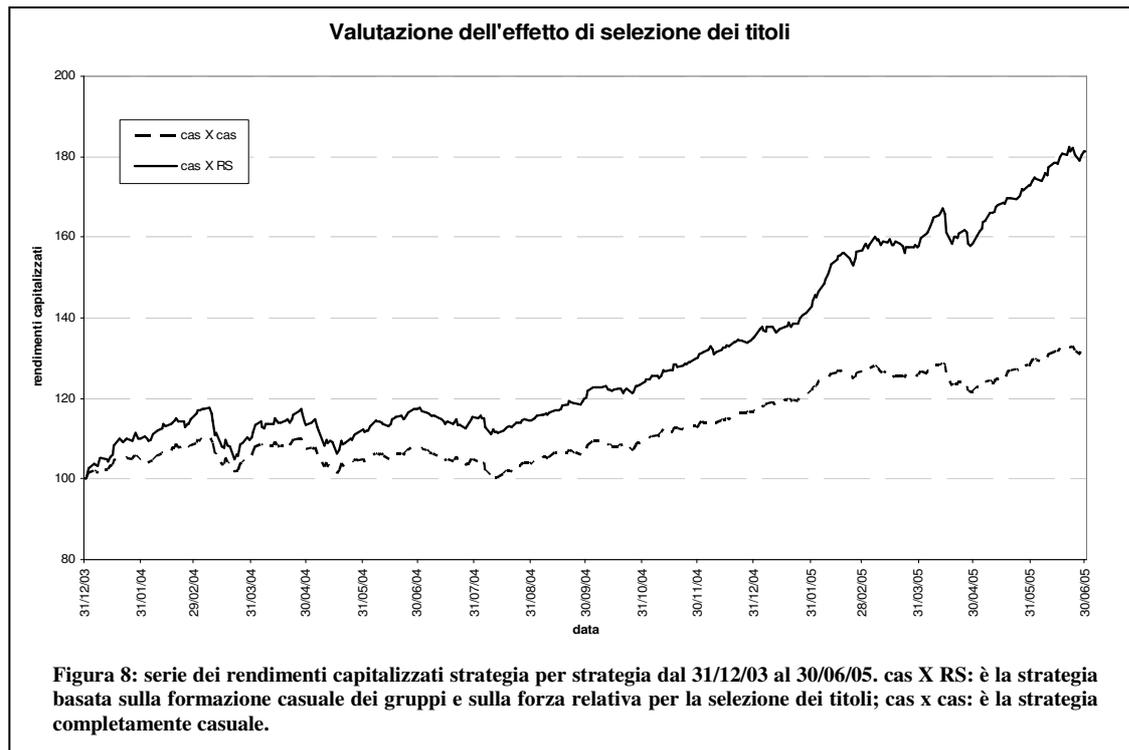


Tabella5.	cas x cas	cas X RS
Max	13,77%	22,69%
3°quartile	8,49%	13,76%
Mediana	3,53%	10,05%
1°quartile	-0,40%	3,02%
Min	-6,32%	-6,31%
Media	3,99%	9,16%

Effetto congiunto raggruppamento - selezione dei titoli.

L'altro effetto che vogliamo considerare è l'effetto congiunto tra la procedura di raggruppamento e selezione dei titoli. Confronteremo, quindi, cinque strategie che si differenziano solamente per il tipo di raggruppamento (i nostri quattro metodi proposti più quello casuale); in questo caso, però, la selezione dei titoli avviene tramite forza relativa e non casualmente. Ovviamente per le quattro strategie

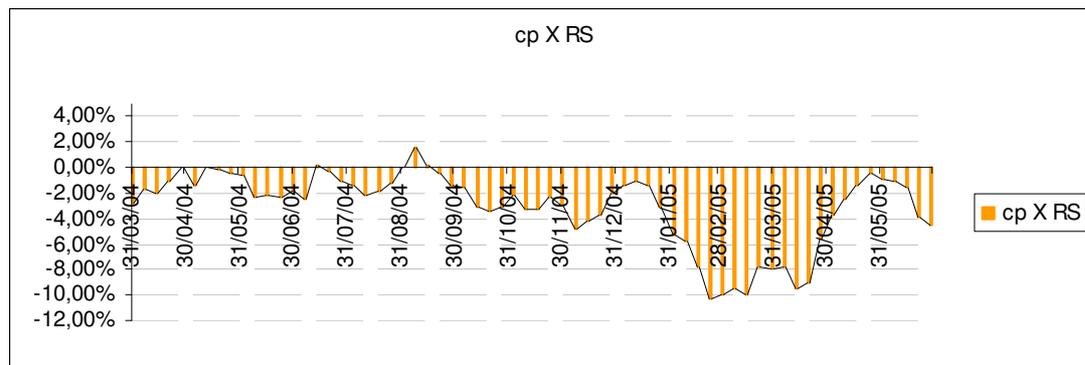
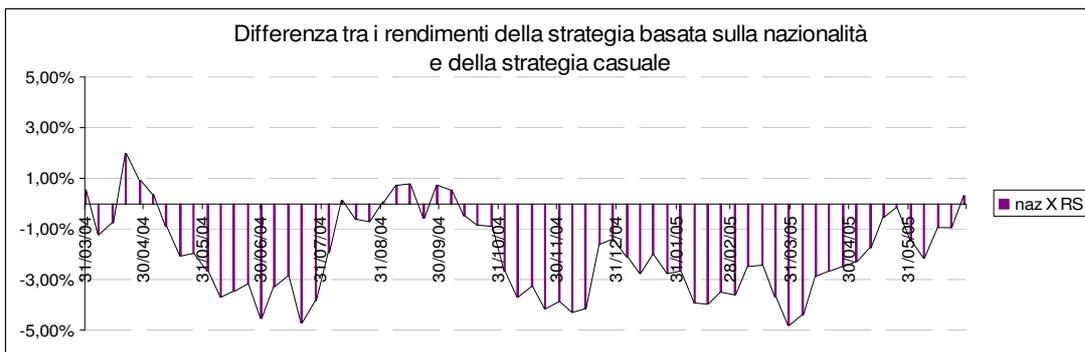
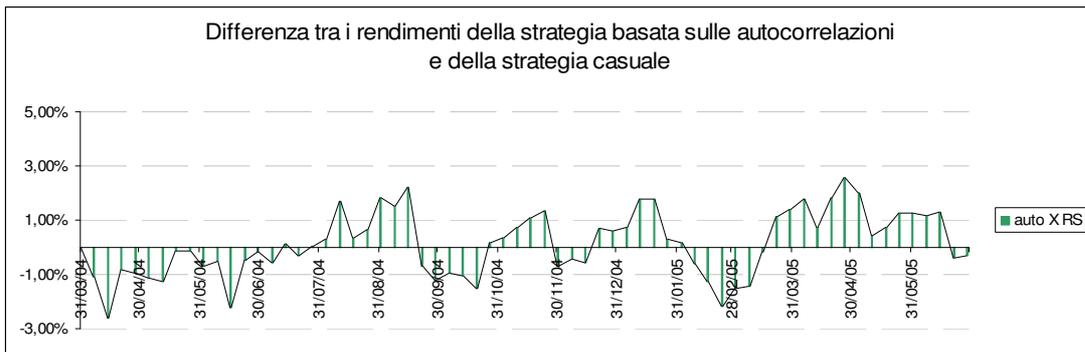
proposte non è necessaria nessuna replicazione dal momento che tutti i parametri sono fissati.

Anche in questo caso si sono calcolati, per ogni strategia, i quartili, il minimo e il massimo della distribuzione dei rendimenti a tre mesi e la media M_3 sono riportati in Tabella6. Guardando alle medie, possiamo rilevare come le due strategie basate l'una sulle autocorrelazioni e l'altra sulle distanze tra modelli GARCH manifestino un effetto congiunto positivo anche se non molto elevato. Lo stesso vale se confrontiamo tra loro le mediane. Per quanto riguarda le strategie con procedura di raggruppamento basata sulla nazionalità o sulle traiettorie l'effetto congiunto è negativo. Quanto detto è visibile graficamente attraverso i grafici delle serie degli extrarendimenti delle strategie rispetto alla strategia casuale solo nel raggruppamento (riportati a partire da fondo pagina).

Le procedure basate sulle traiettorie e sulla nazionalità hanno extrarendimenti quasi sempre negativi mentre le altre due strategie oscillano attorno allo zero e in alcuni periodi, forse condizionate dalle fasi del mercato riescono a fare decisamente meglio con punte di extrarendimento vicino al 4% nel caso della strategia con procedura di raggruppamento basata sulle distanze tra modelli GARCH.

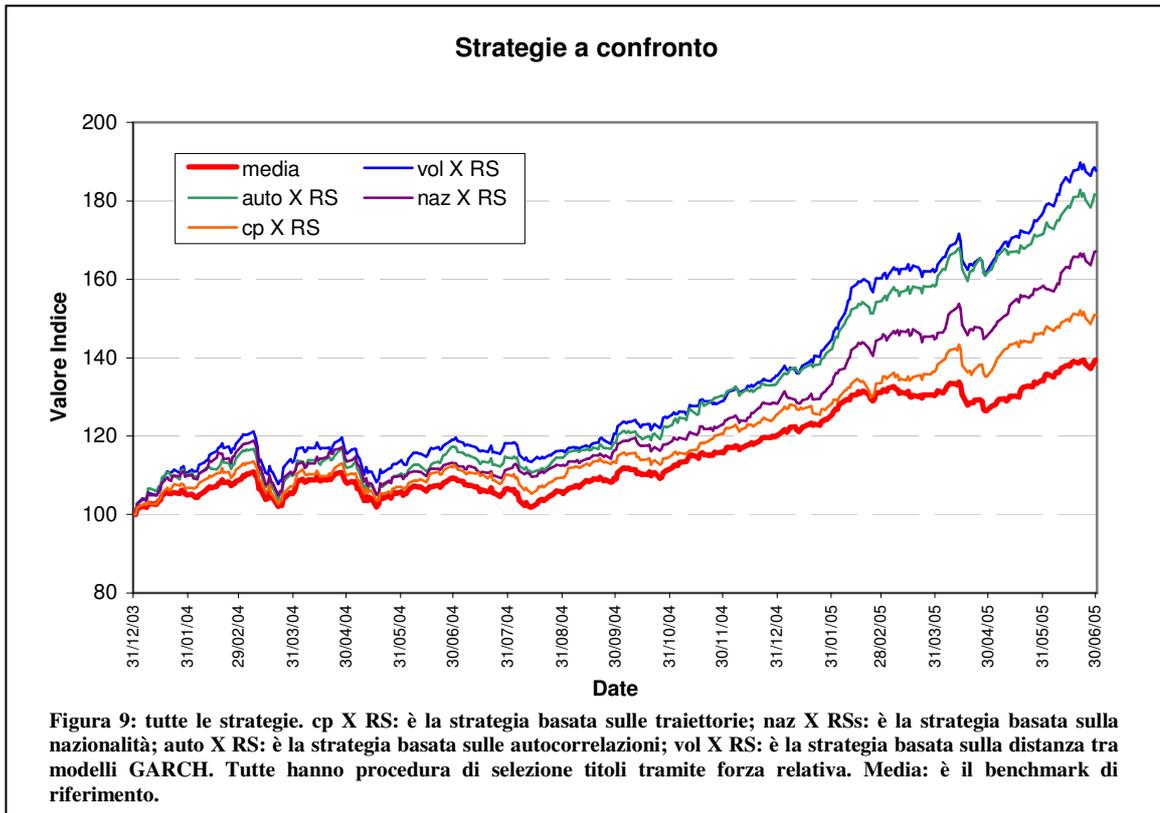
Tabella6.	cas X RS	vol X RS	auto X RS	naz X RS	cp X RS
Max	22,69%	26,08%	23,42%	19,83%	14,27%
3°quartile	14,07%	13,92%	14,71%	11,36%	10,17%
Mediana	10,02%	10,68%	10,40%	8,37%	7,39%
1°quartile	3,23%	2,73%	3,67%	2,70%	2,57%
Min	-4,44%	-6,34%	-5,13%	-7,10%	-5,10%
Media 3 mesi	9,20%	9,47%	9,32%	7,21%	6,13%





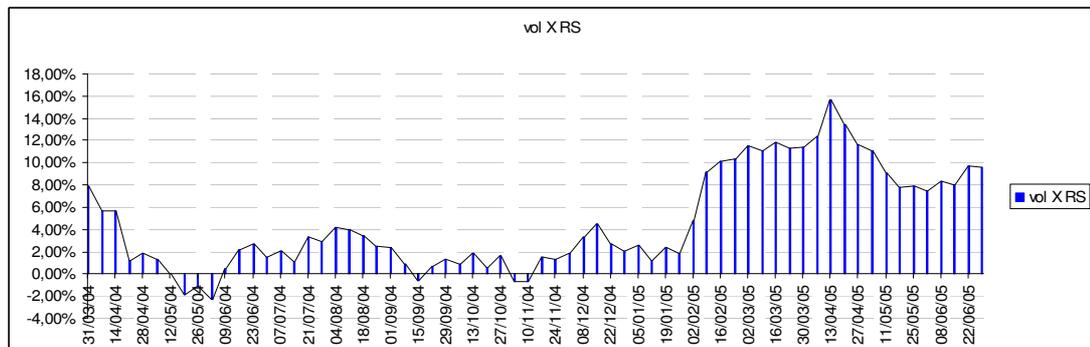
Analisi qualitativa delle strategie proposte.

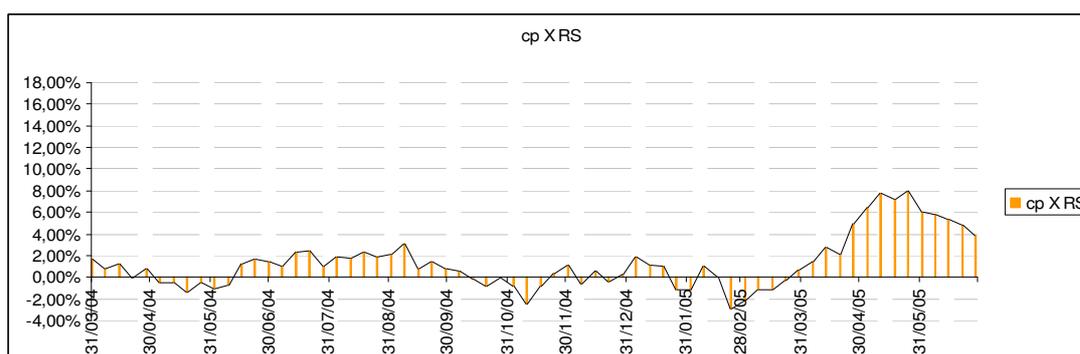
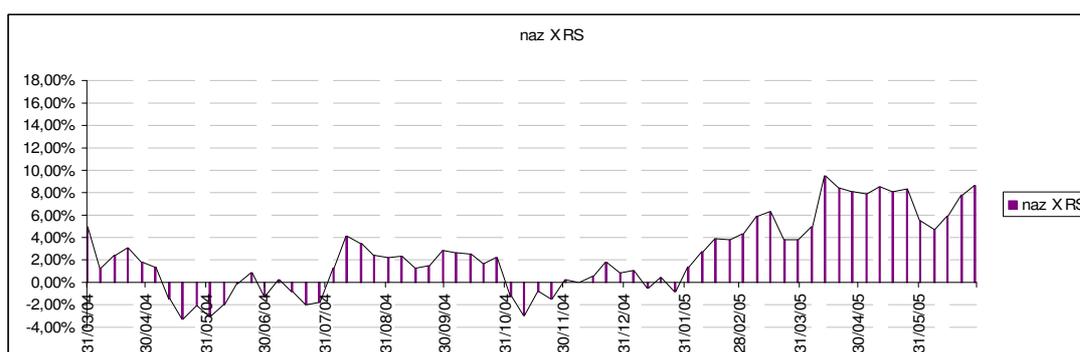
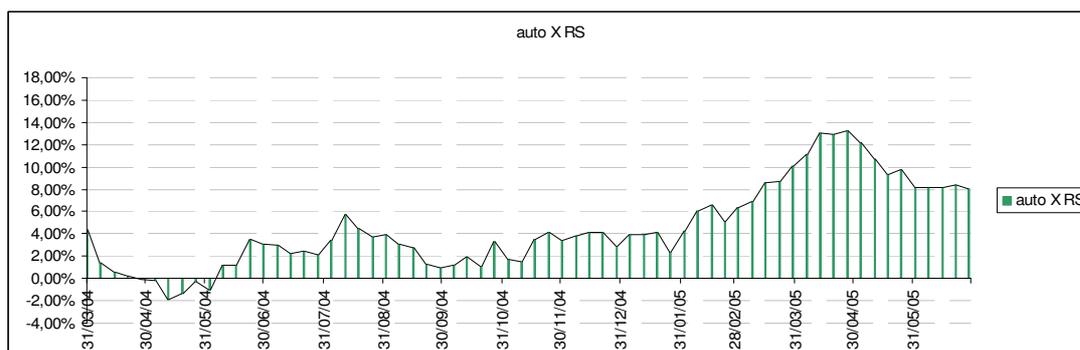
Nella 9 si riporta il grafico dei rendimenti capitalizzati per ognuno dei portafogli gestiti con le diverse strategie dal 31/12/03 al 30/06/05. La prima impressione importante è che tutte le strategie ottengano dei risultati migliori del *benchmark* in termini di rendimento. E questo è molto importante se si considera il fatto che le strategie proposte tengono conto dei costi mentre il portafoglio di riferimento che si è utilizzato come benchmark non ne tiene conto. Alcune strategie in particolare, come quella basata sulle distanze tra modelli GARCH o sulle autocorrelazioni appaiono simili e ottengono un risultato finale in termini di rendimento doppio rispetto al *benchmark*.



Si può notare quanto affermato finora anche dall'analisi delle serie degli extra-rendimenti calcolati sulla serie dei rendimenti a tre mesi dei portafogli relativi alle varie strategie utilizzate rispetto al *benchmark*.

Le serie di questi extra-rendimenti sono riportate nei grafici qui di seguito. Dalla loro analisi emerge l'idea che tali strategie dipendano dall'andamento del mercato e riescano a catturare le migliori performance in presenza di un trend ben definito. Se guardiamo il grafico con rappresentato il *benchmark* in Figura9 notiamo un trend di crescita a partire da agosto 2004 e gli extra-rendimenti maggiori si registrano nel 2005 con punte comprese tra l'8% e il 16% quando il trend è ben definito.





Analisi quantitativa delle strategie proposte.

Riportiamo in tabella gli strumenti che abbiamo sopra descritti. Tutti gli strumenti si riferiscono all'intervallo di tempo che va dal 31/12/03 al 30/06/05.

Tabella7.	vol X RS	auto X RS	naz X RS	cp X RS	benchmark	
Perf.	87,74%	81,54%	67,11%	50,77%	39,33%	Tre periodi di testing
MaxDD	-11,10%	-11,03%	-11,79%	-9,71%	-8,07%	
Recovery	7,33 gg	6,64 gg	8,31 gg	7,48 gg	8,03 gg	
Media 3 mesi	9,47%	9,32%	7,21%	6,13%	4,83%	

L'analisi di questi dati conferma l'analisi qualitativa effettuata sui grafici. La prima osservazione va fatta sui rendimenti: le strategie basate sulla distanza tra modelli GARCH e sulle autocorrelazioni forniscono dei rendimenti doppi rispetto al

benchmark, e questo vale sia nel lungo periodo (vedi performance di periodo) sia mediamente per investimenti di tre mesi. Lo strumento del massimo drawdown unito al tempo medio di recupero evidenzia il fatto che a un rendimento atteso più elevato corrisponde anche una possibile perdita massima più importante. Questo fatto è compensato, però, da un tempo medio di recupero uguale o minore.

L'idea è quindi che le strategie contengano un rischio più elevato rispetto al *benchmark* ma che questo sia il prezzo da pagare per accedere alla possibilità di ottenere rendimenti notevolmente superiori al *benchmark*.

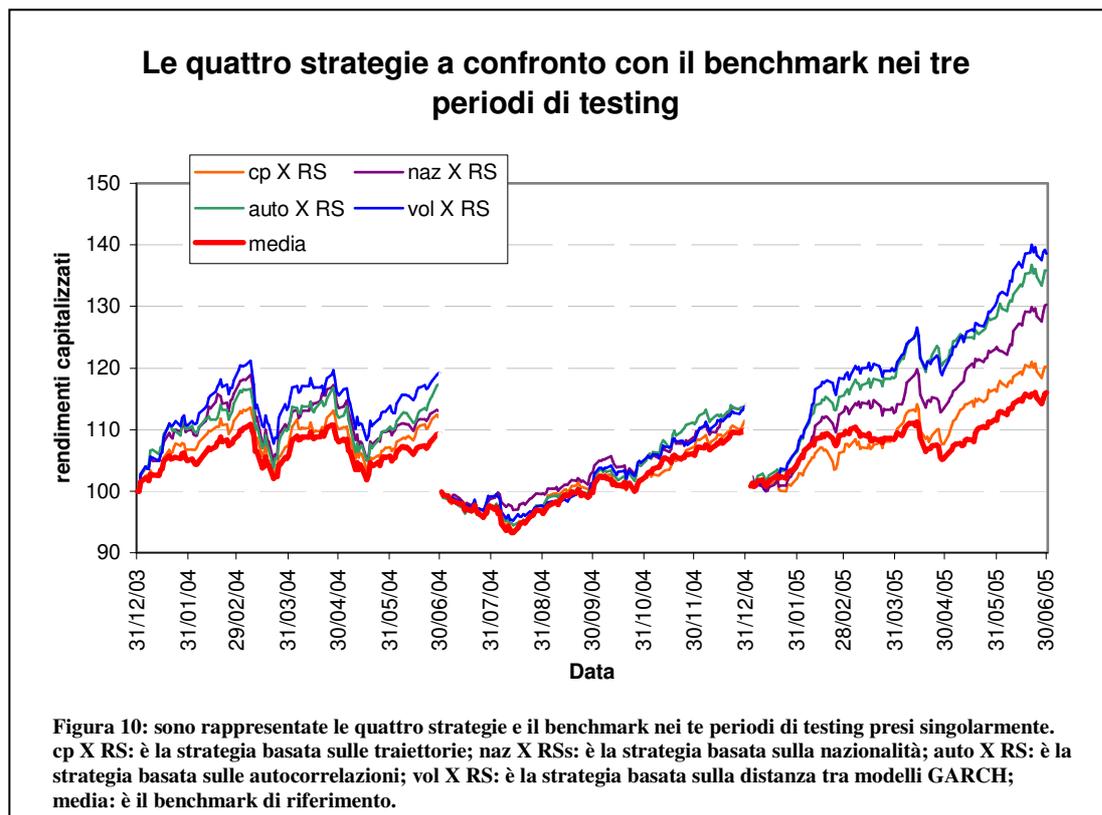
Se guardiamo infine ai quartili della distribuzione dei rendimenti a tre mesi possiamo notare come effettivamente le strategie provochino una dilatazione del campo dei risultati osservati e una traslazione verso l'alto e quindi verso rendimenti positivi maggiori di quelli ottenuti dal *benchmark* negli stessi periodi.

Tabella8.	vol X RS	auto X RS	naz X RS	cp X RS	benchmark
Max	26,08%	23,42%	19,83%	15,45%	14,80%
3°quartile	13,98%	14,49%	11,44%	10,01%	9,47%
Mediana	10,32%	10,61%	8,36%	7,47%	4,33%
1°quartile	2,88%	3,69%	2,77%	2,70%	0,55%
Min	-6,76%	-6,17%	-8,76%	-6,28%	-5,52%

Analisi delle strategie sui singoli periodi di testing.

Per entrare più in profondità nell'analisi è utile verificare se le valutazioni effettuate per tutto il periodo sono valide anche per ciascuno dei sottoperiodi in cui era stata divisa la serie storica. Questo potrebbe darci un indizio ulteriore sull'efficacia dei metodi utilizzati, infatti se in ogni sottoperiodo le serie relative a ciascuna metodologia mantengono caratteristiche costanti di comportamento la fiducia nelle tecniche utilizzate aumenta.

Tabella9.	vol X RS	auto X RS	Naz X RS	cp X RS	Benchmark	
Perf.	19,11%	17,20%	13,05%	12,60%	9,26%	1° periodo di testing
MaxDD	-11,10%	-11,03%	-11,79%	-9,71%	-8,05%	
Recovery	11,78	9,55	10,60	8,92	12,11	
Perf.	13,75%	14,06%	13,48%	11,50%	9,97%	2° periodo di testing
MaxDD	-5,21%	-5,50%	-3,33%	-6,15%	-6,64%	
Recovery	7,36	8,75	7,38	7,54	7,14	
Perf.	38,57%	35,80%	30,26%	20,09%	15,96%	3° periodo di testing
MaxDD	-6,13%	-5,00%	-5,79%	-5,69%	-5,47%	
Recovery	4,35	3,45	4,61	5,00	5,38	



Dall'analisi della Tabella 9 e della Figura 10 appare visibile come in ciascun periodo i portafogli costruiti con le quattro metodologie apportino valore aggiunto rispetto al portafoglio delle medie o a quello costruito casualmente. Il portafoglio costruito in base alle traiettorie e alla forza relativa sembra essere quello meno reattivo e in alcune occasioni rimane al di sotto della curva dei portafogli di riferimento.

I risultati di periodo sono al di sotto degli altri metodi mentre i dati sul drawdown e sul tempo medio di recupero non si discostano molto dal portafoglio di riferimento. I portafogli basati sulla distanza tra i modelli GARCH e la forza relativa o sulle autocorrelazioni e la forza relativa continuano a collocarsi come migliori in termini di performance e sembrano fornire una certa costanza di risultati.

Per tutti sembra valere la regola per la quale se il mercato (inteso come la media dei titoli) ha un trend abbastanza definito la percentuale di massimo drawdown e il tempo medio di recupero sono molto contenuti in linea o inferiori al portafoglio di riferimento mentre se il mercato è caratterizzato da saliscendi, come nel primo periodo, i due valori aumentano più marcatamente e questo incide sulla capacità di ottenere risultati.

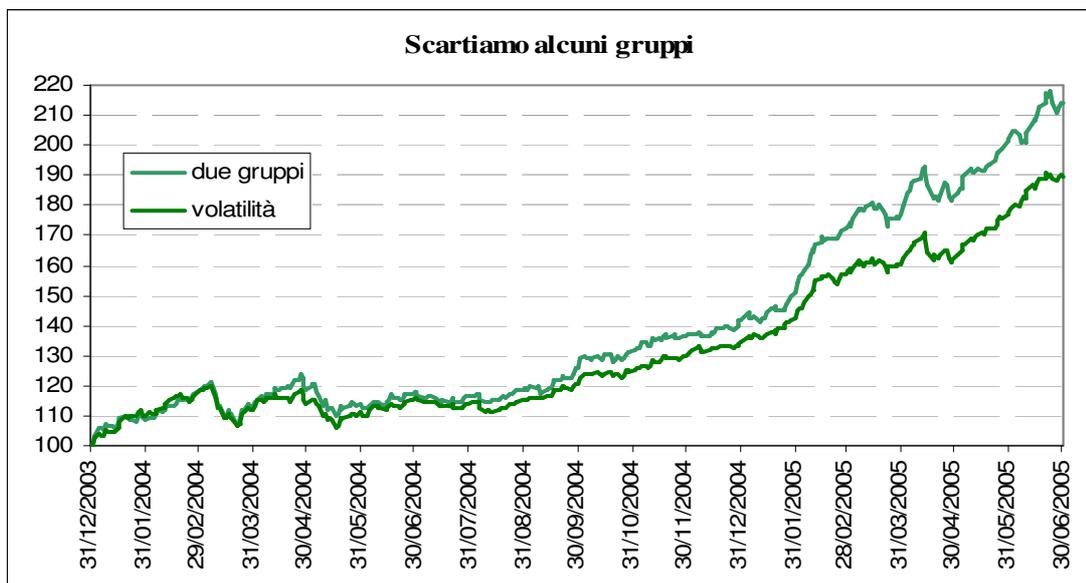
Valutazione della divisione in 5 gruppi.

Un'ultima analisi che si propone è quella di valutare come migliorare eventualmente i già buoni risultati ottenuti con le singole strategie rispetto al benchmark. Abbiamo visto come la selezione dei titoli apporti un effetto considerevole alla strategia e quanto sia importante il tipo di raggruppamento per poter ottenere un effetto congiunto positivo.

Rimane da verificare se le procedure utilizzate per il raggruppamento non possano essere impiegate con uno scopo diverso da quello della divisione dei titoli in 5 gruppi.

Un'ipotesi plausibile potrebbe considerare il fatto che la divisione di tutti i titoli in gruppi possa portare alla suddivisione di titoli con rendimenti superiori alla media e gruppi di titoli con rendimenti minori della media. In questo caso sarebbe opportuno riconoscere i gruppi di titoli che avranno rendimenti inferiori alla media e scartarli prima ancora di applicare la procedura di selezione dei titoli. Ovviamente per far questo abbiamo bisogno di inserire dei vincoli sulle variabili e più sono numerose le variabili più questo diventa difficile.

Si è fatto un tentativo volto a verificare queste ipotesi attraverso una modifica alla strategia basata sulla distanza dei modelli GARCH e la forza relativa. Dal momento che i titoli durante la procedura di raggruppamento vengono sostanzialmente discriminati sulla base dei coefficienti α e β si è ipotizzato di investire solo nei titoli che provenissero da due dei cinque gruppi e in particolare quelli con l' α più elevato. Questo tipo di vincolo è stato applicato a tutti e tre i periodi di testing.



Dalla visione del grafico nella pagina precedente si ha l'impressione che si ha è che l'analisi dei gruppi potrebbe essere utilizzata allo scopo di ridurre il paniere e quindi eliminare quei titoli che rispondono a determinate caratteristiche.

4. Conclusioni.

Lo scopo della tesi è stato valutare le strategie proposte per la gestione automatica dei fondi e in secondo luogo analizzare gli effetti delle singole procedure di raggruppamento e di selezione dei titoli.

Innanzitutto, guardando ai risultati ottenuti complessivamente nei tre periodi di testing esaminati, posso affermare che le strategie proposte ottengono tutte dei buoni risultati. Dal confronto con il *benchmark* di riferimento, nel periodo che va dal 01/01/2004 al 30/06/05, le strategie apportano un incremento, in termini di rendimento finale di periodo, superiore al *benchmark* del 10% - 50% a seconda della strategia utilizzata. Valutando, poi, nello stesso periodo, la media dei rendimenti a tre mesi slittati di una settimana queste considerazioni sono confermate e si può stabilire una certa gerarchia tra le strategie. Per questo strumento la miglior strategia è quella basata sulla formazione dei gruppi attraverso la distanza tra modelli GARCH con selezione dei titoli tramite l'indicatore di forza relativa, che ottiene una media di rendimento a tre mesi pari a 9,47%, quasi il doppio rispetto alla media dei rendimenti del *benchmark*. Di poco inferiore è la strategia basata sulla formazione dei gruppi attraverso la struttura autoregressiva e la selezione dei titoli con la forza relativa con 9,32% contro il 4,83% del *benchmark*, segue la strategia basata sulle informazioni extra-campionarie basata sulle nazionalità (7,21%) e la strategia basata sulle traiettorie (6,13%) entrambe associate alla forza relativa.

Un'analisi che guardi al comportamento delle strategie nei singoli periodi di testing permette di fare un'ulteriore considerazione. L'impressione che si ha guardando ai

periodi separatamente è che le strategie riescano a catturare valore e quindi a fare notevolmente meglio del *benchmark* in determinati periodi quando il trend di mercato è direzionale. Le differenze sono minime, invece, e quindi in linea con il rendimento del *benchmark* nei periodi in cui non c'è un trend ben definito. Questo è visibile sia confrontando in successione le tre performance di periodo sia guardando ai grafici degli extrarendimenti delle strategie rispetto al *benchmark*. Gli extrarendimenti sono più contenuti nel primo e nel secondo periodo, decisamente marcati nel terzo. Ad esempio nel caso della strategia basata sulla distanza tra modelli GARCH e sulla selezione dei titoli tramite forza relativa gli extra-rendimenti in tutto il 2004 non superano quasi mai il 6%, al contrario nella prima parte del 2005 quando il trend è già ben definito sono quasi sempre superiori.

Per valutare gli effetti netti della procedura della selezione dei titoli si è pensato di confrontare le strategie di partenza, opportunamente modificate, con un portafoglio gestito secondo una strategia casuale sia nel raggruppamento che nella selezione dei titoli. Le strategie sono state depurate dell'effetto congiunto al raggruppamento utilizzando una procedura casuale di raggruppamento. I risultati che emergono dal confronto mostrano come ci sia un effetto netto consistente della forza relativa che ha dei rendimenti medi a tre mesi pari a 9,16%, più che doppi rispetto ad una strategia completamente casuale (3,99%). Se guardiamo invece agli effetti congiunti della procedura di raggruppamento con quelli della selezione dei titoli ci accorgiamo che se da un lato alcune procedure di raggruppamento contribuiscono ad un effetto negativo altre invece ne manifestano uno leggermente positivo nel caso, ad esempio della procedura di raggruppamento basata sulla distanza tra modelli GARCH.

5. Riferimenti bibliografici.

- [1] T. Bollerslev, 1986, “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, 31, 307-327.
- [2] F. Cesarini e P. Gualtieri, 2005, *I fondi comuni d’investimento*, Il Mulino.
- [3] D. H. Diggs, R. J. Povinelli, 2003, “A Temporal Pattern Approach for predicting Weekly Financial Time Series”, *Artificial Neural Networks in Engineering5*, St. Louis, Missouri, 707-712.
- [4] T. Di Fonzo, F. Lisi, 2001, *Complementi di statistica economica*, Cleup Editrice.
- [5] C. Dose, S. Cincotti, 2005, “Application to index and enhanced-index tracking portfolio”, *Physica A* 355, 1, 145-151.
- [6] G. Gallo, B. Pacini, 2002, *Metodi quantitativi per i mercati finanziari*, Ed. Carocci.
- [7] M. Gavrilov, D. Anguelov, P. Indyk, R. Motwani, 2000, “Mining The Stock Market: Which Measure is Best?”, *In Proc.of the KDD*, 487-496.
- [8] I. T. Joliffe, 1986, *Principal Component Analysis*, Springer series in statistics.
- [9] K. Kalpakis, D. Gada, V. Puttagunta, 2001, “Distance Measures for Effective Clustering of ARIMA Time-Series”, *In Proceedings of the IEEE International Conference on Data Mining*, 273-280.

- [10] P. J. Kaufman, 2005, *New trading systems and methods*, Wiley Trading.
- [11] M. Liera, 2005, *Capire i mercati finanziari - I Fondi Comuni*, Il Sole 24 ORE.
- [12] S. Micciché, F. Lillo, R. N. Mantenga, 2004, “Correlation Based Hierarchical Clustering In Financial Time Series”, *Proceedings of 31st Workshop of the international School of Solid State Physics*, Erice, Italy.
- [13] E. Otranto, 2004, “Classifying the markets volatility with ARMA distance measures”, *Quaderni di statistica*, 6, 1-19.
- [14] D. Piccolo, 1990, “A distance measure for classifying ARIMA models”, *Journal of Time Series Analysis*, 11, 153-164.
- [15] X. Wang, K. A. Smith, R. Hyndman, D. Alahakoon, 2004, “A Scalable Method for Time Series Clustering”. *Tech. Report Department of Econometrics and Business Statistics Monash University*, Victoria, Australia.
- [16] H. K. Yee, 2005, “Time Series Analysis of Financial Markets”. *Technical Milestone Report*.
- [17] S. Zani, 2000, *Analisi dei dati statistici – Osservazioni multidimensionali*, Giuffré.
- [18] ERRE software: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. R Development Core Team (2005). ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [19] Tseries package: Time series analysis and computational finance. R package version 0.10-0. Compiled by Adrian Trapletti (2005).