

Finanza Quantitativa.

Risk Management e Statistica dei Mercati Monetari e Finanziari

Gianluca Cassese e Matteo Pelagatti

DIPARTIMENTO DI STATISTICA - UNIVERSITÀ MILANO BICOCCA

E-mail address: `gianluca.cassese@unimib.it`

E-mail address: `matteo.pelagatti@unimib.it`

Indice

Calendario delle lezioni	vi
Parte 1. Nozioni elementari	1
Capitolo 1. Gli strumenti finanziari	3
1.1. Strumenti finanziari di base	3
1.2. Strumenti finanziari derivati	5
Capitolo 2. Prezzi e rendimenti	9
2.1. Rendimento uniperiodale	9
2.2. Aggregazione temporale.	10
2.3. Altre definizioni	11
Capitolo 3. Fatti stilizzati	13
3.1. σ domina μ	13
3.2. Mancanza di correlazione	13
3.3. Correlazione quadratica positiva	14
3.4. Non normalità	14
3.5. Esercizi	16
Parte 2. Scelta del portafoglio	17
Capitolo 4. Un'introduzione al modello media varianza	19
4.1. Introduzione	19
4.2. Il caso di due soli titoli.	20
Capitolo 5. Il caso generale.	23
5.1. L'analisi grafica.	23
5.2. Una conclusione generale.	25
5.3. Implicazioni.	27
5.4. L'agente rappresentativo	30
Capitolo 6. Il teorema dei due fondi.	33
6.1. Il Teorema	33
6.2. L'interpretazione finanziaria	35
Capitolo 7. L'esistenza di un titolo privo di rischio	37
7.1. Il CAPM con titolo privo di rischio	37
7.2. La <i>market line</i>	37

Capitolo 8. Il calcolo dei portafogli efficienti.	41
8.1. Matrice Σ invertibile.	41
8.2. Matrice Σ singolare.	43
Capitolo 9. Estensioni	45
9.1. La scelta di portafoglio in generale	45
9.2. La frontiera efficiente	46
Esercizi.	49
Parte 3. Misure di rischio	51
Introduzione	53
Capitolo 10. Le Misure Coerenti del Rischio	55
Capitolo 11. Value-at-Risk	57
11.1. Definizione	57
11.2. Aspetti statistici	59
Capitolo 12. Expected Shortfall	61
12.1. Definizione	61
12.2. La distribuzione normale	62
12.3. La distribuzione t di Student	63
12.4. L'approccio dei quantili	64
Capitolo 13. Extreme Value Theory	67
13.1. Definizione	67
13.2. Stima	68
13.3. Calcolo dei quantili	69
13.4. EVT misure di rischio	69
Capitolo 14. Accordi di Basilea, Misure di Rischio e Coerenza	71
14.1. La Regolamentazione del Rischio Secondo Basilea	71
14.2. Le Misure di Rischio Coerenti	72
Capitolo 15. Tests	75
15.1. Alcuni Tests per il VaR .	75
15.2. Test che utilizzano un insieme allargato delle informazioni.	76
15.3. Un test sulla distribuzione.	77
Esercizi	79
Parte 4. Analisi Empirica dei Mercati Finanziari: Teoria ed Applcazioni	81
Capitolo 16. Richiami di serie storiche stazionarie e integrate	83
16.1. Richiami di probabilità	83
16.2. Previsioni statistiche	89
16.3. Serie storiche e processi stocastici stazionari	93

16.4. Processi ARMA stazionari	95
16.5. Stima di processi ARMA	100
16.6. Processi integrati	104
Capitolo 17. Processi GARCH	111
17.1. Il processo ARCH	112
17.2. Il processo GARCH	113
Capitolo 18. Inferenza per processi GARCH	119
18.1. Stime di Quasi Massima Verosimiglianza	119
18.2. Test di restrizioni sui parametri	123
Capitolo 19. Estensioni della modellistica GARCH	125
19.1. Modelli GARCH asimmetrici	125
19.2. Il premio a rischio ed il modello GARCH-M	128
Capitolo 20. Modelli GARCH multivariati	131
20.1. Processi autoregressivi vettoriali (VAR)	131
20.2. GARCH multivariati	132
20.3. Rappresentazione BEKK del GARCH multivariato	134
20.4. Modelli GARCH multivariati basati sulla stima di modelli GARCH univariati	134
Parte 5. Strumenti derivati sul rischio di credito	139
Capitolo 21. Derivativi sul Rischio di Credito.	141
21.1. Introduzione	141
21.2. Prodotti Basati sul Rischio di <i>Default</i> .	142
21.3. Total Return <i>SWAPS</i> (<i>TRS</i>).	144
21.4. Credit Spread Options, Futures e <i>SWAPS</i> .	144
21.5. Credit Linked Notes.	145
21.6. Obbligazioni Strutturate	145
21.7. Repackaged Notes	145
Parte 6. Modelli binomiali di prezzo.	149
Capitolo 22. Il modello	151
22.1. Alberi	151
22.2. I mercati	152
Capitolo 23. Il principio di valutazione	155
23.1. Il prezzo di replica	155
23.2. Risk-Neutral Pricing	162
Appendici	167
Alcune semplici regole su varianza e covarianza	169
Bibliografia	171

Calendario delle lezioni

Giorno	Data	Orario	Aula	Docente	Argomento
Lunedì	21/11	08:30 - 10:30	Lab 237	Cassese	Introduzione. Le attività finanziarie in genere
Lunedì	21/11	10:30 - 13:30	Lab 237	Pelegatti	Fatti stilizzati
Martedì	22/11	08:30 - 11:30	Lab 237	Cassese	I tassi di rendimento.
Martedì	22/11	11:30 - 13:30	Lab 237	Pelagatti	Introduzione a R
Mercoledì	23/11	08:30 - 11:30	Lab 237	Cassese	Il Modello Media/Varianza
Venerdì	25/11	08:30 - 11:30	Lab 237	Pelegatti	Introduzione a R
Lunedì	28/11	08:30 - 10:30	Lab 237	Cassese	Il Modello Media/Varianza
Lunedì	28/11	10:30 - 13:30	Lab 237	Pelegatti	Introduzione a R - Esercitazione
Martedì	29/11	08:30 - 11:30	Lab 237	Cassese	Il Modello Media/Varianza
Martedì	29/11	11:30 - 13:30	Lab 237	Pelagatti	Econometria della frontiera efficiente
Mercoledì	30/11	08:30 - 11:30	Lab 237	Cassese	Il Modello Media/Varianza
Venerdì	02/12	08:30 - 11:30	Lab 237	Pelegatti	Econometria della frontiera efficiente - Esercitazione
Lunedì	05/12	08:30 - 10:30	Lab 237	Cassese	Il Modello Media/Varianza - Esercitazione
Lunedì	05/12	10:30 - 13:30	Lab 237	Pelegatti	Cenni di serie storiche
Martedì	06/12	08:30 - 11:30	Lab 237	Cassese	Misure di rischio
Martedì	06/12	11:30 - 13:30	Lab 237	Pelagatti	Modelli di volatilità
Venerdì	09/12	08:30 - 11:30	Lab 237	Pelegatti	Modelli di volatilità
Lunedì	12/12	08:30 - 10:30	Lab 237	Cassese	Misure di rischio
Lunedì	12/12	10:30 - 13:30	Lab 237	Pelegatti	Modelli di volatilità - Esercitazione
Martedì	13/12	08:30 - 11:30	Lab 237	Cassese	Misure di rischio
Martedì	13/12	11:30 - 13:30	Lab 237	Pelagatti	Modelli di volatilità e correlazione
Mercoledì	14/12	08:30 - 11:30	Lab 237	Cassese	Misure di rischio - Esercitazione
Venerdì	16/12	08:30 - 11:30	Lab 237	Pelegatti	Modelli di volatilità
Lunedì	19/12	08:30 - 10:30	Lab 237	Cassese	Backtesting
Lunedì	19/12	10:30 - 13:30	Lab 237	Pelegatti	Modelli di volatilità e correlazione
Martedì	20/12	08:30 - 11:30	Lab 237	Cassese	Backtesting
Martedì	20/12	11:30 - 13:30	Lab 237	Pelagatti	Modelli di volatilità e correlazione - Esercitazione
Mercoledì	21/12	08:30 - 11:30	Lab 237	Cassese	Backtesting - Esercitazione
Lunedì	09/01	08:30 - 10:30	Lab 237	Cassese	Derivati e rischio di credito
Lunedì	09/01	10:30 - 13:30	Lab 237	Pelegatti	Modelli di volatilità - Esercitazione
Martedì	10/01	08:30 - 11:30	Lab 237	Cassese	Derivati e rischio di credito
Martedì	10/01	11:30 - 13:30	Lab 237	Pelagatti	Modelli di volatilità e correlazione
Mercoledì	11/01	08:30 - 11:30	Lab 237	Cassese	Esercitazione
Venerdì	13/01	08:30 - 11:30	Lab 237	Pelegatti	Modelli di volatilità

Esame: l'esame si svolgerà in forma di esercizi al computer.

Parte 1

Nozioni elementari

Gli strumenti finanziari

Prima di iniziare ad occuparci dei metodi statistici, è utile dare una definizione dei diversi strumenti e contratti finanziari, la cui valutazione e previsione saranno il principale argomento del corso.

1.1. Strumenti finanziari di base

1.1.1. Azioni. Le *azioni* sono quote di partecipazione nel capitale di una società. Le azioni possono essere scambiate in mercati regolamentati, le borse, o al di fuori di essi. Spesso le piccole aziende nascono dall'idea di un imprenditore, che per realizzarla reperisce finanziamenti attraverso una rete di amici, familiari e amici degli amici, o di *venture capitalists*, che entrano a far parte del capitale dell'azienda in cambio di futuri e incerti profitti. Quando l'azienda è di successo e si ingrandisce, capita spesso che coloro i quali forniscono capitale all'azienda e l'azienda stessa non abbiano più rapporti diretti e le azioni dell'azienda vengano quotate e scambiate in un mercato regolamentato. La valutazione che gli investitori faranno dell'azione di un'azienda sarà basata sulle informazioni e sulle aspettative che essi hanno nei riguardi dell'azienda, oltre che sulle informazioni ed aspettative che essi hanno dello stato generale dell'economia di cui l'azienda è parte.

Spesso il possessore di azioni partecipa al profitto dell'azienda attraverso ai *dividendi*. I dividendi sono somme che la società paga agli azionisti, in proporzione al numero di azioni da essi possedute. In Italia i dividendi vengono distribuiti una volta all'anno e tipicamente vengono resi noti circa un mese prima. Quando si acquista un'azione si può avere (*cum*) o meno (*ex*) diritto alla percezione del successivo dividendo. Ovviamente nel momento in cui l'azione passa da *cum* a *ex*, il suo prezzo scende di una quantità pari al dividendo.

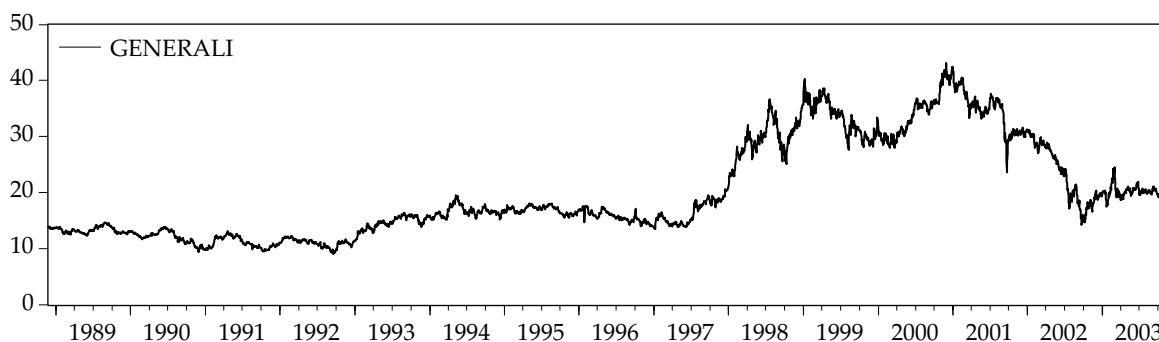


FIGURA 1. prezzo giornaliero del titolo Generali nel periodo 24.11.1988–24.11.2003.

1.1.2. Valute. Le valute ed i relativi tassi di cambio sono un altro strumento finanziario fondamentale. Alcune valute sono ancorate ad altre per mezzo di tassi di cambio fissi, altre sono libere di fluttuare secondo mercato. Ovviamente per i tassi di cambio deve valere la proprietà transitiva, cioè, il tasso di cambio tra la

valuta x e la valuta z (escludendo costi di transazione) deve essere il medesimo che si otterrebbe cambiando la valuta x con la valuta y , e questa con la valuta z . Se così non fosse si creerebbero possibilità di *arbitraggio* (cioè di profitto certo senza bisogno di investimento iniziale) comprando la valuta al tasso più favorevole e vendendola a quello più svantaggioso per l'acquirente.

Il tasso di cambio tra due valute è legato ai tassi di interesse ottenibili con investimenti in ciascuna delle due valute: infatti se si alza il tasso d'interesse per i titoli in valuta x rispetto a quelli in valuta y , a parità di altre condizioni, gli investitori acquisteranno investimenti nella valuta x fino a che l'aggiustamento sul tasso di cambio, che deriverà dall'eccesso di domanda per la valuta x (apprezzamento di x), riporterà i tassi ad un valore di equilibrio. Le banche centrali possono (entro certi limiti) manovrare i tassi opportunamente per influenzare i tassi di cambio.

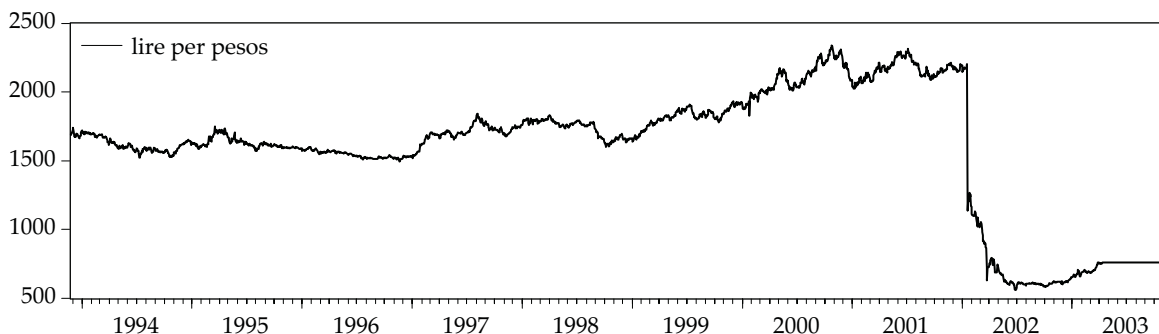


FIGURA 2. tasso di cambio giornaliero pesos–lire nel periodo 24.11.1993–24.11.2003.

1.1.3. Commodities. Alcune merci (*commodities*), quali metalli (preziosi o meno), petrolio, prodotti agricoli e forestali, sono spesso scambiati sui mercati finanziari, non in funzione di un loro consumo o di una loro commercializzazione, ma per fini speculativi o di riduzione del rischio finanziario.

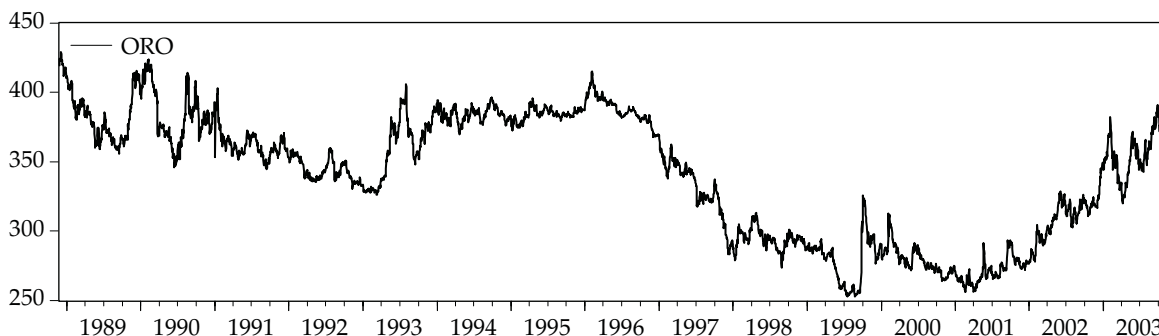


FIGURA 3. prezzo giornaliero dell'oro in dollari per oncia nel periodo 24.11.1988–24.11.2003.

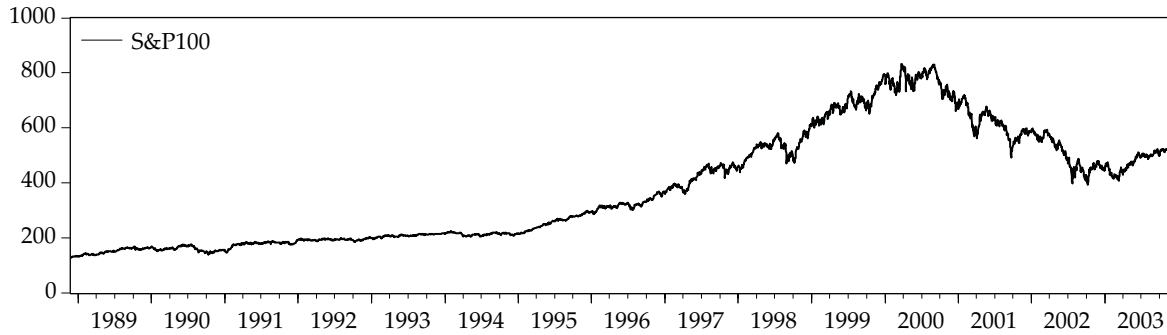


FIGURA 4. indice S&P100 giornaliero nel periodo 24.11.1988–24.11.2003.

1.1.4. Titoli a reddito fisso. Oltre che per mezzo dello strumento azionario, le società, e anche i governi, possono finanziarsi emettendo *obbligazioni*. Una obbligazione è un titolo che, una volta acquistato, frutta un capitale in un istante futuro prestabilito (data di scadenza). Una obbligazione può, inoltre, prevedere il pagamento di *cedole* durante il periodo intercorrente l'acquisto e la scadenza della obbligazione, normalmente con intervalli semestrali o annuali. Una obbligazione senza cedole è detta, con terminologia anglosassone, *zero-coupon bond*, mentre una con cedole prende il nome di *coupon-bearing bond*. In linea di massima, obbligazioni con durata (la differenza temporale tra la data di scadenza e la data di acquisto) più lunga offrono interessi più alti.

Vi sono anche obbligazioni che prevedono che il capitale finale e le cedole dipendano da qualche tasso di interesse ufficiale (per esempio il LIBOR e EURIBOR) e/o dal tasso di inflazione della valuta in cui sono emesse.

1.1.5. Indici. Gli indici sono i valori di panieri (portafogli) di titoli finanziari, utilizzati in genere per valutare la performance di un certo mercato. Gli indici di borsa (per esempio MIB30, S&P500, DAX, CAC, Dow Jones Industrial Average, ecc.) sono costruiti scegliendo i titoli delle società a più alta capitalizzazione oppure appartenenti ad alcuni specifici settori industriali. Esistono anche indici che si riferiscono ad obbligazioni come l'EMBI+ (*Emerging Market Bond Index* di JP Morgan), che misura il valore del debito dei paesi emergenti o il MAE All Bond Index, che include le emissioni obbligazionarie del governo argentino.

Gli indici non sono veri e propri titoli finanziari, essendo dei portafogli virtuali di questi ultimi; tuttavia gli strumenti derivati (si veda oltre) fanno grosso uso di essi, e diversi fondi replicano il loro andamento.

1.2. Strumenti finanziari derivati

Gli strumenti derivati sono contratti basati sull'andamento di uno o più titoli sottostanti (il sottostante può essere anche un indice).

1.2.1. Contratti a termine e futures. Un *contratto a termine* o *forward* è un accordo tra due parti, in cui una parte si impegna ad acquistare dall'altra un'attività finanziaria ad una data futura e ad un prezzo stabilito nel contratto.

Un contratto *future* è come un contratto a termine; tuttavia mentre i forward sono trattati direttamente dalle due parti (*over the counter*), i future sono scambiati in mercati regolamentati. Inoltre, il profitto

derivante da una posizione *future* viene calcolato giornalmente e regolato tra le parti per mezzo di un apposito conto al termine della giornata di contrattazione.

Senza addentrarci troppo nel mondo dei derivati, vediamo come l'applicazione del semplice *principio di non arbitraggio* permetta di valutare il prezzo a termine di un contratto future. Sia $S(t)$ il prezzo del sottostante, supponiamo un'azione, al momento t di stipula del contratto e $S(T)$ il prezzo del sottostante all'istante T , termine del contratto. Sia inoltre F il prezzo a termine, cioè il prezzo a cui dovremo acquistare il sottostante alla scadenza del contratto. Si metta in atto la seguente strategia: si entri in un contratto a termine e contemporaneamente si venda il sottostante. Questa strategia si chiama *andare corti* ed implica la vendita di un'attività che non si detiene. Questa *vendita allo scoperto* è consentita, con talune limitazioni, in diversi mercati. Si può pensare come alla possibilità di individuare qualche d'uno che sia disposto a prestarci il sottostante che ci serve, ovviamente con la promessa che esso sarà poi restituito (al tempo T). Alla scadenza del contratto si riceve il sottostante $S(T)$, con il quale si chiude la posizione corta, e si paga per esso la cifra F . Nel frattempo, nel periodo $(T - t)$, il denaro ricevuto per la vendita allo scoperto di $S(t)$, sarà stato depositato in banca fruttando un interesse R_{T-t} e diventando, all'istante T ,

$$S(t)(1 + R_{T-t}).$$

Dato che il valore del contratto al tempo t è zero, e la nostra posizione a scadenza è $S(t)(1 + R_{T-t}) - F$, escludendo la possibilità di arbitraggio (avere un profitto certo senza esborso), la nostra posizione a scadenza, perveredibile senza incertezza (se l'interesse R_{T-t} è fisso), deve quindi essere nulla, da cui

$$F = S(t)(1 + R_{T-t}).$$

Come vedremo più avanti, in finanza si utilizzando generalmente interessi a tempo continuo per cui $(1 + R_{T-t}) = \exp\{r(T - t)\}$ e

$$F = S(t)e^{r(T-t)}.$$

	tempo t	tempo T
Contratto	0	$S(T) - F$
Sottostante	$-S(t)$	$-S(T)$
Contante	$S(t)$	$S(t)e^{r(T-t)}$
Totale	0	$S(t)e^{r(T-t)} - F$

TABELLA 1. Schema dei valori coinvolti da un contratto future.

1.2.2. Opzioni.

Opzioni europee. Una *opzione europea* è un contratto tra due parti in cui una acquista il diritto (ma non l'obbligo) di comprare (*opzione call*) o vendere (*opzione put*) un particolare sottostante (un titolo, un indice, una commodity, una valuta, ecc.) da o a l'altra parte, ad un prezzo stabilito (*prezzo d'esercizio* o *strike price*), ad una specifica data futura (*data di scadenza*).

Opzioni americane. Una *opzione americana* è un contratto tra due parti in cui una acquista il diritto (ma non l'obbligo) di comprare (*opzione call*) o vendere (*opzione put*) un particolare sottostante da o a l'altra parte, ad un prezzo stabilito (*prezzo d'esercizio* o *strike price*), entro una specifica data futura (*data di scadenza*).

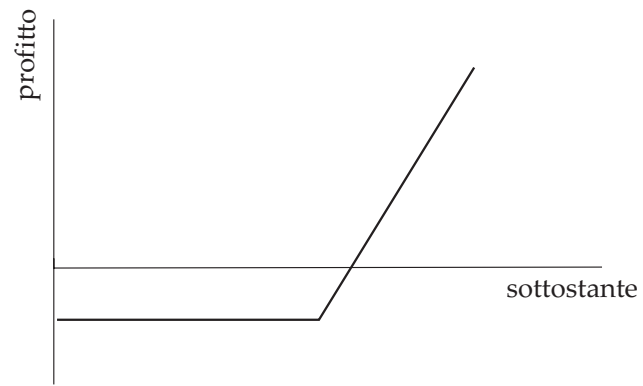


FIGURA 5. diagramma di profitto di una pozione call.

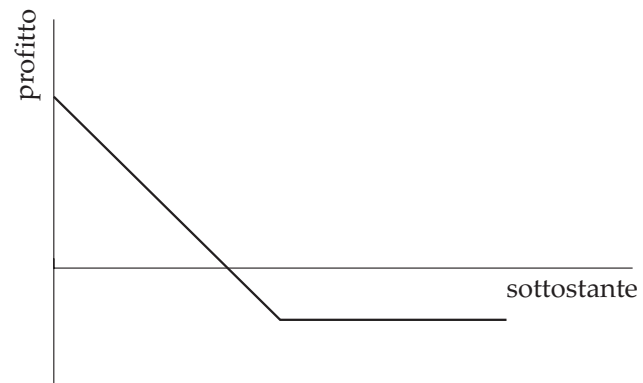


FIGURA 6. diagramma di profitto opzione put.

Prezzi e rendimenti

2.1. Rendimento uniperiodale

Il tasso di rendimento relativo ad una determinata attività finanziaria esprime in modo sintetico il flusso di pagamenti percepiti dal detentore dell'attività nell'arco di un periodo di riferimento. La quantità così calcolata viene espressa in rapporto al valore iniziale dell'attività stessa all'inizio del periodo. La specificazione del periodo è dunque un elemento essenziale.

ESEMPIO 1. Si considerino due date, t e $t + 1$. Allora

$$P_t \vdash \text{-----} \dashv P_{t+1}$$

e scriviamo dunque

$$(2.1.1) \quad R_{t,t+1} = \frac{P_{t+1}}{P_t} \quad r_{t,t+1} = R_{t,t+1} - 1 = \frac{P_{t+1} - P_t}{P_t}$$

La quantità $r_{t,t+1}$ viene in genere indicata come tasso di rendimento netto mentre $R_{t,t+1}$ indica il rendimento lordo.

Molto spesso i titoli offrono anche pagamenti intermedi nella forma di dividendi (per le azioni) o di cedole (per le obbligazioni). Può avere un certo significato tenere distinta la fonte del rendimento, specificando se esso sia di natura capitale o reddituale.

ESEMPIO 2. In termini dell'esempio precedente

$$P_t \vdash \underbrace{\text{-----}}_{D_{t,t+1}} \dashv P_{t+1}$$

Con $D_{t,t+1}$ indichiamo un qualunque pagamento maturato ad una data intermedia tra t e $t + 1$. Scrivere D_t oppure D_{t+1} è una questione convenzionale. Non è invece irrilevante il fatto che in taluni casi il pagamento intermedio potrebbe non essere noto che alla data $t + 1$. Definiamo quindi

$$(2.1.2) \quad r_{t,t+1}^d = \frac{P_{t+1} + D_{t,t+1}}{P_t} - 1 = \frac{P_{t+1}}{P_t} - 1 + \frac{D_{t,t+1}}{P_t} = r_{t,t+1} + \underbrace{\frac{D_{t,t+1}}{P_t}}_{\text{"dividend yield"}}$$

Molto spesso si scrive il dividend/yield come $\delta_{t,t+1} = P_t^{-1} D_{t,t+1}$. Analogamente

$$(2.1.3) \quad R_{t,t+1}^d = r_{t,t+1}^d + 1 = \frac{P_{t+1} + D_{t,t+1}}{P_t}$$

Una questione che spesso si pone è quella relativa alla determinazione del tasso di rendimento di un portafoglio e della relazione che intercorre tra questo ed il rendimento delle attività che compongono il portafoglio stesso. Questo problema si applica anche agli indici azionari, quali il *S&P500* o il *MIB30*.

Indichiamo con a_t^n la quantità acquistata del titolo n -mo alla data t : dunque $a_t^n \geq 0$. Allora il valore di mercato del portafoglio, ovvero il suo prezzo di acquisto alla data t è necessariamente pari a

$$P_t^\pi = \sum_{n=1}^N a_t^n P_t^n$$

dunque il rendimento lordo sarà pari a

$$\begin{aligned} (2.1.4) \quad R_{t,t+1}^\pi &= \frac{P_{t+1}^\pi}{P_t^\pi} = \frac{\sum_{n=1}^N a_t^n P_{t+1}^n}{\sum_{n=1}^N a_t^n P_t^n} \\ &= \sum_{n=1}^N \frac{P_{t+1}^n}{P_t^n} \left(\frac{a_t^n P_t^n}{\sum_{n=1}^N a_t^n P_t^n} \right) \\ &= \sum_{n=1}^N R_{t,t+1}^n \pi_t^n \end{aligned}$$

dove

$$\pi_t^n = \frac{a_t^n P_t^n}{\sum_{n=1}^N a_t^n P_t^n}$$

è il peso del titolo n nel portafoglio. Dunque si conclude

$$(2.1.5) \quad r_{t,t+1}^\pi = \sum_{n=1}^N r_{t,t+1}^n \pi_t^n$$

Se sono inclusi anche dei dividendi, allora

$$R_{t,t+1}^{\pi,d} = \frac{P_{t+1}^\pi + D_{t,t+1}^\pi}{P_t^\pi} = \frac{\sum_{n=1}^N a_t^n (P_{t+1}^n + D_{t,t+1}^n)}{\sum_{n=1}^N a_t^n P_{n,t}} = \sum_{n=1}^N R_{t,t+1}^{n,d} \pi_t^n$$

In altre parole, il rendimento di un portafoglio – netto o lordo – non è che il portafoglio dei rendimenti dei singoli titoli che lo compongono. Dobbiamo però osservare che questa conclusione dipende in modo cruciale dal fatto che la composizione del portafoglio rimanga invariata nel periodo $(t, t+1)$. Dunque in linea di principio se la durata del periodo considerato è relativamente breve si può applicare la (2.1.4) ma una tale relazione è facilmente errata se la durata dell'investimento è sufficientemente lunga perché la composizione del portafoglio sia mutata. Questa semplice constatazione è spesso trascurata nelle applicazioni.

Nel caso generale infatti si ottiene:

$$R_{t,t+1}^\pi = \sum_{n=1}^N R_{t,t+1}^n \pi_{t,t+1}^n$$

dove

$$\pi_{t,t+1}^n = \frac{a_{t+1}^n P_t^n}{\sum_{n=1}^N a_{t+1}^n P_t^n}$$

Dunque in generale $\sum_n \pi_{t,t+1}^n \neq 1$; inoltre i pesi $\pi_{t,t+1}^n$ non possono considerarsi noti alla data t .

2.2. Aggregazione temporale.

Naturalmente la durata del periodo di investimento può variare da un progetto all'altro. Ha dunque importanza tenere conto di questo elemento nella definizione del tasso di rendimento. Idealmente sarebbe opportuno esprimere il tasso di rendimento nella medesima scala temporale in modo da consentire il raffronto tra progetti di investimento con durata differente. Se ad esempio un titolo ci promette un rendimento del 4% su 2 anni ed un altro del 5% su 6 anni non abbiamo modo di determinare in via diretta quale sia il titolo col rendimento maggiore. La convenzione, in generale, è quella di esprimere il tasso di rendimento o di interesse su base annua. Vi sono tuttavia diversi modo di farlo.

Partiamo quindi dalla definizione di rendimento multiperiodale, del tutto analoga alla (2.1.1) (facciamo astrazione dai dividendi):

$$(2.2.1) \quad R_{t,t+k} = \frac{P_{t+k}}{P_t} \quad r_{t,t+k} = R_{t,t+k} - 1$$

Naturalmente, quanto più lunga è la durata k dell'investimento considerato tanto maggiore possiamo aspettarci che sia il rendimento. Per ricavare una misura standardizzata del rendimento e priva di questo difetto possiamo utilizzare il concetto di media geometrica e definire

$$(2.2.2) \quad R_{k,t} = \sqrt[k]{\frac{P_{t+k}}{P_t}} \quad r_{k,t} = \sqrt[k]{\frac{P_{t+k}}{P_t}} - 1$$

ossia il rendimento medio per ciascuno dei periodi nell'intervallo $t \rightarrow t+k$. Evidentemente,

$$(2.2.3) \quad P_{t+k} = P_t(1 + r_{k,t})^k = P_t R_{k,t}^k$$

Il legame che intercorre tra rendimento uniperiodale e multiperiodale è facilmente dato dalla

$$(2.2.4) \quad R_{t,t+k} = \prod_{i=1}^k \frac{P_{t+i}}{P_{t+i-1}} = \prod_{i=1}^k R_{t+i-1,t+i}$$

Questa relazione è importante poiché stabilisce un legame chiaro tra rendimento ad un periodo e rendimento multiperiodale. Sfortunatamente, si tratta di una relazione piuttosto complessa dal punto di vista computazionale. In particolare, se interpretiamo il rendimento uniperiodale come una variabile aleatoria con una propria distribuzione, anche nel caso più semplice in cui il campione $\{R_{t+i-1,t+i}\}_{i=1}^k$ fosse *i.i.d.* e la distribuzione fosse nota (ad esempio una normale) non sarebbe agevole dedurre la distribuzione del rendimento multiperiodale.

La *linearizzazione* della (2.2.4) implica l'utilizzo dei logaritmi. Definiamo le grandezze

$$(2.2.5) \quad r_{t+i-1,t+i}^c = \ln(R_{t+i-1,t+i}), \quad r_{t,t+k}^c = \ln(R_{t,t+k}) \quad \text{e} \quad r_{k,t}^c = \ln(R_{k,t})$$

Dunque $r_{k,t}^c = k^{-1}r_{t,t+k}^c$; inoltre:

$$(2.2.6) \quad r_{t,t+k}^c = \ln\left(\prod_{i=1}^k R_{t+i-1,t+i}\right) = \sum_{i=1}^k \ln(R_{t+i-1,t+i}) = \sum_{i=1}^k r_{t+i-1,t+i}^c$$

Parliamo di $r_{t,t+1}^c$ come del *tasso a capitalizzazione continua* (per le ragioni che vedremo tra breve). Il fatto di maggior rilievo è la relazione *lineare* (2.2.6) che intercorre tra il tasso uniperiodale e quello multiperiodale. Questa proprietà è davvero assai utile. Supponiamo ad esempio, che il campione $\{r_{t+i-1,t+i}^c\}_{i=1}^k$ sia composto da normali indipendenti. Allora la variabile $r_{t,t+k}^c$ sarà anch'essa normalmente distribuita, un fatto di notevole importanza.

2.3. Altre definizioni

La definizione di rendimento che abbiamo data è in realtà una tra le molte, sebbene la principale. Infatti sui diversi mercati si incontrano definizioni differenti delle quali bisogna essere a conoscenza quantomeno per saper interpretare i dati correttamente. Oltre a queste, come abbiamo visto, vi è la definizione del tasso di interesse logaritmico ossia a capitalizzazione continua, usata esclusivamente per la modellistica finanziaria.

Sul mercato dei *bonds* è spesso utilizzato il tasso cosiddetto *di sconto* in cui l'incremento del valore dell'investimento è rapportato al valore finale, anziché a quello iniziale. Questo perché per molti buoni del tesoro, i cosiddetti *T-bills*, il valore a scadenza del titolo è noto fin dall'inizio. Tipicamente questi titoli promettono un rimborso a scadenza $(t+1)$ di 100 dato il quale il prezzo corrente (t) di vendita si interpreta come un valore *scontato* rispetto al valore facciale, soprattutto nelle obbligazioni senza cedola o *Zero Coupon*

Bonds. Dunque lo sconto ammonta a $100 - P_t$ che, in rapporto al valore facciale, 100, corrisponde ad una percentuale

$$(2.3.1) \quad r_t^D = \frac{100 - P_t}{100} \quad \text{e} \quad r_{t,t+k}^D = \frac{1}{k} \frac{100 - P_t}{100}$$

(se il rimborso è alla data $t + 1$ o $t + k$ rispettivamente). Il rapporto che intercorre tra prezzo e rimborso è pertanto

$$P_t = 100(1 - kr_{t,t+k}^D) \quad \text{anziché} \quad P_t = \frac{100}{1 + kr_{t,t+k}^D}$$

Dunque, la relazione che intercorre tra questa definizione particolare di rendimento e quella offerta sopra è la seguente:

$$r_t^D = \frac{r_t}{1 + r_t}$$

Sul mercato monetario (ossia di strumenti con scadenza estremamente breve, quale il mercato interbancario) si utilizza poi un'ulteriore definizione, secondo la quale

$$(2.3.2) \quad r_{t,t+k}^M = \frac{1}{k} \frac{P_{t,t+k} - P_t}{P_t} \quad \text{ossia} \quad P_{t,t+k} = P_t(1 + kr_{t,t+k}^M)$$

Questa formula viene spesso indicata con la denominazione di capitalizzazione semplice: gli interessi maturati ad ogni periodo si sommano tra loro e non generano altri interessi come invece accade nella (2.2.3).

Una menzione a parte tocca alla questione della capitalizzazione. A volte il tasso di interesse uniperiodale si capitalizza più di una volta. Ad esempio un tasso di interesse annuo $r_{t,t+1}^{(2)}$ a capitalizzazione semestrale implica che ogni sei mesi venga percepito (e capitalizzato) un rendimento pari a $\frac{r_{t,t+1}^{(2)}}{2}$. Dunque complessivamente il rendimento lordo ammonta a

$$\frac{P_{t+1}}{P_t} = \left(1 + \frac{r_{t,t+1}^{(2)}}{2}\right)^2$$

Naturalmente la capitalizzazione potrebbe essere di ordine n e si avrebbe

$$\frac{P_{t+1}}{P_t} = \left(1 + \frac{r_{t,t+1}^{(n)}}{n}\right)^n$$

Se si suppone che la successione $\langle r_{t,t+1}^{(n)} \rangle_{n \in \mathbb{N}}$ converga ad un limite, $r_{t,t+1}^c$ allora necessariamente si avrà:

$$\frac{P_{t+1}}{P_t} = \lim_{n \rightarrow \infty} \left(1 + \frac{r_{t,t+1}^{(n)}}{n}\right)^n = \exp(r_{t,t+1}^c)$$

da cui deriva la denominazione di tasso di interesse a capitalizzazione continua.

Fatti stilizzati

Sebbene le proprietà dei rendimenti finanziari tendano ad essere diverse da un dataset all'altro, vi sono fatti che parrebbero essere ricorrenti in tutti i mercati ed in tutti i periodi. Idealmente un modello appropriato dei rendimenti finanziari dovrebbe essere scritto in modo tale da catturare il maggior numero di tali regolarità empiriche. Alcune di queste verranno ora descritte e discusse. Chi fosse interessato ad una analisi più approfondita potrebbe consultare il lavoro di Cont. Per ciascuno dei fatti che ora descriveremo vogliamo indicare in che modo esso si evidenzi e quale sia la sua interpretazione economica. Il tasso di interesse a cui faremo riferimento in questa parte è quello a capitalizzazione continua.

3.1. σ domina μ

Ossia la deviazione standard domina il rendimento atteso. Questa conclusione è davvero semplice e tuttavia cruciale. Naturalmente le grandezze μ e σ non sono note e possono essere approssimate dai corrispondenti stimatori

$$\bar{r} \equiv \frac{\sum_{t=1}^T r_t}{T} \quad \text{e} \quad \bar{s} \equiv \sqrt{\frac{\sum_{t=1}^T (r_t - \bar{r})^2}{T-1}}$$

Sotto l'ipotesi che il campione $\{r_t\}_{t=1}^T$ sia indipendente e normalmente distribuito con valore atteso 0 lo stimatore \bar{r}/\bar{s} è distribuito come una T di student con $T-1$ gradi di libertà. Dunque un valore di tale grandezza campionaria inferiore a 2 in valore assoluto può ritenersi un'indicazione della validità dell'ipotesi che il valore atteso dei rendimenti sia nullo. Nel nostro campione tale rapporto ha un valore di circa 0,028. Questa conclusione è comune a quasi tutti i campioni di rendimenti finanziari ed in genere si interpreta come una buona motivazione per l'ipotesi $E(r) = 0$ comune a molti dei modelli che vedremo. In effetti una conseguenza di questo primo fatto è che l'attenzione principale debba essere concentrata sul fattore della volatilità, un fatto sul quale torneremo.

3.2. Mancanza di correlazione

Prendiamo due insiemi di osservazioni provenienti dal medesimo dataset: $\{r_t\}_{t=1}^{T-\tau}$ e $\{r_{t+\tau}\}_{t=1}^{T-\tau}$. È possibile calcolare la correlazione tra queste due serie calcolando, come consueto, la quantità

$$(3.2.1) \quad \rho_\tau(R) = \frac{\sum_{t=1}^{T-\tau} (r_t - \bar{r})(r_{t+\tau} - \bar{r}')}{\sqrt{\sum_{t=1}^{T-\tau} (r_t - \bar{r})^2 \sum_{t=1}^{T-\tau} (r_{t+\tau} - \bar{r}')^2}}$$

dove $\bar{r} = (T-\tau-1)^{-1} \sum_{t=1}^{T-\tau} r_t$ e $\bar{r}' = (T-\tau-1)^{-1} \sum_{t=1}^{T-\tau} r_{t+\tau}$ sono le medie campionarie delle due serie. La quantità ρ naturalmente può essere calcolata per tutti i valori di $\tau = 1, \dots, T-1$ e si interpreta come un indicatore della quantità aleatoria

$$(3.2.2) \quad \frac{(r_t - \mu)E(r_{t+\tau} - \mu | \mathcal{F}_t)}{\sqrt{(r_t - \mu)^2 E(r_{t+\tau} - \mu)^2 | \mathcal{F}_t}}$$

– dove $E(\cdot | \mathcal{F}_t)$ indica l'aspettativa condizionata e μ è per ipotesi il valore comune del valore atteso di r_t . L'evidenza empirica indica che $\rho_\tau = 0$ è la conclusione più verosimile per tutti i valori di τ . L'interpretazione

di questo risultato, in accordo con la (3.2.2), è che r_t e $r_{t+\tau}$ parrebbero non avere alcun grado di correlazione lineare e che pertanto l'aspettativa condizionata $E(r_{t+1}|\mathcal{F}_t)$ può trattarsi come una costante. Bisogna comunque osservare che la conclusione secondo la quale $r_{t+\tau}$ è distribuito in modo indipendente da r_t non sarà in generale vera in quanto la correlazione lineare è solo un tipo assai particolare di dipendenza statistica. Si può costruire facilmente una funzione che consenta di calcolare la quantità ρ_τ .

3.3. Correlazione quadratica positiva

Possiamo ripetere il medesimo esperimento compiuto più sopra ma dopo aver sostituito i rendimenti con il loro valore quadratico, ossia r_t con r_t^2 . Calcoleremo cioè la quantità $\rho_\tau(r^2)$ definita come nella (3.2.1). Si ottiene in tal caso la conclusione che tale quantità è significativamente positiva, quanto meno quando il parametro τ è sufficientemente piccolo, diciamo per $\tau < 90$ giorni. Questo risultato fornisce una certa evidenza circa l'esistenza di correlazione positiva tra i rendimenti quadratici ossia nella varianza condizionata – della quale i rendimenti quadratici sono una *proxy*. La conclusione appena trovata è senz'altro una delle più importanti sia dal punto di vista statistico che economico. Sotto il profilo economico, essa suggerisce che la volatilità – ovvero la varianza – tende ad avere una certa persistenza nel tempo di modo che periodi di alta volatilità seguono a periodi di alta volatilità ed occorre un certo lasso di tempo perché tale elemento perda importanza. Quanto alle implicazioni statistiche l'evidenza suggerisce che una grande enfasi debba essere posta sulla struttura della volatilità che difatti, in molti modelli, è l'elemento di maggior conto nella spiegazione dei rendimenti.

3.4. Non normalità

Questa è un'altra evidenza assai importante. se ad esempio un *risk manager* volesse effettuare un'analisi statistica dovrebbe certamente formulare una qualche ipotesi circa la distribuzione dei rendimenti. Spesso l'ipotesi maggiormente comoda è proprio quella della distribuzione normale: ma non sempre è possibile. In effetti i rendimenti finanziari sono un esempio molto chiaro del fatto che l'ipotesi della normalità non è sempre quella maggiormente adeguata. Le deviazioni dalla normalità si colgono paragonando la distribuzione empirica dei rendimenti alla distribuzione normale. Si possono fare al riguardo numerosi test vuoi sui diversi momenti, vuoi sull'intera distribuzione.

3.4.1. Non normalità generale. Un modo di verificare se la distribuzione empirica rassomigli a quella normale è quello di paragonare i quantili delle due distribuzioni attraverso quello che in letteratura viene spesso chiamato *QQ plot*. Si tratta di uno strumento importante quanto semplice, che consente di comprendere in modo chiaro il tipo di deviazioni dalla normalità presenti nel campione. È necessario dapprima ordinare le osservazioni in modo crescente: indichiamo con $r_{(i)}$ la i -ma osservazione in ordine di grandezza. In altri termini $r_{(i)}$ rappresenta l' $iT^{-1} \times 100$ -mo percentile della distribuzione empirica (dove con T indichiamo l'ampiezza del campione). Il percentile corrispondente tratto dalla distribuzione normale si calcola semplicemente come $x_i = \Phi_N^{-1}((i - 0,5)T^{-1})$. Il grafico consiste nella rappresentazione congiunta delle due quantità: (x_i, r_i) .

3.4.2. Skewness (asimmetria). La *skewness* o asimmetria fornisce una misura della centratura della distribuzione attorno alla sua media. Una distribuzione perfettamente simmetrica dovrebbe avere un valore nullo della *skewness*; se asimmetrica a favore di valori al di sotto della media il valore sarà negativo, se a

favore di valori al di sopra della media esso sarà positivo. La definizione esatta è

$$(3.4.1) \quad sk = \frac{P[(r_{t+1} - \mu)^3]}{\sigma^3}$$

La distribuzione normale presenta una *skewness* pari a 0 (in effetti lo stesso vale per tutti i momenti di ordine dispari). La maggior parte delle serie di rendimenti finanziari presenta invece una *skewness* negativa, ad indicare una radicale asimmetria tra guadagni e perdite (dato che tipicamente $\mu = 0$) di modo che le perdite sembrano in qualche modo più importanti dei profitti.

3.4.3. Kurtosi. La *kurtosi* è definita come

$$(3.4.2) \quad ku = \frac{P[(r_{t+1} - \mu)^4]}{\sigma^4}$$

ed è pari a 3 nel caso di distribuzione normale. L'interpretazione consueta di questa grandezza è quale misura della massa di probabilità assegnata ad eventi che si trovano nelle code della distribuzione ed un eccesso di kurtosi rispetto al valore di riferimento 3 viene indicata col nome di *code spesse* nella terminologia finanziaria. Anche questo risultato, che i rendimenti finanziari presentino code spesse, è molto importante e pervasivo nella letteratura finanziaria. Esso si applica tanto a modelli di valutazione dei derivati quali le opzioni che ai modelli di gestione del rischio. Un gran numero di contributi alla letteratura scientifica consistono nel tentativo di spiegare questo specifico aspetto della distribuzione dei rendimenti.

3.4.4. Kurtosi e skewness. È stato messo a punto un test statistico allo scopo di analizzare questi ultimi due punti: il test di Jarque-Bera. Esso è definito come

$$(3.4.3) \quad JB = \left[sk^2 + \frac{(ku - 3)^2}{4} \right] \frac{T - 1}{6} \sim \chi_2^2$$

Un valore elevato di questo test indica che lo scostamento congiunto di *skewness* e *kurtosi* dai valori di riferimento della normale è significativo statisticamente.

3.5. Esercizi

Domanda 1. 1 Si rappresenti il grafico dei rendimenti logaritmici giornalieri generati dal *SP500*. Si calcoli media, varianza, skewness e kurtosi. Si disegni l'istogramma dei rendimenti sovrapposto a quello della normale.

Domanda 2. 2 Si tracci l'autocorrelogramma dei rendimenti e dei rendimenti quadratici.

Domanda 3. 3 Si calcoli il rendimento relativo a periodi non sovrapposti di 5, 10 e 15 giorni. Sulla serie così ottenuta si calcoli media, varianza, skewness e kurtosi e si valuti l'ipotesi di distribuzione normale.

Parte 2

Scelta del portafoglio

Un'introduzione al modello media varianza

4.1. Introduzione

Come si è visto, vi sono situazioni particolari in cui si può senz'altro assumere che gli investitori nel decidere la composizione del proprio portafoglio tengano conto esclusivamente del valore atteso e della varianza del rendimento. Come abbiamo detto, questo è solo *un* particolare modello di preferenze dell'investitore, sebbene molto popolare. In particolare la nozione di rischio sottostante l'uso del parametro della varianza è tutt'altro che generale poiché si potrebbe sostenere che il rischio debba misurarsi diversamente ¹. Ulteriori osservazioni su questo tipo di preferenze si trova nei paragrafi seguenti.

La ragione che rende il modello media/varianza così conosciuto sta nel fatto che esso consente di trarre conclusioni che sono al contempo semplici e intuitive. La summa di questo approccio si condensa nel cosiddetto *CAPM* che tratteremo poco più avanti.

Consideriamo dunque il caso in cui il mercato consista di N attività finanziarie, indicizzate da $n = 1, 2, \dots, N$. Con r_n indichiamo il rendimento del titolo n -mo. Dunque, investendo il capitale W nel titolo n se ne ottiene a scadenza un rimborso lordo pari a $W(1 + r_n)$. Immaginiamo ora che il capitale venga investito in tutti i titoli secondo le quote $\alpha_1, \alpha_2, \dots, \alpha_N$. Tali quote, che per definizione soddisfano la proprietà $\sum_{n=1}^N \alpha_n = 1$ formano un portafoglio che indicheremo sinteticamente con α . In tal caso, nel singolo titolo n viene complessivamente investito il capitale $\alpha_n W$ ricavandone un rimborso complessivo pari a $\alpha_n W(1 + r_n)$. Dunque, dall'aver investito nel portafoglio α si ricava

$$\sum_{n=1}^N \alpha_n W(1 + r_n) = W(1 + \sum_{n=1}^N \alpha_n r_n) = W(1 + r_\alpha)$$

dove si è posto implicitamente

$$(4.1.1) \quad r_\alpha = \sum_{n=1}^N \alpha_n r_n$$

Questa formula è assai utile per chiarire la relazione che intercorre tra il rendimento di un portafoglio ed il rendimento dei titoli che ne fanno parte. Naturalmente anche r_α , così come ciascun r_n deve considerarsi come una variabile aleatoria, della quale si può calcolare valore atteso e varianza. In particolare, considerando i pesi alla stregua di costanti, si ha

$$(4.1.2) \quad E(r_\alpha) = E\left(\sum_{n=1}^N \alpha_n r_n\right) = \sum_{n=1}^N \alpha_n E(r_n)$$

¹La moderna teoria del *risk management* in finanza parte proprio da questa osservazione per proporre misure di rischio alternative, quali il *VaR* l'*Expected Shortfall* e via dicendo

ossia $\mu_\alpha = \sum_{n=1}^N \alpha_n \mu_n$. Inoltre,

$$\begin{aligned}
 \text{Var}(r_\alpha) &= E \{ (r_\alpha - \mu_\alpha)^2 \} \\
 &= E \left\{ \left(\sum_{n=1}^N \alpha_n (r_n - \mu_n) \right)^2 \right\} \\
 &= E \left\{ \sum_{n=1}^N \alpha_n (r_n - \mu_n) \sum_{m=1}^N \alpha_m (r_m - \mu_m) \right\} \\
 &= \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n E \{ (r_n - \mu_n)(r_m - \mu_m) \} \\
 &= \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n \text{Cov}(r_n, r_m)
 \end{aligned}$$

ossia

$$(4.1.3) \quad \sigma_\alpha^2 = \sum_{n=1}^N \sum_{m=1}^N \alpha_m \alpha_n \sigma_{n,m}$$

dove si è posto evidentemente $\sigma_{n,m} = \text{Cov}(r_n, r_m)$.

Dalla (4.1.2) e dalla (4.1.3) si comprende come tanto il valore atteso che la varianza del rendimento di un portafoglio dipendono dalla composizione del portafoglio stesso. A seconda di come si modificano i parametri $\alpha_1, \dots, \alpha_N$ varia tanto μ_α che σ_α^2 . Possiamo pertanto congetturare che sussista una relazione precisa tra μ_α e σ_α^2 , qualcosa come una funzione crescente tale per cui il rendimento atteso di un portafoglio è tanto più elevato quanto più alta è la sua varianza. Rendere esplicita tale relazione sarà l'oggetto dell'analisi dei prossimi paragrafi.

4.2. Il caso di due soli titoli.

Consideriamo per il momento un sottoinsieme del mercato formato da tutti i portafogli ottenuti investendo in due sole attività finanziarie, 1 e 2. Indichiamo con r_1 e r_2 i corrispondenti rendimenti. Se il nostro capitale W viene investito in una certa percentuale α nel titolo 1 e la rimanente parte nel titolo 2, allora il rendimento del portafoglio sarà pari a

$$r_\alpha = \alpha r_1 + (1 - \alpha) r_2$$

e dunque

$$(4.2.1) \quad \mu_\alpha = \alpha \mu_1 + (1 - \alpha) \mu_2$$

e anche

$$(4.2.2) \quad \sigma_\alpha^2 = \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2 + 2\alpha(1 - \alpha) \sigma_{1,2}$$

Dalla (4.2.1) ricaviamo facilmente

$$\alpha = \frac{\mu_\alpha - \mu_2}{\mu_1 - \mu_2}$$

che, inserita poi nella (23.2.6), dà luogo alla

$$(4.2.3) \quad \sigma_\alpha^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}}{(\mu_1 - \mu_2)^2} \mu_\alpha^2 - 2 \frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2 - (\mu_1 + \mu_2) \sigma_{1,2}}{(\mu_1 - \mu_2)^2} \mu_\alpha + \frac{\mu_2^2 \sigma_1^2 + \sigma_2^2 \mu_1^2 - 2\mu_1 \mu_2 \sigma_{1,2}}{(\mu_1 - \mu_2)^2}$$

ossia

$$(4.2.4) \quad \sigma_\alpha^2 = A \mu_\alpha^2 - 2B \mu_\alpha + C$$

una relazione che si rappresenta in forma grafica come una parabola a forma di U , come nella Figura 4.2.



Figure 1: La frontiera efficiente con due soli titoli.

Questa semplice espressione consente alcune conclusioni che sarà utile sottolineare.

- (1) Per prima cosa, è implicito nella (4.2.4) il vincolo di non negatività $B^2 \leq AC$ che si traduce nella disuguaglianza $(\mu_1 - \mu_2)^2 \sigma_{1,2}^2 \leq \sigma_1^2 \sigma_2^2 (\mu_1 - \mu_2)^2$ a sua volta equivalente alla disuguaglianza $\sigma_{1,2}^2 \leq \sigma_1^2 \sigma_2^2$ la quale è senz'altro sempre verificata.
- (2) Il portafoglio a varianza minima è caratterizzato da

$$\mu_{\min} = \frac{B}{A} = \frac{\mu_2 \sigma_1^2 + \mu_1 \sigma_2^2 - (\mu_1 + \mu_2) \sigma_{1,2}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}} \quad \text{e} \quad \sigma_{\min}^2 = C - \frac{B^2}{A}$$

e corrisponde al portafoglio

$$\alpha_{\min} = \frac{\sigma_2^2 - \sigma_{1,2}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}}$$

Inoltre, si conclude facilmente che $\sigma_{\alpha}^2 = 0$ se e solo se $\sigma_1 \sigma_2 = |\sigma_{1,2}|$, dunque quando vi è correlazione perfetta (positiva o negativa). Si noti che in tal caso il portafoglio a varianza minima diviene

$$\alpha_{\min} = \frac{\sigma_2}{\sigma_2 - \rho \sigma_1} = \frac{\rho \sigma_2}{\rho \sigma_2 - \sigma_1}$$

dove $\rho = \sigma_{1,2} / \sigma_1 \sigma_2$ assume valore ± 1 (e dunque $\rho^2 = 1$) e

$$\mu_{\min} = \frac{\rho \sigma_2 \mu_1 - \sigma_1 \mu_2}{\rho \sigma_2 - \sigma_1}$$

- (3) Sempre nel caso di perfetta correlazione lineare menzionato al punto precedente si ha anche

$$\sigma_{\alpha}^2 = \left(\frac{(\sigma_1 - \rho \sigma_2) \mu_{\alpha} - (\mu_2 \sigma_1 - \rho \mu_1 \sigma_2)}{\mu_1 - \mu_2} \right)^2$$

ossia

$$\sigma_{\alpha} = \frac{\sigma_1 - \rho \sigma_2}{\mu_1 - \mu_2} \mu_{\alpha} - \frac{\mu_2 \sigma_1 - \rho \mu_1 \sigma_2}{\mu_1 - \mu_2} = \sigma_1 \alpha + \rho \sigma_2 (1 - \alpha)$$

Sussiste quindi una relazione lineare tra σ_{α} e μ_{α} sulla quale spesso la letteratura si è soffermata e che si può rappresentare graficamente come segue:

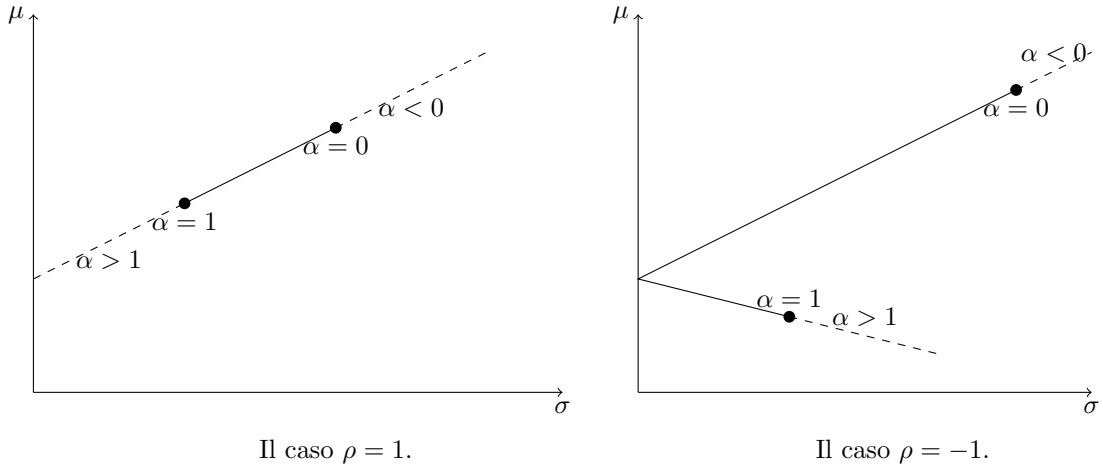


Figure 2: La frontiera efficiente con perfetta correlazione lineare.

- (4) Il caso $A = B = 0$ corrisponde alla situazione in cui tutti i portafogli hanno la medesima varianza e dunque, dalla (23.2.6), al caso $\sigma_1^2 = \sigma_2^2 = \sigma_{1,2}$. Dalla disuguaglianza di Cauchy-Schwartz ricordata più oltre (vedi nota 5.3.2, p. 28) ciò significa che vi è una relazione lineare tra r_1 e r_2 , ossia $r_1 = a + r_2$. Naturalmente, se $a > 0$ (ovvero se $a < 0$) qualunque investitore con preferenze media varianza preferirà investire la massima possibile quantità nel titolo 1 (ovvero nel titolo 2) poiché a parità di rischio consente un rendimento atteso maggiore. Dunque un tale caso è incompatibile con la scelta ottima dell'investitore e con l'equilibrio del mercato tranne che nel caso $r_1 = r_2$. Al di fuori di quest'ultimo caso non esiste alcun portafoglio efficiente poiché è sempre possibile incrementare il rendimento atteso senza accrescere la varianza. Nel seguito supporremo sempre che se r_1 e r_2 rappresentano i rendimenti di due titoli esistenti tali che $r_1 = a + r_2$ allora necessariamente $a = 0$.
- (5) La (4.2.3) è stata ricavata senza imporre alcuna restrizione sul parametro α che può pertanto assumere qualsivoglia valore, anche negativo. Quest'ultimo caso corrisponde alla possibilità da parte dell'investitore di realizzare *vendite allo scoperto* ossia di vendere un determinato titolo anziché acquistarlo. Evidentemente, si tratta di una opportunità assai particolare. L'esempio più chiaro è offerto dal conto corrente sul quale tipicamente il risparmiatore è autorizzato ad accumulare uno *scoperto*, ossia un saldo negativo. In tale circostanza egli sta di fatto acquistando una quantità negativa del titolo corrispondente (ossia il conto corrente bancario) ovvero sta prendendo a prestito anziché prestando i propri risparmi. È chiaro che non tutti i titoli finanziari ammettono questa stessa opportunità e che anche il conto corrente bancario pone comunque alcune limitazioni all'utilizzo dello scoperto. Questa osservazione suggerisce di interpretare con cautela alcune delle conclusioni raggiunte nel seguito.
- (6) Graficamente, si osserva che la parabola è una curva convessa, ossia la regione inclusa tra i due rami della parabole ha la proprietà che qualunque segmento che unisca due punti di questa regione sta all'interno della regione stessa. Questa semplice caratteristica in realtà ci consente, come vedremo, delle importanti generalizzazioni.

Il caso generale.

Le conclusioni ricavate nel paragrafo precedente sono interessanti ma limitate al caso nel quale si investa in due sole attività finanziarie. Intendiamo ora generalizzarle al caso di un numero arbitrario, N , di titoli ossia sostituendo il parametro α con l'insieme $\alpha_1, \dots, \alpha_N$ dei pesi di portafoglio. Naturalmente, in linea di principio potrebbe essere ancora possibile (sotto le opportune condizioni) ricavare una relazione tra rendimento atteso e varianza quale la (4.2.3) sopra ma dal punto di vista matematico questo tentativo si presenta estremamente complicato se non impossibile. Fortunatamente si può procedere in modo assai più semplice.

Anzitutto osserviamo che non tutti i portafogli hanno la stessa importanza in termini economici. Ragionando infatti sotto l'ipotesi che gli investitori abbiano preferenze del tipo media/varianza, vi sono portafogli nei quali nessun investitore con tali caratteristiche si sognerebbe mai di investire. Si tratta dei portafogli *dominati*. Un portafoglio, α , è dominato da un altro portafoglio, β , se questo secondo portafoglio è preferibile al primo tanto sotto il profilo della media che della varianza ossia se vale la duplice disuguaglianza

$$\mu_\alpha \leq \mu_\beta \quad \sigma_\alpha^2 \geq \sigma_\beta^2$$

ed almeno una delle due vale in senso stretto. In tal caso β offre un rendimento atteso almeno altrettanto alto che α ed implica un rischio non superiore a quello insito in α ; inoltre relativamente ad almeno uno di tali criteri (rendimento atteso e rischio) risulta strettamente migliore. In altre parole, un portafoglio è dominato se non vi è investitore con preferenze media/varianza disponibile ad investire in esso. Indichiamo sinteticamente tale situazione scrivendo

$$\beta \succ_M \alpha$$

per dire che il portafoglio β risulta strettamente preferibile al portafoglio α per ogni investitore sul mercato (dunque il suffisso M indica che il criterio si applica a tutti gli investitori presenti sul mercato). È utile introdurre anche la nozione di portafoglio *debolmente dominato*: α è debolmente dominato da β – ovvero $\beta \succeq_M \alpha$ se e solo se $\mu_\beta \geq \mu_\alpha$ e anche $\sigma_\beta^2 \leq \sigma_\alpha^2$ senza che nessuna delle due disuguaglianze debba valere in senso stretto. Evidentemente, tutti i portafogli sono debolmente dominati poiché possiamo sempre scrivere $\alpha \succeq_M \alpha$.

Un portafoglio che non sia dominato da nessun altro portafoglio si dice *efficiente*. È evidente che, limitatamente ai portafogli efficienti, un rendimento atteso più elevato necessariamente si accompagna ad un più alto livello di varianza. In altre parole, se α e β sono entrambi portafogli efficienti, allora necessariamente

$$\mu_\alpha > \mu_\beta \quad \text{se e solo se} \quad \sigma_\alpha^2 > \sigma_\beta^2$$

5.1. L'analisi grafica.

Per comprendere meglio la relazione che intercorre tra rendimento atteso e varianza, osserviamo anzitutto che per ogni coppia di portafogli è possibile tracciare una curva quale quella raffigurata nel paragrafo

precedente. Avremo pertanto un'intera famiglia di tali curve delle quali dobbiamo considerare la parte esterna, ossia *l'involuppo inferiore*. Questo procedimento è illustrato nella Figura 5.1 L'osservazione cruciale è che la curva così costruita avrà anch'essa la medesima forma a U della corrispondente curva con due soli titoli. Più precisamente, l'area all'interno della curva descrive un insieme convesso, ossia ogni coppia di punti è connessa da un segmento che giace tutto all'interno dell'area stessa. Questo argomento è del tutto evidente anche solo in termini grafici.

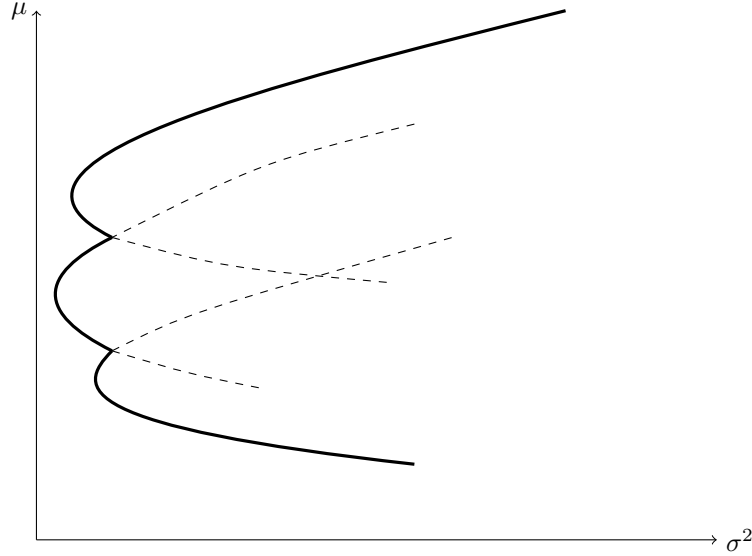


Figure 3: L'involuppo delle frontiere efficienti.

Prendiamo infatti due punti corrispondenti ad altrettanti portafogli e tracciamo il segmento che li unisce. Parimenti possiamo costruire la frontiera efficiente generata da questi due portafogli. Poiché quest'ultima ha la consueta forma a parabola, il segmento giace necessariamente nella regione interna della curva ossia i punti che vi sono rappresentati sono dominati (quanto meno debolmente) da un qualche portafoglio formato combinando i primi due.

Sempre basandoci sulla sola intuizione grafica possiamo concludere che un portafoglio è efficiente se e solo se esso si rappresenta come un punto sulla parte superiore della curva rappresentata nella figura precedente. Questa conclusione apparentemente qualitativa ci consente in realtà delle conclusioni molto importanti.

Osserviamo inoltre che i portafogli efficienti corrispondono solo alla parte superiore della curva tracciata, anche se spesso in modo impreciso, ci si riferisce all'intera curva come alla *frontiera efficiente*.

Questa conclusione qualitativa può essere raggiunta in termini più rigorosi se ragioniamo come segue. Prendiamo due punti (x_1, y_1) e (x_2, y_2) sul piano media/varianza, ciascuno dominato da un corrispondente portafoglio. Ossia vi sono due portafogli, α_1 e α_2 tali che $\mu_{\alpha_i} \geq y_i$ e $\sigma_{\alpha_i}^2 \leq x_i$ per $i = 1, 2$. Poniamo che il punto (x, y) sia intermedio tra i due precedenti, ossia $x = ex_1 + (1 - e)x_2$ e $y = ey_1 + (1 - e)y_2$ per un qualche $0 \leq e \leq 1$, e consideriamo il portafoglio α con rendimento $er_{\alpha_1} + (1 - e)r_{\alpha_2}$. Allora abbiamo

$$\mu_{\alpha} = e\mu_{\alpha_1} + (1 - e)\mu_{\alpha_2} \geq y$$

e

$$\sigma_{\alpha}^2 = \text{VAR}(er_{\alpha_1} + (1 - e)r_{\alpha_2})$$

$$\begin{aligned}
&\leq e\text{VAR}(r_{\alpha_1}) + (1-e)\text{VAR}(r_{\alpha_2}) \\
&= e\sigma_{\alpha_1}^2 + (1-e)\sigma_{\alpha_2}^2 \\
&\leq ey_1 + (1-e)y_2 \\
&= y
\end{aligned}$$

(nella seconda riga abbiamo fatto uso della disuguaglianza di Jensen per la quale, se $0 \leq e \leq 1$ allora $(eX + (1-e)Y)^2 \leq eX^2 + (1-e)Y^2$). Dunque un qualunque punto intermedio tra due punti che ammettono ciascuno un portafoglio dominante, è anch'esso dominato da un qualche portafoglio. In altre parole ancora, l'insieme delle coppie di valori media/varianza dominati da un portafoglio è un insieme convesso.

5.2. Una conclusione generale.

L'analisi precedente ci consente, come abbiamo anticipato, conclusioni inaspettatamente precise. Possiamo infatti utilizzare una proprietà importante degli insiemi convessi per la quale un punto si trova sulla frontiera di un tale insieme se e solo se è possibile far passare per quel punto una *retta tangente*. Ancora una volta, il grafico è piuttosto chiaro – si veda la Figura 5.2.

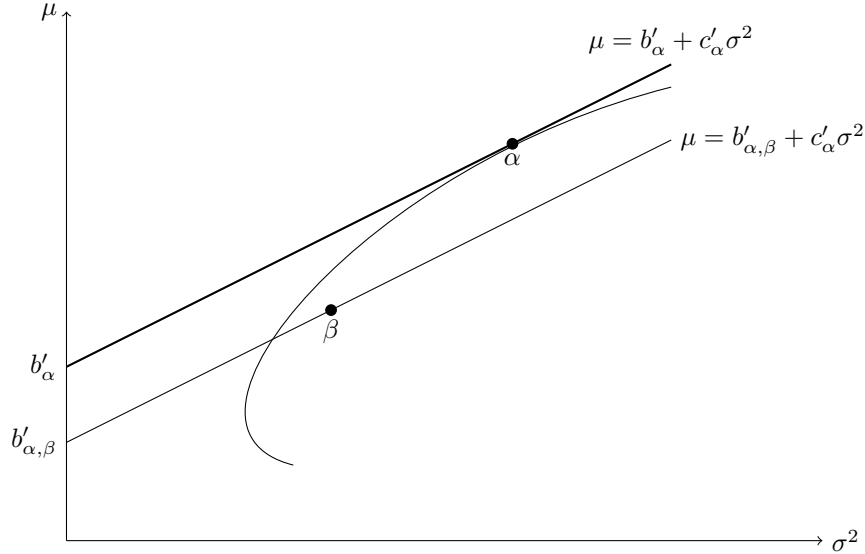


Figure 4: La retta tangente e le sue proprietà.

Naturalmente il punto corrispondente al valore minimo della varianza ammette una retta tangente con inclinazione infinita. L'analisi relativa a questo particolare punto della frontiera rimane pertanto al di fuori delle considerazioni che seguono.

Cosa significa più precisamente una retta tangente e qual è la sua equazione? Prendiamo un punto con coordinate $(\sigma_\alpha^2, \mu_\alpha)$. Una retta è tangente attraverso il punto $(\sigma_\alpha^2, \mu_\alpha)$ se e solo se soddisfa due proprietà:

- (1) la retta passa attraverso il punto stesso;
- (2) la retta con la medesima inclinazione di quella tangente che passi per un qualunque altro punto dell'insieme risulta avere intercetta inferiore.

Ricordiamo che l'equazione di una retta è semplicemente $y = b' + c'x$ dove, nel caso che stiamo considerando, $y = \mu$ e $x = \sigma^2$. Affinché una retta passi attraverso il punto $(\sigma_\alpha^2, \mu_\alpha)$ essa dovrà pertanto avere equazione

$$\mu = (\mu_\alpha - c'_\alpha \sigma_\alpha^2) + c'_\alpha \sigma^2$$

In altri termini una retta di equazione $\mu = b'_\alpha + c'_\alpha \sigma^2$ passa per il punto assegnato se e solo se

$$(5.2.1) \quad b'_\alpha = \mu_\alpha - c'_\alpha \sigma_\alpha^2$$

Evidentemente, vi sono infinite curve con questa proprietà, a seconda della differente inclinazione, ossia del valore del parametro c'_α . Si noti tuttavia che se il portafoglio α è efficiente e se la retta è tangente, allora necessariamente dovrà aversi $c'_\alpha \geq 0$. Osserviamo anche che nel punto della curva corrispondente al portafoglio con varianza minima, la retta tangente è perfettamente verticale ossia dovremmo avere $c'_\alpha = \infty$. Il nostro metodo, pertanto, si applica solo a quei punti per i quali $c'_\alpha < \infty$.

Tra le molte curve passanti per il punto corrispondente al portafoglio α , solo alcune sono tangenti: quelle che posseggono la proprietà (2) precedente. Dunque il parametro $c'_\alpha \geq 0$ dovrà essere tale per cui una retta con questa stessa inclinazione e passante per il punto $(\sigma_\beta^2, \mu_\beta)$ corrispondente ad un generico portafoglio β (efficiente o meno), debba avere intercetta inferiore. Dalla (5.2.1) osserviamo che per passare nel punto corrispondente al portafoglio β l'intercetta deve essere $\mu_\beta - c'_\alpha \sigma_\beta^2$. In altre parole, la retta $\mu = b'_\alpha + c'_\alpha \sigma^2$ è tangente solo se b'_α soddisfa la (5.2.1) e c'_α soddisfa la

$$\mu_\beta - c'_\alpha \sigma_\beta^2 \leq \mu_\alpha - c'_\alpha \sigma_\alpha^2$$

per ogni altro portafoglio β , ossia

$$(5.2.2) \quad \mu_\beta - \mu_\alpha \leq c'_\alpha (\sigma_\beta^2 - \sigma_\alpha^2)$$

Poiché la (5.2.2) si applica a *ogni* portafoglio β , allora quale che sia il numero x essa deve valere anche per quel portafoglio composto investendo una quota x nel portafoglio β e una quota $1 - x$ nel portafoglio α . Un tale portafoglio, che indichiamo per comodità con il suffisso x , ha rendimento atteso e varianza pari rispettivamente a

$$\mu_x = x\mu_\beta + (1-x)\mu_\alpha \quad \text{e} \quad \sigma_x^2 = x^2\sigma_\beta^2 + (1-x)^2\sigma_\alpha^2 + 2x(1-x)\sigma_{\alpha,\beta}$$

dunque, $\mu_x - \mu_\alpha = x(\mu_\beta - \mu_\alpha)$ e $\sigma_x^2 - \sigma_\alpha^2 = x^2\sigma_\beta^2 + (x^2 - 2x)\sigma_\alpha^2 + 2x(1-x)\sigma_{\alpha,\beta}$. Sostituendo tali valori nella (5.2.2) otteniamo dunque

$$(5.2.3) \quad x(\mu_\beta - \mu_\alpha) \leq c'_\alpha \{x^2\sigma_\beta^2 + (x^2 - 2x)\sigma_\alpha^2 + 2x(1-x)\sigma_{\alpha,\beta}\}$$

Consideriamo il caso $x > 0$. Dalla (5.2.3) ricaviamo

$$\mu_\beta - \mu_\alpha \leq c'_\alpha (x\sigma_\beta^2 + (x-2)\sigma_\alpha^2 + 2(1-x)\sigma_{\alpha,\beta})$$

e dunque, per x tendente a 0,

$$\mu_\beta - \mu_\alpha \leq 2c'_\alpha (\sigma_{\alpha,\beta} - \sigma_\alpha^2)$$

Lo stesso procedimento si applica anche al caso $x < 0$ nel quale si ottiene pertanto,

$$\mu_\beta - \mu_\alpha \geq 2c'_\alpha (\sigma_{\alpha,\beta} - \sigma_\alpha^2)$$

e dunque necessariamente:

$$(5.2.4) \quad \mu_\beta = \mu_\alpha + c_\alpha(\sigma_{\alpha,\beta} - \sigma_\alpha^2)$$

dove abbiamo posto per comodità $c_\alpha = 2c'_\alpha$.

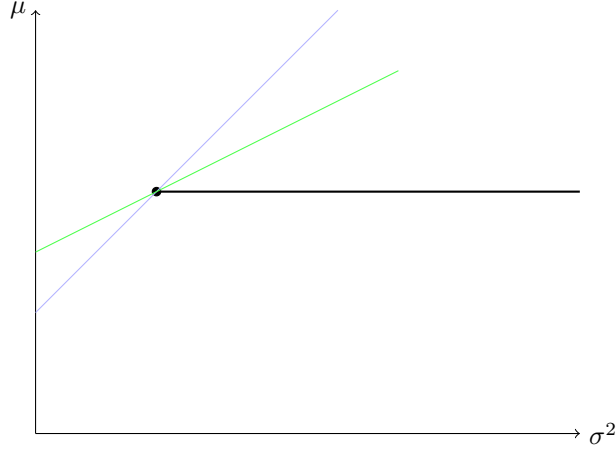


Figure 5: Il caso di rendimenti attesi identici.

5.3. Implicazioni.

La (5.2.4), ottenuta sulla base di semplici intuizioni geometriche e dell'applicazione di alcune proprietà degli insiemi convessi, consente in realtà numerose conclusioni importanti che cercheremo ora di mettere in evidenza. Prima di sviluppare tali implicazioni, soffermiamoci sul caso speciale nel quale, per un qualche portafoglio α efficiente si abbia $c_\alpha = 0$. In tale caso si deduce dalla (5.2.4) che tutti i portafogli hanno necessariamente il medesimo rendimento atteso e che pertanto essi possono differenziarsi solo per il livello della varianza. Dunque i portafogli efficienti si ottengono minimizzando la sola varianza e, conseguentemente, non può che esistere un solo portafoglio efficiente, quello, cioè, a varianza minima. L'insieme di mercato è rappresentato nella Figura ???. Come si nota facilmente dal grafico, quell'unico portafoglio efficiente ammette senz'altro molte rette tangenti ed alcune di esse hanno certamente inclinazione c'_α strettamente positiva. Dunque se α è un portafoglio efficiente possiamo concludere che la (5.2.4) debba valere con $c_\alpha > 0$.

Un altro caso ipotetico si ha qualora due portafogli efficienti, α_1 e α_2 siano tali per cui $c_{\alpha_1} = c_{\alpha_2} = c$. Sempre dalla (5.2.4) ricaviamo, sostituendo $\alpha = \alpha_1$ e $\beta = \alpha_2$ e viceversa, la conclusione secondo cui deve valere

$$0 = \sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2 - 2\sigma_{\alpha_1, \alpha_2} = \text{Var}(r_{\alpha_1} - r_{\alpha_2})$$

ossia $r_{\alpha_1} = a + r_{\alpha_2}$. Dunque uno dei due portafogli avrà un rendimento superiore all'altro, in contraddizione dell'ipotesi che siano entrambi efficienti, salvo che nel caso $a = 0$. Concludiamo pertanto che se α_1 e α_2 sono efficienti allora $c_{\alpha_1} = c_{\alpha_2}$ se e solo se $r_{\alpha_1} = r_{\alpha_2}$.

5.3.1. La caratterizzazione dei portafogli efficienti. Si noti anzitutto che il termine $\mu_\alpha - c_\alpha \sigma_\alpha^2$ che compare nella (5.2.4) dipende esclusivamente da α e può pertanto indicarsi sinteticamente come b_α , ossia, ricordando la (5.2.1),

$$b_\alpha = \mu_\alpha - c_\alpha \sigma_\alpha^2 = b'_\alpha - c'_\alpha \sigma_\alpha^2$$

e dunque la (5.2.4) diviene ¹

$$(5.3.1) \quad \mu_\beta = b_\alpha + c_\alpha \sigma_{\alpha,\beta} \quad \text{per qualunque portafoglio } \beta$$

sulla quale ci concentreremo d'ora in poi e che abbiamo ottenuto sulla base dell'ipotesi che il portafoglio α fosse efficiente. Si tratta pertanto di una condizione *necessaria*: ogni portafoglio α , *se efficiente*, necessariamente soddisfa la (5.3.1). Mostriamo ora che tale condizione è anche sufficiente.

Ammettiamo dunque che il portafoglio α sia tale per cui esistono due costanti b_α e c_α tali che $c_\alpha > 0$ e che vale la (5.2.4). Vogliamo mostrare che sotto queste condizioni α è effettivamente un portafoglio efficiente. Ammettiamo a tale scopo che un altro portafoglio, β , sia tale per cui $\mu_\beta > \mu_\alpha$. Dalla (5.2.4) ciò equivale necessariamente a $\sigma_{\alpha,\beta} > \sigma_\alpha^2$. Ricordiamo una conclusione generale circa la covarianza tra due variabili aleatorie, X e Y ²:

$$(5.3.2) \quad \text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$$

Dunque deve essere

$$\sigma_\alpha^2 \sigma_\beta^2 \geq \sigma_{\alpha,\beta}^2 > \sigma_\alpha^4$$

e pertanto $\sigma_\beta^2 > \sigma_\alpha^2$. Ne concludiamo che non vi è alcun portafoglio che domini α o, in altre parole, che quest'ultimo è un portafoglio efficiente.

Dunque volendo essere un po' pomposi potremmo mettere le nostre conclusioni sotto forma di un vero e proprio teorema

TEOREMA 1. *Un portafoglio α è efficiente se e solo se ha varianza minima ovvero se vi sono due costanti b_α e c_α con $c_\alpha > 0$ che soddisfano la (5.3.1).*

Osserviamo che la disuguaglianza $c_\alpha > 0$ è cruciale e deriva dal fatto che tutte le curve tangenti alla parte *superiore* della curva dei portafogli hanno necessariamente inclinazione positiva. Il caso $c_\alpha = 0$ corrisponde alla speciale situazione in cui tutti i portafogli hanno il medesimo rendimento atteso (si veda più sopra) mentre la $c_\alpha < 0$ contraddistingue invece quei portafogli che giacciono sulla parte inferiore della curva.

Una conclusione che possiamo trarre immediatamente è che due portafogli efficienti non possono avere il medesimo rendimento atteso tranne che siano lo stesso portafoglio. Supponiamo infatti che α_1 e α_2 siano efficienti e che $\mu_{\alpha_1} = \mu_{\alpha_2}$. Che debba aversi anche $\sigma_{\alpha_1}^2 = \sigma_{\alpha_2}^2$ è evidente dal fatto stesso che sono entrambi efficienti; la conclusione $r_{\alpha_1} = r_{\alpha_2}$ è molto più generale. Poiché il rendimento atteso dipende esclusivamente

¹Quest'ultima espressione non è solamente una implicazione della (5.2.4) ma una *reformulazione*, ossia è equivalente ad essa. Infatti sostituendo nella (5.3.1) β con α e sottraendo membro a membro si ricava nuovamente la (5.2.4).

²La (5.3.2) viene spesso chiamata disuguaglianza di *Cauchy-Schwartz*

dalla correlazione con un qualunque portafoglio efficiente, ed essendo tanto α_1 che α_2 efficienti, $\mu_{\alpha_1} = \mu_{\alpha_2}$ implica $\sigma_{\alpha_1, \alpha_2} = \sigma_{\alpha_1}^2 = \sigma_{\alpha_2}^2$. Da questa duplice uguaglianza non è difficile concludere che

$$\frac{\sigma_{\alpha_1, \alpha_2}}{\sigma_{\alpha_1} \sigma_{\alpha_2}} = 1 \quad \text{e} \quad \frac{\sigma_{\alpha_1, \alpha_2}}{\sigma_{\alpha_1}^2} = 1$$

ossia necessariamente $r_{\alpha_1} = r_{\alpha_2}$ (si veda la (23.2.9)).

5.3.2. Interpretazione economica. L'aspetto più significativo del Teorema 1, ossia della (5.3.1), sta nel fatto che, dato un qualunque portafoglio efficiente³, il rendimento atteso di un titolo risulta essere esclusivamente funzione della covarianza tra questo titolo ed il portafoglio stesso. Questa conclusione è tutt'altro che scontata e ben difficilmente poteva essere raggiunta sulla base della sola intuizione.

In particolare, il rendimento di un titolo è tanto più elevato quanto maggiore è la covarianza con il portafoglio efficiente. Questo implica che se la covarianza del titolo 1 con il portafoglio efficiente α è superiore a quella del titolo 2, allora la stessa conclusione vale per qualunque portafoglio efficiente. In simboli, se β e γ sono due portafogli qualunque e α_1 e α_2 due portafogli efficienti, allora $\sigma_{\beta, \alpha_1} > \sigma_{\gamma, \alpha_1}$ se e solo se $\sigma_{\beta, \alpha_2} > \sigma_{\gamma, \alpha_2}$.

L'intercetta b_α che figura nella (5.3.1) rappresenta conseguentemente il rendimento atteso di un qualunque portafoglio β che sia *ortogonale* al portafoglio efficiente α , ossia tale che $\sigma_{\alpha, \beta} = 0$. Se ad esempio esiste un titolo il cui rendimento, r_f è deterministico, se vi è cioè un titolo privo di rischio, allora esso è necessariamente ortogonale a qualunque altro portafoglio (efficiente o no) di modo che $b_\alpha = r_f$: in tal caso il valore del parametro b_α è il medesimo per tutti i portafogli efficienti.

Non è difficile, sostituendo β con α nella (5.3.1), concludere che $c_\alpha = \frac{\mu_\alpha - b_\alpha}{\sigma_\alpha^2}$ di modo che la (5.3.1) diviene

$$(5.3.3) \quad \mu_\beta = b_\alpha + \frac{\sigma_{\alpha, \beta}}{\sigma_\alpha^2} (\mu_\alpha - b_\alpha)$$

Nel caso speciale, poi, in cui esista un titolo privo di rischio, si ha

$$(5.3.4) \quad \mu_\beta = r_f + \frac{\sigma_{\alpha, \beta}}{\sigma_\alpha^2} (\mu_\alpha - r_f)$$

Il termine $\sigma_{\alpha, \beta} / \sigma_\alpha^2$ che compare nella precedente espressione riveste un'importanza particolare. Infatti, nella teoria della regressione lineare il rapporto

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

indica la misura della dipendenza lineare tra la variabile aleatoria Y e la X e viene spesso indicato con il termine *Beta* di Y rispetto a X . La (5.3.3) e la (5.3.4) si trovano dunque spesso scritte nella forma

$$\mu_\beta = b_\alpha + B_{\beta, \alpha} (\mu_\alpha - b_\alpha) \quad \text{e} \quad \mu_\beta = r_f + B_{\beta, \alpha} (\mu_\alpha - r_f)$$

³Ad eccezione del portafoglio a varianza minima

⁴Nel nostro caso questa terminologia è un pò infelice dato che spesso abbiamo indicato con la lettera β un generico portafoglio.

dove $B_{\beta,\alpha} = \sigma_{\alpha,\beta} / \sigma_\alpha^2$ è per l'appunto il *Beta* di β rispetto al portafoglio efficiente α^5 . Si noti infine che, laddove esista un titolo privo di rischio, la (5.3.4) si può equivalentemente riformulare in termini di *extra rendimenti* o, per usare la terminologia anglosassone, di *equity premium* osservando che dalla

$$\mu_\beta - r_f = B_{\beta,\alpha}(\mu_\alpha - r_f)$$

l'extra rendimento $\mu_\beta - r_f$ del portafoglio β rispetto al titolo privo di rischio risulta proporzionale all'extra rendimento del portafoglio efficiente e che tale fattore di proporzionalità è per l'appunto il *Beta*. Una volta noti i *Beta* dei vari titoli (i quali potrebbero calcolarsi ad esempio tramite una stima econometrica) sarà sufficiente valutare l'*equity premium* del portafoglio efficiente per ottenere una valutazione del rendimento di qualunque altro portafoglio.

5.4. L'agente rappresentativo

Vi è un modo ulteriore di determinare i portafogli efficienti sfruttando il fatto che questi sono i soli portafogli che verrebbero mai scelti da un investitore con preferenze del tipo media/varianza.

Consideriamo pertanto un agente che debba decidere le quote $\alpha_1, \dots, \alpha_N$ del proprio portafoglio, avendo a disposizione un capitale iniziale pari a W . Si noti che il rendimento *assoluto* dell'investimento, scegliendo un portafoglio α ammonta a $W r_\alpha$. Se le preferenze sono del tipo ipotizzato, allora il problema dell'investitore si può scrivere come

$$(5.4.1) \quad \max_{\{\alpha \in \mathbb{R}^N : \sum_{n=1}^N \alpha_n = 1\}} U(W \mu_\alpha, W^2 \sigma_\alpha^2)$$

Le condizioni del primo ordine sono

$$(5.4.2) \quad W U'_\mu \mu_n + W^2 U'_{\sigma^2} 2 \sum_{m=1}^N \sigma_{m,n} \alpha_m = -\lambda \quad n = 1, \dots, N$$

dove abbiamo indicato con U'_μ e U'_σ l'utilità marginale relativa al rendimento atteso e alla varianza, rispettivamente. Tali condizioni possono essere riscritte in forma matriciale come segue:

$$(5.4.3) \quad W U'_\mu E(r) + 2W^2 U'_{\sigma^2} \Sigma \alpha^* = -\lambda \mathbf{1}$$

dove $E(r)$ è il vettore dei rendimenti attesi (μ_1, \dots, μ_N) e Σ la matrice varianza/covarianza.

Applicando α^* e β alla (5.4.3) si ricava

$$(5.4.4) \quad \mu_\beta = \mu_{\alpha^*} - 2W \frac{U'_{\sigma^2}}{U'_\mu} (\sigma_{\alpha^*,\beta} - \sigma_{\alpha^*}^2)$$

Graficamente si ha la situazione raffigurata nella Figura 5.4, nella quale le linee con concavità verso l'alto rappresentano le curve di indifferenza.

⁵Poiché spesso il portafoglio efficiente α è considerato dato, si parla comunemente di Beta del portafoglio intendendo rispetto ad α .

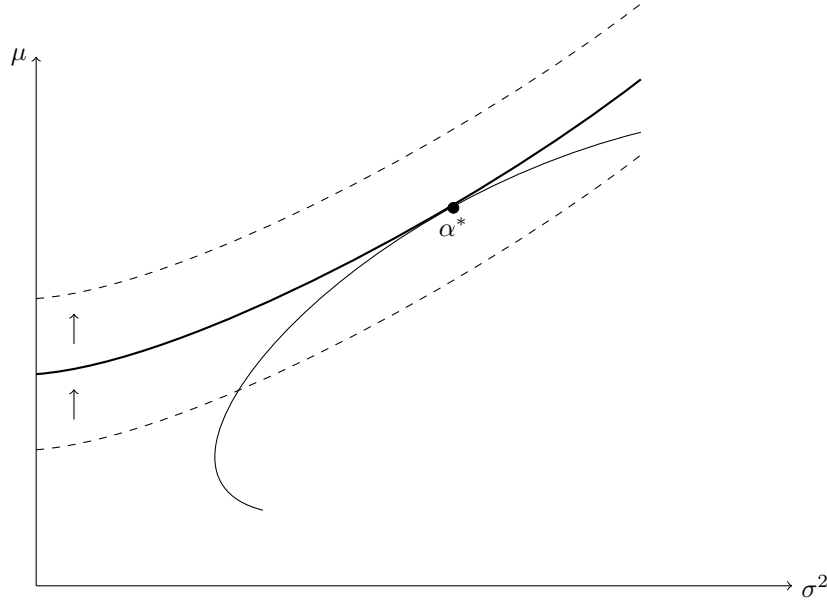


Figure 6: La scelta ottima dell'agente rappresentativo.

Raffrontando la (5.3.4) con la (5.4.4) si osserva anzitutto che, al variare del rapporto $c_{\alpha^*} = -2W \frac{U'_{\sigma^2}}{U'_\mu}$ da un agente all'altro, si determina un diverso portafoglio ottimale. Secondariamente, è del tutto evidente che il portafoglio α^* rappresentato nella (5.4.4) è senz'altro efficiente in quanto si ha $U'_{\sigma^2} > 0$. Il caso $c_{\alpha^*} = 0$ si ha laddove l'investitore sia indifferente rispetto al rischio. Si noti che in tal caso la scelta di portafoglio ammette una soluzione ottimale se e solo se $\mu_\beta = \mu_{\alpha^*}$ per tutti i portafogli β . In effetti, trattandosi di un agente neutrale al rischio ossia interessato solo al rendimento atteso, e potendo prendere posizioni scoperte, se i rendimenti attesi dei titoli fossero differenti, allora egli acquisterebbe i titoli col rendimento più elevato e venderebbe quelli col rendimento minore, in quantità infinita. Il caso $c_{\alpha^*} = \infty$ è anche interessante poiché corrisponde ad una retta tangente verticale, ossia passante per il portafoglio a varianza minima. In effetti questo investitore avrebbe vuoi $U'_{\sigma^2} = \infty$ oppure $U'_\mu = 0$. In entrambi i casi, vuoi perché totalmente avverso al rischio, vuoi perché disinteressato al rendimento atteso, la sua scelta verrebbe dominata dal desiderio di ridurre il rischio quanto più possibile. In terzo luogo, la scelta di portafoglio sarà in genere differente a seconda del capitale disponibile da investire. Infatti il rapporto c_{α^*} dipenderà in generale sia dalle preferenze che dall'ammontare del capitale. Scegliendo diverse preferenze e diverse dotazioni, si otterranno scelte di portafoglio diverse. Infine, ricollegandoci a quest'ultimo punto, è ben chiaro che facendo variare la funzione di utilità ed anche il capitale disponibile si può avere un qualunque valore del rapporto c_{α^*} , ossia un qualunque portafoglio, purché efficiente, può essere considerato l'investimento ottimale per un corrispondente agente economico.

5.4.1. Una nota sulle preferenze. Come già detto ripetutamente, le preferenze media varianza sono solo un tipo molto particolare di preferenze. Vogliamo tuttavia mostrare che esso non così special come si sarebbe indotti a credere. Consideriamo infatti due casi nei quali questo modello rappresenta senz'altro un'approssimazione adeguata del caso più generale.

Un criterio molto popolare per valutare le scelte caratterizzate da incertezza è quello dell'utilità attesa, ossia, data una funzione di utilità $U : \mathbb{R} \rightarrow \mathbb{R}$ ed una variabile aleatoria X , la quantità $U(X)$ è senza dubbio

essa stessa aleatoria e se ne può calcolare talvolta il valore atteso come

$$E\{U(X)\}$$

Parliamo in tal caso di utilità attesa della quantità X .

Poniamo che la funzione di utilità sia del tipo $U(x) = e^{tx}$ e che X sia una variabile normalmente distribuita. In tal caso

$$E(U(X)) = E(\exp(tx)) = \exp(tE(x) - \frac{1}{2}t^2 \text{Var}(x))$$

poiché l'utilità attesa coincide di fatto con la funzione generatrice dei momenti e dipende necessariamente solo dal valore atteso e dalla varianza che sono gli unici due parametri della distribuzione normale.

Per ottenere un'approssimazione di funzioni differenziabili a piacere è spesso utile ricorrere alla formula di Taylor. La funzione di utilità si può scrivere come

$$U(X) = U(E(X)) + \frac{dU}{dx}(a)(X - E(X)) + \frac{(X - E(X))^2}{2} \frac{d^2U}{dX^2}(a) + o((X - a)^3)$$

Passando al valore atteso si ottiene quindi

$$E(U(X)) = U(E(X)) + \frac{\text{Var}(X)}{2} \frac{d^2U}{dX^2}(E(X)) + \dots$$

di modo che, se i termini di ordine superiore al secondo sono trascurabili, allora l'utilità attesa si approssima con una funzione di media e varianza. Si noti che se la funzione U è concava (il che denota avversione al rischio), allora la varianza entra nella funzione approssimata con segno negativo.

Il teorema dei due fondi.

Non è affatto detto che i titoli che formano un portafoglio efficiente siano essi stessi efficienti. Vi sono anzi numerosi esempi numerici (alcuni compaiono sotto forma di esercizi) nei quali titoli che singolarmente sarebbero dominati possono essere combinati tra loro formando un portafoglio efficiente. Verrebbe tuttavia da supporre che combinando tra loro due portafogli efficienti se ne ottenga uno necessariamente efficiente. Proviamo a verificare se questa intuizione è corretta.

6.1. Il Teorema

6.1.1. La condizione necessaria. Poniamo che α_1 e α_2 siano due portafogli efficienti e che il portafoglio α sia ottenuto investendo una porzione ϵ nel portafoglio α_1 e la rimanente parte $1 - \epsilon$ nel portafoglio α_2 . Dunque scriviamo

$$(6.1.1) \quad r_\epsilon = \epsilon r_{\alpha_1} + (1 - \epsilon) r_{\alpha_2}$$

Naturalmente, essendo α_1 e α_2 efficienti vi saranno costanti b_{α_1} , b_{α_2} , c_{α_1} e c_{α_2} tali che $c_{\alpha_1} > c_{\alpha_2} > 0$ e che per qualunque portafoglio β

$$\mu_\beta = b_{\alpha_1} + c_{\alpha_1} \sigma_{\alpha_1, \beta} = b_{\alpha_2} + c_{\alpha_2} \sigma_{\alpha_2, \beta}$$

Dato che

$$\sigma_{\alpha_i, \beta} = \frac{\mu_\beta - b_{\alpha_i}}{c_{\alpha_i}} \quad \text{per} \quad i = 1, 2$$

allora

$$\begin{aligned} \sigma_{\epsilon, \beta} &= \epsilon \sigma_{\alpha_1, \beta} + (1 - \epsilon) \sigma_{\alpha_2, \beta} \\ &= \epsilon \frac{\mu_\beta - b_{\alpha_1}}{c_{\alpha_1}} + (1 - \epsilon) \frac{\mu_\beta - b_{\alpha_2}}{c_{\alpha_2}} \\ &= \mu_\beta \left(\frac{\epsilon}{c_{\alpha_1}} + \frac{1 - \epsilon}{c_{\alpha_2}} \right) - \left(\frac{\epsilon b_{\alpha_1}}{c_{\alpha_1}} + \frac{(1 - \epsilon) b_{\alpha_2}}{c_{\alpha_2}} \right) \\ &= \left(\frac{\epsilon}{c_{\alpha_1}} + \frac{1 - \epsilon}{c_{\alpha_2}} \right) \left(\mu_\beta - \frac{\epsilon b_{\alpha_1} c_{\alpha_2} + (1 - \epsilon) b_{\alpha_2} c_{\alpha_1}}{\epsilon c_{\alpha_2} + (1 - \epsilon) c_{\alpha_1}} \right) \\ &= \frac{\mu_\beta - b_\epsilon}{c_\epsilon} \end{aligned}$$

ossia

$$(6.1.2) \quad \mu_\beta = b_\epsilon + c_\epsilon \sigma_{\epsilon, \beta}$$

dove evidentemente abbiamo posto

$$b_\epsilon = \frac{\epsilon b_{\alpha_1} c_{\alpha_2} + (1 - \epsilon) b_{\alpha_2} c_{\alpha_1}}{\epsilon c_{\alpha_2} + (1 - \epsilon) c_{\alpha_1}} \quad \text{e} \quad c_\epsilon = \left(\frac{\epsilon}{c_{\alpha_1}} + \frac{1 - \epsilon}{c_{\alpha_2}} \right)^{-1}$$

Naturalmente, questa espressione ha un senso matematico compiuto solo se

$$\frac{\epsilon}{c_{\alpha_1}} + \frac{1 - \epsilon}{c_{\alpha_2}} \neq 0 \quad \text{ossia} \quad \epsilon c_{\alpha_2} + (1 - \epsilon) c_{\alpha_1} \neq 0$$

In particolare $c_\epsilon > 0$ se e solo se

$$(6.1.3) \quad \epsilon < \frac{c_{\alpha_1}}{c_{\alpha_1} - c_{\alpha_2}}$$

Dunque il portafoglio formato come nella (6.1.1) è efficiente se e solo se il parametro ϵ soddisfa la (6.1.3).

6.1.2. La condizione sufficiente. Possiamo anche stabilire una conclusione per certi aspetti inversa. Si considerino a tale scopo tre portafogli efficienti, α , α_1 e α_2 tali che $\mu_{\alpha_1} > \mu_{\alpha_2}$ ¹. Si noti che, essendo α efficiente dalla (5.3.1) discende che per qualche parametro $c_\alpha > 0$ si ha

$$0 < \mu_{\alpha_1} - \mu_{\alpha_2} = c_\alpha(\sigma_{\alpha, \alpha_1} - \sigma_{\alpha, \alpha_2})$$

di modo che la disuguaglianza precedente equivale alla $\sigma_{\alpha, \alpha_1} > \sigma_{\alpha, \alpha_2}$. Poniamo $\varepsilon = \frac{\mu_\alpha - \mu_{\alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}}$ e $r_\varepsilon = \varepsilon r_{\alpha_1} + (1 - \varepsilon)r_{\alpha_2}$. Evidentemente, per costruzione, $\mu_\varepsilon = \mu_\alpha$. Inoltre si noti che

$$\varepsilon = \frac{\mu_\alpha - \mu_{\alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}} = c_{\alpha_1} \frac{\sigma_{\alpha, \alpha_1} - \sigma_{\alpha_1, \alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}}$$

e parimenti

$$1 - \varepsilon = \frac{\mu_{\alpha_1} - \mu_\alpha}{\mu_{\alpha_1} - \mu_{\alpha_2}} = c_{\alpha_2} \frac{\sigma_{\alpha_1, \alpha_2} - \sigma_{\alpha, \alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}}$$

di modo che

$$\begin{aligned} \frac{\varepsilon}{c_{\alpha_1}} + \frac{1 - \varepsilon}{c_{\alpha_2}} &= \frac{\sigma_{\alpha, \alpha_1} - \sigma_{\alpha_1, \alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}} + \frac{\sigma_{\alpha_1, \alpha_2} - \sigma_{\alpha, \alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}} \\ &= \frac{\sigma_{\alpha, \alpha_1} - \sigma_{\alpha, \alpha_2}}{\mu_{\alpha_1} - \mu_{\alpha_2}} \\ &> 0 \end{aligned}$$

Di conseguenza, ponendo

$$c_\varepsilon = \frac{\mu_{\alpha_1} - \mu_{\alpha_2}}{\sigma_{\alpha, \alpha_1} - \sigma_{\alpha, \alpha_2}}$$

concludiamo che la disuguaglianza (6.1.3) è rispettata e pertanto che il portafoglio ϵ è anch'esso efficiente. Inoltre, essendo per costruzione $\mu_\alpha = \mu_\epsilon$ segue che $\sigma_{\alpha, \alpha_0} = \sigma_{\epsilon, \alpha_0}$ per ogni portafoglio α_0 efficiente. Sostituendo $\alpha_0 = \alpha$ e $\alpha_0 = \epsilon$ otteniamo la conclusione $\sigma_\alpha^2 = \sigma_{\epsilon, \alpha} = \sigma_\epsilon^2$. Dunque, dalle proprietà del modello lineare, $r_\alpha = r_\epsilon$.

Concludiamo dunque che dati due portafogli efficienti, ogni altro portafoglio efficiente può essere replicato esattamente investendo opportunamente nei due portafogli dati. Ovvero, riassumendo, abbiamo dimostrato il risultato seguente:

TEOREMA 2. *Siano dati due portafogli efficienti, α_1 e α_2 con $c_{\alpha_1} > c_{\alpha_2}$. Un portafoglio α è efficiente se e solo può scriversi come*

$$r_\alpha = \epsilon r_{\alpha_1} + (1 - \epsilon)r_{\alpha_2}$$

$$\text{con } \epsilon < \frac{c_{\alpha_1}}{c_{\alpha_1} - c_{\alpha_2}}$$

¹Infatti, come osservato più sopra, nel caso $\mu_{\alpha_1} = \mu_{\alpha_2}$ i due portafogli sarebbero di fatto identici.

6.2. L'interpretazione finanziaria

Questo risultato è molto importante ed è noto in letteratura come *teorema dei due fondi*. Il suo significato in termini economici è chiaro. Un investitore può accedere a qualunque opportunità efficiente di investimento purché possa investire in due fondi di investimento i cui portafogli siano efficienti. In altri termini non è affatto necessario che il singolo risparmiatore compia complesse (e talvolta costose) operazioni di investimento per ottenere il portafoglio che egli considera ottimale, poiché basta che si rivolga ad un intermediario (una banca ad esempio) che dispone di due fondi di investimento efficienti limitandosi a decidere quanto investire nel primo e quanto nel secondo.

Per comprendere meglio la portata di questo risultato, si osservi che, qualora il numero N dei titoli trattati sul mercato sia ampio, la scelta ottimale di portafoglio può implicare l'acquisto di un gran numero di titoli. Questo, nel modo reale, potrebbe non essere agevole per almeno due distinti motivi. Anzitutto perché l'acquisto o la vendita di titoli implica un costo di transazione tale per cui, quanto maggiore è il numero delle transazioni tanto più elevato il costo che si sostiene. Secondariamente, e a differenza delle nostre ipotesi, la possibilità di frazionare l'acquisto di un titolo è limitata dunque chi dispone di un capitale esiguo non riesce ad acquistare che un numero assai ridotto di titoli.

Pertanto, l'opportunità offerta al singolo investitore da un intermediario che disponga di due fondi di investimento efficienti è davvero preziosa e può leggersi come razionalizzazione del ruolo degli intermediari finanziari.

Si noti inoltre che non sarà difficile calcolare la composizione del portafoglio a varianza minima (il quale per definizione è senz'altro efficiente). Infatti se α_1 e α_2 sono efficienti e se $\alpha = \epsilon\alpha_1 + (1 - \epsilon)\alpha_2$, allora

$$\sigma_\alpha^2 = \epsilon^2(\sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2 - 2\sigma_{\alpha_1\alpha_2}) - 2\epsilon(\sigma_{\alpha_2}^2 - \sigma_{\alpha_1,\alpha_2}) + \sigma_{\alpha_2}^2$$

dalla quale, minimizzando rispetto ad ϵ , si ottiene

$$(6.2.1) \quad \alpha_{\min} = \frac{\sigma_{\alpha_2}^2 - \sigma_{\alpha_1,\alpha_2}}{\sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2 - 2\sigma_{\alpha_1\alpha_2}}\alpha_1 + \frac{\sigma_{\alpha_1}^2 - \sigma_{\alpha_1,\alpha_2}}{\sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2 - 2\sigma_{\alpha_1\alpha_2}}\alpha_2 \quad \sigma_{\min}^2 = \frac{\sigma_{\alpha_1}^2 \sigma_{\alpha_2}^2 - \sigma_{\alpha_1,\alpha_2}^2}{\sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2 - 2\sigma_{\alpha_1\alpha_2}}$$

L'esistenza di un titolo privo di rischio

Abbiamo già incontrato l'ipotesi che esista un titolo il cui rendimento, r_f , risulti privo di rischio. Naturalmente, nella realtà questa assunzione andrebbe discussa anche se spesso si identifica tale titolo con un buono del Tesoro con scadenza breve, diciamo tre mesi. Altre volte, è il tasso *LIBOR* che nelle indagini empiriche ricopre questo ruolo. Le note vicende della crisi finanziaria recente indicano tuttavia che tale scelta sia quanto meno discutibile.

7.1. Il CAPM con titolo privo di rischio

Si ricordi la (5.3.4): se α è un portafoglio efficiente e β un qualunque altro portafoglio,

$$\mu_\beta = r_f + \frac{\sigma_{\alpha,\beta}}{\sigma_\alpha^2}(\mu_\alpha - r_f)$$

che possiamo anche riscrivere come

$$\frac{\mu_\beta - r_f}{\sigma_\beta} = \frac{\sigma_{\alpha,\beta}}{\sigma_\alpha \sigma_\beta} \frac{\mu_\alpha - r_f}{\sigma_\alpha}$$

Dunque il migliore rapporto *rischio/rendimento* si ottiene formando quel portafoglio che abbia la correlazione massima con il portafoglio efficiente, spesso identificato con il *portafoglio di mercato*. Implicitamente, nel fare questo ragionamento abbiamo identificato il rischio del portafoglio β con lo scarto quadratico σ_β . Naturalmente, non cambia granché nel misurare il rischio attraverso lo scarto quadratico anziché la varianza. Tuttavia, nel caso specifico in cui vi è un titolo privo di rischio, l'analisi della frontiera efficiente risulta particolarmente facile.

7.2. La *market line*

Supponiamo di investire una percentuale ϵ in un portafoglio rischioso α e la rimanente parte nel titolo privo di rischio. Se ne otterrà un rendimento atteso pari a $\mu = \epsilon\mu_\alpha + (1 - \epsilon)r_f$ ed una varianza pari a $\sigma^2 = \epsilon^2\sigma_\alpha^2$. Queste sono le stesse formule viste più sopra per il caso di un mercato popolato da due soli titoli (si veda la (4.2.1) e la (23.2.6)). Per sostituzione si ottiene

$$(7.2.1) \quad \sigma = \sigma_\alpha \frac{\mu - r_f}{\mu_\alpha - r_f} = \frac{\sigma_\alpha}{\mu_\alpha - r_f}(\mu - r_f)$$

dove il fattore

$$(7.2.2) \quad \frac{\mu_\alpha - r_f}{\sigma_\alpha}$$

viene spesso indicato in letteratura come *Sharpe ratio*. Dunque vi è una relazione precisa tra varianza di un portafoglio ed il suo extra rendimento e questa dipende, tramite il reciproco dello *Sharpe ratio*, dal portafoglio α considerato. Ecco che la scelta di investimento di un agente con preferenze *mean/variance* si può descrivere semplicemente attraverso due semplici stadi. (i) si tratta anzitutto di trovare il portafoglio rischioso α che abbia il valore minimo dello *Sharpe ratio* e, dato questo, (ii) combinare in modo ottimale il portafoglio α col titolo privo di rischio in modo da ottenere la combinazione di rendimento atteso e varianza più confacente alle proprie preferenze.

Questo argomento, originariamente proposto dallo stesso Sharpe, parrebbe suggerire che vi sia un unico titolo in corrispondenza del quale il valore dello *Sharpe ratio* è massimo. Se così fosse, le conclusioni del lemma dei due fondi viste più sopra sarebbero false, in quanto da esse si deduce che qualunque portafoglio combinato con il titolo privo di rischio (senz'altro efficiente) fornisce tutte le migliori opportunità di investimento. Questa apparente contraddizione si risolve osservando che tutti i portafogli efficienti hanno il medesimo *Sharpe ratio* il cui valore corrisponde al massimo tra tutti i portafogli esistenti sul mercato.

Per raggiungere questa conclusione osserviamo, dalla (5.2.4), che se α_1 e α_2 sono entrambi efficienti e se esiste un titolo privo di rischio allora

$$\frac{\mu_{\alpha_1} - r_f}{\sigma_{\alpha_1}} = c_{\alpha_1} \sigma_{\alpha_1} = c_{\alpha_2} \frac{\sigma_{\alpha_2, \alpha_1}}{\sigma_{\alpha_1}} = \frac{\sigma_{\alpha_2, \alpha_1}}{\sigma_{\alpha_1} \sigma_{\alpha_2}} c_{\alpha_2} \sigma_{\alpha_2} = \frac{\sigma_{\alpha_2, \alpha_1}}{\sigma_{\alpha_1} \sigma_{\alpha_2}} \frac{\mu_{\alpha_2} - r_f}{\sigma_{\alpha_2}} \leq \frac{\mu_{\alpha_2} - r_f}{\sigma_{\alpha_2}}$$

Ovviamente, anche la disuguaglianza inversa deve valere. In altre parole, si ottiene una ulteriore caratterizzazione dei portafogli efficienti: i portafogli efficienti sono i portafogli ai quali è associato il massimo valore dello *Sharpe ratio*. Non solo ma, come è facile vedere dai passaggi precedenti, deve necessariamente valere

$$\frac{\sigma_{\alpha_2, \alpha_1}}{\sigma_{\alpha_1} \sigma_{\alpha_2}} = 1$$

Dunque i rendimenti dei portafogli efficienti presentano una correlazione lineare positiva perfetta.

In linea di principio il problema dell'investimento presenta una forma molto semplice come il problema della scelta tra un solo titolo (portafoglio) rischioso e un titolo privo di rischio, quale ad esempio la moneta. Questa conclusione conforta in parte l'approccio macroeconomico alle scelte finanziarie, compendiato nella curva *LM* dei manuali e secondo il quale il problema potrebbe semplicemente vedersi come la scelta tra *moneta e titoli*, nonostante le attività finanziarie alternative alla moneta siano moltissime. Dal punto di vista grafico, l'analisi diviene semplicissima:

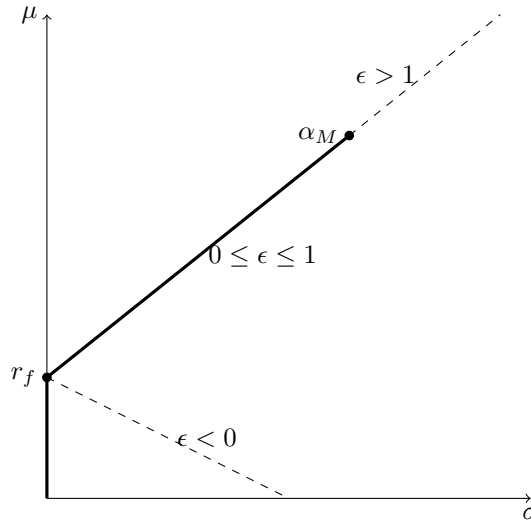


Figure 7: La frontiera efficiente in presenza di un titolo privo di rischio.

Dal grafico si vede con chiarezza che non è mai efficiente prendere una posizione corta sul portafoglio efficiente ($\epsilon < 0$) per investire nel titolo privo di rischio. Infatti, investire più del 100% nel titolo privo di

rischio non può contribuire a ridurre ulteriormente la varianza mentre produce un abbassamento del rendimento atteso in quanto il portafoglio efficiente, essendo rischioso, deve necessariamente offrire un rendimento atteso strettamente superiore a r_f .

Il calcolo dei portafogli efficienti.

La condizione (5.3.1) è dunque densa di implicazioni importanti, come abbiamo visto. Tuttavia fino ad ora non disponiamo di un metodo per poter calcolare in modo esplicito la composizione di tutti i portafogli efficienti. Vogliamo pertanto utilizzare ora la (5.3.1) per ottenere una formula esplicita per calcolare i portafogli efficienti.

8.1. Matrice Σ invertibile.

Sostituiamo nella (5.3.1) β con un portafoglio che consiste esclusivamente nel titolo n . Si avrà in tal caso $\mu_n = b_\alpha + b_\alpha \sigma_{\alpha,n}$ ossia, per $n = 1, \dots, N$:

$$\begin{aligned}\mu_1 &= b_\alpha + c_\alpha \sum_{n=1}^N \alpha_n \sigma_{n,1} \\ \mu_2 &= b_\alpha + c_\alpha \sum_{n=1}^N \alpha_n \sigma_{n,2} \\ &\vdots \\ \mu_N &= b_\alpha + c_\alpha \sum_{n=1}^N \alpha_n \sigma_{n,N}\end{aligned}$$

ovvero, in termini matriciali,

$$(8.1.1) \quad \mu = b_\alpha \mathbf{1} + c_\alpha \Sigma \alpha$$

dove con Σ abbiamo indicato la matrice con elemento generico dato dalla covarianza $\sigma_{n,m}$, con μ il vettore (colonna) con elemento generico μ_n , con α il vettore degli elementi α_n e con $\mathbf{1}$ un vettore di tutti elementi unitari. Dalla (8.1.1) si ricava la soluzione $\Sigma^{-1}(\mu - b_\alpha \mathbf{1}) = c_\alpha \alpha$ dunque

$$(8.1.2) \quad \alpha = \frac{\Sigma^{-1}(\mu - b_\alpha \mathbf{1})}{\mathbf{1}^T \Sigma^{-1}(\mu - b_\alpha \mathbf{1})}$$

la quale naturalmente presuppone che la matrice Σ sia invertibile e che $c_\alpha > 0$.

Possiamo esprimere la (8.1.1) in funzione di c_α ponendo

$$\alpha = \Sigma^{-1}(\mu - b_\alpha \mathbf{1}) \frac{1}{c_\alpha}$$

e, tenendo conto che $\mathbf{1}^T \alpha = 1$,

$$b_\alpha = \frac{\mathbf{1}^T \Sigma^{-1} \mu - c_\alpha}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$$

da cui infine si ricava per sostituzione

$$\begin{aligned}(8.1.3) \quad \alpha &= \frac{1}{c_\alpha} \left\{ \Sigma^{-1} \mu - \frac{\mathbf{1}^T \Sigma^{-1} \mu - c_\alpha}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \Sigma^{-1} \mathbf{1} \right\} \\ &= \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} + \frac{1}{c_\alpha} \left\{ \Sigma^{-1} \mu - \frac{\mathbf{1}^T \Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right\} \\ &= \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \left(1 - \frac{\mathbf{1}^T \Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \frac{1}{c_\alpha} \right) + \frac{\Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mu} \left(\frac{\mathbf{1}^T \Sigma^{-1} \mu}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \frac{1}{c_\alpha} \right) \\ &= \alpha_{\min}(1 - \varepsilon_\alpha) + v_* \varepsilon_\alpha\end{aligned}$$

dove abbiamo posto

$$\alpha_{\min} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma^{-1}\mathbf{1}} \quad v_* = \frac{\Sigma^{-1}\mu}{\mathbf{1}^T \Sigma^{-1}\mu} \quad \varepsilon_\alpha = \frac{\mathbf{1}^T \Sigma^{-1}\mu}{\mathbf{1}^T \Sigma^{-1}\mathbf{1}} \frac{1}{c_\alpha}$$

Osserviamo infatti che il vettore α_{\min} corrisponde al portafoglio ottenuto sostituendo $c_\alpha = \infty$ nella (8.1.1). Esso corrisponde pertanto al portafoglio con varianza minima e la somma dei suoi elementi da in effetti 1. Il vettore v_* corrisponde ad un ulteriore portafoglio che corrisponde al parametro

$$c_\alpha = \frac{\mathbf{1}^T \Sigma^{-1}\mu}{\mathbf{1}^T \Sigma^{-1}\mathbf{1}}$$

Questo valore corrisponde al limite cui tende b_α allorché $c_\alpha \rightarrow 0$. Come si vede sul grafico, il valore dell'intercetta, b_α , appunto è crescente al decrescere della pendenza, cioè c_α . Dunque, il suddetto valore è senz'altro positivo. Abbiamo dunque una scomposizione del tutto simile a quella del Lemma dei due fondi, con l'eccezione che i due portafogli α_{\min} e v_* sono fissati e non arbitrari. Si vede tuttavia che il Lemma dei due fondi segue in via diretta dalla scomposizione appena ottenuta.

Consideriamo a titolo di applicazione il seguente esempio:

ESEMPIO 3. *Sul mercato si ha $\mu = [1 \ 3 \ 5 \ 2]^T$ e*

$$\Sigma = \begin{bmatrix} 1 & -1 & 0,7 & 0,2 \\ -1 & 4 & 0,2 & 2 \\ 0,7 & 0,2 & 2 & 0,6 \\ 0,2 & 2 & 0,6 & 0,8 \end{bmatrix} \quad \Sigma^{-1} = \begin{bmatrix} 0,50566 & 0,71092 & 0,04531 & -1,1735 \\ 0,71092 & 0,39465 & -0,31264 & 0,097412 \\ 0,04531 & -0,31264 & 0,36248 & 0,61169 \\ -1,1735 & 0,097412 & 0,61169 & -0,84277 \end{bmatrix}$$

Ponendo $b_\alpha = 0,5$ si ha $\mu - 0,5 \times \mathbf{1} = [0,5 \ 2,5 \ 4,5 \ 1,5]^T$ e quindi

$$\Sigma^{-1}[\mu - 0,5\mathbf{1}_4] = [0,47378 \ 0,08132 \ 1,7898 \ 1,1452]^T$$

Di conseguenza,

$$\alpha = [0,1357 \ 0,0233 \ 0,5128 \ 0,3281]^T$$

Un'altra possibilità, molto opportuna dal punto di vista dei calcoli, è $b_\alpha = 0$ dalla quale segue

$$\alpha = [0,1408 \ 0,1431 \ 0,5825 \ 0,1336]^T$$

Tornando al problema dell'investitore rappresentativo esaminato più sopra, non è difficile vedere che dalla (5.4.3) si ricava agevolmente

$$\alpha^* = -\frac{WU'_\mu}{2W^2U'_{\sigma^2}} \Sigma^{-1}(\lambda/WU'_\mu \mathbf{1} + \mu)$$

da cui, posto $\bar{\lambda} = -\lambda/WU'_\mu$, si ha

$$\mathbf{1}^T \Sigma^{-1}(\mu - \bar{\lambda}\mathbf{1}) = -\frac{2W^2U'_{\sigma^2}}{WU'_\mu}$$

e dunque

$$(8.1.4) \quad \alpha^* = \frac{\Sigma^{-1}(\mu - \bar{\lambda}\mathbf{1})}{\mathbf{1}^T \Sigma^{-1}(\mu - \bar{\lambda}\mathbf{1})}$$

In altri termini, dal paragone con la (8.1.2), si conclude che variando le preferenze e la disponibilità del capitale, si ottengono tutti i diversi portafogli efficienti, come già si era intuito nei paragrafi precedenti.

8.2. Matrice Σ singolare.

Vale la pena considerare anche il caso in cui Σ non sia invertibile, condizione che si incontra senz'altro se uno dei titoli esistenti è privo di rischio. Torniamo a tale scopo sulla (8.1.1) la quale è comunque una condizione necessaria e sufficiente affinché α sia un portafoglio efficiente.

Indichiamo con $\bar{\Sigma}$ la matrice $N \times (N+1)$ ottenuta aggiungendo a Σ una prima colonna di tutti elementi unitari. Inoltre se α è un portafoglio allora con $\bar{\alpha}$ indichiamo il vettore ottenuto aggiungendo ad α una prima componente, α_0 . Ossia

$$\bar{\Sigma} = [\mathbf{1}_N | \Sigma] \quad \bar{\alpha} = [\alpha_0 | \alpha]$$

Non è difficile rendersi conto che la (8.1.1) risulta equivalente alla

$$(8.2.1) \quad \mu = c_\alpha \bar{\Sigma} \bar{\alpha}$$

Come facilmente si comprende, la (8.2.1) ammette infinite soluzioni se solo la matrice $\bar{\Sigma}$ ha rango N . Un caso particolare in cui questa condizione ricorre è quello in cui la matrice Σ è invertibile, ma altri casi possono essere considerati. Se ad esempio tra le attività finanziarie ve n'è una priva di rischio, allora una colonna di $\bar{\Sigma}$ sarà composta da elementi nulli: il sistema continua ad ammettere una soluzione se, eliminando tale colonna, si ottiene un minore di $\bar{\Sigma}$ di ordine N .

Vi possono essere alcuni tra i titoli esistenti tali per cui, eliminando la corrispondente colonna, si ottiene una sottomatrice invertibile di $\bar{\Sigma}$. In tal caso si ottiene la soluzione

$$(\bar{\Sigma}^n)^{-1} (\mu - c_\alpha \alpha_n \bar{\Sigma}_n) = \bar{\alpha}^n$$

dove $\bar{\Sigma}^n$ e $\bar{\alpha}^n$ sono stati ottenuti da $\bar{\Sigma}$ e $\bar{\alpha}$ eliminando la colonna n , indicata rispettivamente da $\bar{\Sigma}_n$ e α_n . In tal caso una soluzione, ossia un portafoglio efficiente, può essere trovato per ogni valore di $\bar{\alpha}_n$, ad esempio $\bar{\alpha}_n = 0$.

Estensioni

Naturalmente, l'ipotesi che le preferenze degli individui siano del tipo media/varianza sottostante tutta la teoria della frontiera efficiente, è discutibile. Il fatto stesso che il *CAPM* non abbia avuto conferme empiriche definitive induce a pensare che un più generale modello di preferenze andrebbe forse esaminato.

9.1. La scelta di portafoglio in generale

Immaginiamo allora che vi siano due soli periodi, $t = 0$ e $t = 1$ e che, alla data $t = 1$ si aprano $k = 1, \dots, K$ possibili scenari¹. Indichiamo con c_0 il consumo nel primo periodo e con c_k il consumo nel secondo periodo nel caso si verifichi lo stato ω_k . Utilizziamo gli stessi indici anche per descrivere la ricchezza W disponibile all'agente (ad esempio il reddito da lavoro) in ogni circostanza. Le preferenze dell'agente rappresentativo potrebbero allora essere del tipo $U(c_0, c_1, \dots, c_K)$ ed il suo vincolo di bilancio

$$(9.1.1) \quad c_0 = W_0 - \sum_{n=1}^N \theta_n S_0^n \quad c_k = W_k + \sum_{n=1}^N \theta_n S_{1,k}^n \quad k = 1, \dots, K$$

Naturalmente con S_0^n indichiamo il prezzo di acquisto dell'attività finanziaria n -ma, con θ_n il numero di unità acquistate (dunque $\alpha_n = \theta_n S_0^n / W_0$) e con $S_1^n = S_{1,1}^n, \dots, S_{1,K}^n$ il valore finale del titolo n , comprensivo delle cedole eventualmente maturate.

Le condizioni del primo ordine di questo problema (in assenza di vincoli ulteriori) sono le seguenti

$$-\frac{\partial U}{\partial c_0} S_0^n + \sum_{k=1}^K \frac{\partial U}{\partial c_k} S_{1,k}^n = 0 \quad n = 1, \dots, N$$

od anche, supponendo $S_0^n \neq 0$ e ponendo $r_k^n = S_{1,k}^n / S_0^n - 1$,

$$(9.1.2) \quad 1 = \sum_{k=1}^K \frac{\partial U / \partial c_k}{\partial U / \partial c_0} (1 + r_k^n) \quad n = 1, \dots, N$$

Introduciamo ora un ulteriore titolo, $n = 0$, tale che

$$r_k^0 = \frac{1}{\sum_{k=1}^K \frac{\partial U / \partial c_k}{\partial U / \partial c_0}} - 1$$

ossia privo di rischio e tale che la (9.1.2) è soddisfatta per definizione. Naturalmente il titolo $n = 0$ potrebbe benissimo non esistere sul mercato². In ogni caso, questo porta alla conclusione

$$(9.1.3) \quad 0 = \sum_{k=1}^K \bar{Z}_k (r_k^n - r_s^0) \quad n = 0, 1, \dots, N$$

¹Di questa ipotesi non c'è bisogno se le preferenze sono del tipo media/varianza

²Visto che la (9.1.2) si estende anche al titolo $n = 0$, la scelta di portafoglio dell'agente non viene modificata dall'eventuale introduzione di questo ulteriore titolo.

dove abbiamo posto

$$(9.1.4) \quad \bar{Z}_k = \frac{\frac{\partial U}{\partial c_k}}{\frac{\partial U}{\partial c_0} \sum_{k=1}^K \frac{\partial U}{\partial c_k}}$$

La variabile aleatoria \bar{Z} ha due caratteristiche interessanti: (i) $\bar{Z}_k > 0$ per $k = 1, \dots, K$, (ii) $\sum_{k=1}^K \bar{Z}_k = 1$. In particolare la (i) discende dall'ipotesi che il saggio marginale di sostituzione sia positivo, ovvero che l'agente rappresentativo non sia disponibile ad accrescere a qualunque costo il consumo alla data $t = 0$. Tuttavia, queste proprietà ci consentono di definire una nuova misura di probabilità, Q , data implicitamente da

$$(9.1.5) \quad Q(F) = \sum_{\{1 \leq k \leq K : k \in F\}} \bar{Z}_k$$

In termini di questa nuova misura si ha

$$(9.1.6) \quad r^0 = E_Q(r^n) \quad \text{ossia} \quad S_0^n = E_Q\left(\frac{S_1^n}{1+r^0}\right) \quad n = 0, 1, \dots, N$$

In pratica, stando alla (9.1.6), il prezzo delle attività finanziarie si determinerebbe come il valore scontato atteso del suo valore futuro. Naturalmente questa conclusione sarebbe scontata se l'agente fosse neutrale al rischio poiché con preferenze siffatte, come sappiamo, conta solo il valore atteso delle quantità incerte. Tuttavia, l'agente che consideriamo qui non è necessariamente di questo tipo e in effetti la (9.1.6) vale solo dopo aver sostituito la misura di probabilità iniziale con la Q che per tale ragione viene denominata, *probabilità aggiustata per il rischio* o *probabilità neutrale al rischio*. Ovviamente, se la probabilità iniziale viene indicata con P e se poniamo $Z_k = \bar{Z}_k/P(\omega_k)$ e $\bar{S}_1 = S_1/(1+r^0)$, allora possiamo anche scrivere la (9.1.6) come

$$(9.1.7) \quad r^0 = E_P(Zr^n) \quad \text{ossia} \quad S_0^n = E_P(Z\bar{S}_1^n) \quad n = 0, 1, \dots, N$$

Naturalmente $\bar{S}_1^0 = S_0^0$ per cui possiamo porre a piacere $S_0^0 = 1$ senza perdita di generalità. Il titolo $n = 0$ si interpreta come un *bond* privo di rischio.

9.2. La frontiera efficiente

Alcuni dei risultati visti nei capitoli precedenti possono essere riprodotti qui con alcune semplici manipolazioni.

Anzitutto, la (9.1.7) si può facilmente modificare per ottenere

$$(9.2.1) \quad \mu_n = E_P(r^n) = E_P(r^n)E_P(Z) = E_P(Zr^n) - \text{Cov}(Z, r^n) = r^0 - \text{Cov}(Z, r^n) = r^0 + \frac{\text{Cov}(-Z, r^n)}{\sigma_Z^2} \sigma_Z^2$$

Si ricordi inoltre che, dalla (9.1.4), Z è, in ultima istanza, legato al saggio marginale di sostituzione. Dunque la (9.2.1) indica che il rendimento atteso di un titolo si discosta dal rendimento privo di rischio per un fattore che potremmo chiamare *correzione per il rischio*. Quest'ultimo termine, tuttavia, non riflette tanto la rischiosità in sé del titolo quanto la correlazione col consumo. Vi è una componente aggregata di rischio, data dal termine σ_Z^2 , che conta nella determinazione del rendimento atteso dei titoli mentre la componente *idiosincratICA*, ossia ortogonale a Z , non ha importanza.

Dalla (9.2.1) ricaviamo inoltre la disuguaglianza

$$(9.2.2) \quad |\mu_n - r^0| \leq \sigma_n \sigma_Z$$

nella quale si ha uguaglianza se e solo se vi è correlazione lineare perfetta tra r^n e Z . Questa disuguaglianza stabilisce una regione all'interno della quale verranno necessariamente a trovarsi le coppie di valori μ, σ per tutti i portafogli esistenti sul mercato. Questa regione può essere rappresentata facilmente in termini grafici come segue

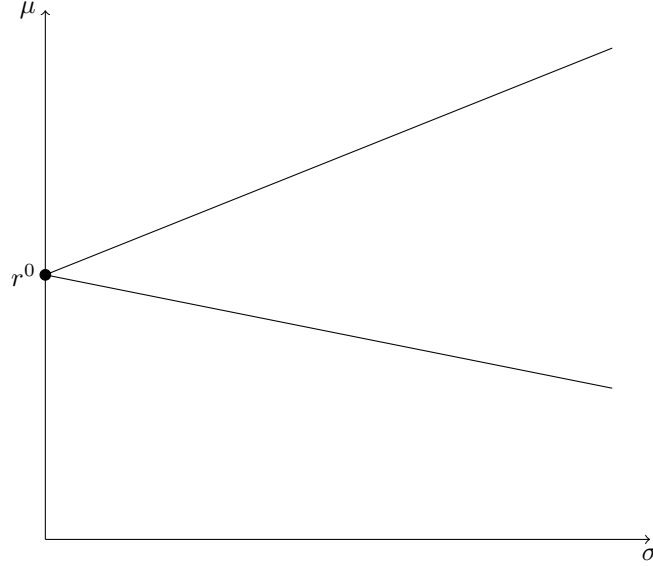


Figure 8: La frontiera efficiente in termini di σ .

Naturalmente, tutti i titoli che si trovano sulla frontiera sono perfettamente correlati, vuoi positivamente vuoi negativamente. Dunque, se α indica un portafoglio corrispondente ad un titolo lungo la frontiera, qualunque altro portafoglio β che si trovi lungo la frontiera potrà scriversi ha rendimento tale che

$$r_\beta = r^0 + c(r_\alpha - r^0)$$

la quale espressione corrisponde al teorema dei due fondi visto in precedenza.

Inoltre, poiché ogni portafoglio α sulla parte superiore della frontiera è perfettamente positivamente correlato con $-Z$ otteniamo dalla (9.2.1),

$$\mu_\alpha = r^0 + \sigma_Z \sigma_\alpha$$

e dunque, sostituendo

$$-Z = A + \frac{\text{Cov}(-Z, r_\alpha)}{\sigma_\alpha^2} r_\alpha = A + \frac{\sigma_Z}{\sigma_\alpha} r_\alpha = A + \frac{\mu_\alpha - r^0}{\sigma_\alpha^2} r_\alpha$$

si ottiene infine

$$(9.2.3) \quad \mu_\beta = r^0 + \frac{\text{Cov}(r_\alpha, r_\beta)}{\sigma_\alpha^2} (\mu_\alpha - r^0)$$

per ogni portafoglio β ed ogni portafoglio efficiente α , ossia una versione della (5.3.4).

Esercizi.

I seguenti esercizi sono basati sul dataset `Portfolio`.

Domanda 4.

- (1) Si costruisca e rappresenti graficamente la frontiera efficiente.
- (2) Si determini quali tra i titoli individuali sia il più lontano dalla frontiera rispetto al criterio del rendimento atteso.
- (3) Si determini quali tra i titoli individuali sia il più lontano dalla frontiera rispetto al criterio della varianza.

Domanda 5.

- (1) Si scriva l'equazione che caratterizza i portafogli efficienti.
- (2) Sulla base della risposta al punto precedente, è possibile formare un portafoglio efficiente nel quale non si investe nel primo titolo?
- (3) Se la risposta al punto precedente è stata affermativa, si calcoli un tale portafoglio.

Domanda 6.

- (1) Si calcoli un portafoglio efficiente tale che ogni portafoglio ortogonale ad esso abbia un rendimento atteso pari a 1%. Possiamo concludere che esso è effettivamente efficiente?
- (2) Possiamo dire con certezza che il suo rendimento atteso non potrà mai essere inferiore a 1%?
- (3) Si calcoli il rendimento atteso del portafoglio ottenuto al punto precedente.

Domanda 7. Si usino i rendimenti simulati dei 30 titoli, ottenuti col comando

```
X<-0.01*matrix(rnorm(3000),100,30)
```

- (1) Si costruisca e rappresenti la frontiera efficiente.
- (2) Si calcoli il portafoglio a varianza minima ed il suo livello effettivo di varianza.
- (3) Qual è la composizione di un portafoglio con rendimento atteso pari a 3%?

Domanda 8. Due portafogli efficienti α_1 e α_2 possono essere tra loro ortogonali? E se sì, in quali casi?

Domanda 9. Considerate un investitore con preferenze del tipo media/varianza quali

$$U(W\mu_\alpha, W^2\sigma_\alpha^2) = W\mu_\alpha - \frac{b}{2}W^2\sigma_\alpha^2 \quad b > 0$$

ed indicate con $E(r)$ il vettore dei rendimenti attesi e con Σ la matrice varianza/covarianza.

- (1) Calcolate analiticamente l'espressione della scelta ottima di portafoglio ed indicate quale relazione debba sussistere tra avversione al rischio e ricchezza affinché la scelta del portafoglio sia invariante rispetto alla ricchezza;

- (2) Calcolate esplicitamente la scelta del portafoglio nel caso $b = 5$ con i dati contenuti nel file `Portfolio.csv`;
- (3) Sulla base dei dati forniti, quale sarebbe la scelta del portafoglio se ci si proponesse di avere un rendimento atteso del 3%?

Domanda 10.

- (1) Data la generale relazione che caratterizza i portafogli efficienti, si illustri in che modo il parametro dell'intercetta, b_α , possa determinarsi dato il parametro della pendenza, c_α ;
- (2) Si consideri un investitore con preferenze del tipo

$$U(W\mu_\alpha, W^2\sigma_\alpha^2) = W\mu_\alpha - \frac{1}{2}W^2\sigma_\alpha^2$$

Sulla base del risultato del punto precedente, calcolate il portafoglio ottimale di questo investitore, il suo rendimento atteso e la sua varianza con $W = 0.1$;

- (3) Cosa accade al valore atteso ed alla varianza del rendimento del portafoglio ottimale se la ricchezza investita raddoppia.

Domanda 11. Si consideri il dataset contenuto nel file `Portfolionew.csv`.

- (1) Determinate la composizione e la varianza di un portafoglio efficiente α con $\mu_\alpha = 0.2\%$.
- (2) Supponete che gli investitori siano tenuti ad investire una medesima quota pari ad $1/30$ inciascuno dei primi 10 titoli e possano scegliere a piacere quanto investire in tutti gli altri. Quale sarebbe la forma della frontiera efficiente relativamente a tale insieme ristretto di portafogli?

Domanda 12.

- (1) Immaginate che esista un portafoglio con varianza nulla. Cosa può dirsi della correlazione lineare tra due portafogli efficienti rischiosi?
- (2) Quale forma particolare ha la frontiera efficiente tracciata in funzione dello scarto quadratico anziché della varianza?
- (3) dato il dataset `portfolio.csv`, è possibile formare un portafoglio efficiente nel quale non si hanno posizioni scoperte nei primi 4 titoli?
- (4) Più in generale, è possibile formare un portafoglio efficiente privo di posizioni scoperte?

Domanda 13. Si scriva la funzione `portfolio.optim` che, per un dato campione ed una data funzione di utilità del tipo media/varianza resituisca la scelta ottima dell'investitore.

Parte 3

Misure di rischio

Introduzione

In una qualche misura la teoria del portafoglio considerata più sopra contiene numerose indicazioni per l'analisi del rischio. In effetti la frontiera efficiente non è che una descrizione di tutti quei portafogli che consentono di minimizzare il rischio per un dato livello di rendimento atteso. Naturalmente in questa conclusione è implicito che il rischio debba essere misurato con la varianza. Come abbiamo già detto questa interpretazione della varianza ha un adeguato supporto nella teoria statistica ma non risulta necessariamente adeguata da un punto di vista finanziario. In effetti, la moderna teoria del *risk management* prende le mosse dalla critica all'uso della varianza come unica misura corretta del rischio.

Per illustrare il punto, si immagini un'opportunità di investimento che offra un rendimento certo pari a 2% ed un'altra il cui rendimento sia distribuito in modo uniforme tra 4% e 6%. In quest'ultimo caso il rendimento medio è pari a 5% e gli scostamenti da tale valore sono interpretati come un rischio la cui misura complessiva è per l'appunto la varianza. Se giudicassimo la rischiosità utilizzando tale criterio dovremmo senz'altro concludere che il secondo investimento è più rischioso del primo. Tuttavia, sotto il profilo finanziario il secondo dei due investimenti è certamente meno rischioso del primo poiché promette un rendimento che, senza alcuna incertezza, è superiore al rendimento del primo di almeno un punto percentuale.

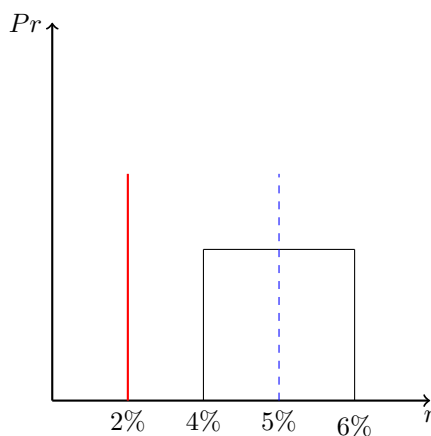


Figure 9: Rischiosità degli investimenti.

In effetti la varianza interpreta il rischio come la possibilità che si verifichi uno scostamento dalla media tanto positivo che negativo. Tuttavia, dal punto di vista strettamente finanziario, le due circostanze hanno un'interpretazione del tutto differente, se non opposta. Infatti scostamenti al di sopra della media significano profitti inattesi; al disotto, perdite inattese. Ed è certamente solo a queste ultime che l'investitore guarda laddove si preoccupa del rischio. Rendimenti eccezionalmente elevati non rappresentano, comprensibilmente, una preoccupazione.

Partendo da queste osservazioni, la moderna teoria del rischio finanziario considera soprattutto quella parte della distribuzione di probabilità che si trova al di sotto di una certa soglia ossia si interessa alla coda sinistra della distribuzione piuttosto che alla distribuzione nella sua interezza. Le misure di rischio che sono state proposte sono dunque necessariamente misure *asimmetriche*, contrariamente alla varianza. Naturalmente, se la distribuzione sottostante è simmetrica, tale aspetto perde parzialmente di importanza.

Vediamo quindi se è possibile formulare alcune proprietà astratte che una misura del rischio dovrebbe auspicabilmente possedere e verifichiamo se esse siano soddisfatte ad esempio dalla varianza.

Le Misure Coerenti del Rischio

Anzitutto, sembra ragionevole, sulla base di quanto detto prima, che una misura del rischio $\rho(r_\alpha W_t)$ sia associata solo ed esclusivamente alle possibili perdite generate dall'investimento del capitale W_t alla data t nel portafoglio α . Dunque se $r_\alpha \geq 0$ allora dovrà aversi $\rho(r_\alpha W_t) = 0$. Questa proprietà non è certamente verificata se il rischio è misurato tramite la varianza. Inoltre, se un investimento α rende più di un altro, β , essendo le perdite generate da quest'ultimo inferiori si debba avere $\rho(r_\alpha W_t) \leq \rho(r_\beta W_t)$ ossia il rischio diminuisce all'aumentare delle prospettive di rendimento.

In secondo luogo, essendo il capitale investito differente nei diversi progetti è desiderabile che la misura del rischio considerata possa facilmente esprimersi in termini unitari. Questo è certamente il caso se richiediamo che $\rho(r_\alpha W_t) = \rho(r_\alpha)W_t$ per ogni $W_t \geq 0$. In altre parole richiediamo che il rischio sia direttamente proporzionale al capitale investito, a parità di portafoglio. Evidentemente anche questa proprietà non è soddisfatta dalla varianza, mentre certamente lo è dallo scarto quadratico medio.

Nonostante i molti difetti dell'indicatore di rischio fornito dalla varianza, esso cattura un aspetto intuitivo ed importante ossia quello della diversificazione. La disuguaglianza $\text{Var}(aX + (1-a)Y) \leq a \text{Var}(X) + (1-a) \text{Var}(Y)$ che vale per ogni $0 \leq a \leq 1$ indica infatti che la suddivisione del rischio tra X e Y riduce la rischiosità complessiva. La proprietà corrispondente si può formulare nel modo seguente: $\rho(r_\alpha W_a + r_\beta W_b) \leq \rho(r_\alpha W_a) + \rho(r_\beta W_b)$. L'idea generale è che il rischio ottenuto sommando due diversi investimenti non può eccedere la somma dei rischi individuali in quanto i rischi potrebbero compensarsi l'uno con l'altro.

Infine, una misura adeguata del rischio dovrebbe inoltre fornire indicazioni operative. Ad esempio, sarebbe opportuno nella prospettiva di un'autorità garante del mercato che la misura del rischio associato ad un investimento potesse tradursi direttamente in una prescrizione di *policy* in termini ad esempio delle riserve che è necessario accantonare a fronte di un determinato investimento per poter compensare l'esposizione sottostante. In altre parole dovremmo intendere $\rho(r_\alpha W_t)$ come l'ammontare di riserve richiesto per azzerare il rischio. La proprietà corrispondente consiste nel richiedere che $\rho(r_\alpha W_t + a) = \rho(r_\alpha W_t) - a$ per ogni $a \in \mathbb{R}_+$. Per comprendere meglio questa proprietà si immagini che ad una banca che, per conto di un proprio cliente, investe il capitale W_t nel portafoglio r_α venga richiesto di accantonare riserve liquide (che immagineremo per semplicità infruttifere) pari a $\rho(r_\alpha W_t)$. In tal caso il rischio complessivo dell'investimento sarà $\rho(r_\alpha W_t + \rho(r_\alpha W_t)) = \rho(r_\alpha W_t) - \rho(r_\alpha W_t) = 0$.

Da questa breve analisi concludiamo che quattro proprietà desiderabili di una misura di rischio sono le seguenti:

- (1) $\rho(X) \leq \rho(Y)$ se $X \geq Y$;
- (2) $\rho(\lambda X) = \lambda \rho(X)$ se $\lambda > 0$;
- (3) $\rho(X + a) = \rho(X) - a$ se $a \in \mathbb{R}$;
- (4) $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Una misura di rischio con tali proprietà si chiama in letteratura una *misura coerente di rischio*. Evidentemente, la varianza non è coerente, così come lo scarto quadratico medio.

Tra le misure di rischio che analizzeremo ci interessano soprattutto:

- *Value-at-Risk*, ossia *VaR*;
- *Expected Shortfall*, ossia *ES*;
- *Extreme Value Theory*, ossia *EVT*.

Prima di passare ai dettagli di questo argomento, definiamo in modo più formale la distribuzione di probabilità troncata ad α . Il modo più semplice di descrivere tale distribuzione è per il tramite della funzione di ripartizione. Se F_Z è la distribuzione cumulativa della variabile aleatoria A and $\alpha \in \mathbb{R}$ – in genere $\alpha < 0$ – allora definiamo

$$F_Z^\alpha(t) = P(z \leq t | z \leq \alpha)$$

Utilizzando le formule consuete si ottiene

$$F_Z^\alpha(t) = \frac{P(Z \leq t \text{ and } z \leq \alpha)}{P(z \leq \alpha)} = \frac{P(z \leq t \wedge \alpha)}{P(z \leq \alpha)} = \frac{F_Z(t \wedge \alpha)}{F_Z(\alpha)} = \frac{F_Z(t)}{F_Z(\alpha)} \wedge 1$$

Di conseguenza, se indichiamo con f_z^α la corrispondente funzione di densità, abbiamo

$$f_Z^\alpha = \begin{cases} \frac{f_Z(t)}{F_Z(\alpha)} & \text{se } t \leq \alpha \\ 0 & \text{altrimenti} \end{cases}$$

Analogamente, possiamo definire

$$F_{\alpha,Z}(t) = P(z \leq t | z > \alpha)$$

ottenendo

$$(10.0.4) \quad F_{\alpha,z}(t) = \frac{P(\alpha < z \leq t)}{P(z > \alpha)} = \frac{P(z \leq t) - P(z \leq \alpha)}{P(z > \alpha)} \vee 0 = \frac{F_Z(t) - F_z(\alpha)}{1 - F_Z(\alpha)} \vee 0$$

e

$$(10.0.5) \quad f_{\alpha,z}(t) = \begin{cases} \frac{f_Z(t)}{1 - F_z(\alpha)} & \text{se } t > \alpha \\ 0 & \text{altrimenti} \end{cases}$$

Queste distribuzioni *troncate* saranno lo strumento di analisi principale nei paragrafi seguenti.

Value-at-Risk

11.1. Definizione

Nonostante i suoi grossi limiti, il VaR è divenuta una misura di rischio assai diffusa sui mercati finanziari e, conseguentemente è divenuta un punto di riferimento per tutti gli operatori, probabilmente in ragione della sua estrema e forse ingannevole semplicità dovuta al fatto che utilizza solo i quantili della distribuzione.

Alla data t un investitore potrebbe preoccuparsi del fatto che le perdite legate ad un certo investimento π potrebbero nel periodo successivo $t + k$ rivelarsi superiori ad una certa soglia x . Indicando con $L_{t,t+k}$ e $G_{t,t+k}$ perdite e guadagni (dall'acronimo inglese *loss* e *gain*) nel periodo che va da t a $t + k$, possiamo esprimere tale preoccupazione scrivendo

$$L_{t,t+k} > x$$

– con $x > 0$ – o, in altri termini,

$$(11.1.1) \quad -x > G_{t,t+k} = W_{t+k}^\pi - W_t^\pi = \frac{W_{t+k}^\pi - W_t^\pi}{W_t^\pi} W_t^\pi = r_{t,t+k}^\pi W_t^\pi$$

Se p è il corrispondente livello di probabilità,

$$P(r_{t,t+k}^\pi W_t^\pi < -x) = p$$

Naturalmente una delle due grandezze, p o x , deve essere endogena. Possiamo cioè fissare alternativamente x ed ottenere il livello p della probabilità oppure possiamo fissare p e determinare il valore della soglia x . Il VaR è la risposta a questa seconda domanda e cioè si calcola come soluzione del problema:

$$P(r_{t,t+k}^\pi W_t^\pi < -VaR_{p,k}^\pi) = p$$

Il valore dell'incognita $VaR_{p,k}^\pi$ dipenderà in linea di principio dal valore dell'investimento iniziale W_t^π . Ponendo

$$VaR_{p,k}^\pi = \frac{\$VaR_{p,k}^\pi}{W_t}$$

(e assumendo, ragionevolmente, $W_t^\pi > 0$) si conclude

$$(11.1.2) \quad P(r_{t,t+k}^\pi < -VaR_{p,k}^\pi) = p$$

Gli indici p e k si riferiscono al valore preassegnato della probabilità ed alla durata del periodo dell'investimento considerata. Nel seguito ometteremo l'indice π , per semplicità.

Alcune osservazioni relative alla formula (11.1.2) sembrano opportune:

- (1) il tasso di rendimento utilizzato, $r_{t,t+k}$, è quello aritmetico e non quello logaritmico, che peraltro potrebbe utilizzarsi altrettanto bene. Utilizzando quest'ultimo la (11.1.1) diverrebbe

$$\ln \left(1 - \frac{x}{W_t} \right) > r_{t,t+k}^{c,\pi}$$

Questa espressione ha senso compiuto solo per $x < W_t$. Indicando con $\widetilde{VaR}_{p,k}$ e $VaR_{p,k}$ il VaR relativamente al rendimento aritmetico e logaritmico rispettivamente si conclude

$$VaR_{p,k} = -\ln \left(1 - \widetilde{VaR}_{p,k} \right) = -\ln \left(1 - \frac{\$VaR_{p,k}}{W_t} \right)$$

La maggior parte degli autori ed anche degli operatori di mercato trascurano questa sottile distinzione ed approssimano semplicemente $VaR_{p,k}$ con $W_t VaR_{p,k}$ anziché con $W_t [1 - \exp(-VaR_{p,k})]$.

- (2) ad essere pignoli bisognerebbe preoccuparsi che la (11.1.2) ammetta una soluzione per ogni livello di p . La risposta è certamente positiva se la densità cumulata è continua (nel qual caso la disuguaglianza stretta può essere rimpiazzata dalla disuguaglianza debole). Tuttavia, quando la distribuzione è discontinua, la definizione fornita più sopra deve essere sostituita dalla seguente

$$VaR_{p,k} = \arg \inf_x P(r_{t,t+k} < -x) \geq p$$

Una volta fatta questa osservazione, nel resto della trattazione assumeremo che la distribuzione sia senza eccezione continua.

Restringendo l'attenzione al caso di distribuzioni continue, è chiaro che la (11.1.2) equivale alla

$$p = F_{z_{t,t+k}} \left(-\frac{VaR_{p,k} + E(r_{t,t+k})}{\sigma_{r_{t,t+k}}} \right)$$

nella quale $\sigma_{t,t+k}$ è la volatilità di $r_{t,t+k}$ e $z_{t,t+k}$ è il rendimento standardizzato

$$z_{t,t+k} = \frac{r_{t,t+k} - E(r_{t,t+k})}{\sigma_{t,t+k}}$$

Indichiamo con $F_{z_{t,t+k}}^{-1}(p)$ il p -mo quantile. Allora

$$(11.1.3) \quad VaR_{p,k} = -[F_{z_{t,t+k}}^{-1}(p)\sigma_{t,t+k} + E(r_{t,t+k})]$$

Questo risultato merita i seguenti commenti

- $VaR_{p,k} > 0$ se e solo se $\text{if } E(r_{t,t+k}) < -F_{z_{t,t+k}}^{-1}(p)\sigma_{t,t+k}$ ossia se e solo se

$$p < F_{z_{t,t+k}} \left(-\frac{E(r_{t,t+k})}{\sigma_{t,t+k}} \right)$$

In altre parole, $VaR_{p,k}$ è un indicatore delle perdite potenziali solo per valori sufficientemente piccoli di p . Nella maggior parte delle applicazioni è adeguato porre $E(r_{t,t+k}) = 0$. In tal caso l'indicatore richiede semplicemente di scegliere $p < F_{z_{t,t+k}}(0)$;

- laddove il rendimento atteso sia nullo, il membro di destra della (11.1.3) si riduce a $-F_{z_{t,t+k}}^{-1}(p)\sigma_{t,t+k}$. In tal caso non si può dire che il VaR sia una misura di rischio significativamente diversa dalla varianza dato che tra le due grandezze sussiste una relazione diretta.
- VaR è certamente una funzione decrescente di p e di $E(r_{t,t+k})$ mentre è crescente rispetto alla volatilità laddove $F_{z_{t,t+k}}^{-1}(p) < 0$.
- Per calcolare il VaR l'unica possibile difficoltà è quella di identificare la distribuzione dei rendimenti standardizzati, $F_{z_{t,t+k}}$.

11.2. Aspetti statistici

Qual è la scelta corretta per la distribuzione di $z_{t,t+k}$? La risposta a questa domanda ha un'importanza cruciale per calcolare il VaR così come altre misure di rischio. Due alternative possibili sono: costruire e stimare un modello statistico dei rendimenti finanziari ovvero di utilizzare la distribuzione empirica. La metodologia del $GARCH$ illustrata più sopra è un candidato naturale ed un esempio appropriato del primo approccio.

11.2.1. Distribuzione normale. Una della ipotesi di fondo che abbiamo introdotto allo scopo di stimare il $GARCH$ è stata che i rendimenti standardizzati, z_{t+1} , fossero condizionalmente normalmente distribuiti con valore atteso nullo e varianza unitaria. Inoltre, in quel modello $E(r_{t,t+k}) = 0$. Pertanto nella (11.1.3) la quantità $F_{z_{t,t+k}}^{-1}(p)$ dovrebbe essere sostituita con il p -mo quantile della distribuzione normale standard, ϕ_N^p .

ESEMPIO 4. *Supponiamo ad esempio che $\sigma_{t+1} = 3\%$ e che $W_t = 100,000$. Si ottengono allora i risultati seguenti:*

p	1%	3%	5%	10%
	ϕ_N^p	ϕ_N^p	ϕ_N^p	ϕ_N^p
(11.2.1) $VaR_{p,1}$	-2,3263	-1,8808	-1,6449	-1,2816
$W_t VaR_{p,1}$	0,0698	0,0564	0,0493	0,0384
$W_t [1 - \exp(-VaR_{p,1})]$	6979	5642	4935	3845
	6741	5486	4815	3772

Questo esempio fornisce una conferma delle nostre precedenti conclusioni. Il raffronto tra le ultime due righe della (11.2.1) mostra come l'approssimazione al VaR che spesso viene adottata non è sempre del tutto trascurabile.

11.2.2. La distribuzione t di student. Nella discussione circa il modello $GARCH$ abbiamo raggiunto la conclusione che la distribuzione dei rendimenti standardizzati è in genere tutt'altro che normale, soprattutto per quel che concerne le code della distribuzione. Una possibile alternativa a tale ipotesi è rappresentata dalla distribuzione t di student *standardizzata*, la quale può essere descritta per il tramite della funzione di densità seguente:

$$(11.2.2) \quad \bar{t}_d(z) = \frac{\Gamma\left(\frac{1+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \frac{1}{\sqrt{\pi(d-2)}} \left(1 + \frac{z^2}{d-2}\right)^{-\frac{1+d}{2}} = \bar{t}_d(0) \left(1 + \frac{z^2}{d-2}\right)^{-\frac{1+d}{2}}$$

definita per $d > 2$.

Si osservi che la distribuzione t di student convenzionale, la cui densità viene indicata con $t_d(z)$ e della quale si trovano facilmente i quantili già tabulati, è differente da quella standardizzata. In effetti si ha

$$\bar{t}_d(z) = t_d(z) \sqrt{\frac{d}{d-2}} \quad \text{e} \quad F_{\bar{t}_d}^{-1}(p) = \sqrt{\frac{d-2}{d}} F_{t_d}^{-1}(p)$$

dove la quantità $\frac{d}{d-2}$ corrisponde alla varianza di una variabile distribuita secondo una t di student con d gradi di libertà. Inoltre la kurtosi ammonta a

$$kurtosis(\bar{t}_z(d)) = 3 + \frac{6}{d-4}$$

una quantità maggiore di 3 quando $d > 4$. La proprietà maggiormente conveniente di questa distribuzione è la sua relativa semplicità di impiego dovuta al fatto che essa dipende in realtà da un solo parametro e i suoi quantili sono facilmente calcolabili.

ESEMPIO 5. Calcoliamo ora il VaR facendo uso della distribuzione t di student con $d = 10$ (dunque con un eccesso di kurtosi pari a 1) e con gli stessi parametri dell'esempio 4. Si ottiene

(11.2.3)	p	1%	3%	5%	10%
	$\phi_{t_d}^p$	-2,4720	-1,8964	-1,6211	-1,2273
	$VaR_{p,1}$	0,0742	0,0569	0,0486	0,0368
	$W_t VaR_{p,1}$	7416	5689	4863	3682
	$W_t [1 - \exp(-VaR_{p,1})]$	7148	5530	4747	3615

Come si vede facilmente dal raffronto, il valore del VaR è maggiore nel caso di una distribuzione t di student.

Un aspetto negativo di questa distribuzione consiste nel fatto che essa è simmetrica, ossia la skewness è 0, Presentiamo quindi un ulteriore approccio con o scopo di superare questo limite.

11.2.3. Cornish-Fisher. La formula di *Cornish Fisher* descrive una funzione dei quantili relativa ad una distribuzione ignota ma dipendente da due parametri, skewness e kurtosi. I quantili vengono calcolati secondo al formula

$$\phi_{CF}^p = \phi_{\mathcal{N}}^p + \frac{z_1}{6} [(\phi_{\mathcal{N}}^p)^2 - 1] + \frac{z_2}{24} [(\phi_{\mathcal{N}}^p)^3 - 3\phi_{\mathcal{N}}^p] - \frac{z_1}{36} [2(\phi_{\mathcal{N}}^p)^3 - 5\phi_{\mathcal{N}}^p]$$

nella quale i parametri z_1 e z_2 corrispondono a skewness e eccesso di kurtosi rispettivamente. Nelle applicazioni tali parametri potrebbero essere posti pari al valore dei corrispondenti parametri del campione.

EXAMPLE 1. Calcoliamo il VaR secondo la formula CF usando gli stessi parametri dell'Esempio 4 e assumendo che la skewness sia pari a -1% e l'eccesso di kurtosi pari a 1 – il che rende tale parametro uguale a quello utilizzato nell'esempio 5.

(11.2.4)	p	1%	3%	5%	10%
	ϕ_{CF}^p	-2,6713	-1,9760	-1,6550	-1,2137
	$VaR_{p,1}$	0,0801	0,0593	0,0496	0,0364
	$W_t^\pi VaR_{p,1}^\pi$	8014	5928	4965	3641
	$W_t [1 - \exp(-VaR_{p,1})]$	7701	5756	4844	3575

Risulta istruttivo paragonare questi risultati con quelli della tabella (11.2.3).

Expected Shortfall

12.1. Definizione

Una ragionevole alternativa al VaR è rappresentata dal *Expected Shortfall*, o più brevemente ES – talvolta anche denominato *Conditional VaR*, ossia $CVaR$. Lo scopo principale di questa misura è quello di rispondere all'ultima tra le critiche al VaR avanzate più sopra. Si definisce nel modo seguente:

$$(12.1.1) \quad ES_{x,k} = -E(r_{t,t+k} | r_{t,t+k} \leq -x)$$

Pertanto,

$$ES_{x,k} = -\frac{\int_{-\infty}^{-x} r dF_{r_{t,t+k}}(r)}{F_{r_{t,t+k}}(-x)}$$

In termini di rendimenti standardizzati, dei quali si conosce meglio la distribuzione, (12.1.1) diviene

$$\begin{aligned} ES_{x,k} &= -\left\{ E\left(z_{t,t+k} \mid z_{t,t+k} \leq -\frac{x + E(r_{t,t+k})}{\sigma_{t,t+k}}\right) + E(r_{t,t+k}) \right\} \\ &= -\left\{ \frac{\sigma_{t,t+k}}{F_{z_{t,t+k}}\left(-\frac{x + E(r_{t,t+k})}{\sigma_{t,t+k}}\right)} \int_{-\infty}^{-\frac{x + E(r_{t,t+k})}{\sigma_{t,t+k}}} z dF_{z_{t,t+k}}(z) + E(r_{t,t+k}) \right\} \end{aligned}$$

Questa misura fornisce un'indicazione più precisa rispetto al VaR di quel che potrebbe accadere qualora $r_{t+1} \leq -t$. Infatti essa prende in considerazione tutti i possibili valori di r_{t+1} in tale caso speciale. Al fine di paragonare ES e VaR è utile calcolare i valori di ES condizionatamente a un dato livello di VaR . Calcoliamo dunque la quantità

$$\begin{aligned} ES_{p,k} &\equiv ES_{VaR_{p,k},k} \\ (12.1.2) \quad &= -\left\{ \frac{\sigma_{t,t+k}}{F_{z_{t,t+k}}\left(-\frac{VaR_{p,k} + E(r_{t,t+k})}{\sigma_{t,t+k}}\right)} \int_{-\infty}^{-\frac{VaR_{p,k} + E(r_{t,t+k})}{\sigma_{t,t+k}}} z dF_{z_{t,t+k}}(z) + E(r_{t,t+k}) \right\} \\ &= -\left\{ \frac{\sigma_{t,t+k}}{p} \int_{-\infty}^{F_{z_{t,t+k}}^{-1}(p)} z dF_{z_{t,t+k}}(z) + E(r_{t,t+k}) \right\} \end{aligned}$$

Così come per il VaR , anche per l' ES è possibile valutare tale misura in relazione all'investimento, quantità che indichiamo per analogia come $\$ES_{t,1}$. Questo si può fare in due modi alternativi, a seconda se r_{t+1} indica il rendimento aritmetico o logaritmico. Nel primo caso usiamo il simbolo $\widetilde{ES}_{t,1}$ ottenendo $\$ES_{t,1} = W_t \widetilde{ES}_{t,1}$. Nel caso logaritmico, tuttavia, non vi è un modo esplicito per porre in relazione l'aspettativa del logaritmo con il logaritmo dell'aspettativa. In tal caso è dunque inevitabile accettare l'approssimazione $\$ES_{t,1} = W_t ES_{t,1}$ – nonostante sia in realtà inesatta.

La relazione che intercorre tra VaR e ES diviene chiara se si scrive $\hat{x} = \frac{t + E(r_{t+1})}{\sigma_{t+1}}$ e $p_x = F_{z_{t+1}}(-\hat{x})$. In tal caso si ricava

$$E(r_{t,t+k}) = -\{VaR_{p_x,k} - x_t \sigma_{t,t+k}\}$$

ossia

$$\begin{aligned}
ES_{x,k} &= - \left\{ \frac{\sigma_{t,t+k}}{p_x} \int_{-\infty}^{-\hat{x}} z dF_{z_{t,t+k}}(z) + E(r_{t,t+k}) \right\} \\
&= - \left\{ \frac{\sigma_{t,t+k}}{p_x} \int_{-\infty}^{-\hat{x}} z dF_{z_{t,t+k}}(z) - VaR_{p_x,k} + \hat{x} \sigma_{t,t+k} \right\} \\
&= VaR_{p_x,k} - \sigma_{t,t+k} \left(\hat{x} + \frac{\int_{-\infty}^{-\hat{x}} z dF_{z_{t,t+k}}(z)}{p_x} \right)
\end{aligned}$$

e, nel caso $t = -VaR_{p,k}$ (e dunque $p_x = p$),

$$ES_{p,k} = VaR_{p,k} - \sigma_{t,t+k} \left(F_{z_{t,t+k}}^{-1}(p) + \frac{\int_{-\infty}^{-F_{z_{t,t+k}}^{-1}(p)} z dF_{z_{t,t+k}}(z)}{p} \right)$$

Si noti che in ogni caso il termine in parentesi è certamente non positivo e dunque l' *Expected Shortfall* è certamente maggiore del *VaR*. Inoltre osserviamo che anche l'*ES* risulta essere una misura legata in modo piuttosto stretto alla volatilità.

Le espressione precedenti si semplificano ulteriormente nel caso in cui si assuma $E(r_{t,t+k}) = 0$, nel quale si ha

$$(12.1.3) \quad ES_{x,k} = - \frac{\sigma_{t+1}}{F_{z_{t,t+k}} \left(-\frac{x}{\sigma_{t+1}} \right)} \int_{-\infty}^{-\frac{x}{\sigma_{t+1}}} z dF_{z_{t,t+k}}(z)$$

e

$$(12.1.4) \quad ES_{p,k} = - \frac{\sigma_{t,t+k}}{p} \int_{-\infty}^{F_{z_{t,t+k}}^{-1}(p)} z dF_{z_{t,t+k}}(z)$$

12.2. La distribuzione normale

Consideriamo il caso in cui i rendimenti standardizzati siano normalmente distribuiti, come nel modello *GARCH* visto più sopra. In questo caso speciale la funzione di densità della normale $f_{\mathcal{N}}$ soddisfa

$$\frac{d}{dx} f_{\mathcal{N}}(x) = \frac{d}{dx} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) = -x \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) = -x f_{\mathcal{N}}(x)$$

e dunque

$$\begin{aligned}
- \int_a^b x dF_{z_{t+1}}(x) &= - \int_a^b x f_{z_{t+1}}(x) dx \\
&= \int_a^b f'_{\mathcal{N}}(x) dx \\
&= f_{\mathcal{N}}(b) - f_{\mathcal{N}}(a)
\end{aligned}$$

– il caso $a = -\infty$ or $b = \infty$ si tratta col passaggio al limite. Concludiamo pertanto

$$ES_{x,1} = \frac{f_{\mathcal{N}} \left(-\frac{x}{\sigma_{t+1}} \right)}{F_{\mathcal{N}} \left(-\frac{x}{\sigma_{t+1}} \right)} \sigma_{t+1}$$

in cui $F_{\mathcal{N}}$ è la densità cumulata della distribuzione normale standard. Nel caso in cui $x = VaR_{p,1}$ otteniamo

$$ES_{p,1} = \frac{f_{\mathcal{N}}(\phi_{\mathcal{N}}^p)}{p} \sigma_{t+1}$$

ESEMPIO 6. Torniamo all'esempio di $p = 59$, cioè $\sigma_{t+1} = 3\%$, $W_t = 100,000$ con z_{t+1} distribuito normalmente. Vogliamo calcolare l' ES in relazione ai diversi valori del VaR ottenuti in quell'esempio. In altri termini,

p	1%	3%	5%	10%
$VaR_{p,1}$	0,0698	0,0564	0,0493	0,0384
$\phi_{\mathcal{N}}^p$	-2,3263	-1,8808	-1,6449	-1,2816
$f_{\mathcal{N}}(\phi_{\mathcal{N}}^p)$	0,0267	0,0680	0,1031	0,1755
$ES_{t,1}$	0,0800	0,0680	0,0619	0,0526
$\$ES_{t,1}$	7996	6804	6188	5265

Si raffrontino le colonne, ad esempio la prima, delle tabelle (11.2.1) e (12.2.1). Con un livello di confidenza dell'1%, il VaR è pari a 0,0698 ossia lo scenario peggiore è rappresentato da un rendimento inferiore a -6,98%. Tuttavia il rendimento atteso in questo stesso scenario è pari a -8%, ossia molto inferiore. In particolare, con questo stesso livello di confidenza il VaR di portafoglio ammonta a 6,741 in termini esatti e 6,979 in termini approssimati mentre il corrispondente ammontare atteso è pari a 7,996: le perdite attese superano la soglia indicata dal VaR di più di 1,000.

12.3. La distribuzione t di Student

Possiamo esplicitamente calcolare l' ES anche nel caso di una distribuzione del tipo t di student. Ricordando la (11.2.2), si ponga $d > 2$ e si consideri la funzione

$$(12.3.1) \quad G_d(z) = \frac{d-2}{d-1} c_d \left(1 + \frac{z^2}{d-2}\right)^{-\frac{d-1}{2}}$$

Si noti che

$$G_d(z) = \frac{d-2}{d-1} \bar{t}_d(0) \left(\frac{\bar{t}_d(z)}{\bar{t}_d(0)}\right)^{\frac{d-1}{d+1}} = \frac{d-2}{d-1} \bar{t}_d(0)^{\frac{2}{d+1}} \bar{t}_d(z)^{\frac{d-1}{d+1}} = \gamma_d \bar{t}_d(z)^{\frac{d-1}{d+1}}$$

Inoltre, la sua derivata si calcola come

$$dG_d(z)/dz = -c_d \left(1 + \frac{z^2}{d-2}\right)^{-\frac{d+1}{2}} z = -\bar{t}_d(z)z$$

dove con \bar{t}_d si è indicata la funzione di densità (11.2.2). In altre parole,

$$\begin{aligned} - \int_a^b z f_{\bar{t}_d}(z) dz &= \int_a^b \frac{d}{dz} G_{\bar{t}_d}(z) dz = [G_{\bar{t}_d}(b) - G_{\bar{t}_d}(a)] \\ &= \frac{d-2}{d-1} c_d \left[\left(1 + \frac{b^2}{d-2}\right)^{-\frac{d-1}{2}} - \left(1 + \frac{a^2}{d-2}\right)^{-\frac{d-1}{2}} \right] \end{aligned}$$

Tenendo conto del fatto che $G_d(-\infty) = 0$ otteniamo quindi

$$- \int_{-\infty}^b z \bar{t}_d(z) dz = G_d(b) = \gamma_d \bar{t}_d(b)^{\frac{d-1}{d+1}}$$

una espressione che, nonostante l'apparente complessità, è interamente esplicita e può essere calcolata molto facilmente. Otteniamo quindi

$$ES_{x,k} = \frac{G_d\left(-\frac{x+E(r_{t+1})}{\sigma_{t,t+k}}\right)}{F_{\bar{t}_d}\left(-\frac{x+E(r_{t+1})}{\sigma_{t,t+k}}\right)} \sigma_{t,t+k} - E(r_{t+1})$$

dove F_{t_d} indica la densità cumulata di una t di student non standardizzata. Inoltre

$$ES_{p,k} = \frac{G_d \left(F_{t_d}^{-1}(p) \right)^{\frac{d-1}{d+1}}}{p} - E(r_{t+1})$$

ESEMPIO 7. Si torni all'esempio di p. 60. Vogliamo calcolare l'Expected Shortfall relativamente ai diversi valori del VaR ottenuti in quell'esempio, come nella (12.1.2). In altre parole,

p	1%	3%	5%	10%
$VaR_{p,1}$	0,0742	0,0569	0,0486	0,0368
$\phi_{\mathcal{N}}^p$	-2,4720	-1,8964	-1,6211	-1,2273
$G(\phi_{t_d}^p)$	0,0301	0,0728	0,1077	0,1779
$ES_{t,1}$	0,0902	0,0728	0,0646	0,0534
$\$ES_{t,1}$	9024	7275	6462	5337

Si raffrontino le colonne delle tabelle (11.2.3) e (12.3.2).

12.4. L'approccio dei quantili

Lo stesso approccio seguito sopra può, *in linea di principio*, essere applicato con qualunque ipotesi circa la distribuzione, inclusa la formula *Cornish-Fisher* per i quantili. Non è difficile calcolare il valore atteso dati i quantili, benché sia un fatto poco noto, basandosi sulla formula del cambio di variabile. Infatti, se F_X è la densità cumulata della variabile aleatoria X si osserva che la sua inversa, F_X^{-1} , ossia la funzione quantile, è sempre ben definita in quanto la F_X è crescente. Ma allora

$$\int_a^b x dF_X(x) = \int_{F_X^{-1}(a)}^{F_X^{-1}(b)} F_X^{-1}(z) dz$$

Pertanto, nel calcolo dell' ES possiamo sostituire la quantità $\int_{-\infty}^{-\frac{t}{\sigma_{t+1}}} x dF_X(x)$ con $\int_0^{F_X^{-1}\left(-\frac{t}{\sigma_{t+1}}\right)} F_X^{-1}(z) dz$ e, nel caso particolare in cui $t = VaR_{p,1}$ (si veda la (12.1.2)), otteniamo

$$ES_{VaR_{p,1},1} = \frac{\int_0^p F_X^{-1}(z) dz}{p} \sigma_{t+1}$$

Questa formula può essere impiegata nel caso della formula *Cornish-Fisher* nel qual caso l'integrale che vi figura potrebbe essere calcolato numericamente. Sfortunatamente la procedura numerica tende ad essere piuttosto instabile il che rende la questione troppo delicata ai fini dell'esposizione che intendiamo farne.

ESEMPIO 8. Si torni ancora una volta all'Esempio 8. Vogliamo ora calcolare l'Expected Shortfall in corrispondenza dei diversi valori del VaR ottenuti in quell'esempio, utilizzando la (12.1.2). In altri termini,

p	1%	3%	5%	10%
$VaR_{p,1}$	0,0801	0,0593	0,0496	0,0364
ϕ_{CF}^p	-2,6713	-1,9760	-1,6550	-1,2137
$ES_{t,1}$				
$\$ES_{t,1}$				

Un'altra formula utile nei calcoli è la seguente:

$$\begin{aligned}
 \int_a^b x dF_X(x) &= \int_{a \wedge 0}^{b \wedge 0} x dF_X(x) + \int_{a \vee 0}^{b \vee 0} x dF_X(x) \\
 &= F(b \wedge 0)(b \wedge 0) - F(a \wedge 0)(a \wedge 0) + \int_{a \wedge 0}^{b \wedge 0} F_X(x) dx \\
 &\quad [1 - F(a \vee 0)](a \vee 0) - [1 - F(b \vee 0)](b \vee 0) + \int_{a \vee 0}^{b \vee 0} [1 - F_X(x)] dx
 \end{aligned}$$

dove si ha in genere $a = -\infty$ e $b < 0$ e quindi

$$\int_a^b x dF_X(x) = F(b)b + \int_a^b F_X(x) dx$$

Extreme Value Theory

13.1. Definizione

Un ulteriore approccio al *risk management* si basa sulla cosiddetta *Extreme Value Theory, EVT*. Questa ha il vantaggio, rispetto alle tradizionali ipotesi circa la distribuzione, di essere essenzialmente non parametrica e di basarsi piuttosto su di un risultato di convergenza simile a quello del *Teorema Limite Centrale* ma relativo alla coda della distribuzione.

Sia X una variabile aleatoria e consideriamo la funzione di densità cumulata condizionata $F_{u,X}$ definita nella (10.0.4)

$$F_{u,X}(u+z) = \frac{F_X(u+z) - F_X(u)}{1 - F_X(u)}$$

dove $z, u > 0$. Nel nostro caso X rappresenta alternativamente le perdite, i rendimenti negativi o i rendimenti negativi standardizzati e, tradotta in questi termini, la funzione $F_{u,X}$ fornisce indicazione precise circa la coda sinistra del rendimento di un investimento. Il risultato di convergenza al quale si faceva riferimento è il seguente: quando $u \rightarrow \infty$ la distribuzione converge alla cosiddetta distribuzione di Pareto generalizzata

$$G(z; \xi, \beta) = \begin{cases} 1 - \left(1 + z \frac{\xi}{\beta}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp\left(-\frac{z}{\beta}\right) & \xi = 0 \end{cases}$$

con $\beta > 0$ e

$$\begin{cases} z \geq 0 & \xi \geq 0 \\ 0 \leq z \leq -\frac{\beta}{\xi} & \xi < 0 \end{cases}$$

L'unico caso che a noi interessa è quello in cui $\xi > 0$ che corrisponde al caso delle cosiddette *fat tails*.

Si osservi che se $F_{u,X}(z) \simeq G(z; \xi, \beta)$ allora

$$\begin{aligned} F_X(u+z) &\simeq G(z; \xi, \beta)(1 - F_X(u) + F_X(u)) \\ &= 1 + (1 - F_X(u))[G(z; \xi, \beta) - 1] \\ &= 1 - (1 - F_X(u)) \left(1 + z \frac{\xi}{\beta}\right)^{-\frac{1}{\xi}} \end{aligned}$$

Spesso conviene scrivere $y = z + u$ - dunque $y \geq u$ - ottenendo

$$F_X(y) \simeq 1 - (1 - F_X(u)) \left(1 + (y - u) \frac{\xi}{\beta}\right)^{-\frac{1}{\xi}}$$

In che modo viene utilizzato questo risultato nel *risk management*? Procediamo in tre passi:

- (1) Si stimano i parametri ξ e β ;
- (2) Si usano queste stime per inferire il valore di F_X^u ;
- (3) Si applica F_X^u al calcolo di *VaR* e *ES*.

13.2. Stima

La procedura di stima che consideriamo è estremamente semplice ed è basata sull'approssimazione seguente:

$$\left(1 + (y - u) \frac{\xi}{\beta}\right)^{-\frac{1}{\xi}} \simeq c'_u y^{-\frac{1}{\xi}}$$

ossia

$$F_X(y) = 1 - (1 - F_X(u)) \left(1 + (y - u) \frac{\xi}{\beta}\right)^{-\frac{1}{\xi}} \simeq 1 - c_u y^{-\frac{1}{\xi}}$$

Implicitamente si ricava

$$c_u = (1 - F_X(u)) u^{\frac{1}{\xi}} \quad F_X(y) = 1 - (1 - F_X(u)) \left(\frac{y}{u}\right)^{-\frac{1}{\xi}}$$

e

$$F_{u,X}(y) = \frac{F_X(y) - F_X(u)}{1 - F_X(u)} = 1 - \left(\frac{y}{u}\right)^{-\frac{1}{\xi}} \quad y \geq u$$

Da queste formule si deduce l'espressione della funzione di densità

$$f_X(y) = (1 - F_X(u)) \frac{y^{\frac{1}{\xi}-1}}{\xi u^{-\frac{1}{\xi}}}$$

e della funzione di densità condizionata definita dalla (10.0.5)

$$f_{u,X}(y) = \frac{f_X(y)}{1 - F_X(u)} = \frac{1}{\xi u} \left(\frac{y}{u}\right)^{-\frac{1}{\xi}-1}$$

Sotto l'ipotesi di distribuzione i.i.d., il logaritmo della funzione di verosimiglianza condizionata è data dalla

$$\begin{aligned} \ln(L(y_1, \dots, y_T)) &= \ln \left(\prod_{\{1 \leq t \leq T: y_t \geq u\}} \frac{y_t^{-\frac{1}{\xi}-1}}{\xi u^{-\frac{1}{\xi}}} \right) \\ &= \sum_{\{1 \leq t \leq T: y_t \geq u\}} \ln \left(\frac{y_t^{-\frac{1}{\xi}-1}}{\xi u^{-\frac{1}{\xi}}} \right) \\ &= - \left\{ T_u \ln(\xi) + \left(\frac{1}{\xi} + 1\right) \sum_{\{1 \leq t \leq T: y_t \geq u\}} \ln(y_t) - \frac{T_u}{\xi} \ln(u) \right\} \end{aligned}$$

dove con T_u indichiamo non la numerosità del campione che sarà T , bensì il numero di quelle osservazioni y_t che si trovano al di sopra della soglia u . La derivata di questa espressione rispetto a ξ è data da

$$-\frac{T_u}{\xi} + \frac{1}{\xi^2} \sum_{\{1 \leq t \leq T: y_t \geq u\}} \ln(y_t) - \frac{T_u}{\xi^2} \ln(u) = -\frac{T_u}{\xi} + \frac{1}{\xi^2} \sum_{\{1 \leq t \leq T: y_t \geq u\}} \ln\left(\frac{y_t}{u}\right)$$

da cui si ricava la stima di massima verosimiglianza seguente:

$$\hat{\xi} = \frac{1}{T_u} \sum_{\{1 \leq t \leq T: y_t \geq u\}} \ln\left(\frac{y_t}{u}\right)$$

Si osservi che la quantità $1 - F_X(u)$ trova un corrispettivo campionario nella quantità $T_u T^{-1}$, il rapporto tra il numero delle osservazioni oltre la soglia u e le osservazioni totali. Pertanto sarebbe ragionevole imporre la restrizione $T_u/T = 1 - F_X(u)$. Infine, sebbene tale restrizione non sia una implicazione matematica, è utile concentrarsi sul caso in cui $\hat{\xi} < 1$.

13.3. Calcolo dei quantili

Siamo chiaramente interessati alla distribuzione della coda $F_{u,X}$ ed alla sua densità $f_{u,X}$. Otteniamo

$$F_X(y) = 1 - \frac{T_u}{T} \left(\frac{y}{u} \right)^{-\frac{1}{\xi}} \quad \text{per } y \geq u$$

e dunque

$$F_X^{-1}(p) = \left[\frac{T_u}{(1-p)T} \right]^{\hat{\xi}} u$$

Dato che la variabile aleatoria X è legata ai rendimenti negativi, è giustificato il nostro interesse per la distribuzione di probabilità della grandezza $-X$. A questo fine è bene tenere in mente che

$$F_X(y) = P(X \leq y) = P(-X \geq -y) = 1 - P(-X < -y) = 1 - F_{-X}(-y)$$

e

$$F_{-X}^{-1}(p) = -F_X^{-1}(1-p)$$

ossia nel nostro caso

$$(13.3.1) \quad F_{-X}(y) = \frac{T_u}{T} \left(\frac{y}{u'} \right)^{-\frac{1}{\xi}} \quad y \leq u' = -u$$

e

$$(13.3.2) \quad F_{-X}^{-1}(p) = \left[\frac{T_u/T}{p} \right]^{\hat{\xi}} u' \quad p < \frac{T_u}{T}$$

Si noti che la (13.3.2) non ha senso compiuto al di fuori della restrizione specificata in quanto il valore ottenuto sarebbe altrimenti maggiore di $-u$, cioè cadrebbe in un intervallo in cui la funzione F_{-X} non è definita.

13.4. EVT misure di rischio

Passando alle applicazioni in tema di misure di rischio, nel caso in esame si ha $X = -r_{t,t+k}$ di modo che tanto lo stimatore $\hat{\xi}$ introdotto più sopra che la soglia u divengono implicitamente funzione di k . Avremo pertanto

$$VaR_{p,k} = \left[\frac{T_u/T}{p} \right]^{\hat{\xi}} u$$

e, dalla (13.3.1) e dalla (13.3.2)

$$\begin{aligned} ES_{x,k} &= - \frac{\int_{-\infty}^{-x} r dF_{r_{t,t+k}}(r)}{F_{r_{t,t+k}}(-x)} \\ &= - \frac{\int_{-\infty}^{-x} (-r) dF_X(-r)}{1 - F_X(x)} \\ &= \frac{\int_x^{\infty} x dF_X(x)}{1 - F_X(x)} \\ &= \frac{T_u}{\hat{\xi} T} u^{\frac{1}{\hat{\xi}}} \frac{\int_x^{\infty} x^{-\frac{1}{\hat{\xi}}} dx}{\frac{T_u}{T} \left(\frac{x}{u} \right)^{-\frac{1}{\hat{\xi}}}} \\ &= \frac{u^{\frac{1}{\hat{\xi}}}}{1 - \hat{\xi}} \frac{x^{1-\frac{1}{\hat{\xi}}}}{\left(\frac{x}{u} \right)^{-\frac{1}{\hat{\xi}}}} \\ &= \frac{x}{1 - \hat{\xi}} \end{aligned}$$

e

$$ES_{p,k} = \frac{VaR_{p,k}}{1 - \hat{\xi}}$$

dove si è fatto uso della disuguaglianza $\hat{\xi} < 1$, la quale in genere è soddisfatta.

Accordi di Basilea, Misure di Rischio e Coerenza

Le misure di rischio viste più sopra, tra le più diffuse, sono state esaminate sia sotto il profilo della loro utilità pratica che della loro coerenza logica. Il primo punto di vista è quello che viene attribuito al mercato ed ai diversi agenti che vi operano, tra cui le autorità di regolamentazione. Il secondo aspetto riguarda piuttosto l'ambito accademico

14.1. La Regolamentazione del Rischio Secondo Basilea

La gestione del rischio non è solamente una preoccupazione dei singoli investitori, privati risparmiatori o istituzioni finanziarie. Un aspetto importante, specie alla luce degli eventi dell'ultimo anno, riguarda la stabilità del sistema finanziario nel suo complesso e della supervisione che su di esso esercitano le autorità preposte. Già negli anni '80 una lunga scia di fallimenti bancari ha costellato la storia finanziaria di molti paesi. Già in un rapporto stilato per il *FMI* nel 1997 si paragonavano le potenziali conseguenze macroeconomiche di tali crisi agli effetti della *Grande Depressione*.

In realtà fin dal 1974 la *Banca dei regolamenti Internazionali* aveva creato il cosiddetto *Comitato di Basilea* per la supervisione bancaria. Negli Stati Uniti, sin dagli anni '30 già esisteva la *Federal Deposit Insurance Corporation (FDIC)* con lo scopo di offrire protezione ai depositanti. La Comunità Europea ha dal canto suo adottato un'apposita direttiva sui requisiti di capitale. Queste diverse fonti di regolamentazione hanno registrato negli anni alcune convergenze in particolare sulla importanza dei cosiddetti requisiti minimi di capitale. Nel 1988 tutti e 12 i paesi membri del Comitato di Basilea hanno firmato il primo accordo in tema di supervisione bancaria, meglio noto come *Basilea I*. In seguito questo è stato riformato dal cosiddetto emendamento del 1996. Infine nel 2005 il comitato di Basilea ha interamente rivisto l'argomento promulgando un secondo testo, noto come *Basilea II*, la cui applicazione ha avuto inizio solo a partire dall'anno 2007.

L'attenzione posta sui cosiddetti requisiti minimi di capitale si può comprendere assai semplicemente guardando ad una stilizzazione dell'attivo di una generica banca. Il risultato economico di una banca sarà:

$$Dk(1 + \rho) + D(1 - k) \sum_{n=1}^N \alpha_n(1 + r_n) - D(1 + r_D)$$

dove D sono i depositi raccolti presso il pubblico, k il coefficiente di riserva obbligatoria ed α_n la quota dei depositi (al netto della riserva obbligatoria) investita nell'attività finanziaria n -ma. Indicando con una barra il rendimento al netto del tasso di interesse pagato ai depositanti, si ottiene

$$Dk\bar{\rho} + D(1 - k) \sum_{n=1}^N \alpha_n \bar{r}_n$$

Immaginiamo che la banca disponga anche di un certo capitale proprio remunerato ad un tasso ρ_0 che immaginiamo privo di rischio. Il tal caso l'espressione precedente si trasforma nella

$$(14.1.1) \quad K\bar{\rho}_0 + Dk\bar{\rho} + D(1 - k) \sum_{n=1}^N \alpha_n \bar{r}_n$$

Intuitivamente, l'evento insolvenza bancaria si verifica allorché questa espressione diviene negativa, il che significa che il reddito percepito dalla banca tramite i suoi investimenti non è sufficiente a coprire la remunerazione dei depositi. Si ha dunque insolvenza solo laddove $K\bar{\rho}_0 + Dk\bar{\rho} + D(1-k)\sum_{n=1}^N \alpha_n \bar{r}_n \leq 0$ ossia

$$(14.1.2) \quad D(1-k) \sum_{n=1}^N \alpha_n \bar{r}_n \leq - \left\{ \frac{K}{D(1-k)} \bar{\rho}_0 + \frac{k}{1-k} \bar{\rho} \right\}$$

Evidentemente, quanto più alto è il rapporto K/D tanto più remota, a parità di altre condizioni, si fa la possibilità che la banca vada incontro a fallimento (purché $\bar{\rho}_0 > 0$). Poiché al quota $D(1-k)$ di depositi non tenuta a riserva viene investita, essa è spesso indicata come il totale degli *assets* e si rappresenta con la lettera A .

L'accordo noto come *Basilea I* introduceva due criteri in proposito: anzitutto un rapporto tra capitale proprio e assets non inferiore al 5%; secondo, un *Cooke ratio* non inferiore al 8%. Il *Cooke ratio* è una quantità simile al rapporto K/A in cui tuttavia si utilizza un diverso sistema per calcolare A . Infatti, i titoli in cui la banca investe devono essere pesati per un opportuno coefficiente di rischio. Ad esempio, gli accantonamenti liquidi hanno coefficiente nullo mentre le obbligazioni emesse da singole imprese hanno un coefficiente massimo pari a 100%. Inoltre, alcune voci importanti che non rientrano nei bilanci bancari poiché si riferiscono ad accordi assai specifici, devono essere esplicitamente calcolate utilizzando il criterio del valore di mercato più un coefficiente applicato al valore nominale dell'asset. Questo modo di valutare le componenti dell'attivo bancario è di grande importanza soprattutto per quel che riguarda i titoli derivati che, pur potendo dare luogo a perdite anche ingenti hanno spesso un valore di mercato piuttosto ridotto (si pensi al *future*). In sostanza questo criterio di Cooke implica che i pesi di portafoglio α_n introdotti più sopra debbano essere aggiustati per il rischio, α_n^* . La quantità $A^* = D(1-k) \sum_{n=1}^N \alpha_n^*$ viene denominata *risk-weighted assets*. Si prescrive $K/A^* \geq 8\%$.

L'emendamento a Basilea I era volto a superare alcune difficoltà insite in quell'accordo. Anzitutto il rischio considerato era esclusivamente il rischio di credito e non ad esempio il rischio di mercato. Inoltre la classificazione dei rischi era piuttosto netta e non teneva conto di criteri di mercato quali il *rating*. Per valutare il rischio di mercato venne proposto un coefficiente basato sul $VaR_{1\%,10}$. In particolare, il rischio di mercato veniva valutato secondo la formula

$$aVaR_{1\%,10} + b$$

dove a e b sono coefficienti positivi e $a \geq 3$.

L'accordo *Basilea II* ha ampliato molto l'approccio al rischio bancario introducendo la possibilità per le banche di scegliere tra diverse forme di calcolo dei requisiti patrimoniali. L'approccio standard è essenzialmente una revisione di Basilea I volta a tenere conto di tutti i dettagli dei crediti erogati: il rating, l'esistenza di garanzie collaterali, etc... Questi elementi modificano i coefficienti di rischiosità visti sopra.

14.2. Le Misure di Rischio Coerenti

Come si è visto, sotto le ipotesi distributive più comuni come la normale o, più in generale, se la distribuzione utilizzata è simmetrica, le misure di rischio viste più sopra non portano ad un risultato strutturalmente differente da quello che si otterrebbe utilizzando il concetto di varianza o di scarto quadratico medio.

Un elemento più specifico di critica proviene dal cosiddetto approccio delle *misure di rischio coerenti*. Si tratta di un approccio che, anziché proporre una specifica misura e verificarne poi le proprietà più o meno desiderabili, prende le mosse da un elenco di proprietà che ogni misura di rischio dovrebbe avere per essere,

per l'appunto, *coerente*. Il VaR e l'*Expected Shortfall* non sarebbero coerenti nel senso chiarito da quella letteratura. La letteratura su questo argomento ha preso le mosse dal contributo di Artzner, Delbaen, Heath e Eber [1]

L'idea di fondo è che una misura di rischio sia da interpretarsi come una indicazione delle riserve che un investitore, ad esempio una banca, dovrebbe accantonare per far fronte ai rischi a cui si espone con le proprie scelte finanziarie. In questi termini, l'accantonamento richiesto dovrebbe anzitutto essere *proporzionale* al capitale investito. Secondariamente esso dovrebbe essere crescente all'aumentare delle prospettive di perdite, ossia *inversamente monotona*. Infine, se l'investitore sottoscrive più di un investimento, la misura di rischio relativa al portafoglio così composto dovrebbe essere inferiore (o meglio non superiore) alla somma delle misure relative a ciascun investimento individuale in altri termini essa dovrebbe essere *subadditiva*. In modo più formale, se \mathcal{X} raffigura l'insieme dei payoff di tutti gli investimenti possibili e $\rho : \mathcal{X} \rightarrow \mathbb{R}$ la misura di rischio, abbiamo:

- (i) $\rho(\lambda X) = \lambda \rho(X)$ per ogni $X \in \mathcal{X}$ e $\lambda \geq 0$;
- (ii) $\rho(X) \geq \rho(Y)$ per $X, Y \in \mathcal{X}$ tali che $X \leq Y$;
- (iii) $\rho(X + Y) \leq \rho(X) + \rho(Y)$ per $X, Y \in \mathcal{X}$;
- (iv) $\rho(X + \alpha) = \rho(X) - \alpha$ per $X \in \mathcal{X}$, $\alpha \in \mathbb{R}$

Talvolta a questi assiomi si aggiunge anche il seguente

- (v) $\rho(X) > 0$ per tutti i $X \in \mathcal{X}$ con $X \leq 0$, $X \neq 0$.

Una misura di rischio che certamente *non* soddisfa questi requisiti è il VaR , quanto meno non in generale. Consideriamo il caso di due rischi indipendenti e normalmente distribuiti, di modo che anche la loro somma sia distribuita in modo normale. Allora se $p < 0.5$ allora $\Phi_{\mathcal{N}}^{-1}(p) < 0$ e dunque

$$\begin{aligned} VaR_p(X + Y) &= -[E(X + Y) + \Phi_{\mathcal{N}}^{-1}(p)\sigma(X + Y)] \\ &\leq -[E(X + Y) + \Phi_{\mathcal{N}}^{-1}(p)(\sigma(X) + \sigma(Y))] \\ &= VaR_p(X) + VaR_p(Y) \end{aligned}$$

Tuttavia, se X e Y sono distribuite in modo diverso o se non sono indipendenti, allora la distribuzione della somma non sarà più normale e dunque il quantile che appare nel calcolo del VaR sarà modificato. Non è difficile costruire esempi in cui ciò implica che la misura non sia subadditiva.

Tests

Proviamo a proporre una serie di tests delle misure di rischio proposte precedentemente. Implicitamente questi tests consentono di valutare la bontà del modello statistico dei rendimenti utilizzando tuttavia informazioni relative agli eventi rischiosi. Come sempre in finanza, anche i modelli di rischio consistono in buona sostanza in predizioni circa il valore atteso di certe grandezze. Pertanto le strategie seguite per i test hanno in genere la medesima impostazione.

15.1. Alcuni Tests per il VaR .

Indichiamo con $VaR_{p,k}(t)$ la corrispondente misura di rischio calcolata alla data t e consideriamo la seguente variabile aleatoria:

$$\chi_t^{p,k} = \begin{cases} 1 & \text{se } r_{t,t+k} \leq -VaR_{p,k}(t) \\ 0 & \text{se } r_{t,t+k} > -VaR_{p,k}(t) \end{cases}$$

Evidentemente, la variabile $\chi_t^{p,k}$ ha una distribuzione binomiale, ossia

$$f(\chi_t^{p,k}) = \pi^{\chi_t^{p,k}} (1 - \pi)^{1 - \chi_t^{p,k}}$$

L'evento $\chi_t^{p,k} = 1$ ha cioè probabilità pari a π . Secondo il nostro modello $P(\chi_t^{p,k} = 1) = P(r_{t,t+k} \leq -VaR_{p,k}(t)) = p$ e pertanto l'ipotesi nulla sul parametro π è $\pi = p$. Assumendo che il campione $\{\chi_t^{p,k} : t = 1, \dots, T\}$ sia *i.i.d* possiamo calcolare la funzione di verosimiglianza

$$\log(L) = \sum_{t=1}^T \log(f(\chi_t^{p,k})) = \sum_{t=1}^T \chi_t^{p,k} \log(\pi) + \sum_{t=1}^T (1 - \chi_t^{p,k}) \log(1 - \pi) = T_1 \log(\pi) + T_0 \log(1 - \pi)$$

(dove abbiamo posto $T_1 = \sum_{t=1}^T \chi_t^{p,k}$ e $T_0 = T - T_1$) dalla quale si ricava la stima di massima verosimiglianza $\pi^* = T_1/T$. Il rapporto di verosimiglianza $L(p)/L(\pi^*)$ è pertanto inferiore all'unità ma, se le nostre predizioni sono corrette non troppo. Dunque la statistica

$$LR(p, \pi) = -2 \ln \left(\frac{L(p)}{L(\pi^*)} \right)$$

dovrebbe risultare non significativamente diversa da 0. Con un numero sufficiente di osservazioni la distribuzione del test si approssima ad una χ_1^2 e su questa base è facile decidere se l'ipotesi sia da rigettare oppure no.

Ovviamente altri test si possono fare seguendo il medesimo approccio. Ad esempio si può sottoporre a verifica l'assunzione di indipendenza supponendo, in alternativa, che la variabile $\chi_t^{p,k}$ sia di tipo markoviano ossia che vi sia una matrice di transizione

$$\Pi = \begin{bmatrix} \pi_1 & 1 - \pi_1 \\ \pi_0 & 1 - \pi_0 \end{bmatrix}$$

dove

$$\pi_0 = P(\chi_t^{p,k} = 1 \mid \chi_{t-1}^{p,k} = 0) \quad \pi_1 = P(\chi_t^{p,k} = 1 \mid \chi_{t-1}^{p,k} = 1)$$

L'ipotesi che vogliamo sottoporre a verifica è che $\pi_0 = \pi_1$. Per effettuare questo test dividiamo il nostro campione nei due sotto campioni $\{(\chi_t^{p,k} | \chi_{t-1}^{p,k} = 0) : t = 2, \dots, T\}$ e $\{(\chi_t^{p,k} | \chi_{t-1}^{p,k} = 1) : t = 2, \dots, T\}$, il primo distribuito indipendentemente secondo una Bernoulli del parametro π_0 , il secondo con parametro π_1 . Poiché tra i due campioni vi è perfetta indipendenza la funzione di verosimiglianza sarà del tipo

$$\log(L(\Pi)) = T_{1,0} \log(\pi_0) + T_{0,0} \log(1 - \pi_0) + T_{1,1} \log(\pi_1) + T_{0,1} \log(1 - \pi_1)$$

dove

$$\begin{aligned} T_{1,0} &= \sum_{t=2}^T \chi_t^{p,k} | (\chi_{t-1}^{p,k} = 0) & T_{0,0} &= \sum_{t=2}^T (1 - \chi_t^{p,k}) | (\chi_{t-1}^{p,k} = 0) \\ T_{1,1} &= \sum_{t=2}^T \chi_t^{p,k} | (\chi_{t-1}^{p,k} = 1) & T_{0,1} &= \sum_{t=2}^T (1 - \chi_t^{p,k}) | (\chi_{t-1}^{p,k} = 1) \end{aligned}$$

Il criterio della massima verosimiglianza fornisce dunque il risultato seguente

$$\pi_0^* = \frac{T_{1,0}}{T_{1,0} + T_{0,0}} \quad \pi_1^* = \frac{T_{1,1}}{T_{1,1} + T_{0,1}}$$

Il test del rapporto di verosimiglianza diviene ora:

$$LR(\pi, \Pi) = -2 \ln \left(\frac{L(\pi^*)}{L(\Pi^*)} \right) \sim \chi_1^2$$

Eventualmente, un'ulteriore verifica si potrebbe effettuare testando p contro Π e calcolare dunque la quantità:

$$LR(p, \Pi) = -2 \ln \left(\frac{L(p)}{L(\Pi^*)} \right) \sim \chi_2^2$$

Chiaramente $LR(p, \Pi) = LR(p, \pi) + LR(\pi, \Pi)$.

15.2. Test che utilizzano un insieme allargato delle informazioni.

A ben vedere il nostro primo esercizio del paragrafo precedente potrebbe essere riassunto nella stima della equazione

$$\chi_t^{p,k} = p + \epsilon_t$$

dove ϵ_t è una variabile con valori $1 - p, -p$ e valore atteso nullo. Il test condotto più sopra è stato ottenuto stimando il modello

$$\chi_t^{p,k} = b_0 + \epsilon_t$$

e valutando l'ipotesi $b_0 = p$. Questa strategia econometrica è un caso davvero particolare che si può facilmente estendere utilizzando il modello

$$\chi_t^{p,k} = b_0 + b_1 x_t + \epsilon_t$$

nel quale x è un vettore di regressori opportunamente selezionati. In tal caso l'ipotesi da sottoporre a verifica è $b_0 = p$ e $b_1 = 1$.

Questa metodologia *allargata* si applica anche alla verifica circa l' ES , infatti, $ES_{t,k}(t) = E(r_{t,t+k} | r_{t,t+k} \leq -t)$. Si consideri allora la variabile $y_{t+k} = r_{t,t+k} | (r_{t,t+k} \leq -t)$ ed il campione corrispondente. Osserviamo che l'ampiezza di tale campione è di gran lunga inferiore a quella del campione originale poiché molte osservazioni sono state scartate, il che genera ovviamente più di un problema.

Evidentemente secondo il nostro modello

$$y_{t+k} = ES_{t,k}(t) + \epsilon_{t+k}$$

Dunque una verifica può essere condotta stimando il modello allargato

$$y_{t+k} = b_0 + b_1 ES_{t,k}(t) + b_2 x_t + \epsilon_{t+k}$$

e verificando l'ipotesi nulla $b_0 = b_2 = 1 - b_1 = 0$.

15.3. Un test sulla distribuzione.

Il modello di riferimento fornisce indicazioni sulla distribuzione dei rendimenti, ad esempio consente di specificare la distribuzione cumulata $F_{r_{t,t+k}}$. Una volta osservato il valore effettivo del rendimento, cioè la sua realizzazione $\bar{r}_{t,t+k}$, si può calcolare la probabilità *ex-post*, ossia la quantità:

$$p_t^k = F_{r_{t,t+k}}(\bar{r}_{t,t+k})$$

della quale si può conseguentemente ottenere il campione $\{p_t^k : t = 1, \dots, T\}$. Tali quantità corrispondono alle previsioni formulate alla data t circa il rendimento registrato sul periodo $t, t+k$ ed in qualche misura racchiudono tutte le informazioni di cui disponiamo per descrivere l'andamento di tale variabile nel senso che la parte che rimane inspiegata dovrebbe idealmente essere una componente puramente aleatoria. Pertanto non possiamo formulare una previsione circa l'evoluzione della quantità p_t^k o, in altri termini, non possiamo spiegare con un modello le nostre stesse previsioni. Dunque se il modello $F_{r_{t,t+k}}$ è corretto, allora le quantità p_t^k dovrebbero essere distribuite in modo uniforme sull'intervallo $[0, 1]$. Questa ipotesi può essere sottoposta a verifica, utilizzando, ad esempio, il diagramma dei quantili del campione e della distribuzione uniforme ed osservando gli eventuali scostamenti. Naturalmente questa stessa procedura si può applicare anziché all'intera distribuzione, solamente alla sua coda sinistra, selezionando solo quelle osservazioni che si trovano al di sotto di una certa soglia.

Esercizi

Domanda 14. Si risponda facendo uso del dataset `Portfolio.csv`.

- (1) Utilizzando le prime 300 osservazioni, si tracci la frontiera efficiente e si calcoli il portafoglio a varianza minima e i portafogli con rendimento atteso 1% e 2%.
- (2) Sulla rimanente parte del campione si calcoli il VaR e l' ES giornalieri per ciascuno dei tre portafogli del punto precedente con coefficiente $p = 4\%$ ed utilizzando il metodo storico e storico con pesi ($\lambda = 0.98$).
- (3) Tracciate su tre grafici distinti l'andamento dei rendimenti negativi, del VaR e dell' ES per ciascuno dei tre portafogli.
- (4) Effettuate il Backtest per tutti e tre i portafogli e commentate le conclusioni.

Domanda 15. Si consideri il dataset `Portfolio.csv`.

- (1) Si calcolino i portafogli ottimali per un investitore con preferenze del tipo:

$$U(W\mu_\alpha, W^2\sigma_\alpha^2) = W\mu_\alpha - \frac{b}{2}W^2\sigma_\alpha^2$$

con $W = 1$ e $b = 1$ oppure $b = 1.5$, rispettivamente;

- (2) Si costruiscano le serie storiche dei due portafogli calcolati al punto 1.;
- (3) Fissate le prime 100 osservazioni, si calcoli, per ciascuna delle osservazioni successive il $VaR_{1\%}$ sotto l'ipotesi che la distribuzione dei rendimenti sia normale con media e varianza pari ai corrispondenti momenti campionari calcolati su di una finestra mobile delle ultime 100 osservazioni;
- (4) Eseguiti un BackTest.

Domanda 16. Si consideri il dataset `Portfolio.csv`.

- (1) Si raffiguri la frontiera efficiente;
- (2) Si calcoli la composizione, il rendimento atteso e la varianza di un portafoglio efficiente nel quale il 5% è investito nel primo titolo e si verifichi che esso sia effettivamente efficiente;
- (3) Dato il portafoglio del punto precedente, si costruisca una serie storica dei rendimenti e sulla base di questa si calcoli il $VaR_{1,3\%}$ utilizzando il metodo storico con pesi ($\lambda = 0.8$) ed una finestra mobile di 100 osservazioni passate;
- (4) rappresentate sul medesimo grafico la serie storica del $VaR_{1,3\%}$ e delle perdite generate dal portafoglio.

Domanda 17. Si consideri il dataset `SP` e si fissino le prime 50 osservazioni.

- (1) Si costruisca una matrice con in rendimenti negativi nella prima colonna e, in quelle seguenti, il $VaR_{1,0.03}$ giornaliero (usando un campione mobile di 50 osservazioni) calcolato secondo gli approcci campionari: (i) storico, (ii) storico con pesi, (iii) storico simulato (bootstrapping)

- (2) si rappresentino graficamente le 4 serie sul medesimo grafico e si commenti il risultato;
- (3) per ciascuna delle tre serie calcolate più sopra si effettui un backtest.

Domanda 18. Si consideri il dataset **SP** e lo si suddivida in due campioni, il primo contenente 1000 osservazioni.

- (1) Si utilizzi il primo sottocampione per stimare i parametri di un modello GARCH(1,1) con valori iniziali $\alpha = 0.1$, $\beta = 0.8$ ed adottando il *variance targeting*.
- (2) Dati i valori dei parametri ottenuti al punto precedente, si costruisca una serie storica per la volatilità condizionata del secondo sottocampione.
- (3) Per ciascun giorno del secondo sottocampione si calcoli il $Var_{1,0.03}$ giornaliero
- (4) Si rappresentino graficamente i rendimenti negativi *versus* il Var come calcolato più sopra.
- (5) Si esegua un *Backtest*.

Domanda 19. Si rifaccia l'esercizio precedente utilizzando ora un modello GARCH(2,2) con valori iniziali $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\beta_1 = \beta_2 = 0.3$. Si confrontino i due grafici ottenuti.

Parte 4

Analisi Empirica dei Mercati Finanziari: Teoria ed Applcazioni

Richiami di serie storiche stazionarie e integrate

16.1. Richiami di probabilità

In quanto segue si farà spesso uso di concetti con cui lo studente dovrebbe essere già a proprio agio, tuttavia si riassumono nella presente sezione diverse definizioni e risultati che verranno utilizzati nel seguito del corso.

16.1.1. Spazi di probabilità e variabili casuali.

DEFINIZIONE 1. *Uno spazio di probabilità (Ω, \mathcal{F}, P) consiste in tre oggetti:*

- (1) Ω , un insieme non vuoto, detto spazio campionario, che contiene i possibili esiti elementari di un ipotetico esperimento casuale,
- (2) \mathcal{F} , una σ -algebra di Ω , ossia un insieme di sottoinsiemi di Ω tali che
 - (a) l'insieme vuoto \emptyset appartiene a \mathcal{F} ,
 - (b) se $A \in \mathcal{F}$ allora anche il complementare $A^c \in \mathcal{F}$,
 - (c) se A_1, A_2, A_3, \dots sono in \mathcal{F} , allora anche $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$
- (3) una misura di probabilità P definita su (Ω, \mathcal{F}) , ossia un'applicazione da \mathcal{F} all'intervallo reale $[0, 1]$, tale che
 - (a) $P(\Omega) = 1$
 - (b) se A_1, A_2, A_3, \dots è una sequenza di insiemi disgiunti di \mathcal{F} , allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

La coppia (Ω, \mathcal{F}) viene detta spazio misurabile.

Una σ -algebra molto utile, quando $\Omega = \mathbb{R}^m$ è quella di Borel¹.

DEFINIZIONE 2. *La σ -algebra di Borel su \mathbb{R}^m è la più piccola σ -algebra contenente tutti gli intervalli (iper-rettangoli) aperti di \mathbb{R}^m .*

La σ -algebra di Borel conterrà quindi anche gli intervalli chiusi, essendo essi complementi di intervalli aperti, conterrà gli intervalli semi-aperti del tipo $[a, b)$ e $(a, b]$ (per $\Omega = \mathbb{R}$), che si possono ottenere come intersezioni numerabili di insiemi aperti

$$[a, b) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right),$$

conterrà singoli punti, che sono intervalli chiusi, e le loro unioni numerabili, come per esempio l'insieme dei numeri razionali, ecc.

¹In realtà si può definire la σ -algebra di Borel per ogni spazio topologico Ω , ossia per ogni spazio in cui siano definiti gli insiemi aperti.

ESEMPIO 1. Si consideri l'esperimento lancio di tre monete. L'insieme Ω sarà dato dai possibili esiti (T =testa, C =croce)

$$\{(TTT), (TTC), (TCT), (TCC), (CTT), (CTC), (CCT), (CCC)\}.$$

La σ -algebra più piccola che si possa costruire (ma anche la meno interessante) è data dagli insiemi $\{\emptyset, \Omega\}$. La σ -algebra più grande che si possa costruire su tale insieme è

$$\begin{aligned} \{ & \emptyset, \Omega, \\ & (TTT), (TTC), (TCT), (TCC), (CTT), (CTC), (CCT), (CCC), \\ & \text{\textit{tutte le unioni degli ultimi 8 insiemi}} \} \end{aligned}$$

Un'altra σ -algebra interessante è generata a partire dagli insiemi che danno informazioni solo sul primo lancio. Definiamo, per abbreviare la notazione, i seguenti insiemi

$$\begin{aligned} A_T &= \{(TTT), (TTC), (TCT), (TCC)\} = \{\text{testa al primo lancio}\} \\ A_C &= \{(CTT), (CTC), (CCT), (CCC)\} = \{\text{croce al primo lancio}\}. \end{aligned}$$

È semplice verificare che $\mathcal{F}_1 = \{\emptyset, \Omega, A_T, A_C\}$ è una σ -algebra, infatti contiene l'insieme vuoto e tutti i complementi e unioni di A_T e A_C .

ESERCIZIO 1. Si consideri l'insieme Ω definito nell'esempio 1 e si costruisca la σ -algebra \mathcal{F}_2 a partire dagli insiemi che danno informazioni solo sul primo e sul secondo lancio.

Come appare dall'esempio e dall'esercizio precedenti, una σ -algebra può essere pensata come una collezione di informazioni: nell'esempio del lancio di tre monete, dire a quale insieme di \mathcal{F}_1 appartiene il risultato dell'esperimento, corrisponde a informare sull'esito del primo lancio della sequenza. Allo stesso modo, dire a quale insieme di \mathcal{F}_2 appartiene il risultato dell'esperimento, corrisponde a informare sull'esito del primo e del secondo lancio della sequenza. Supponendo che le monete vengano lanciate una dopo l'altra, si può dire che \mathcal{F}_1 contiene l'informazione fino all'istante 1 e analogamente, che \mathcal{F}_2 contiene l'informazione fino all'istante 2. Spesso torna utile inizializzare la sequenza $\mathcal{F}_1, \mathcal{F}_2, \dots$ con la σ -algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$ che non contiene alcuna informazione sull'esito dei lanci.

Una sequenza di σ -algre con le proprietà di quella appena descritta viene detta *filtrazione* e gioca un ruolo importante nell'analisi delle serie storiche.

DEFINIZIONE 3. Sia (Ω, \mathcal{F}) uno spazio misurabile. Una sequenza $(\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots)$ di σ -algre su Ω è detta *filtrazione* in (Ω, \mathcal{F}) se

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \dots \subseteq \mathcal{F}.$$

Un altro concetto fondamentale nello studio dei fenomeni stocastici è quello di *funzione misurabile*.

DEFINIZIONE 4. Siano (Ω, \mathcal{F}) e (Ψ, \mathcal{G}) spazi misurabili. Una funzione misurabile da (Ω, \mathcal{F}) a (Ψ, \mathcal{G}) è una funzione $X : \Omega \mapsto \Psi$ tale che² $X^{-1}(B) \in \mathcal{F}$ per ogni $B \in \mathcal{G}$.

Per rendere più palpabile la definizione di funzione misurabile si consideri il seguente esempio.

²Con la scrittura $X^{-1}(B)$, si intende $\{\omega : X(\omega) \in B\}$.

ESEMPIO 2. Si prenda in considerazione il lancio di due monete e l'associato spazio misurabile

$$\begin{aligned}\Omega &= \{TT, TC, CT, CC\} \\ \mathcal{F} &= \{\emptyset, TT, TC, CT, CC, \\ &\quad TT \cup TC, TT \cup CT, TT \cup CC, TC \cup CT, TC \cup CC, CT \cup CC, \\ &\quad TC \cup CT \cup CC, TT \cup CT \cup CC, TT \cup TC \cup CC, TT \cup TC \cup CT, \Omega\}.\end{aligned}$$

Si definisca la funzione $X : \Omega \mapsto \{0, 1, 2\}$ che associa ad ogni lancio delle due monete la somma delle teste uscite:

$$X(TT) = 2, X(TC) = X(CT) = 1, X(CC) = 0.$$

La σ -algebra più fine che si possa costruire sull'insieme $\Psi = \{0, 1, 2\}$ è

$$\mathcal{G} = \{\emptyset, 0, 1, 2, 0 \cup 1, 0 \cup 2, 1 \cup 2, \Psi\}.$$

Mostriamo che $X(\omega)$ è \mathcal{F} -misurabile:

$$\begin{aligned}X^{-1}(\{\emptyset\}) &= \{\emptyset\} \\ X^{-1}(\{0\}) &= \{CC\} \\ X^{-1}(\{1\}) &= \{CT \cup TC\} \\ X^{-1}(\{2\}) &= \{TT\} \\ X^{-1}(\{0 \cup 1\}) &= \{CC \cup CT \cup TC\} \\ X^{-1}(\{0 \cup 2\}) &= \{CC \cup TT\} \\ X^{-1}(\{1 \cup 2\}) &= \{CT \cup TC \cup TT\} \\ X^{-1}(\{0 \cup 1 \cup 2\}) &= \{\Omega\}.\end{aligned}$$

Dato che ogni altra sigma algebra su Ψ è meno fine (ha meno sottoinsiemi) di \mathcal{G} , si è dimostrato che X è \mathcal{F} -misurabile qualunque sia la σ -algebra su Ψ .

La σ -algebra definita nel precedente esempio per mezzo dell'inversione $X^{-1}(A)$, fatta per ogni $A \in \mathcal{G}$, è detta σ -algebra generata da X .

Alcune funzioni misurabili di particolare interesse hanno dei nomi propri.

- quando $X : \Omega \mapsto \mathbb{R}$, X viene detta *variabile casuale*,
- quando $X : \Omega \mapsto \mathbb{R}^m$, X prende il nome *vettore casuale*,
- quando $X : \Omega \mapsto \mathbb{R}^\infty$, X è detta *processo stocastico a tempo discreto*,
- quando $X : \Omega \mapsto \mathbf{C}[a, b]$, dove $\mathbf{C}[a, b]$ indica l'insieme delle funzioni reali continue definite sull'intervallo $[a, b]$, X è detta *processo stocastico a tempo continuo*.

Nei quattro casi sopra elencati si intende che la σ -algebra associata all'insieme codominio di X è la σ -algebra di Borel \mathcal{B} .

16.1.2. Aspettativa condizionata e processo differenza di Martingala. Si suppone che il lettore abbia già familiarità con una definizione di *aspettativa condizionata* o *valore atteso condizionato*, tuttavia per completezza riportiamo la seguente definizione formale.

DEFINIZIONE 5. Sia Y una variabile casuale definita sullo spazio $\{\Omega, \mathcal{F}\}$ tale che $\mathbb{E}[|Y|] < \infty$, e sia $\mathcal{G} \subset \mathcal{F}$ una σ -algebra. Esiste una variabile casuale $\mathbb{E}[Y|\mathcal{G}]$ chiamata *aspettativa condizionata* (o *valore atteso condizionato*), tale che

- (i) $\mathbb{E}[Y|\mathcal{G}]$ è \mathcal{G} -misurabile e $\mathbb{E}[|\mathbb{E}(Y|\mathcal{G})|] < \infty$,
- (ii) $\mathbb{E}[Y|\mathcal{G}]$ soddisfa

$$\mathbb{E}[\mathbb{I}_G \mathbb{E}(Y|\mathcal{G})] = \mathbb{E}[\mathbb{I}_G Y]$$

per tutti gli insiemi G in \mathcal{G} , dove \mathbb{I}_G è la funzione indicatrice pari a 1 in G e zero altrove.

Si noti che la comune notazione $\mathbb{E}[Y|Z]$ è da interpretarsi come $\mathbb{E}[Y|\sigma(Z)]$, dove $\sigma(Z)$ è la σ -algebra generata da Z .

Si elencano senza dimostrarle le principali proprietà del valore atteso condizionato.

TEOREMA 3 (Proprietà dell'aspettativa condizionata). *Siano Y, Y_1, \dots, Y_k variabili casuali con momento primo finito, a_1, \dots, a_k costanti finite e Z variabile casuale \mathcal{G} -misurabile e tale che $\mathbb{E}[|ZY|] < \infty$. Le seguenti affermazioni sono vere:*

- (i) *Linearità:* $\mathbb{E}\left[\sum_{j=1}^k a_j Y_j | \mathcal{G}\right] = \sum_{j=1}^k a_j \mathbb{E}[Y_j | \mathcal{G}]$
- (ii) $\mathbb{E}[ZY | \mathcal{G}] = Z \mathbb{E}[Y | \mathcal{G}]$,
- (iii) $\mathbb{E}[Z(Y - \mathbb{E}[Y|\mathcal{G}])] = 0$,
- (iv) *Disuguaglianza di Jensen:* sia g funzione misurabile convessa sul supporto di Y con aspettativa finita, allora $g(\mathbb{E}[Y|\mathcal{G}]) \leq \mathbb{E}[g(Y)|\mathcal{G}]$. Se g è concava, la disuguaglianza è invertita.
- (v) *Legge della aspettative iterate:* siano $\mathcal{F}_1 \subset \mathcal{F}_2$ σ -algebre sullo spazio campionario di Y , Ω , allora

$$\mathbb{E}[\mathbb{E}(Y|\mathcal{F}_2)|\mathcal{F}_1] = \mathbb{E}[Y|\mathcal{F}_1].$$

- (vi) *Minimizzazione dell'errore quadratico medio:* si supponga che $\mathbb{E}[Y^2] < \infty$, allora

$$\mathbb{E}[(Y - \mathbb{E}(Y|\mathcal{G}))^2] \leq \mathbb{E}[(Y - Z)^2]$$

per qualunque variabile casuale \mathcal{G} -misurabile Z .

Un processo a tempo discreto che comparirà di frequente nel resto di questo testo è il processo *differenza di martingala*.

DEFINIZIONE 6 (Sequenza stocastica adattata a una filtrazione). *Sia $\{Y_t\}$ una sequenza di variabili casuali, e sia $\{\mathcal{F}_t\}$ una filtrazione. Se Y_t è misurabile rispetto a \mathcal{F}_t per ogni t , allora $\{Y_t\}$ si dice adattata alla sequenza $\{Y_t\}$, e $\{Y_t, \mathcal{F}_t\}$ è chiamata sequenza stocastica adattata.*

Intuitivamente, se pensiamo al parametro t come al tempo, possiamo vedere \mathcal{F}_t come ad una σ -algebra contenente informazioni sulla sequenza $\{Y_t\}$ accumulate fino al tempo t . Se \mathcal{F}_{t-1} è strettamente contenuta in \mathcal{F}_t , allora Y_t non sarà \mathcal{F}_{t-1} -misurabile dato che in genere esisteranno valori di Y_t a cui non corrisponde alcun elemento in \mathcal{F}_{t-1} .

DEFINIZIONE 7 (Processo differenza di martingala). *Sia $\{Y_t, \mathcal{F}_t\}$ una sequenza stocastica adattata. $\{Y_t\}$ si dice processo differenza di martingala rispetto alla filtrazione $\{\mathcal{F}_t\}$ se*

$$\mathbb{E}[Y_t | \mathcal{F}_{t-1}] = 0 \quad \forall t.$$

Una differenza di martingala è quindi un processo per cui la migliore previsione basata sull'insieme informativo \mathcal{F}_{t-1} , nel senso della minimizzazione dell'errore quadratico medio (si veda la proprietà (vi) del valore atteso condizionato), è sempre nulla.

Ovviamente è sempre possibile costruire una differenza di martingala partendo da una sequenza adattata $\{X_t, \mathcal{F}_t\}$ sottraendo a X_t la sua aspettativa condizionata a \mathcal{F}_{t-1} :

$$Y_t = X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}].$$

ESERCIZIO 2 (Proprietà elementari delle differenze di martingala). *Dimostrare che se $\{X_t\}$ è una differenza di martingala rispetto a \mathcal{F}_t , sono vere le seguenti affermazioni:*

- ✓ $\mathbb{E}[X_{t+k}|\mathcal{F}_t] = 0$ per k intero positivo e t arbitrario;
- ✓ $\mathbb{E}[X_t] = 0$ per ogni t ;
- ✓ $\mathbb{E}[X_t X_s] = 0$ per qualunque scelta di t e s interi tali che $t \neq s$.

Le differenze di martingala giocano un ruolo importante sia per la modellazione dei rendimenti finanziari, sia per derivare le proprietà degli stimatori e dei test per i parametri incogniti di modelli di serie storiche. Questo ultimo utilizzo è legato all'esistenza di leggi dei grandi numeri e teoremi centrali del limite per differenze di martingala.

16.1.3. Convergenze stocastiche. Molto spesso non è possibile ottenere la distribuzione esatta di uno stimatore o statistica test per ogni grandezza campionaria e pertanto è necessario ricorrere ad approssimazioni *asintotiche*. Data la natura particolare degli spazi di probabilità e delle funzioni misurabili su essi definite, il concetto di limite dell'analisi, con cui il lettore è sicuramente familiare, va opportunamente modificato ed esteso.

I più rilevanti modi di convergenza di una sequenza di variabili (o vettori) casuali X_1, X_2, \dots a una variabile (o vettore) casuale limite X , tutte definite sul medesimo spazio di probabilità (Ω, \mathcal{F}, P) , sono

convergenza in probabilità:

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1 \quad \forall \varepsilon \in [0, \infty)$$

le simbologie sintetiche generalmente utilizzate sono $\text{plim } X_n = X$, $X_n \xrightarrow{P} X$, oppure $X_n \xrightarrow{\mathcal{P}} X$;
convergenza in media quadratica o in norma L_2 :

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^2] = 0$$

e viene in genere indicata (in Italiano) con $X_n \xrightarrow{mq} X$ oppure $X_n \xrightarrow{L_2} X$;

convergenza quasi certa:

$$P\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\} = 1$$

e viene indicata (in Italiano) con $X_n \xrightarrow{qc} X$ oppure facendo seguire il simbolo classico di limite da q.c.

Un'ulteriore convergenza che coinvolge solamente le funzioni di ripartizione delle variabili casuali, e che pertanto non richiede che esse siano definite sul medesimo spazio di probabilità è la

convergenza in distribuzione: sia $F_n(\cdot)$ la sequenza di funzioni di ripartizione delle variabili casuali X_n , e $F(\cdot)$ la funzione di ripartizione di X , allora X_n si dice convergere in distribuzione (o in legge o debolmente) a X se

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{per ogni valore di } x \text{ per il quale } F \text{ è continua;}$$

tale convergenza è indicata con $X_n \xrightarrow{d} X$, $X_n \xrightarrow{\mathcal{D}} X$, $\mathcal{L}(X_n) \rightarrow \mathcal{L}(X)$ oppure $X_n \Rightarrow X$ e più raramente con $X_n \rightsquigarrow X$.

Riportiamo senza dimostrarle le implicazioni tra le varie modalità di convergenza.

TEOREMA 4 (Implicazioni tra convergenze stocastiche). *Valgono le seguenti implicazioni:*

$$(i) \quad X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X,$$

- (ii) $X_n \xrightarrow{qc} X \Rightarrow X_n \xrightarrow{p} X$,
- (iii) $X_n \xrightarrow{mq} X \Rightarrow X_n \xrightarrow{p} X$,
- (iv) $X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{p} c$ se c è una costante,
- (v) $X_n \xrightarrow{d} X, |X_n - Y_n| \xrightarrow{p} 0 \Rightarrow Y_n \xrightarrow{d} X$,
- (vi) $X_n \xrightarrow{p} X, Y_n \xrightarrow{p} Y \Rightarrow (X_n, Y_n) \xrightarrow{p} (X, Y)$,
- (vii) $X_n \xrightarrow{d} X, Y_n \xrightarrow{d} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c)$ se c è una costante.

Le seguenti implicazioni sono invece note come Teorema di Slutsky.

TEOREMA 5 (Teorema di Slutsky). *Siano X_n, X, Y_n variabili, vettori o matrici casuali. Se $X_n \xrightarrow{d} X$ e $Y \xrightarrow{p} c$, con c costante, allora*

- (i) $X_n + Y_n \xrightarrow{d} X + c$,
- (ii) $X_n Y_n \xrightarrow{d} cX$,
- (iii) $Y_n^{-1} X_n \xrightarrow{d} c^{-1} X$ se c^{-1} esiste finito.

Una classe di funzioni misurabili particolarmente importante è quella delle funzioni continue (quasi certamente). Il seguente teorema mostra a che cosa convergono funzioni continue di sequenze convergenti di variabili casuali.

TEOREMA 6 (delle funzioni continue). *Siano X_1, X_2, \dots e X variabili (vettori) casuali, e sia $g(\cdot)$ una funzione continua a meno di un insieme di probabilità nulla rispetto alla distribuzione di X (cioè, se D_g è l'insieme dei punti di discontinuità di g deve valere $\Pr(X \in D_g) = 0$), allora*

- (i) $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$,
- (ii) $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$,
- (iii) $X_n \xrightarrow{qc} X \Rightarrow g(X_n) \xrightarrow{qc} g(X)$.

Ora, rinfrescate le idee sul significato dei diversi modi di convergenza stocastica, si presentano i due teoremi di convergenza fondamentali, cioè la legge dei grandi numeri e il teorema centrale limite, per i processi differenza di martingala.

TEOREMA 7 (Legge forte dei grandi numeri). *Sia $\{Y_t, \mathcal{F}_t\}$ una processo differenza di martingala. Se per qualche $r \geq 1$,*

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[|Y_t|^{2r}]}{t^{1+r}} < \infty,$$

allora al divergere di n

$$\bar{Y}_n = \frac{1}{n} \sum_{t=1}^n Y_t \rightarrow 0 \quad q.c.$$

La condizione sui momenti, più forte rispetto a quella del caso i.i.d., è comunque abbastanza generale da essere assolta dalla gran parte dei modelli che presenteremo nelle prossime pagine. Ovviamente tale condizione è rispettata da tutti i processi differenza di martingala a varianza costante (finita) o a varianza non costante ma uniformemente limitata.

TEOREMA 8 (Teorema centrale del limite). *Sia $\{Y_t, \mathcal{F}_t\}$ un processo differenza di martingala e valgano le seguenti assunzioni:*

- ✓ $\mathbb{E}[Y_t^2] = \sigma_t^2 > 0$ per ogni t ,
- ✓ $\frac{1}{n} \sum_{t=1}^n \sigma_t^2 = \sigma^2 > 0$,

✓ $\mathbb{E}[|Y_t|^r] < \infty$ per qualche $r > 2$ e ogni t .

Allora $\sqrt{n}\bar{Y}_n$ converge in distribuzione a una normale $N(0, \sigma^2)$.

16.2. Previsioni statistiche

16.2.1. Il problema della previsione. Da un punto di vista statistico *prevedere* significa determinare con il minore errore possibile la realizzazione di una variabile casuale per mezzo della realizzazione di altre variabili casuali. Pertanto per potere risolvere il problema è necessario scegliere una *funzione di perdita* o di *costo* e determinare il *previsore ottimo*, cioè la funzione (misurabile) delle variabili osservabili, che minimizza la *perdita attesa*.

Formalmente, si supponga di volere prevedere Y per mezzo di X_1, \dots, X_k . Sia $\ell(\cdot) : \mathbb{R} \mapsto [0, \infty)$ la funzione di perdita scelta e \mathcal{G} la classe delle funzioni di X_1, \dots, X_k all'interno della quale cercare il previsore. Si noti che \mathcal{G} può consistere nella classe di tutte le funzioni misurabili nei k argomenti oppure in una sottoclasse di essa come, ad esempio, quella di tutte le funzioni lineari. Un previsore ottimo è una funzione g che risolve il problema di minimo

$$(16.2.1) \quad \min_{g \in \mathcal{G}} \mathbb{E}\ell(Y - g(X_1, \dots, X_k)),$$

dove si suppone che il valore atteso in formula esista. Non necessariamente il previsore ottimo è unico; tuttavia ciò sarà vero (con probabilità 1) nei casi concreti che prenderemo in considerazione nei prossimi paragrafi.

Nel prevedere, vi sono situazioni in cui un errore per difetto implica costi molto più alti di un errore per eccesso. Per esempio, si supponga di dovere prevedere la piena di una lago vicino ad una grande città. Quando la piena supera un certo livello, il centro della città viene allagato danneggiando negozi, cantine e abitazioni, e pertanto il costo di un errore di previsione positivo (realizzazione maggiore della previsione) è piuttosto ingente. Al contrario, quando si prevede per difetto, il costo consiste solamente nell'apertura di alcune chiuse di sfogo, nell'allagamento di bacini e campi, e nella predisposizione di barriere in città. La funzione di costo per questo problema di previsione potrebbe avere la forma in Figura 1.

Si noti che una funzione di costo di questo tipo implica un previsore ottimale *distorto*. In altre parole, il valore atteso dell'errore di previsione non è zero, perché questa funzione di perdita implica un minore costo atteso nel prevedere valori alti del livello del lago (cioè valori negativi dell'errore di previsione più frequenti).

Spesso, tuttavia, non esiste una funzione di costo esplicita, specialmente quando si producono previsioni che devono essere utilizzate da una gran massa di utenti, come, ad esempio, le previsioni di crescita del prodotto interno lordo o del tasso d'inflazione. In questi casi si usa preferibilmente una funzione di perdita simmetrica, in modo tale che la previsione non sia distorta (errore atteso di previsione nullo). Per motivi legati soprattutto alla semplicità matematica di cui si renderà conto il lettore più avanti, la funzione di perdita più utilizzata è l'errore quadratico, $\ell(E) = E^2$, dove $E := Y - g(X_1, \dots, X_k)$. Altre funzioni talvolta usate sono l'errore assoluto, $\ell(E) = |E|$ o, quando Y è strettamente positiva (con probabilità 1), l'errore assoluto relativo $\ell(E) = |E|/Y$. Il valore atteso di queste funzioni di perdita ha un nome preciso in letteratura³:

Mean Square Error:: $\text{MSE}(E) = \mathbb{E}[E^2]$,

Mean Absolute error:: $\text{MAE}(E) = \mathbb{E}[|E|]$,

Mean Absolute Percentage Error:: $\text{MAPE}(E) = 100 \cdot \mathbb{E}[|E|/y]$.

³Dato che la letteratura statistica è prevalentemente in lingua inglese, in questo testo si preferisce rendere il lettore familiare con termini e sigle comuni in tale letteratura.

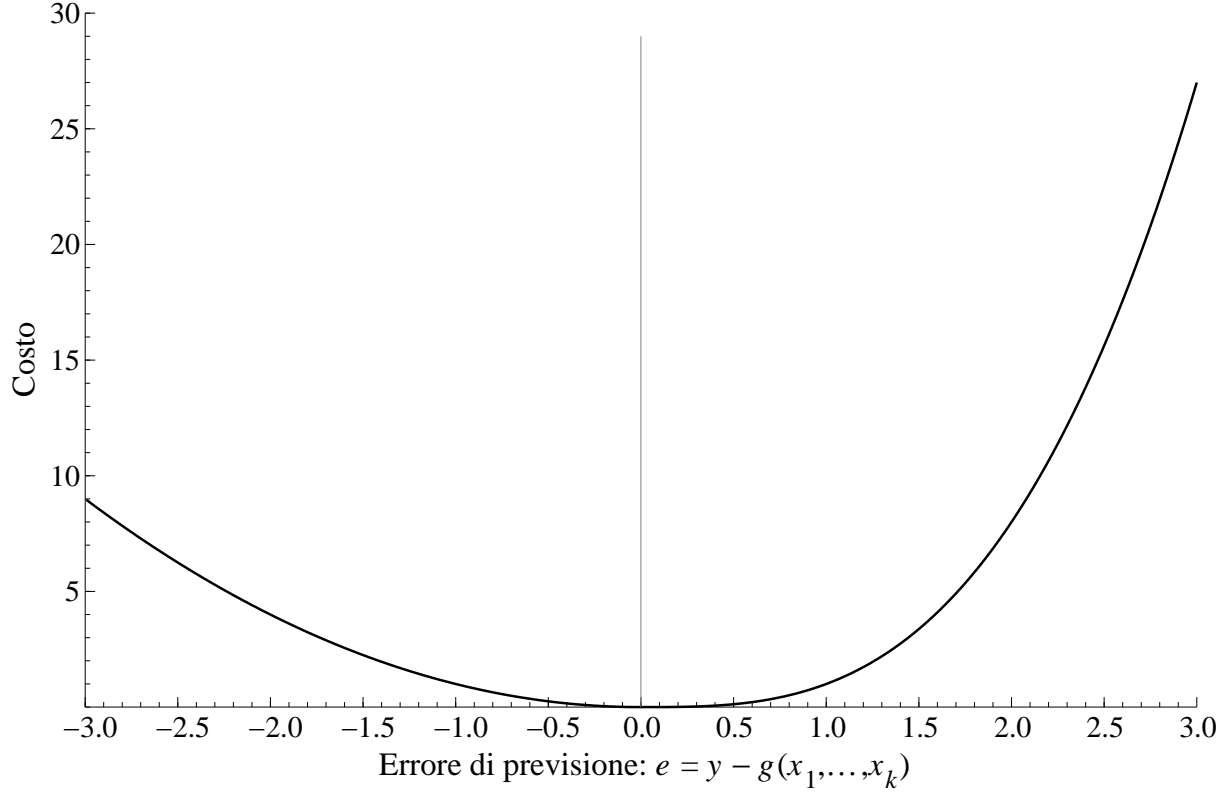


FIGURA 1. Una possibile funzione di costo asimmetrica.

In questo testo si considerano solamente previsori ottimi rispetto all'errore quadratico medio (MSE = Mean Square Error).

16.2.2. Il previsore ottimo. Da questo momento in poi quando si parlerà in questo testo di *previsore ottimo* senza ulteriori aggettivazioni, si intenderà, *ottimo rispetto al MSE*, cioè ottimo rispetto ad una funzione di perdita quadratica $\ell(E) = E^2$.

TEOREMA 9 (del previsore ottimo). *Siano Y, X_1, \dots, X_k variabili casuali con varianza finita, $\ell(E) = E^2$ e*

$$\mathcal{G} = \{\text{tutte le funzioni misurabili di } X_1, \dots, X_k\}.$$

Allora l'unica⁴ funzione $g \in \mathcal{G}$ che risolve il problema di previsione in equazione (16.2.1) è il valore atteso condizionato

$$g(X_1, \dots, X_k) = \mathbb{E}[Y|X_1, \dots, X_k].$$

⁴Qui per *unica* si intende *unica a meno di insiemi di probabilità nulla*. Cioè $g(\mathbf{X})$ è soluzione unica del problema di previsione se per tutte le funzioni $f \in \mathcal{G}$ che risolvono il problema vale $\Pr\{g(\mathbf{X}) = f(\mathbf{X})\} = 1$.

DIMOSTRAZIONE. Sia $g \in \mathcal{G}$ una generica funzione misurabile nei suoi argomenti e per alleggerire la scrittura si ponga $\mathbf{X} = (X_1, \dots, X_k)'$. L'errore quadratico medio di previsione è dato da

$$\begin{aligned} \mathbb{E}(Y - g(\mathbf{X}))^2 &= \\ \mathbb{E}(Y - \mathbb{E}[Y|\mathbf{X}] + \mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))^2 &= \\ \mathbb{E}(Y - \mathbb{E}[Y|\mathbf{X}])^2 + \mathbb{E}(\mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))^2 + 2\mathbb{E}[(Y - \mathbb{E}[Y|\mathbf{X}])(\mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))]. \end{aligned}$$

Ora, se si condiziona il valore atteso dell'ultimo addendo a \mathbf{X} si ottiene

$$\begin{aligned} \mathbb{E}[(Y - \mathbb{E}[Y|\mathbf{X}])(\mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))|\mathbf{X}] &= \\ (\mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))\mathbb{E}(Y - \mathbb{E}[Y|\mathbf{X}]|\mathbf{X}) &= 0. \end{aligned}$$

Prendendo il valore atteso di quest'ultima quantità rispetto alla distribuzione di \mathbf{X} , si ottiene nuovamente zero, e sfruttando la legge dei valori attesi iterati possiamo concludere che tale prodotto incrociato è nullo. Pertanto, l'errore quadratico medio di previsione è minimo quando $\mathbb{E}[g(\mathbf{X})] = \mathbb{E}[Y|\mathbf{X}]$, dato che in questo caso la quantità non-negativa $\mathbb{E}(\mathbb{E}[Y|\mathbf{X}] - g(\mathbf{X}))^2$ si annulla. \square

Una semplice applicazione della legge dei valori attesi iterati dimostra che il valore atteso condizionato di Y è un previsore *corretto* (o non distorto):

$$\mathbb{E}[Y - \mathbb{E}(Y|\mathbf{X})] = \mathbb{E}[Y] - \mathbb{E}[\mathbb{E}(Y|\mathbf{X})] = \mathbb{E}[Y] - \mathbb{E}[Y] = 0.$$

Si noti che per applicare questo risultato è necessario avere un modello direttamente per il valore atteso condizionato di Y , tipo un modello di regressione, oppure la distribuzione congiunta delle variabili casuali Y, X_1, \dots, X_k .

16.2.3. Il previsore lineare ottimo. Se nella ricerca della funzione ottima ci si limita alla classe delle funzioni lineari in $1, X_1, \dots, X_k$, allora le informazioni indispensabili alla costruzione del previsore ottimo si riducono ai primi due momenti del vettore $(Y, X_1, \dots, X_k)'$.

Pertanto, definiamo i primi due momenti di $(Y, X_1, \dots, X_k)'$ come segue:

$$\mu_Y := \mathbb{E}[Y], \quad \mu_{\mathbf{X}} := \mathbb{E}[\mathbf{X}],$$

e

$$\Sigma_{Y\mathbf{X}} := \mathbb{E}[(Y - \mu_Y)(\mathbf{X} - \mu_{\mathbf{X}})'], \quad \Sigma_{\mathbf{X}\mathbf{X}} := \mathbb{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})'], \quad \Sigma_{\mathbf{X}Y} := \Sigma'_{Y\mathbf{X}}.$$

TEOREMA 10 (del previsore lineare ottimo). *Siano Y, X_1, \dots, X_k variabili casuali con varianza finita, $\ell(E) = E^2$ e*

$$\mathcal{G} = \{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k, \forall \beta := (\beta_0, \dots, \beta_k) \in \mathbb{R}^{k+1}\}.$$

Allora:

(i) *l'unica funzione $g \in \mathcal{G}$ che risolve il problema di previsione in equazione (16.2.1) è la proiezione lineare*

$$(16.2.2) \quad g(X_1, \dots, X_k) = \mathbb{P}[Y|\mathbf{X}] := \mu_Y + \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}(\mathbf{X} - \mu_{\mathbf{X}}),$$

con $\Sigma_{\mathbf{X}\mathbf{X}}^{-1}$ inversa generalizzata nel caso $\Sigma_{\mathbf{X}\mathbf{X}}$ non abbia rango pieno,

(ii) *il suo MSE è dato da*

$$\mathbb{E}(Y - \mathbb{P}[Y|\mathbf{X}])^2 = \text{Var}[Y] - \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{X}\mathbf{X}}^{-1}\Sigma_{\mathbf{X}Y},$$

(iii) *$\mathbb{P}[Y|\mathbf{X}]$ è un previsore corretto*

$$\mathbb{E}[Y - \mathbb{P}[Y|\mathbf{X}]] = 0$$

(iv) e l'errore di previsione è ortogonale (incorrelato) a \mathbf{X}

$$\mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])\mathbf{X}'] = \mathbf{0}'.$$

Prima di dimostrare il teorema, è utile fare un paio di osservazioni. Si noti che dalla (16.2.2) è semplice derivare la formula per calcolare il vettore dei coefficienti β :

$$\beta'_1 := \begin{bmatrix} \beta_1 & \dots & \beta_k \end{bmatrix} = \Sigma_{Y\mathbf{X}} \Sigma_{\mathbf{X}\mathbf{X}}^{-1}, \quad \beta_0 = \mu_Y - \Sigma_{Y\mathbf{X}} \Sigma_{\mathbf{X}\mathbf{X}}^{-1} \mu_{\mathbf{X}}.$$

Inoltre, mentre sotto le condizioni del teorema il previsore lineare è sempre unico, il vettore dei coefficienti β è unico solo se $\Sigma_{\mathbf{X}\mathbf{X}}$ è a rango pieno.

DIMOSTRAZIONE. Iniziamo a dimostrare il punto (iii). Per comodità si ponga $\beta'_1 := \Sigma_{Y\mathbf{X}} \Sigma_{\mathbf{X}\mathbf{X}}^{-1}$. Allora il valore atteso dell'errore di previsione è

$$\mathbb{E}[Y - \mu_Y - \beta'_1(\mathbf{X} - \mu_{\mathbf{X}})] = \mathbb{E}[Y - \mu_Y] - \beta'_1 \mathbb{E}[\mathbf{X} - \mu_{\mathbf{X}}] = 0.$$

Per dimostrare il punto (iv) si noti che essendo $\mathbb{P}[Y|\mathbf{X}]$ previsore corretto, risulta (mostrare per esercizio)

$$\mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])\mathbf{X}'] = \mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])(\mathbf{X} - \mu_{\mathbf{X}})'].$$

Allora

$$\begin{aligned} & \mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])(\mathbf{X} - \mu_{\mathbf{X}})'] = \\ & \mathbb{E}[(Y - \mu_Y - \beta'_1(\mathbf{X} - \mu_{\mathbf{X}}))(\mathbf{X} - \mu_{\mathbf{X}})'] = \\ & \mathbb{E}[(Y - \mu_Y)(\mathbf{X} - \mu_{\mathbf{X}})'] - \beta'_1 \mathbb{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})'] = \\ & \Sigma_{Y\mathbf{X}} - \beta'_1 \Sigma_{\mathbf{X}\mathbf{X}} = \\ & \Sigma_{Y\mathbf{X}} - \Sigma_{Y\mathbf{X}} = 0. \end{aligned}$$

Dimostriamo ora il punto (i) facendo vedere che non esistono funzioni lineari di $(1, X_1, \dots, X_k)$ che comportano un MSE più piccolo di quello del previsore lineare. Sia $g(\mathbf{X}) := \delta + \gamma' \mathbf{X}$ una qualunque funzione lineare di \mathbf{X} , allora il suo MSE è dato da

$$\begin{aligned} \text{MSE}_g &= \mathbb{E}(Y - \delta - \gamma' \mathbf{X})^2 \\ &= \mathbb{E}(Y - \mathbb{P}[Y|\mathbf{X}] + \mathbb{P}[Y|\mathbf{X}] - \delta - \gamma' \mathbf{X})^2 \\ &= \mathbb{E}(Y - \mathbb{P}[Y|\mathbf{X}])^2 + \mathbb{E}(\mathbb{P}[Y|\mathbf{X}] - \delta - \gamma' \mathbf{X})^2 \\ &\quad + 2\mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])(\mathbb{P}[Y|\mathbf{X}] - \delta - \gamma' \mathbf{X})] \\ &= \text{MSE}_{\mathbb{P}[Y|\mathbf{X}]} + \mathbb{E}[\mathbb{P}[Y|\mathbf{X}] - g(\mathbf{X})]^2 \\ &\quad + 2\mathbb{E}[(Y - \mathbb{P}[Y|\mathbf{X}])(\mu_Y - \beta_1 \mu_{\mathbf{X}} - \delta + (\beta'_1 - \gamma')\mathbf{X})]. \end{aligned}$$

Ora, per la correttezza di $\mathbb{P}[Y|\mathbf{X}]$ e per la sua ortogonalità a \mathbf{X} , il doppio prodotto nell'ultima riga è sempre nullo. Pertanto il MSE del previsore lineare g è minimo quando $g(\mathbf{X}) = \mathbb{P}[Y|\mathbf{X}]$ con probabilità 1.⁵

Si lascia al lettore la dimostrazione del punto (ii). □

⁵Si noti che questo non implica necessariamente che $\delta = \beta_0$ e $\gamma = \beta_1$. Tali identità sono vere solo quando $\Sigma_{\mathbf{X}\mathbf{X}}$ ha rango pieno.

16.2.4. Il caso gaussiano. Le proprietà di gaussianità (o normalità) e linearità sono intimamente legate. Infatti, ogni combinazione lineare di variabili casuali gaussiane è a sua volta gaussiana. Inoltre, come visto nel paragrafo precedente, per costruire il previsore lineare ottimo è sufficiente conoscere i primi due momenti del vettore casuale, e tali momenti caratterizzano completamente la distribuzione normale.

Come risulterà evidente dal seguente teorema, previsore ottimo e previsore lineare ottimo coincidono nel caso gaussiano.

TEOREMA 11 (distribuzione condizionata di una normale multivariata). *Sia \mathbf{Z} un vettore casuale gaussiano così ripartito*

$$\mathbf{Z} := \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix},$$

e con vettore delle medie e matrice di covarianza, conformemente ripartiti,

$$\boldsymbol{\mu} := \mathbb{E}[\mathbf{Z}] = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \quad \boldsymbol{\Sigma} := \mathbb{E}[(\mathbf{Z} - \boldsymbol{\mu})(\mathbf{Z} - \boldsymbol{\mu})'] = \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{bmatrix}.$$

Allora la variabile casuale $\mathbf{Y}|\mathbf{X}$ è a sua volta normale con momenti

$$\boldsymbol{\mu}_{Y|X} = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\mathbf{X} - \boldsymbol{\mu}_X), \quad \boldsymbol{\Sigma}_{Y|X} = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}.$$

Si invita il lettore a dimostrare il teorema sopra enunciato utilizzando il seguente lemma.

LEMMA 1 (determinante e inversa di una matrice a blocchi). *Valgono le seguenti identità:*

(i)

$$\left| \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix} \right| = |\mathbf{T}| |\mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}|.$$

(ii)

$$\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & -\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & \mathbf{Q}^{-1} \end{bmatrix},$$

con $\mathbf{Q} = \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}$.

Confrontando il previsore lineare in equazione (16.2.2) con l'enunciato del Teorema 11 risulta chiaro che nel caso di dati congiuntamente normali risulta

$$\mathbf{P}[\mathbf{Y}|\mathbf{X}] = \mathbb{E}[\mathbf{Y}|\mathbf{X}],$$

cioè il previsore ottimo coincide con il previsore lineare ottimo.

Inoltre, nel caso gaussiano, e *solo in questo caso*, la varianza condizionata non dipende dal valore del vettore casuale rispetto a cui si sta condizionando, e pertanto il MSE coincide con la varianza condizionata: $\text{MSE} := \mathbb{E}[\text{Var}(\mathbf{Y}|\mathbf{X})] = \text{Var}(\mathbf{Y}|\mathbf{X})$.

16.3. Serie storiche e processi stocastici stazionari

Una *serie storica* è una collezione di misurazioni riguardanti un certo fenomeno (per es. il PIL di una paese, un indice dei prezzi, il valore di un titolo azionario, ecc.), raccolte nel tempo. In particolare, in questa sede, ci si occuperà solamente di serie storiche, le cui osservazioni sono raccolte a periodi equidistanti (o comunque a situazioni assimilabili a questa condizione). Da un punto di vista statistico, una serie storica è vista come il risultato di un meccanismo di generazione di dati, che prende il nome di *processo stocastico*, e sul quale è necessario fare inferenza per poter capire e prevedere l'andamento futuro della serie stessa.

Per serie storica x_1, x_2, \dots, x_T si intenderà, in quanto segue, una parte finita della realizzazione di un processo stocastico a tempo discreto $\{X_t\}_{t=-\infty}^{\infty}$. Il processo stocastico è quindi visto come un meccanismo generatore di dati, che segue leggi stocastiche piuttosto che deterministiche e le cui funzioni di ripartizione marginali verranno indicate con $F_{X_{t_1}, \dots, X_{t_k}}(x_{t_1}, \dots, x_{t_k})$, con $t_1, \dots, t_k \in \mathbb{Z}$.

Dato che una serie storica ci permette di conoscere una sola realizzazione x_t della rispettiva variabile casuale X_t , rendendo pressoché impossibile ogni tipo di inferenza tradizionale (basata su realizzazioni indipendenti di variabili casuali identicamente distribuite), è utile considerare alcune classi di processi stocastici che, rispettando alcune condizioni di regolarità, consentono di costruire opportune tecniche inferenziali.

DEFINIZIONE 8 (Stazionarietà forte). *Il processo stocastico $\{X_t\}$ è stazionario in senso forte se per ogni valore in \mathbb{N} di k e h vale*

$$(16.3.1) \quad F_{X_{t_1}, \dots, X_{t_k}}(x_{t_1}, \dots, x_{t_k}) = F_{X_{t_1+h}, \dots, X_{t_k+h}}(x_{t_1+h}, \dots, x_{t_k+h}).$$

In altre parole, un processo è stazionario se la funzione di ripartizione congiunta di ciascuna k -upla di variabili casuali del processo è invariante a traslazioni temporali.

DEFINIZIONE 9 (Stazionarietà debole). *Il processo stocastico $\{X_t\}$ è stazionario in senso debole se per ogni t e $h \in \mathbb{Z}$ valgono*

$$(16.3.2) \quad \mathbb{E}(X_t) = \mu < \infty$$

$$(16.3.3) \quad \text{Cov}(X_t, X_{t-h}) = \gamma_h < \infty.$$

La stazionarietà in senso debole corrisponde quindi alla indipendenza dal tempo della media, della varianza e della covarianza a ritardo h del processo. Naturalmente, se i primi due momenti di un processo esistono, vale l'implicazione

$$\text{Stazionarietà forte} \Rightarrow \text{Stazionarietà debole}.$$

La funzione di h , $\gamma_h = \text{Cov}(X_t, X_{t-h})$ è detta *funzione di autocovarianza* del processo $\{X_t\}$, ed è semplice verificare che valgono le proprietà

$$\begin{aligned} \gamma_h &= \gamma_{-h} \\ -\gamma_0 &\leq \gamma_h \leq \gamma_0 \quad (\text{Disuguaglianza di Cauchy-Schwarz}). \end{aligned}$$

Spesso è più conveniente utilizzare una versione normalizzata della funzione di autocovarianza, ovvero la *funzione di autocorrelazione* $\rho_h = \gamma_h / \gamma_0$, che gode della proprietà di essere compresa nell'intervallo $[-1, 1]$, oltre che $\rho_h = \rho_{-h}$.

Partendo dalle due distinte definizioni di stazionarietà si possono dare due definizioni del processo stocastico più semplice: il *white noise*.

DEFINIZIONE 10 (White noise in senso forte). *Il processo $\{\epsilon_t\}$ a media nulla, varianza finita, e tale che ϵ_t è indipendente da ϵ_s , con $t \neq s$, è detto white noise in senso forte.*

DEFINIZIONE 11 (White noise in senso debole). *Il processo $\{\epsilon_t\}$ a media nulla, varianza finita, e funzioni di autocovarianza*

$$\gamma_h = \begin{cases} \sigma^2 & \text{per } h = 0 \\ 0 & \text{per } h = 1, 2, \dots \end{cases}$$

è detto white noise in senso debole.

Ogni white noise in senso forte è anche white noise in senso debole, mentre in genere non è vera l'implicazione opposta. Tuttavia per una classe importante di white noise, i white noise gaussiani, vale la doppia implicazione.

DEFINIZIONE 12 (White noise gaussiano). *Il processo white noise $\{\varepsilon_t\}$, per cui ogni ε_t ha densità di probabilità normale è detto white noise gaussiano.*

Dato che per variabili casuali normali incorrelazione e indipendenza si implicano vicendevolmente, il white noise gaussiano è contemporaneamente white noise in senso forte ed in senso debole.

Per i processi debolmente stazionari vale un risultato molto potente dovuto a H. Wold.

TEOREMA 12 (di Wold). *Ogni processo debolmente stazionario a media nulla X_t può essere decomposto addittivamente in due componenti tra loro incorrelate*

$$(16.3.4) \quad X_t = D_t + Y_t,$$

tali che D_t è un processo deterministico, cioè prevedibile senza errore sulla base di una funzione lineare del proprio passato, e

$$(16.3.5) \quad Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

con ε_t white noise in senso debole, è una componente puramente stocastica del processo.

Il processo descritto nella formula (16.3.5) è detto *processo lineare generale*, ed è facile verificare che è un processo stocastico (debolmente) stazionario con

$$\begin{aligned} \mathbb{E}(Y_t) &= 0 \\ \gamma_h &= \text{Cov}(Y_t, Y_{t+h}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \end{aligned}$$

dove σ^2 è la varianza del white noise ε_t .

Generalmente per le serie storiche finanziarie stazionarie (o rese tali attraverso opportune trasformazioni) si assume $D_t = 0$. Ovviamente non è possibile condurre inferenza statistica sugli infiniti parametri del processo lineare generale, tuttavia si possono trovare approssimazioni di tale processo parsimoniose per quanto riguarda il numero di parametri da stimare. La classe dei processi ARMA stazionari consente di approssimare il processo lineare generale, per mezzo di modelli con un numero finito (e contenuto) di parametri da stimare.

16.4. Processi ARMA stazionari

In quanto segue si farà ampio uso dell'operatore lineare *ritardo* B , che applicato al processo X_t produce $BX_t = X_{t-1}$, e gode delle medesime regole algebriche delle variabili complesse⁶. Con la simbologia $\alpha_r(B)$ si indicherà il polinomio di r -esimo grado in B , $\alpha_r(B) = 1 - \alpha_1 B - \dots - \alpha_r B^r$ (a volte $\alpha_r(B) = 1 + \alpha_1 B + \dots + \alpha_r B^r$).

Utilizzando la simbologia appena introdotta, il processo lineare generale in equazione (16.3.5) può essere riscritto equivalentemente

$$Y_t = \psi_{\infty}(B)\varepsilon_t = (1 + \psi_1 B + \psi_2 B^2 + \dots)\varepsilon_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

⁶Per esempio, siano a e b due numeri reali (o complessi): $B^r = B^{r-1} \cdot B$, $(a - bB)^2 = a^2 - 2abB + b^2 B^2$, $1 + aB + a^2 B^2 + a^3 B^3 + \dots = (1 - aB)^{-1}$ se $|a| < 1$, ecc.

16.4.1. Processi media mobile. Un primo modo per approssimare il processo lineare generale utilizzando un numero finito di parametri può essere ottenuto troncando il processo stesso a ritardo q :

$$\begin{aligned} Y_t &= \theta_q(B)\varepsilon_t \\ &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}. \end{aligned}$$

Tale processo prende il nome di *processo media mobile* (MA) ed è stazionario (debolmente se il white noise lo è in senso debole, fortemente se il white noise lo è in senso forte) con i primi due momenti dati da

$$\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}) = 0,$$

$$\begin{aligned} \gamma_h &= \mathbb{E}(Y_t Y_{t+h}) = \\ &= \mathbb{E}[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q})(\varepsilon_{t+h} + \theta_1\varepsilon_{t+h-1} + \dots + \theta_q\varepsilon_{t+h-q})] = \\ &= \begin{cases} \sigma^2[1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2] & h = 0 \\ \sigma^2[\theta_h + \theta_{h+1}\theta_1 + \theta_{h+2}\theta_2 + \dots + \theta_q\theta_{q-h}] & h = 1, 2, \dots, q \\ 0 & h = q+1, q+2, \dots \end{cases} \end{aligned}$$

Se le radici dell'equazione di q -esimo grado $\theta_q(z) = 0$ sono tutte, in modulo, maggiori di 1, allora vale l'espansione in serie

$$\theta_q^{-1}(B) = 1 + \pi_1 B + \pi_2 B + \dots,$$

con $\{\pi_i\}_{i=1,2,\dots}$ funzioni note dei coefficienti $\theta_1, \dots, \theta_q$, e il processo MA ammette una rappresentazione autoregressiva di ordine infinito

$$Y_t = \varepsilon_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots$$

ed è detto *invertibile*.

16.4.2. Processi autoregressivi. Il processo lineare generale può essere approssimato anche per mezzo del *processo autoregressivo* (AR) di ordine p

$$\begin{aligned} \phi_p(B)Y_t &= \varepsilon_t \\ (16.4.1) \quad Y_t &= \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \end{aligned}$$

ma solo se l'inversa del polinomio $\phi_p(B)$ esiste finita. Tale condizione, che è anche condizione di stazionarietà per il processo AR, è verificata quando le p radici (reali e/o complesse) dell'equazione caratteristica in z , $\phi_p(z) = 0$, o per esteso

$$1 - \phi_1 z - \dots - \phi_p z^p = 0$$

sono tutte in modulo strettamente maggiori di 1 (o come si suol dire utilizzando la rappresentazione cartesiana dei numeri complessi, esterne al cerchio di raggio unitario). In tal caso per il reciproco del polinomio autoregressivo in B vale l'espansione

$$\phi_p(B)^{-1} = 1 + \pi_1 B + \pi_2 B^2 + \pi_3 B^3 + \dots$$

Per esempio, per il processo AR(1),

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

la radice dell'equazione caratteristica $1 - \phi z = 0$ è $1/\phi$, che è in modulo maggiore di uno per $|\phi| < 1$. Quando tale condizione è rispettata vale la seguente espansione

$$(1 - \phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \dots,$$

da cui segue che il processo gode della rappresentazione MA

$$Y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots$$

Sfruttando la rappresentazione MA di un processo AR stazionario è semplice calcolare il valore medio del processo:

$$\mathbb{E}(Y_t) = \mathbb{E}[\phi_p(B)^{-1} \varepsilon_t] = 0.$$

Le autocovarianze possono essere calcolate moltiplicando l'equazione (16.4.1) per Y_{t-h} e prendendone il valore atteso (si ricordi che $\gamma_{-h} = \gamma_h$),

$$\begin{aligned} \gamma_h &= \mathbb{E}[Y_t Y_{t-h}] \\ &= \mathbb{E}[\phi_1 Y_{t-1} Y_{t-h} + \phi_2 Y_{t-2} Y_{t-h} + \dots + \phi_p Y_{t-p} Y_{t-h} + \varepsilon_t Y_{t-h}] = \\ &= \begin{cases} \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2 & \text{per } h = 0 \\ \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2} + \dots + \phi_p \gamma_{h-p} & \text{per } h = 1, 2, \dots \end{cases} \end{aligned}$$

16.4.3. Processi ARMA. Un migliore approssimazione del processo lineare generale può essere ottenuta combinando i due processi stazionari visti in un unico *processo autoregressivo a media mobile* di ordine (p, q) , o più in breve ARMA(p, q),

$$(16.4.2) \quad Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

o più compattamente

$$\phi_p(B)Y_t = \theta_q(B)\varepsilon_t$$

Il processo ARMA è stazionario quando le radici dell'equazione caratteristica della parte AR, $\phi_p(z) = 0$, sono tutte in modulo maggiori di uno.

Il momento primo di un processo ARMA stazionario può essere facilmente ricavato prendendo il valore atteso della rappresentazione puramente MA del processo stesso

$$\mathbb{E}(Y_t) = \mathbb{E}[\phi(B)^{-1} \theta(B) \varepsilon_t] = 0,$$

mentre la funzione di autocovarianza può essere calcolata moltiplicando entrambi i membri della equazione (16.4.2) per Y_{t-h} e prendendone il valore atteso. L'andamento delle autocorrelazioni per un processo misto è più complicata rispetto a quella di un processo puramente AR o MA.

Naturalmente è possibile includere una media non nulla nella definizione di un processo ARMA. Un processo ARMA con media μ può essere rappresentato come

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + \dots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

o alternativamente come

$$Y_t = \nu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

dove $\mu = \mathbb{E}(Y_t) = \phi(1)^{-1} \nu$ e quindi $\nu = \phi(1)\mu$.

16.4.4. Previsione di un processo ARMA. Dato un processo Y_t , un aspetto molto rilevante nell'analisi delle serie storiche è fare inferenza sulla futura realizzazione della serie storica (previsione), utilizzando l'informazione disponibile in un dato istante. Si supponga di aver osservato la realizzazione del processo (serie storica) (y_1, y_2, \dots, y_t) , e si indichi con \mathcal{F}_t l'informazione disponibile al tempo t (cioè la filtrazione rispetto a cui Y_t è adattato). In quanto segue si supporrà che l'unica informazione rilevante per la formazione di previsioni sulle future realizzazioni del processo sia la serie storica stessa. Il previsore della serie al tempo $T+h$, con $h \in \mathbb{N}$, sarà quindi una funzione $f_{T+h}(\mathcal{F}_T)$ dell'informazione disponibile al tempo T . Per individuare il previsore "ottimale" è necessario stabilire una funzione di perdita (o costo) opportuna, per poi cercare la forma funzionale del previsore che ne minimizza il valore atteso (perdita attesa). Il previsore che minimizza una certa funzione di perdita è detto previsore ottimale rispetto a quella funzione di perdita. La funzione di perdita più comunemente utilizzata è l'errore quadratico medio di previsione (**MSE**, *Mean Square Error*) definito da

$$\text{MSE}(Y_{T+h} - f_{T+h}(\mathcal{F}_T)) = \mathbb{E}[(f_{T+h}(\mathcal{F}_T) - Y_{T+h})^2],$$

che, come si è già visto, è minimo quando il previsore è il valore atteso condizionato del processo

$$f_{T+h}(\mathcal{F}_T) = \mathbb{E}(Y_{T+h} | \mathcal{F}_T).$$

Si supponga per il momento di conoscere l'infinito passato della serie storica, da cui $\mathcal{F}_T = \sigma(Y_T, Y_{T-1}, Y_{T-2}, \dots)$. Per un processo ARMA la previsione un periodo in avanti ($h=1$) è data da

$$\hat{Y}_{T+1|T} = \mathbb{E}(Y_{T+1} | \mathcal{F}_T) = \nu + \phi_1 Y_T + \dots + \phi_p Y_{T-p+1} + \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T-q+1},$$

infatti le innovazioni $\varepsilon_T, \varepsilon_{T-1}, \dots$ sono \mathcal{F}_t -misurabili, essendo funzioni lineari di Y_t, Y_{t-1}, \dots . Il MSE di previsione è dato da

$$\mathbb{E}(Y_T - \hat{Y}_{T|T-1})^2 = \mathbb{E}(\varepsilon_T^2) = \sigma^2.$$

Per ottenere le previsioni $h=2$ periodi in avanti si sfrutta la legge dei valori attesi iterati

$$\begin{aligned} \hat{Y}_{T+2} &= \mathbb{E}(Y_{T+2} | \mathcal{F}_T) = \\ &= \mathbb{E}[\mathbb{E}_{Y_{T+2}}(Y_{T+2} | \mathcal{F}_{T+1}) | \mathcal{F}_T] = \\ &= \mathbb{E}[\nu + \phi_1 Y_{T+1} + \phi_2 Y_T + \dots + \phi_p Y_{T+2-p} + \\ &\quad + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T + \dots + \theta_q \varepsilon_{T+2-q} | \mathcal{F}_T] = \\ &= \nu + \phi_1 \hat{Y}_{T+1|T} + \phi_2 Y_T + \dots + \phi_p Y_{T-p+2} + \theta_2 \varepsilon_T + \dots + \theta_q \varepsilon_{T+2-q}. \end{aligned}$$

In generale, utilizzando la stessa tecnica per un generico h positivo si perviene a

$$(16.4.3) \quad \hat{Y}_{T+h|T} = \nu + \phi_1 \hat{Y}_{T+h-1|T} + \dots + \phi_p \hat{Y}_{T+h-p|T} + \theta_1 \hat{\varepsilon}_{T+h-1|T} + \dots + \theta_q \hat{\varepsilon}_{T+h-q|T},$$

dove l'attesa condizionata del processo innovazione è banalmente

$$\hat{\varepsilon}_{t|T} := \mathbb{E}[\varepsilon_t | \mathcal{F}_T] = \begin{cases} \varepsilon_t, & \text{per } t \leq T; \\ 0, & \text{t.c.T} \end{cases}.$$

ESERCIZIO 3. *Mostrare che per processi ARMA stazionari la previsione converge alla media del processo per $h \rightarrow \infty$.*

Per calcolare l'errore di previsione è conveniente utilizzare la rappresentazione puramente MA del processo ARMA,

$$Y_{T+h} = \mu + \varepsilon_{T+h} + \sum_{i=1}^{\infty} \psi_i \varepsilon_{T+h-i}.$$

Applicando la previsione in formula (16.4.3) alla forma puramente MA, si ottiene

$$\hat{Y}_{T+h|T} = \mu + \sum_{i=h}^{\infty} \psi_i \varepsilon_{T+h-i},$$

da cui si evince che l'errore di previsione è dato da

$$Y_{T+h} - \hat{Y}_{T+h|T} = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1}.$$

Il valore atteso dell'errore di previsione è quindi nullo per ogni h , mentre la sua varianza vale

$$\mathbb{E}[(Y_{T+h} - \hat{Y}_{T+h|T})^2] = \sigma^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2),$$

e cresce con l'aumentare di h fino a convergere alla varianza del processo

$$\lim_{h \rightarrow \infty} \mathbb{E}(\hat{Y}_{T+h|T} - Y_{T+h}) = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \dots).$$

Se, invece, solo una realizzazione finita di una traiettoria è disponibile (come nella realtà d'altronde), ossia quando $\mathcal{F}_{1:t} = \sigma(Y_t, \dots, Y_1)$, calcolare il valore atteso condizionato del processo non è così semplice quando una componente MA è presente nel processo. Infatti, mentre il valore atteso condizionato di un AR(p) rimane, per $t > p$,

$$\mathbb{E}(Y_t | \mathcal{F}_{1:t-1}) = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$$

con MSE pari a σ^2 , ε_t non è più misurabile rispetto alla σ -algebra generata dal presente e finito passato di Y_t . Infatti usando la rappresentazione AR(∞) di un processo MA, ci si accorge che ε_t è funzione lineare di tutta la storia passata di Y_t , mentre $\mathcal{F}_{1:t}$ contiene solamente informazioni sugli ultimi t periodi del processo.

In tal caso ci si può limitare a prevedere per mezzo alla proiezione lineare di Y_{T+h} sul proprio passato noto al tempo T .

$$(16.4.4) \quad \hat{Y}_{T+h|T} = \mathbb{P}[Y_{T+h} | Y_T, \dots, Y_1] = \mu + \gamma'_{T+h} \mathbf{\Gamma}_T^{-1} (\mathbf{Y} - \boldsymbol{\mu}),$$

dove

$$\boldsymbol{\gamma}_{T+h} := [\text{Cov}(Y_{T+h}, Y_T), \text{Cov}(Y_{T+h}, Y_{T-1}), \dots, \text{Cov}(Y_{T+h}, Y_1)]'$$

e $\mathbf{\Gamma}_T = \{\gamma_{|j-i|}\}_{i,j \in \{1, \dots, T\}}$ è la matrice di autocovarianza del vettore $\mathbf{Y} = [Y_T, \dots, Y_1]'$; lo scalare μ e il vettore $\boldsymbol{\mu}$ contengono la media (ripetuta) del processo Y_t . Si rammenti che per processi ARMA gaussiani, la proiezione lineare coincide con il valore atteso condizionato, e quindi con il previsore ottimale rispetto al MSE, che è pari a

$$(16.4.5) \quad v_{T+h|T} := \gamma_0 - \gamma'_{T+h} \mathbf{\Gamma}_T^{-1} \boldsymbol{\gamma}_{T+h}.$$

L'inversione della matrice $\mathbf{\Gamma}_T$, che dipende in maniera non banale dai coefficienti del processo ARMA è un'operazione onerosa anche per un calcolatore, specialmente quando il numero di osservazioni T è grande. Il problema viene in genere risolto numericamente per mezzo di alcuni algoritmi quali l'*algoritmo di Durbin-Levinson*, l'*algoritmo delle innovazioni* o il *filtro di Kalman*.

Tuttavia, un semplice modo per avere una soluzione approssimata al problema della previsione è la sostituzione nella (16.4.3) della serie delle innovazioni ε_t con la sua approssimazione e_t , calcolata come segue:

$$e_t = \begin{cases} 0, & \text{per } t = 1, 2, \dots, \max(p, q); \\ Y_t - \nu - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, & \text{altrementi.} \end{cases}$$

Il MSE di previsione viene approssimato invece con la varianza delle innovazioni σ^2 . È semplice dimostrare che se il processo ARMA Y_t è stazionario e invertibile, allora

$$\mathbb{E}(\varepsilon_t - e_t)^2 \rightarrow 0, \quad \text{per } t \rightarrow \infty,$$

e il MSE della previsione approssimata converge a quello della previsione basata sull'infinito passato, ovvero σ^2 . In altre parole, se t è sufficientemente grande, la distanza tra la stima delle innovazioni e le vere innovazioni del processo sarà piccola e la stessa cosa avviene tra la distanza delle previsioni approssimate e quelle basate sull'infinito passato.

ESEMPIO 3 (Approssimazione delle innovazioni di un MA(1) sulla base di una traiettoria finita). *Applichiamo quanto appena visto al processo $Y_t = \varepsilon_t - \theta\varepsilon_{t-1}$ per una traiettoria che ha origine in $t = 1$ ⁷. Ponendo $e_0 = 0$ e sostituendo ricorsivamente, si ha*

$$e_t = Y_t + \theta e_{t-1} = \sum_{i=0}^{t-1} \theta^i Y_{t-i}$$

per il processo delle innovazioni approssimate. Compiendo le medesime sostituzioni ricorsive avendo a disposizione il processo delle innovazioni ε_t si ottiene invece

$$\varepsilon_t = Y_t + \theta\varepsilon_{t-1} = \sum_{i=0}^{t-1} \theta^i Y_{t-i} + \theta^t \varepsilon_0$$

Pertanto la differenza tra i due processi si riduce a

$$\varepsilon_t - e_t = \theta^t \varepsilon_0,$$

da cui $\mathbb{E}(\varepsilon_t - e_t)^2 = \sigma^2 \theta^{2t}$, che, per $|\theta| < 1$ (condizione di invertibilità) converge a 0 al crescere di t .

L'approssimazione risulta essere anche una stima corretta di ε_t , infatti

$$\mathbb{E}(\varepsilon_t - e_t) = \mathbb{E}(\theta^t \varepsilon_0) = 0.$$

La varianza di e_t è invece maggiore di quella di ε_t , includendo l'incertezza sul valore iniziale, ma anch'essa converge a velocità geometrica a σ^2 :

$$\text{Var}(e_t) = \text{Var}(\varepsilon_t + (e_t - \varepsilon_t)) = \text{Var}(\varepsilon_t + \theta^t \varepsilon_0) = \sigma^2(1 + \theta^{2t}).$$

Pertanto, se il processo ARMA è invertibile e t è sufficientemente grande, i risultati ottenuti per previsioni basate sull'infinito passato rimangono approssimativamente validi anche quando si utilizzi il processo e_t in luogo di ε_t .

16.5. Stima di processi ARMA

16.5.1. Stima ai minimi quadrati condizionati di processi puramente autoregressivi. Si supponga di volere stimare i parametri incogniti del modello AR(p)

$$(16.5.1) \quad Y_t = \nu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t,$$

con $\{\varepsilon_t\}$ processo differenza di martingala adattato alla filtrazione generata da Y_t, Y_{t-1}, \dots , non necessariamente gaussiano. L'aspetto della (16.5.1) è quello di una regressione lineare con regressori stocastici (essendo

⁷Si noti che, senza perdita di generalità, ma per semplicità nei conti si è parametrizzato il coefficiente MA per mezzo di $-\theta$ in luogo di θ .

realizzazioni passate del processo stesso). Perché allora non chiedersi come si comporta lo stimatore ai minimi quadrati (d'ora in poi LS) di $\beta = (\nu, \phi_1, \dots, \phi_p)'$? Si ponga $\mathbf{X}_t = (1, Y_{t-1}, \dots, Y_{t-p})'$, in modo tale da poter riscrivere la (16.5.1) come

$$(16.5.2) \quad Y_t = \mathbf{X}_t' \beta + \varepsilon_t.$$

Supponendo di avere p osservazioni $(Y_{t-p+1}, Y_{t-p+2}, \dots, Y_0)$ pre-campionarie (per comodità notazionale stiamo semplicemente supponendo di avere $T+p$), e minimizzando la somma dei quadrati degli errori di regressione (nel nostro caso errori di previsione un periodo in avanti)

$$\sum_{t=1}^T (Y_t - \mathbf{X}_t' \beta)^2 = \min_{\beta},$$

si ottiene

$$\hat{\beta}_T = \left[\sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right]^{-1} \left[\sum_{t=1}^T (\mathbf{X}_t Y_t) \right],$$

che, sostituendo Y_t con $\mathbf{X}_t' \beta + \varepsilon_t$, diventa

$$(16.5.3) \quad \hat{\beta}_T = \beta + \left[\sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right]^{-1} \left[\sum_{t=1}^T (\mathbf{X}_t \varepsilon_t) \right].$$

Prendendo il valore atteso di entrambe le parti della precedente equazione, si nota che $\hat{\beta}_T$ non è uno stimatore corretto dato che

$$\mathbb{E} \left\{ \left[\sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right]^{-1} \left[\sum_{t=1}^T (\mathbf{X}_t \varepsilon_t) \right] \right\}$$

non è, in genere, nullo. Le proprietà di $\hat{\beta}_T$ per campioni finiti non sono facilmente determinabili, vediamo quindi come si comporta asintoticamente

$$(16.5.4) \quad \sqrt{T}(\hat{\beta}_T - \beta) = \left[(1/T) \sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right]^{-1} \left[(1/\sqrt{T}) \sum_{t=1}^T (\mathbf{X}_t \varepsilon_t) \right].$$

Il primo fattore a destra dell'uguale è

$$\left[(1/T) \sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right] = \begin{pmatrix} 1 & (1/T) \sum Y_{t-1} & (1/T) \sum Y_{t-2} & \dots & (1/T) \sum Y_{t-p} \\ (1/T) \sum Y_{t-1} & (1/T) \sum Y_{t-1}^2 & (1/T) \sum Y_{t-1} Y_{t-2} & \dots & (1/T) \sum Y_{t-1} Y_{t-p} \\ (1/T) \sum Y_{t-2} & (1/T) \sum Y_{t-2} Y_{t-1} & (1/T) \sum Y_{t-2}^2 & \dots & (1/T) \sum Y_{t-2} Y_{t-p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ (1/T) \sum Y_{t-p} & (1/T) \sum Y_{t-p} Y_{t-1} & (1/T) \sum Y_{t-p} Y_{t-2} & \dots & (1/T) \sum Y_{t-p}^2 \end{pmatrix}.$$

Per gli elementi della prima riga e colonna vale, al divergere di T ⁸

$$\frac{1}{T} \sum_{t=1}^T Y_{t-j} \xrightarrow{p} \mathbb{E}(Y_{t-j}) = \mu,$$

⁸La convergenza in probabilità è garantita dall'ergodicità dei processi AR stazionari con distribuzione assolutamente continua.

mentre per gli altri elementi della matrice

$$\frac{1}{T} \sum_{t=1}^T Y_{t-i} Y_{t-j} = \mathbb{E}(Y_{t-i} Y_{t-j}) \xrightarrow{p} \gamma_{|i-j|} + \mu^2.$$

Pertanto, applicando il teorema delle funzioni continue all'inversa (che è funzione continua dato che la matrice in oggetto è invertibile con probabilità 1)

$$(16.5.5) \quad \left[\frac{1}{T} \sum_{t=1}^T (\mathbf{X}_t \mathbf{X}_t') \right]^{-1} \xrightarrow{p} \mathbf{Q}^{-1},$$

con

$$\mathbf{Q} = \begin{pmatrix} 1 & \mu & \mu & \dots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \dots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \dots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \dots & \gamma_0 + \mu^2 \end{pmatrix}.$$

Per quanto riguarda il secondo fattore della (16.5.4), si può facilmente vedere che $\{\mathbf{X}_t \varepsilon_t\}$ è un processo differenza di martingala, infatti, ponendo $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, Y_{t-2}, \dots)$, per ogni t vale

$$\mathbb{E}(\mathbf{X}_t \varepsilon_t | \mathcal{F}_{t-1}) = \mathbf{X}_t \mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0.$$

Notando che la matrice di covarianza di $\mathbf{X}_t \varepsilon_t$ è

$$\mathbb{E}(\mathbf{X}_t \varepsilon_t \varepsilon_t' \mathbf{X}_t') = \mathbb{E}(\varepsilon_t^2) \mathbb{E}(\mathbf{X}_t \mathbf{X}_t') = \sigma^2 \mathbf{Q},$$

e sfruttando il teorema centrale del limite per differenze di martingala (8) si ricava

$$(16.5.6) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^T (\mathbf{X}_t \varepsilon_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}).$$

Mettendo assieme i risultati (16.5.5) e (16.5.6) si ottiene

$$(16.5.7) \quad \sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}).$$

ESERCIZIO 4. Si mostri che lo stimatore ai minimi quadrati del coefficiente autoregressivo di un $AR(1)$ senza costante converge in distribuzione a una normale con varianza $(1 - \phi^2)$.

16.5.2. Stima di massima verosimiglianza gaussiana di processi ARMA e sue approssimazioni. Se y_1, \dots, y_T è una traiettoria campionaria finita di un processo ARMA gaussiano, la sua funzione di verosimiglianza è data dalla densità congiunta vista come funzione dei parametri incogniti

$$(16.5.8) \quad \mathcal{L}(\psi) = f_\psi(y_1, \dots, y_T) = \frac{1}{(2\pi)^{T/2} |\mathbf{\Gamma}_T|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \mathbf{\Gamma}_T^{-1} (\mathbf{y} - \boldsymbol{\mu}), \right\}$$

dove $\psi = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$ è il vettore dei parametri del modello, $\boldsymbol{\mu}$ è un vettore con i T elementi uguali alla media del processo μ e $\mathbf{\Gamma}_T$ è la matrice di autocovarianza del processo ARMA che è funzione unicamente dai coefficienti del modello.

Come si è già notato nella sezione delle previsioni, la costruzione e l'inversione di $\mathbf{\Gamma}_T$ per un generico modello ARMA è una operazione onerosa anche per un calcolatore, in particolare quando T è grande. Pertanto si preferisce utilizzare la fattorizzazione della densità congiunta dei dati nel prodotto delle densità condizionate

$$f_\psi(y_1, \dots, y_T) = f_\psi(y_1) \prod_{t=2}^T f_\psi(y_t | y_1, \dots, y_{t-1}).$$

Si noti che nel caso gaussiano la media e la varianza di ciascuna distribuzione condizionata sono date dalla proiezione (16.4.4) e dalla sua varianza (16.4.5), da cui possiamo scrivere

$$f_{\psi}(y_t|y_1, \dots, y_{t-1}) = \frac{1}{\sqrt{2\pi v_{t|t-1}}} \exp \left\{ -\frac{1}{2} \frac{(y_t - \hat{y}_{t|t-1})^2}{v_{t|t-1}} \right\}.$$

Come si è già detto per le previsioni, esistono diversi algoritmi che forniscono le sequenze $\hat{y}_{t|t-1}$ e $v_{t|t-1}$. La distribuzione iniziale $f_{\psi}(y_1)$ è semplicemente la densità marginale

$$f_{\psi}(y_1) = \frac{1}{\sqrt{2\pi\gamma_0}} \exp \left\{ -\frac{1}{2} \frac{(y_1 - \mu)^2}{\gamma_0} \right\}.$$

Gli stimatori di massima verosimiglianza (ML) dei parametri ψ sono quei valori che risolvono

$$\hat{\psi}_T := \arg \max_{\psi} \mathcal{L}(\psi)$$

o, equivalentemente

$$\hat{\psi}_T := \arg \max_{\psi} \ell(\psi)$$

dove $\ell(\psi) := \log \mathcal{L}(\psi)$ è la funzione di *log-verosimiglianza*.

Per un processo ARMA gaussiano stazionario e invertibile gli stimatori di massima verosimiglianza sono consistenti, e asintoticamente normali:

$$\sqrt{T}(\hat{\psi}_T - \psi_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1}(\psi_0)),$$

dove ψ_0 è il vettore dei veri parametri e

$$\mathbf{I}(\psi_0) = -\mathbb{E} \left[\frac{\partial^2}{\partial \psi \partial \psi'} \log f_{\psi}(Y_t|Y_{t-1}, Y_{t-2}, \dots) \right]_{\psi=\psi_0}$$

è la matrice d'informazione di Fisher, che può essere stimata consistentemente se al valore atteso si sostituisce la media campionaria e al valore vero ψ_0 la sua stima $\hat{\psi}_T$:

$$\begin{aligned} \hat{\mathbf{I}}_T(\hat{\psi}_T) &= -\frac{1}{T} \sum_{t=1}^n \left[\frac{\partial^2}{\partial \psi \partial \psi'} \log f_{\psi}(y_t|y_{t-1}, y_{t-2}, \dots) \right]_{\psi=\hat{\psi}_T} \\ &= -\frac{1}{T} \frac{\partial^2}{\partial \psi \partial \psi'} \left[\sum_{t=1}^n \log f_{\psi}(y_t|y_{t-1}, y_{t-2}, \dots) \right]_{\psi=\hat{\psi}_T} \\ &= -\frac{1}{T} \frac{\partial^2}{\partial \psi \partial \psi'} [\ell(\psi)]_{\psi=\hat{\psi}_T}. \end{aligned}$$

Si può dimostrare che $\hat{\mathbf{I}}_T(\hat{\psi}_T) \xrightarrow{p} \mathbf{I}(\psi_0)$.

Pertanto, se T è sufficientemente grande si può approssimare la distribuzione dello stimatore di massima verosimiglianza con

$$\hat{\psi}_T \approx \mathcal{N}(\psi_0, \hat{\mathbf{I}}_T^{-1}(\hat{\psi}_T)/T).$$

Utili generalizzazioni della stima di massima verosimiglianza verranno trattate approfonditamente più avanti, in un apposito capitolo.

Uno stimatore alternativo alla massima verosimiglianza, più semplice da calcolare, ma asintoticamente equivalente allo stimatore ML è lo stimatore ai minimi quadrati condizionati CLS (conditional least squares).

Lo stimatore CLS è dato da quei valori dei coefficienti ARMA che risolvono il problema di minimo:

$$S(\psi) := \sum_{t=p+1}^T e_t^2 = \min_{\psi},$$

dove $\psi = (\nu, \psi_1, \dots, \psi_p, \theta_1, \dots, \theta_q)$ è il vettore dei coefficienti (si noti l'assenza della varianza σ^2 e la presenza della costante ν in luogo della media) mentre e_t è il processo delle innovazioni approssimate, già utilizzato nella sezione sulle previsioni,

$$e_t = \begin{cases} 0, & \text{per } t = 1, 2, \dots, \max(p, q); \\ Y_t - \nu - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, & \text{altreimenti.} \end{cases}$$

La varianza delle innovazioni può essere stimata con

$$s_T^2 = \frac{S(\tilde{\psi}_T)}{n - 2p - q - 1}.$$

dove si è indicato con $\tilde{\psi}_T$ il vettore delle stime CLS.

Nel caso vi sia una componente MA, sia la stima ML, sia quella CLS sono da ottenere per mezzo di algoritmi di ottimizzazione numerica, che approssimano il massimo (o minimo) di una funzione, partendo da un punto dato e usando l'informazione contenuta nel gradiente e, a volte, nell'Hessiana cercano di trovare punti sempre più alti (o più bassi) della funzione. Se il valore iniziale dei parametri passato all'ottimizzatore è vicino a quello che massimizza (o minimizza) la funzione obiettivo, allora la procedura sarà in genere più rapida e vi saranno meno problemi legati alla presenza di massimi (minimi) locali⁹.

Un modo semplice ed efficace per ottenere valori iniziali dei parametri in ψ è il seguente.

- (1) Si stimi con i minimi quadrati un modello puramente autoregressivo di ordine k sufficientemente grande e si calcolino gli errori di (auto)regressione $e_t = y_t - \hat{y}_t$. In particolare, deve valere $k > p + q$ (per esempio si può fissare $k = \max(p + q + 1, \sqrt{T})$).
- (2) Si regredisca la serie y_t su p ritardi di se stessa e su q ritardi di e_t .
- (3) Si usino i coefficienti dei ritardi di y_t come pre-stima dei coefficienti autoregressivi, ϕ_i , quelli dei ritardi di e_t come stima preliminare dei coefficienti MA, θ_j , e, se necessario (stime ML) si utilizzi la varianza dell'errore di regressione per inizializzare σ^2 .

Le pre-stime descritte sfruttano il fatto che un modello AR(k) può essere usato per approssimare un modello ARMA stazionario e invertibile arbitrariamente bene, al crescere di k e, pertanto, una prima stima delle innovazioni può essere ottenuta dai residui del modello AR. Dato che la stima LS di un modello lineare esiste in forma chiusa, il tempo di calcolo di una regressione lineare col metodo LS è enormemente più breve di quello richiesto da un ottimizzatore numerico.

16.6. Processi integrati

Come si è visto in precedenza un processo ARMA è stazionario solo se tutte le radici dell'equazione caratteristica $\phi(z) = 0$ sono in modulo maggiori di uno. I processi ARMA che non rispettano tali proprietà possono avere almeno una radice (in modulo) uguale all'unità e/o avere almeno una radice minore (in modulo) di uno. Questi ultimi processi non rivestono un grosso interesse nello studio di serie storiche economiche e finanziarie in quanto esplosivi. Si consideri, per esempio, il processo

$$Y_t = 1.1Y_{t-1} + \varepsilon_t$$

⁹Si può dimostrare che nel caso di processi ARMA stazionari e invertibili la funzione ottimizzanda, sia essa la verosimiglianza o la somma dei quadrati condizionata, possiede unico punto di ottimo.

con valore iniziale $Y_0 = y_0$. La radice di tale processo è $z_1 = 1.1^{-1}$, in modulo minore di uno. Il valore atteso al tempo t di tale processo è

$$\begin{aligned}\mathbb{E}(Y_1) &= 1.1y_0 \\ \mathbb{E}(Y_2) &= 1.1^2y_0 \\ &\dots \quad \dots \\ \mathbb{E}(Y_t) &= 1.1^ty_0\end{aligned}$$

che cresce con un ritmo esponenziale: per $t = 100$ e $y_0 = 1$ si ha $\mathbb{E}(Y_{100}) = 13780.6$.

Più interessanti sono invece i processi con una o più radici unitarie (radici che, in modulo, sono esattamente uguali a uno). Si consideri il processo

$$(16.6.1) \quad \varphi_{p+d}(B)Y_t = \theta_q(B)\varepsilon_t,$$

con $\varphi_{p+d}(B) = 1 - \varphi_1B - \varphi_2B^2 - \dots - \varphi_{p+d}B^{p+d}$, polinomio con p radici esterne al cerchio unitario e d radici unitarie, allora, per note regole algebriche, è possibile fattorizzare $\varphi_{p+d}(B)$ in

$$\varphi_{p+d}(B) = \phi_p(B)(1 - B)^d,$$

dove il polinomio $\phi_p(B)$ contiene solo radici esterne al cerchio unitario. Il processo (16.6.1), più comunemente scritto fattorizzato

$$(16.6.2) \quad \phi_p(B)(1 - B)^dY_t = \theta_q(B)\varepsilon_t,$$

prende il nome di processo autoregressivo integrato media mobile, o più in breve $\text{ARIMA}(p, d, q)$. Inoltre, qualunque processo non stazionario, la cui differenza d -esima sia stazionaria è detto *processo integrato di ordine d* . Il processo $\text{ARIMA}(p, d, q)$ è evidentemente un processo integrato di ordine d , dato che il processo $W_t = (1 - B)^dY_t$ è un processo $\text{ARMA}(p, q)$ stazionario.

16.6.1. Il processo *random walk*. Il processo integrato di ordine uno più semplice è

$$(16.6.3) \quad Y_t = Y_{t-1} + \varepsilon_t,$$

detto *random walk* (RW), detto anche, in Italiano, *passeggiata aleatoria*. In realtà, in letteratura si trovano versioni diverse di RW a seconda di come viene definito il processo ε_t . Per adesso si supporrà che ε_t sia un white noise in senso debole con varianza σ^2 .

La definizione in equazione (16.6.3) deve essere completata con la fissazione di un valore iniziale y_0 , dato che ciò che avviene durante il sentiero di un processo integrato, viene “ricordato” per sempre dal processo stesso, ossia ogni singolo *shock* influenza tutte le realizzazioni future del processo.

Sostituendo a se stessa, ricorsivamente, l'equazione (16.6.3) si può scrivere il RW nella forma

$$(16.6.4) \quad Y_t = y_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t.$$

In questo modo si è reso evidente come ogni realizzazione del white noise, influenzi definitivamente il livello del processo. Utilizzando la (16.6.4) è facile calcolare i primi due momenti non condizionati del RW al tempo t :

$$\begin{aligned}\mathbb{E}(Y_t) &= y_0 \\ \mathbb{E}[(Y_t - y_0)^2] &= t\sigma^2\end{aligned}$$

La previsione che minimizza il MSFE è il momento primo condizionato

$$\hat{y}_{T+h} = \mathbb{E}(Y_{T+h}|y_T, y_{T-1}, \dots) = \mathbb{E}(Y_{T+h}|y_T) = y_T,$$

mentre, essendo il RW riscrivibile come $Y_{T+h} = Y_T + \varepsilon_{T+1} + \varepsilon_{T+2} + \dots + \varepsilon_{T+h}$, l'errore di previsione è

$$Y_{T+h} - \hat{y}_{T+h} = \varepsilon_{T+1} + \dots + \varepsilon_{T+h},$$

e quindi l'errore atteso di previsione e la varianza di previsione sono

$$\begin{aligned} \mathbb{E}(Y_{T+h} - \hat{y}_{T+h}) &= 0 \\ \mathbb{E}[(Y_{T+h} - \hat{y}_{T+h})^2] &= h\sigma^2. \end{aligned}$$

La varianza di previsione diverge al crescere dell'intervallo di previsione.

Una generalizzazione interessante del RW è il processo *random walk with drift* (RWD), o passeggiata aleatoria con deriva, definito da

$$(16.6.5) \quad Y_t = \delta + Y_{t-1} + \varepsilon_t,$$

con valore iniziale y_0 . Sostituendo ricorsivamente a se stessa la (16.6.5) si ottiene la seguente rappresentazione del RWD

$$(16.6.6) \quad Y_t = y_0 + \delta t + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t.$$

I primi due momenti del processo sono quindi dati da

$$\begin{aligned} \mathbb{E}(Y_t) &= y_0 + \delta t \\ \mathbb{E}[(Y_t - y_0 - \delta t)^2] &= t\sigma^2. \end{aligned}$$

e la previsione che minimizza il MSFE è

$$\mathbb{E}(Y_{T+h}|y_T, y_{T-1}, \dots) = \mathbb{E}(Y_{T+h}|y_T) = y_T + \delta h.$$

L'errore di previsione è identico a quello visto in precedenza per il RW, e quindi anche i relativi valore atteso e varianza.

La differenza prima di un processo RW è un processo white noise,

$$Y_t - Y_{t-1} = \varepsilon_t,$$

mentre la differenza prima di un processo RWD è un processo stazionario auto-incorrelato (i.i.d., se ε_t è white noise in senso forte) con media δ e varianza σ^2 ,

$$Y_t - Y_{t-1} = \delta + \varepsilon_t.$$

16.6.2. Test di radice unitaria. Per mettere sotto verifica l'ipotesi di presenza di una radice unitaria in un processo ARMA è stata proposta una grossa quantità di test. Tuttavia i test che tuttora riscuotono più successo sono i primi apparsi in letteratura, ossia il *test Augmented Dickey-Fuller* (ADF) ed il test di Phillips-Perron. Qui si illustrerà in breve solamente il test ADF.

16.6.2.1. *Comportamento dello stimatore ai minimi quadrati di un AR(1) in presenza di radice unitaria.*

La forma dello stimatore ai minimi quadrati del parametro ϕ di un AR(1)

$$(16.6.7) \quad Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

è,

$$\hat{\phi}_T = \frac{\sum_{t=1}^T Y_t Y_{t-1}}{\sum_{t=1}^T Y_{t-1}^2}.$$

Sfruttando risultati noti della stima ai minimi quadrati, si ha anche

$$\hat{\phi}_T - \phi = \frac{\sum_{t=1}^T Y_{t-1} \varepsilon_t}{\sum_{t=1}^T Y_{t-1}^2},$$

che, quando $|\phi| < 1$, ha distribuzione asintotica normale:

$$\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{D} \mathcal{N}(0, \sigma^2 / \mathbb{E}(Y_{t-1}^2)).$$

Poiché

$$\mathbb{E}(Y_{t-1}^2) = \mathbb{E}\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1-i}\right)^2 = \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \sigma^2 / (1 - \phi^2),$$

allora

$$(16.6.8) \quad \sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{D} \mathcal{N}(0, (1 - \phi^2)).$$

Normalmente, sfruttando tale risultato si può costruire un test di significatività per ipotesi su ϕ ; tuttavia quando si vuole verificare l'ipotesi $H_0 : \phi = 1$, la varianza in (16.6.8) è nulla. Per far convergere $\hat{\phi}_T - 1$ in distribuzione (e non in probabilità), è necessario moltiplicarlo per T , piuttosto che per \sqrt{T} . Per vedere perchè, si indagheranno le proprietà distribuzionali del numeratore e del denominatore del rapporto

$$(16.6.9) \quad \hat{\phi}_T - 1 = \frac{\sum_{t=1}^T Y_{t-1} \varepsilon_t}{\sum_{t=1}^T Y_{t-1}^2}.$$

Quando $\phi = 1$ il processo (16.6.7) è un RW, che come già visto può essere scritto come

$$Y_t = \sum_{s=1}^t \varepsilon_s,$$

e nel caso di white noise gaussiano, che si sta ora considerando, ha distribuzione $\mathcal{N}(0, \sigma^2 t)$. Elevando al quadrato la (16.6.7), per $\phi = 1$ si ottiene

$$Y_t^2 = Y_{t-1}^2 + 2Y_{t-1}\varepsilon_t + \varepsilon_t^2,$$

che risolvendo per $Y_{t-1}\varepsilon_t$ fornisce il risultato

$$Y_{t-1}\varepsilon_t = \frac{1}{2}(Y_t^2 - Y_{t-1}^2 - \varepsilon_t^2).$$

Quindi il numeratore della (16.6.9) può essere riscritto nella forma

$$\sum_{t=1}^T Y_{t-1}\varepsilon_t = \frac{1}{2}(Y_T^2 - y_0^2) - \frac{1}{2} \sum_{t=1}^T \varepsilon_t^2.$$

Si supponga ora per comodità, e senza perdere di generalità dei risultati, che $y_0 = 0$, e si divida l'ultima espressione per $\sigma^2 T$:

$$\frac{\sum_{t=1}^T Y_{t-1}\varepsilon_t}{\sigma^2 T} = \frac{1}{2} \left(\frac{Y_T}{\sigma\sqrt{T}} \right)^2 - \frac{1}{2\sigma^2} \frac{\sum_{t=1}^T \varepsilon_t^2}{T}.$$

$Y_T/(\sigma\sqrt{T})$ è una variabile casuale normale standardizzata, e quindi il suo quadrato è una variabile casuale χ_1^2 , mentre $T^{-1} \sum_{t=1}^T \varepsilon_t^2$ converge in probabilità a σ^2 . Mettendo insieme i risultati, si ha che il numeratore della (16.6.9) ha distribuzione asintotica, al divergere di T ,

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T Y_{t-1} \varepsilon_t \xrightarrow{D} \frac{1}{2}(X - 1),$$

dove X è una variabile casuale chi-quadrato con un grado di libertà.

Si prenda ora in considerazione il denominatore del rapporto (16.6.9). Dato che $Y_{t-1} \sim \mathcal{N}(0, \sigma^2(t-1))$, allora il valore atteso del denominatore è

$$\mathbb{E} \left(\sum_{t=1}^T Y_{t-1}^2 \right) = \sigma^2 \sum_{t=1}^T (t-1) = \sigma^2 (T-1)T/2,$$

che diverge per $T \rightarrow \infty$. Per far convergere tale valore atteso è necessario dividere per T^2 :

$$\mathbb{E} \left(\frac{1}{T^2} \sum_{t=1}^T Y_{t-1}^2 \right) = \frac{\sigma^2}{2} \left(1 - \frac{1}{T} \right) \rightarrow \frac{\sigma^2}{2}, \quad \text{per } T \rightarrow \infty.$$

Quindi

$$(16.6.10) \quad T(\hat{\phi}_T - 1) = \frac{T^{-1} \sum_{t=1}^T Y_{t-1} a_t}{T^{-2} \sum_{t=1}^T Y_{t-1}^2}$$

ha una distribuzione rapporto di una variabile casuale chi-quadrato (traslata e riscalata) ed una variabile casuale con distribuzione non standard. Ricavare tale distribuzione è tutt'altro che banale, ed esula dagli scopi di questi appunti. Tale distribuzione, spesso detta di Dickey-Fuller, non è comunque scrivibile in forma chiusa e va quindi simulata. Esistono nella letteratura econometrica numerose tavole dei valori critici per i valori di probabilità (della coda sinistra) più utilizzati.

16.6.2.2. I test di Dickey-Fuller e Augmented Dickey-Fuller. La statistica $\tau_0 = T(\hat{\phi}_T - 1)$ può essere utilizzata per testare l'ipotesi $H_0 : \phi = 1$, contro l'alternativa $H_1 : \phi < 0$. A volte in luogo della statistica (16.6.10) viene utilizzata la statistica

$$(16.6.11) \quad \tau_1 = \frac{(\hat{\phi}_T - 1)}{\hat{\sigma}_{\hat{\phi}_T}},$$

dove $\hat{\sigma}_{\hat{\phi}_T}$ è l'errore standard dello stimatore ai minimi quadrati di ϕ , per la sua analogia con il classico test t . Naturalmete la distribuzione di (16.6.11) sotto $H_0 : \phi = 1$, non è una t di Student, ma è legata alla distribuzione di Dickey-Fuller, ed è stata solo simulata e organizzata in tavole.

Spesso nei software statistici (EViews compreso), viene implementato un modello autoregressivo leggermente diverso, seppur equivalente, al modello (16.6.7). Sottraendo Y_{t-1} nell'equazione (16.6.7) a sinistra e a destra dell'uguale si ottiene

$$\Delta Y_t = \eta Y_{t-1} + \varepsilon_t,$$

dove $\Delta = (1 - B)$ è l'operatore differenza e $\eta = (1 - \phi)$. Le statistiche test (16.6.10) diventano quindi

$$\begin{aligned} \tau_0 &= T\hat{\eta}_T \\ \tau_1 &= \frac{\hat{\eta}_T}{\hat{\sigma}_{\hat{\eta}_T}}. \end{aligned}$$

Le ststisiche τ_1 e τ_2 hanno distribuzioni diverse a seconda del modello stimato e dell'ipotesi testata. Di seguito si elencano i quattro casi presi solitamente in considerazione.

Caso 1.

Modello stimato: $Y_t = \phi Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = Y_{t-1} + \varepsilon_t$

Distribuzioni di τ_0 e τ_1 tabulate.

Caso 2.

Modello stimato: $Y_t = \alpha + \phi Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = Y_{t-1} + \varepsilon_t$

Distribuzioni di τ_0 e τ_1 tabulate.

Caso 3.

Modello stimato: $Y_t = \alpha + \phi Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = \alpha + Y_{t-1} + \varepsilon_t$

τ_1 converge in distribuzione alla normale standardizzata.

Caso 4.

Modello stimato: $Y_t = \alpha + \delta t + \phi Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = \alpha + Y_{t-1} + \varepsilon_t$

Distribuzioni di τ_0 e τ_1 tabulate.

Per generalizzare i test al caso in cui i disturbi $\{\varepsilon_t\}$ non siano indipendenti, ma solo stazionari, Dickey e Fuller propongono di stimare in luogo di un processo AR(1), un processo AR(p), approssimando quindi l'eventuale struttura di autocovarianza del processo ε_t . Per concentrare in un unico parametro la statistica sufficiente per testare l'ipotesi di presenza di una radice unitaria, si ponga

$$\begin{aligned}\rho &\equiv \phi_1 + \phi_2 + \dots + \phi_p \\ \zeta_j &\equiv -[\phi_{j+1} + \phi_{j+2} + \dots + \phi_p].\end{aligned}$$

Il polinomio autoregressivo $\phi_p(B)$ può essere riscritto

$$(1 - \rho B) - (\zeta_1 B + \zeta_2 B^2 + \dots \zeta_{p-1} B^{p-1})(1 - B),$$

infatti sviluppando i prodotti e raccogliendo,

$$\begin{aligned}&(1 - \rho B) - (\zeta_1 B + \zeta_2 B^2 + \dots \zeta_{p-1} B^{p-1})(1 - B) = \\&= 1 - \rho B - \zeta_1 B + \zeta_1 B^2 - \zeta_2 B^2 + \zeta_2 B^3 - \dots - \zeta_{p-1} B^{p-1} + \zeta_{p-1} B^p = \\&= 1 - (\rho + \zeta_1)B - (\zeta_2 - \zeta_1)B^2 - (\zeta_3 - \zeta_2)B^3 - \dots - (\zeta_{p-1} - \zeta_{p-2})B^{p-1} \\&\quad - (-\zeta_{p-1})B^p = \\&= 1 - [(\phi_1 + \dots + \phi_p) - (\phi_2 + \dots + \phi_p)]B \\&\quad - [-(\phi_3 + \dots + \phi_p) + (\phi_2 + \dots + \phi_p)]B^2 - \dots \\&\quad - [-(\phi_p) + (\phi_{p-1} + \phi_p)] = \\&= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.\end{aligned}$$

Sostituendo la nuova rappresentazione dell'operatore $AR(p)$ si ottiene la seguente rappresentazione di un processo $AR(p)$:

$$(16.6.12) \quad Y_t = \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t.$$

Ora, se il processo contiene una radice unitaria, il polinomio $\phi(z)$ si deve annullare per $z = 1$, ossia,

$$1 - \phi_1 - \dots - \phi_p = 0,$$

e ciò implica che il parametro ρ sia uguale a 1. Un'altra forma, leggermente diversa, spesso utilizzata per rappresentare il processo $AR(p)$ è ottenuta sottraendo ad entrambi i membri della (16.6.12) Y_{t-1} :

$$\Delta Y_t = \eta Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

con $\eta = (\rho - 1) = \phi_1 + \dots + \phi_p - 1$, che nel caso di processo integrato è uguale a zero.

I quattro casi del test Augmented Dickey-Fuller possono ora essere così riassunti (le stime dei parametri si intendono sempre fatte col metodo dei minimi quadrati):

Caso 1.

Modello stimato: $Y_t = \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$

Le distribuzioni dei test τ_0 e τ_1 (definiti per ρ) sono le stesse del caso 1. di cui sopra. Sono validi i test t o F sugli altri parametri del modello.

Caso 2.

Modello stimato: $Y_t = \alpha + \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$

Le distribuzioni dei test τ_0 e τ_1 (definiti per ρ) sono le stesse del caso 2. di cui sopra. Sono validi i test t o F sugli altri parametri del modello.

Caso 3.

Modello stimato: $Y_t = \alpha + \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = \alpha + Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$ con $\alpha \neq 0$

La statistica $\hat{\rho}_T$ converge con velocità $T^{3/2}$ ad una gaussiana. Tutti gli altri parametri convergono ad una distribuzione gaussiana con velocità $T^{1/2}$. Rimangono validi i test t ed F .

Caso 4.

Modello stimato: $Y_t = \alpha + \delta t + \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\mathcal{N}(0, \theta)$

$H_0 : Y_t = \alpha + Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \zeta_2 \Delta Y_{t-2} + \dots \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$

Le distribuzioni dei test τ_0 e τ_1 (definiti per ρ) sono le stesse del caso 4. di cui sopra.

Processi GARCH

Una caratteristica comune a quasi tutte le serie storiche di rendimenti di attività finanziarie è il *volatility clustering*. Per volatility clustering si intende una forma di eteroschedasticità in cui, per istanti di tempo vicini le varianze $\sigma_{\varepsilon_t}^2$ del processo ε_t hanno valori simili, mentre per istanti di tempo lontani tali varianze sono dissimili, o con le parole di Mandelbrot

...large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes...

Questa persistenza della volatilità è illustrata empiricamente in figura 1, attraverso la deviazione standard mobile, che mostra come si susseguano periodi di alta volatilità e periodi di bassa volatilità.

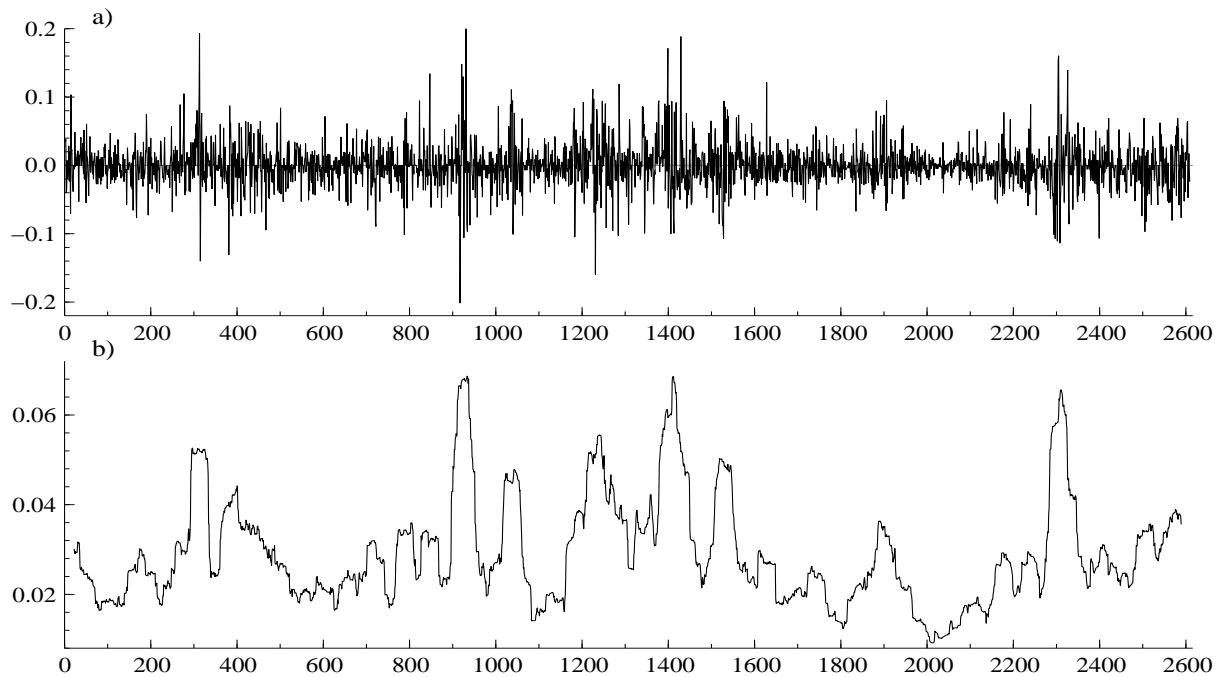


FIGURA 1. a) serie dei rendimenti giornalieri e b) deviazione standard mobile su 41 termini del titolo Alitalia

Si illustreranno nei prossimi paragrafi alcuni processi generatori di serie storiche eteroschedastiche con persistenza di volatilità, proposti negli anni '80 da Engle e Bollerslev e fino ad oggi oggetto di intensa ricerca e numerose applicazioni.

17.1. Il processo ARCH

Un semplice processo che genera serie storiche con il comportamento descritto è il processo ARCH(q) (*AutoRegressive Conditional Heteroskedasticity*), definito da

$$(17.1.1) \quad \begin{aligned} \varepsilon_t &= \xi_t \sigma_t & \xi_t &\sim \text{WN}(0, 1) \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2. \end{aligned}$$

Sotto opportune condizioni sui parametri, che indagheremo in seguito, valori passati di ε_t grandi in valore assoluto producono una varianza (condizionata a tali valori passati) di ε_t , al tempo t alta, e viceversa.

La definizione (17.1.1) va completata ponendo dei vincoli sui parametri che garantiscano la positività della varianza (condizionata). È facile vedere che condizione sufficiente e necessaria a garantire $\sigma_t^2 > 0$ per ogni t è

$$\omega > 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, q.$$

Si noti che almeno un parametro α_i deve essere maggiore di zero per produrre eteroschedasticità.

Riscrivendo il processo (17.1.1) come

$$\varepsilon_t = \xi_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2},$$

è evidente che la relazione che intercorre tra ε_t e ε_{t-h} , $h=1, \dots, q$, non è lineare.

Spesso la distribuzione condizionata del processo $\{\varepsilon_t\}$ è assunta normale: $(\varepsilon_t | \varepsilon_t, \varepsilon_{t-1}, \dots) \sim \mathcal{N}(0, \sigma_t^2)$, o equivalentemente $\xi_t \sim \mathcal{N}(0, 1)$. Nei casi in cui tale ipotesi potrà essere ritenuta vera, sarà possibile ottenere stimatori di massima verosimiglianza dei parametri del modello, alternativamente saranno comunque possibili stime di quasi-massima verosimiglianza (vedere oltre).

Il processo ARCH (17.1.1), $\{\varepsilon_t\}$ può essere anche pensato come disturbo di un modello di regressione

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t$$

o di un processo ARIMA

$$\phi(B)Y_t = \nu + \theta(B)\varepsilon_t.$$

In tal modo i menzionati modelli produrranno eteroschedasticità condizionata, ed il processo ARIMA-ARCH esibirà sia dipendenza lineare (parte ARIMA), sia dipendenza dei momenti di ordine superiore (parte ARCH).

In linea di principio, nella (17.1.1), in luogo del quadrato delle osservazioni ritardate, si potrebbe utilizzare una qualunque funzione $f(\cdot)$ monotona crescente in $|\varepsilon_{t-h}|$, e si otterrebbe un processo ad eteroschedasticità condizionata autoregressiva:

$$\sigma_t^2 = \omega + \alpha_1 f(|\varepsilon_1|) + \dots + \alpha_q f(|\varepsilon_{t-q}|).$$

Tuttavia il fatto che per il processo (17.1.1) valga

$$\mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2, \quad \mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$$

rende la scelta del quadrato in qualche modo naturale, oltre a semplificare molto la derivazione di alcune proprietà del processo stesso.

Prima di passare all'analisi delle proprietà di un processo ARCH, si introduce una classe di processi, che include, come caso particolare, il processo ARCH, e quindi, per non ripeterci, deriveremo le proprietà di questa classe più ampia.

17.2. Il processo GARCH

Nelle serie storiche finanziarie i *volatility clusters* hanno spesso una durata tale da rendere necessario un modello ARCH(q) con un q piuttosto grande, quindi con un numero di parametri da stimare elevato. Per contenere il numero di parametri da stimare in situazioni in cui la “memoria” del processo non è brevissima, è stata proposta una generalizzazione del processo ARCH, il processo GARCH (*Generalized ARCH*), che, utilizzando la stessa logica dei modelli ARMA, riduce il numero di parametri normalmente da stimare.

Il processo GARCH(q, p) è definito dalle seguenti relazioni

$$\begin{aligned} \varepsilon_t &= \xi_t \sigma_t & \xi_t &\sim \text{WN}(0, 1) \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \\ (17.2.1) \quad &+ \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \dots + \beta_p \sigma_{t-p}^2. \end{aligned}$$

Anche in questo caso, affinché σ_t^2 sia sempre positiva, si devono imporre le seguenti condizioni sufficienti¹:

$$\omega > 0, \quad \alpha_i, \beta_j \geq 0, \quad i = 1, \dots, q, \quad j = 1, \dots, p.$$

Analogamente a quanto visto per i modelli ARMA, se l'equazione caratteristica

$$1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p = 0,$$

è verificata solo per radici $|z_i| > 1$, allora il processo GARCH ammette una rappresentazione come processo ARCH di ordine infinito:

$$\begin{aligned} \sigma_t^2 &= \beta(B)^{-1} \omega + \beta(B)^{-1} (\alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2) \\ &= \omega^* + \sum_{i=1}^{\infty} \alpha_i^* \varepsilon_{t-i}^2. \end{aligned}$$

17.2.1. Rappresentazione ARMA del quadrato di un processo GARCH. Per evidenziare alcune analogie del processo $\{\varepsilon_t^2\}$ definito nella (17.2.1) con il processo ARMA si ponga

$$\eta_t = \varepsilon_t^2 - \sigma_t^2$$

da cui $\sigma_t^2 = \varepsilon_t^2 - \eta_t$. Sostituendo nella (17.2.1), si ottiene

$$\varepsilon_t^2 - \eta_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_{t-q} \varepsilon_{t-q}^2 + \beta_1 (\varepsilon_{t-1}^2 - \eta_{t-1}) + \dots + \beta_p (\varepsilon_{t-p}^2 - \eta_{t-p}),$$

risolvendo per ε_t^2 e raccogliendo i vari ε_{t-i}^2 si ha

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{\max\{p, q\}} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \eta_t - \sum_{i=1}^p \beta_i \eta_{t-i},$$

con

$$\alpha_i = 0 \text{ per } i > q \text{ e } \beta_i = 0 \text{ per } i > p,$$

che è la rappresentazione ARMA($\max\{p, q\}, p$) del processo GARCH(q, p). Naturalmente tale rappresentazione è di qualche utilità se il disturbo η_t ha le proprietà di un white noise, almeno in senso debole. In effetti è semplice vedere che $\mathbb{E}(\eta_t) = 0$, e si può mostrare che il processo $\{\eta_t\}$ non è autocorrelato. Tuttavia, pur essendo $\{\eta_t\}$ white noise in senso debole, non lo è in senso forte, in quanto esiste dipendenza nei momenti di ordine superiore al secondo.

¹In realtà in Nelson e Cao (1992), *Econometrica* 59, 347–370, vengono individuate delle condizioni sui parametri, meno stringenti, che sono sufficienti e necessarie per la positività (quasi certa) di σ_t^2 . Tuttavia per il modello GARCH(1,1), che è il più utilizzato nelle applicazioni tali condizioni coincidono con quelle qui date.

Utilizziamo ora la legge dei valori attesi iterati per calcolare i momenti del processo $\{\eta_t\}$. Il momento di primo ordine di $\{\eta_t\}$ è

$$\begin{aligned}\mathbb{E}(\eta_t) &= \mathbb{E}(\varepsilon_t^2 - \sigma_t^2) \\ &= \mathbb{E}[\mathbb{E}(\varepsilon_t^2 - \sigma_t^2 | \mathcal{F}_{t-1})] \\ &= \mathbb{E}[\sigma_t^2 - \sigma_t^2] \\ &= 0.\end{aligned}$$

Per ottenere la varianza di η_t (limitatamente al processo GARCH(1,1)), si calcoli prima la varianza condizionata, che, essendo il valore atteso condizionato nullo, è data da

$$\begin{aligned}\mathbb{E}(\eta_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}(\varepsilon_t^4 + \sigma_t^4 - 2\varepsilon_t^2 \sigma_t^2 | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\xi_t^4 \sigma_t^4 | \mathcal{F}_{t-1}) + \sigma_t^4 - 2\mathbb{E}(\xi_t^2 \sigma_t^2 \sigma_t^2 | \mathcal{F}_{t-1}) \\ &= \sigma_t^4 m_{\xi(4)} - \sigma_t^4 = (m_{\xi(4)} - 1)\sigma_t^4,\end{aligned}$$

dove $m_{\xi(4)}$ è il momento quarto di ξ_t , che si suppone finito. Nel caso ξ_t sia supposta normale, il suo momento quarto è pari a 3 e $\mathbb{E}(\eta_t^2 | \mathcal{F}_{t-1}) = 2\sigma_t^4$. Per calcolare $\mathbb{E}(\eta_t^2)$ basta quindi ottenere $\mathbb{E}(\sigma_t^4)$:

$$\begin{aligned}\mathbb{E}(\sigma_t^4) &= \mathbb{E}(\omega + \alpha\varepsilon_{t-1} + \beta\sigma_{t-1}^2)^2 \\ &= \omega^2 + \alpha^2\mathbb{E}(\varepsilon_{t-1}^4) + \beta^2\mathbb{E}(\sigma_{t-1}^4) + 2\omega\alpha\mathbb{E}(\varepsilon_{t-1}^2) + 2\omega\beta\mathbb{E}(\sigma_{t-1}^2) \\ &\quad + 2\alpha\beta\mathbb{E}(\varepsilon_{t-1}^2\sigma_{t-1}^2) \\ &= \omega^2 + \alpha^2 m_{\xi(4)} \mathbb{E}(\sigma_{t-1}^4) + \beta^2 \mathbb{E}(\sigma_{t-1}^4) + 2\omega\alpha\sigma^2 + 2\omega\beta\sigma^2 + 2\alpha\beta\mathbb{E}(\sigma_{t-1}^4)\end{aligned}$$

dove si è posto $\sigma^2 = \mathbb{E}(\sigma_t^2)$ (la varianza marginale del processo), le cui condizioni di esistenza ed il cui valore si ricaveranno in seguito. La soluzione stazionaria si trova ponendo $\tau = \mathbb{E}(\sigma_t^4) = \mathbb{E}(\sigma_{t-1}^4)$ e risolvendo rispetto a τ :

$$\tau = \frac{\omega^2 + 2\omega\sigma^2(\alpha + \beta)}{1 - \alpha^2 m_{\xi(4)} - \beta^2 - 2\alpha\beta},$$

per i valori di α e β per cui il denominatore è maggiore di zero. Mettendo insieme i risultati si ottiene

$$\mathbb{E}(\sigma_t^4) = (m_{\xi(4)} - 1)\tau.$$

La funzione di covarianza di $\{\eta_t\}$ è, per $k \neq 0$

$$\begin{aligned}\gamma_\eta(k) &= \mathbb{E}(\eta_t \eta_{t-k}) \\ &= \mathbb{E}[(\varepsilon_t^2 - \sigma_t^2)(\varepsilon_{t-k}^2 - \sigma_{t-k}^2)] \\ &= \mathbb{E}[\mathbb{E}(\varepsilon_t^2(\varepsilon_{t-k}^2 - 1) | \mathcal{I}_{t-1})] \\ &= \mathbb{E}[\sigma_t^2 \eta_{t-k} \mathbb{E}(\xi_t^2 - 1 | \mathcal{I}_{t-1})] \\ &= \mathbb{E}[\sigma_t^2 \eta_{t-k} (1 - 1)] \\ &= 0,\end{aligned}$$

dove il passaggio dalla terza alla quarta linea è giustificato dal fatto che σ_t^2 , σ_{t-k}^2 e η_t sono indirettamente contenute nell'insieme informativo \mathcal{F}_{t-1} , essendo funzioni deterministiche di $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)^2$. Si è quindi

²Ciò è evidente per σ_t^2 quando si considera la rappresentazione ARCH(∞) del processo GARCH, e quindi anche per $\eta_{t-k} = \varepsilon_{t-k}^2 - \sigma_{t-k}^2$.

dimostrato che η_t è un WN in senso debole, ma che non lo è in senso forte in quanto la sua varianza condizionata $\mathbb{E}(\eta_t^2 | \mathcal{F}_{t-1})$ non è costante.

La rappresentazione ARMA del quadrato di un processo GARCH è di notevole utilità quando si voglia studiare la relazione tra i processi GARCH ed il correlogramma del quadrato del processo GARCH stesso. La stima dei parametri di un processo GARCH attraverso la sua rappresentazione ARMA è possibile ma sub-ottimale in quanto non tiene conto della dipendenza nei momenti di ordine superiore al primo del disturbo η_t .

La rappresentazione ARMA del processo GARCH è anche utile per ricavare la varianza non condizionata di $\{\varepsilon_t\}$ come media non condizionata di un processo ARMA:

$$\mathbb{E}(\varepsilon_t^2) = (1 - \alpha_1 - \beta_1 - \alpha_2 - \beta_2 - \dots - \alpha_{\max\{q,p\}} - \beta_{\max\{q,p\}})^{-1} \omega.$$

La varianza non condizionata di $\{\varepsilon_t\}$ esiste finita, quindi, solo quando

$$\alpha_1 + \dots + \alpha_q + \beta_1 + \dots + \beta_p < 1.$$

17.2.2. Momenti del processo GARCH. Un modo per ottenere il valore atteso non condizionato di ε_t utilizzando la legge dei valori attesi iterati è

$$\mathbb{E}[\varepsilon_t] = \mathbb{E}[\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})] = \mathbb{E}(0) = 0.$$

In modo simile si calcolano le autocovarianze del processo $\{\varepsilon_t\}$. Dato che il valore atteso di $\{\varepsilon_t\}$ è zero, la funzione di autocovarianza a intervallo $k \in \mathbb{N}$ è data da

$$\begin{aligned} \gamma_\varepsilon(k) &= \mathbb{E}[\varepsilon_t \varepsilon_{t-k}] \\ &= \mathbb{E}[\mathbb{E}(\varepsilon_t \varepsilon_{t-k} | \mathcal{F}_{t-1})] \\ &= \mathbb{E}[\varepsilon_{t-k} \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})] \\ &= \mathbb{E}[\varepsilon_{t-k} \cdot 0] \\ &= 0 \end{aligned}$$

La varianza non condizionata di $\{\varepsilon_t\}$, che è stata già calcolata utilizzando la rappresentazione ARMA del processo $\{\varepsilon_t^2\}$, può essere ottenuta sfruttando nuovamente la legge dei valori attesi iterati:

$$\begin{aligned} \gamma_\varepsilon(0) &= \mathbb{E}[\varepsilon_t^2] \\ &= \mathbb{E}[\mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1})] \\ &= \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[\omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2] \\ &= \omega + \alpha_1 \mathbb{E}(\varepsilon_{t-1}^2) + \dots + \alpha_q \mathbb{E}(\varepsilon_{t-q}^2) + \\ &\quad + \beta_1 \mathbb{E}[\mathbb{E}(\varepsilon_{t-1}^2 | \mathcal{F}_{t-2})] + \dots + \beta_q \mathbb{E}[\mathbb{E}(\varepsilon_{t-q}^2 | \mathcal{F}_{t-q-1})] \\ &= \omega + \alpha_1 \mathbb{E}(\varepsilon_{t-1}^2) + \dots + \alpha_q \mathbb{E}(\varepsilon_{t-q}^2) + \beta_1 \mathbb{E}(\varepsilon_{t-1}^2) + \dots + \beta_q \mathbb{E}(\varepsilon_{t-q}^2) \\ &= \omega + (\alpha_1 + \beta_1) \gamma_\varepsilon(0) + \dots + (\alpha_{\max\{p,q\}} + \beta_{\max\{p,q\}}) \gamma_\varepsilon(0) \\ &= \omega (1 - \alpha_1 - \beta_1 - \dots - \alpha_{\max\{p,q\}} - \beta_{\max\{p,q\}})^{-1}. \end{aligned}$$

Si potrebbe inoltre mostrare che, se la distribuzione condizionata di $(\varepsilon_t | \mathcal{F}_{t-1})$ è simmetrica (come per esempio la normale), allora anche la distribuzione non condizionata di ε_t è simmetrica, e quindi tutti i momenti di ordine dispari sono nulli: $\mathbb{E}(\varepsilon_t^3) = \mathbb{E}(\varepsilon_t^5) = \mathbb{E}(\varepsilon_t^7) = \dots = 0$.

Inoltre se ξ_t ha distribuzione normale standardizzata ed $\{\varepsilon_t\}$ è un processo GARCH(1,1), il momento quarto (necessario per calcolare la curtosi) è³

$$\begin{aligned}
m_{\varepsilon(4)} &= \mathbb{E}(\varepsilon_t^4) \\
&= \mathbb{E}[\mathbb{E}(\varepsilon_t^4 | \mathcal{F}_{t-1})] \\
&= \mathbb{E}[3\sigma_t^4] \\
&= 3\mathbb{E}[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2)^2] \\
&= 3\mathbb{E}[\omega^2 + \alpha^2\varepsilon_{t-1}^4 + \beta^2\sigma_{t-1}^4 + 2\omega\alpha\varepsilon_{t-1}^2 + 2\omega\beta\sigma_{t-1}^2 + 2\alpha\beta\varepsilon_{t-1}^2\sigma_{t-1}^2] \\
&= 3[\omega^2 + \alpha^2\mathbb{E}(\varepsilon_{t-1}^4) + \beta^2\mathbb{E}(\sigma_{t-1}^4) + 2\omega\alpha\mathbb{E}(\varepsilon_{t-1}^2) + 2\omega\beta\mathbb{E}(\sigma_{t-1}^2) + \\
&\quad + 2\alpha\beta\mathbb{E}(\varepsilon_{t-1}^2\sigma_{t-1}^2)] \\
&= 3\omega^2 + 3\alpha^2m_{\varepsilon(4)} + 3\beta^2\frac{1}{3}m_{\varepsilon(4)} + 6\omega\alpha\frac{\omega}{1-\alpha-\beta} + 2\omega\beta\frac{\omega}{1-\alpha-\beta} + \\
&\quad + 6\alpha\beta\mathbb{E}[\mathbb{E}(\varepsilon_{t-1}^2\sigma_{t-1}^2 | \mathcal{I}_{t-2})],
\end{aligned}$$

dove

$$\mathbb{E}[\mathbb{E}(\varepsilon_{t-1}^2\sigma_{t-1}^2 | \mathcal{F}_{t-2})] = \mathbb{E}[\sigma_{t-1}^2\mathbb{E}(\varepsilon_{t-1}^2 | \mathcal{F}_{t-2})] = \mathbb{E}[\sigma_{t-1}^4] = m_{\varepsilon(4)}/3$$

e risolvendo rispetto a $m_{\varepsilon(4)}$,

$$\begin{aligned}
m_{\varepsilon(4)} \cdot (1 - 3\alpha^2 - \beta^2 - 2\alpha\beta) &= 3\omega^2 + \frac{6\omega^2(\alpha + \beta)}{(1 - \alpha - \beta)} \\
m_{\varepsilon(4)} &= \frac{3\omega^2 - 3\omega^2(\alpha + \beta) + 6\omega^2(\alpha + \beta)}{(1 - \alpha - \beta)(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)} \\
m_{\varepsilon(4)} &= \frac{3\omega^2(1 + \alpha + \beta)}{(1 - \alpha - \beta)(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)}.
\end{aligned}$$

Il momento quarto esiste quindi solo se

$$(\beta^2 + 2\alpha\beta + 3\alpha^2) < 1,$$

ovvero per

$$-\alpha - \sqrt{-2\alpha^2 + 1} < \beta < -\alpha + \sqrt{-2\alpha^2 + 1},$$

infatti i momenti pari non possono essere negativi, in quanto valori attesi di variabili casuali positive. È ora possibile calcolare la curtosi della distribuzione non condizionata di $\{\varepsilon_t\}$:

$$\begin{aligned}
\kappa &= \frac{\mathbb{E}(\varepsilon_t^4)}{[\mathbb{E}(\varepsilon_t^2)]^2} \\
&= \frac{3\omega^2(1 + \alpha + \beta)}{(1 - \alpha - \beta)(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)} \frac{(1 - \alpha - \beta)^2}{\omega^2} \\
(17.2.2) \quad &= \frac{3(1 - \alpha - \beta)(1 + \alpha + \beta)}{(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)}.
\end{aligned}$$

È facile a questo punto vedere che se le condizioni di esistenza dei momenti secondo e quarto

$$\{\alpha \geq 0\} \cap \{\beta \geq 0\} \cap \{\alpha + \beta < 1\} \cap \{-\alpha - \sqrt{-2\alpha^2 + 1} < \beta < -\alpha + \sqrt{-2\alpha^2 + 1}\}$$

³Si ricordano i seguenti risultati. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, allora $\mathbb{E}[(X - \mu)^4] = 3\sigma^4$, e quindi la varianza può essere espressa in funzione del momento quarto centrato, $\sigma^2 = \mathbb{E}[(X - \mu)^4]/3$. Se $\{\varepsilon_t\}$ è un processo GARCH(1,1), allora $\mathbb{E}(\varepsilon_t^2) = \omega(1 - \alpha - \beta)^{-1}$, ma anche $\mathbb{E}(\sigma_t^2) = \omega(1 - \alpha - \beta)^{-1}$ (dimostrare per esercizio).

sono verificate, per $\alpha > 0$ la curtosi della distribuzione non condizionata di un processo GARCH(1,1) è maggiore di 3, ovvero ha code più pesanti di una normale, infatti

$$\begin{aligned}\kappa &= \frac{3(1-\alpha-\beta)(1+\alpha+\beta)}{(1-\beta^2-2\alpha\beta-3\alpha^2)} > 3 \\ (1-\alpha-\beta)(1+\alpha+\beta) &> (1-\beta^2-2\alpha\beta-3\alpha^2) \\ 1-\alpha^2-\beta^2-2\alpha\beta-1+\beta^2+2\alpha\beta+3\alpha^2 &> 0 \\ \alpha^2 &> 0.\end{aligned}$$

Per $\alpha = 0$, invece, la (17.2.2) si semplifica a 3, che è la curtosi di una normale; e per $\beta = 0$, la (17.2.2) si riduce a

$$\kappa = 3 \frac{1-\alpha^2}{1-3\alpha^2},$$

che, se da un lato rende evidente la leptocurtosi della distribuzione non condizionata di un processo ARCH, dall'altro pone un ulteriore vincolo sul parametro α : affinché la curtosi esista positiva, fermo restando $\alpha \geq 0$, il denominatore deve essere maggiore di zero, e quindi $\alpha > \sqrt{1/3} \approx 0.577$.

In figura 2 è rappresentato graficamente l'insieme dei vincoli sui parametri di un GARCH(1,1) e una serie di curve di iso-curtosi,

$$\beta = \frac{\sqrt{(-2\kappa\alpha^2 + \kappa - 3)(\kappa - 3)}}{(\kappa - 3)}.$$

ottenute risolvendo la (17.2.2) rispetto a β , per diversi valori di κ (in figura $\kappa = 5, 7, 10, 20, 50$).

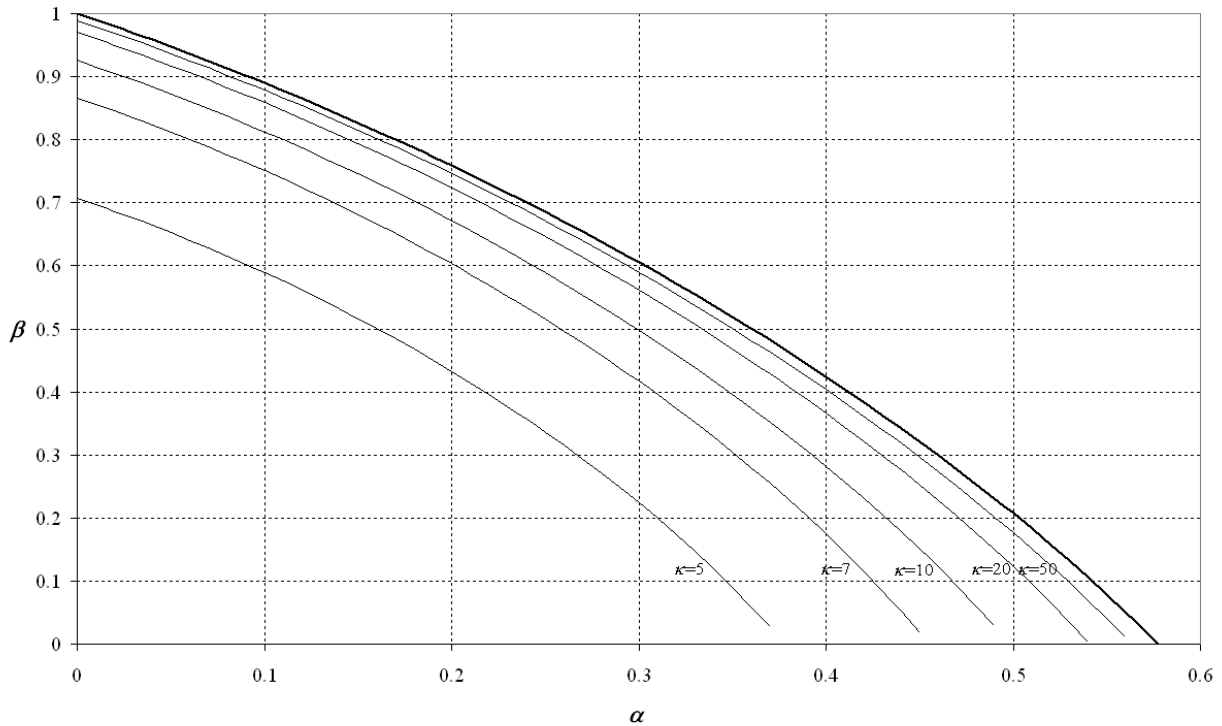


FIGURA 2. il luogo dei punti compresi tra gli assi cartesiani (inclusi) e la curva in grassetto (esclusa) è l'insieme delle coppie di valori dei parametri α e β di un processo GARCH(1,1) con momenti fino al quarto esistenti. Le altre curve sono iso-curtosi, cioè coppie di valori di α e β che producono un determinato valore di curtosi della distribuzione del processo.

17.2.3. Previsione di un processo GARCH. Si è visto che il previsore ottimo dal punto di vista del MSFE è il valore atteso condizionato $\mathbb{E}[\varepsilon_{T+h}|\mathcal{F}_T]$, $h \in \mathbb{N}$. La previsione un periodo in avanti di $\{\varepsilon_t\}$ si ottiene facilmente dalla definizione del processo GARCH stesso,

$$\hat{\varepsilon}_{T+1} = \mathbb{E}[\varepsilon_{T+1}|\mathcal{F}_T] = 0.$$

Le previsioni successive si ottengono applicando la legge dei valori attesi iterati

$$\hat{\varepsilon}_{T+h} = \mathbb{E}[\varepsilon_{T+h}|\mathcal{F}_T] = \mathbb{E}[\mathbb{E}(\varepsilon_{T+h}|\mathcal{F}_{T+h-1})|\mathcal{I}_T] = \mathbb{E}[0|\mathcal{F}_T] = 0.$$

Per calcolare la varianza condizionata, cioè la previsione della volatilità, si noti che

$$(17.2.3) \quad \mathbb{E}[\varepsilon_{T+h}^2|\mathcal{F}_T] = \mathbb{E}[\mathbb{E}(\varepsilon_{T+h}^2|\mathcal{F}_{T+h-1})|\mathcal{F}_T] = \mathbb{E}(\sigma_{T+h}^2|\mathcal{F}_T);$$

Per un GARCH(1,1) si ha

$$\begin{aligned} \hat{\sigma}_{T+1|T}^2 &= \mathbb{E}(\sigma_{T+1}^2|\mathcal{F}_T) = \sigma_{T+1}^2 = \omega + \alpha\varepsilon_T^2 + \beta\sigma_T^2, \\ \hat{\sigma}_{T+2|T}^2 &= \mathbb{E}(\sigma_{T+2}^2|\mathcal{F}_T) = \omega + \alpha\mathbb{E}(\varepsilon_{T+1}^2|\mathcal{F}_T) + \beta\sigma_{T+1}^2 = \omega + (\alpha + \beta)\sigma_{T+1}^2, \\ \hat{\sigma}_{T+3|T}^2 &= \mathbb{E}(\sigma_{T+3}^2|\mathcal{F}_T) = \omega + \alpha\mathbb{E}(\varepsilon_{T+2}^2|\mathcal{F}_T) + \beta\mathbb{E}(\sigma_{T+2}^2|\mathcal{F}_T) = \omega + (\alpha + \beta)\hat{\sigma}_{T+2|T}^2, \\ \hat{\sigma}_{T+h|T}^2 &= \mathbb{E}(\sigma_{T+h}^2|\mathcal{F}_T) = \omega + \alpha\mathbb{E}(\varepsilon_{T+h-1}^2|\mathcal{F}_T) + \beta\mathbb{E}(\sigma_{T+h-1}^2|\mathcal{F}_T) = \omega + (\alpha + \beta)\hat{\sigma}_{T+h-1|T}^2, \end{aligned}$$

dove dalla terza riga in poi si è utilizzato il risultato (17.2.3). Sostituendo iterativamente, la previsione della varianza h periodi in avanti può essere riscritta come

$$\hat{\sigma}_{T+h|T}^2 = \omega[1 + (\alpha + \beta) + (\alpha + \beta)^2 + \dots + (\alpha + \beta)^{h-2}] + (\alpha + \beta)^{h-1}\sigma_{T+1}^2,$$

che, se $\alpha + \beta < 1$ (quindi escludendo l'IGARCH) per $t \rightarrow \infty$ converge alla varianza marginale (o non condizionale)

$$\hat{\sigma}_{T+\infty|T}^2 = \frac{\omega}{1 - \alpha - \beta}.$$

Quando il processo è l'IGARCH(1,1) (integrated GARCH)

$$\sigma_t^2 = \alpha\varepsilon_{t-1}^2 + (1 - \alpha)\sigma_{t-1}^2$$

come si vede dall'equazione di previsione sostituendo $\omega = 0$ e $\beta = (1 - \alpha)$, la previsione della varianza diventa costante

$$\hat{\sigma}_{T+h|T}^2 = \hat{\sigma}_{T+h-1|T}^2 = \dots = \sigma_{T+1}^2.$$

Inferenza per processi GARCH

18.1. Stime di Quasi Massima Verosimiglianza

Un'assunzione implicita che si fa quando si utilizza il metodo della massima verosimiglianza (ML) per stimare i parametri di un modello statistico è che il modello sia stato correttamente specificato. Più precisamente, si suppone che all'interno della famiglia di distribuzioni di probabilità utilizzata come modello parametrico sia contenuto il modello “vero”, che ha generato le osservazioni del campione a disposizione. Tuttavia, il più delle volte si è lungi dal pretendere che il modello utilizzato sia il “vero” modello che la natura ha utilizzato per generare i dati. Halbert White, in un articolo¹ di fondamentale importanza per la teoria della stima, indaga sulle proprietà degli stimatori di massima verosimiglianza, quando il modello utilizzato è erroneamente specificato. Nelle prossime sezioni saranno esposti i risultati di White e poi applicati ai modelli GARCH.

Per poter dare una diversa interpretazione delle stime di massima verosimiglianza e del concetto di consistenza di uno stimatore, in presenza di un modello mal specificato si introduce uno strumento noto come criterio di informazione (o anche discrepanza) di Kullback-Leiber.

18.1.1. Criterio di informazione di Kullback-Leibler. Sia \mathbf{X} un vettore di variabili casuali² X_t , $t = 1, \dots, n$, indipendenti con funzione di ripartizione congiunta $G(\mathbf{x})$ e densità congiunta $g(\mathbf{x})$.

Sia $\{F(\mathbf{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, con Θ sottoinsieme compatto di uno spazio euclideo p -dimensionale, la famiglia di funzioni di ripartizione in cui si ritiene (erroneamente o meno) sia inclusa la vera distribuzione $G(\mathbf{x})$. Si supponga che esista per ogni $\boldsymbol{\theta}$ la funzione di densità $f(\mathbf{x}|\boldsymbol{\theta})$ relativa a $F(\mathbf{x}|\boldsymbol{\theta})$.

Per poter motivare l'utilizzo della massima verosimiglianza anche nel caso in cui $G(\mathbf{x})$ non sia in $\{F(\mathbf{x}|\boldsymbol{\theta})\}_{\boldsymbol{\theta} \in \Theta}$ si introduce un nuovo strumento, *il criterio di informazione di Kullback-Leibler* (KLIC):

$$(18.1.1) \quad \text{KLIC}(g : f) = \mathbb{E}_g \left[\ln \frac{g(\mathbf{X})}{f(\mathbf{X}|\boldsymbol{\theta})} \right]$$

$$(18.1.2) \quad = \mathbb{E}_g[\ln g(\mathbf{X})] - \mathbb{E}_g[\ln f(\mathbf{X}|\boldsymbol{\theta})].$$

Il KLIC è una misura della discrepanza tra la distribuzione vera G e la distribuzione $F(\mathbf{x}; \boldsymbol{\theta})$. Si può dimostrare che

$$\text{KLIC}(g : f) \geq 0,$$

$$\text{KLIC}(g : f) = 0 \iff f = g \text{ quasi ovunque.}$$

Per rendere minima la (18.1.1), dato che il primo addendo della (18.1.2) è una costante (ignota), è necessario trovare quel valore $\boldsymbol{\theta}^*$, che rende massimo $\mathbb{E}_g[\ln f(\mathbf{X}|\boldsymbol{\theta})]$. Ora, quando il modello parametrico è correttamente specificato, cioè quando $g(\mathbf{x}) \in \{f(\mathbf{x}|\boldsymbol{\theta})\}_{\boldsymbol{\theta} \in \Theta}$, allora il KLIC raggiunge il suo valore minimo assoluto (0, discrepanza nulla), per quel valore $\boldsymbol{\theta}^*$ per cui $f(\mathbf{x}|\boldsymbol{\theta}^*) = g(\mathbf{x})$ (quasi ovunque); altrimenti $\boldsymbol{\theta}^*$ è

¹White H. (1982). Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1.

²Quanto verrà detto è valido anche se le X_t , in luogo di variabili casuali scalari, sono vettori casuali, e \mathbf{X} una matrice.

semplicemente il valore di θ che rende minima la discrepanza tra il modello vero e quello usato, sebbene mai annullandola.

18.1.2. Proprietà degli stimatori di quasi massima verosimiglianza. A questo punto è necessario generalizzare alcuni concetti fondamentali di teoria della stima in modo da adattarli alla situazione di possibile errata specificazione del modello. Un modello statistico verrà detto *identificato* se il KLIC raggiunge il minimo assoluto, rispetto a θ , per un unico valore θ^* . Uno stimatore di θ verrà detto *consistente* se esso converge in probabilità al valore quasi-vero θ^* , che minimizza il KLIC.

Un naturale stimatore di $\mathbb{E}_g[\ln f(\mathbf{X}|\theta)]$ è³

$$l_n(\theta|\mathbf{X}) = \frac{1}{n} \sum_{t=1}^n \ln f(X_t|\theta),$$

che è la funzione di log-verosimiglianza⁴ del modello (possibilmente) erroneamente specificato, anche detta *funzione di quasi log-verosimiglianza*. Lo stimatore di quasi massima verosimiglianza

$$\hat{\theta}_n = \arg \max_{\theta} l_n(\theta|\mathbf{X}),$$

è quindi uno stimatore di θ^* , e, alla luce di quanto detto, può essere interpretato come stimatore di “minima ignoranza” rispetto al modello vero.

TEOREMA 13 (Consistenza dello stimatore di quasi massima verosimiglianza). *Se il KLIC esiste ben definito ed il modello è identificato, allora*⁵

$$\text{p} \lim_{n \rightarrow \infty} \hat{\theta}_n = \theta^*.$$

Si noti che la corretta specificazione del modello statistico è condizione solo sufficiente e per nessun motivo necessaria alla consistenza di $\hat{\theta}_n$ rispetto al vero parametro d'interesse. Per esempio, se il modello statistico ipotizzato è gaussiano, gli stimatori di quasi massima verosimiglianza dei parametri media e varianza

$$\begin{aligned} \hat{\mu}_n &= \frac{1}{n} \sum_{t=1}^n X_t \\ \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{t=1}^n (X_t - \hat{\mu}_n)^2 \end{aligned}$$

sono consistenti per la media e la varianza vere, anche se il modello gaussiano è errato.

Imponendo alcune condizioni di regolarità aggiuntive è possibile dimostrare che lo stimatore di quasi massima verosimiglianza ha distribuzione asintotica normale. Si supponga che esistano il gradiente e l'Hessiana di $f(X_t|\theta)$ rispetto a θ e si definiscano le seguenti matrici

$$\begin{aligned} \mathbf{A}_n(\theta) &= -\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ln f(X_t|\theta)}{\partial \theta \partial \theta'}, \\ \mathbf{B}_n(\theta) &= \frac{1}{n} \sum_{t=1}^n \frac{\partial \ln f(X_t|\theta)}{\partial \theta} \frac{\partial \ln f(X_t|\theta)}{\partial \theta'}. \end{aligned}$$

³Si utilizza la seguente simbologia: mentre $f(\mathbf{x})$ indica la densità congiunta di \mathbf{X} , $f(x_t)$ denota la densità marginale di X_t .

⁴Si noti che usualmente $l(\theta) = \sum_{t=1}^n \ln f(X_t|\theta)$, ma si ricordi che moltiplicare la funzione per una costante positiva non modifica il valore del parametro θ nel punto di massimo.

⁵In realtà il risultato di White è più forte di quello qui esposto, infatti White dimostra che la convergenza è quasi certa; tuttavia per semplicità, in questa sede ci si limiterà a trattare di convergenze in probabilità e in distribuzione.

Si supponga che esitino anche i valori attesi necessari al calcolo delle seguenti matrici

$$\begin{aligned}\mathbf{A}(\boldsymbol{\theta}) &= -\mathbb{E} \left[\frac{\partial^2 \ln f(X_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right], \\ \mathbf{B}(\boldsymbol{\theta}) &= \mathbb{E} \left[\frac{\partial \ln f(X_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln f(X_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right].\end{aligned}$$

Si supponga che esistano le inverse di $\mathbf{A}_n(\boldsymbol{\theta})$ e di $\mathbf{A}(\boldsymbol{\theta})$ e siano

$$\begin{aligned}\mathbf{C}_n(\boldsymbol{\theta}) &= \mathbf{A}_n(\boldsymbol{\theta})^{-1} \mathbf{B}_n(\boldsymbol{\theta}) \mathbf{A}_n(\boldsymbol{\theta})^{-1} \\ \mathbf{C}(\boldsymbol{\theta}) &= \mathbf{A}(\boldsymbol{\theta})^{-1} \mathbf{B}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta})^{-1}.\end{aligned}$$

TEOREMA 14 (Normalità asintotica degli stimatori di quasi massima verosimiglianza). *Se oltre alle condizioni supposte valide per la consistenza, valgono le condizioni per l'esistenza delle matrici sopra definite, allora per $n \rightarrow \infty$*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(0, \mathbf{C}(\boldsymbol{\theta}^*)).$$

Inoltre

$$\text{p} \lim_{n \rightarrow \infty} \mathbf{C}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{C}(\boldsymbol{\theta}^*).$$

Il teorema 14, oltre a fornire la distribuzione asintotica degli stimatori di quasi massima verosimiglianza, indica che la matrice $\mathbf{C}_n(\hat{\boldsymbol{\theta}}_n)$ è uno stimatore consistente della matrice di covarianza ignota $\mathbf{C}(\boldsymbol{\theta}^*)$. Per costruire test di significatività asintotici per ipotesi del tipo $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$, si può utilizzare la statistica

$$t = \frac{\hat{\theta}_i - \theta_0}{\sqrt{c_{ii}}},$$

con $\hat{\theta}_i$ i -esima riga del vettore $\hat{\boldsymbol{\theta}}_n$, e c_{ii} i -esimo elemento sulla diagonale della matrice $\mathbf{C}_n(\hat{\boldsymbol{\theta}}_n)$, confrontandola con una distribuzione normale standard.

Quando il modello è correttamente specificato ($G(\mathbf{x}) \in \{F(\mathbf{x}|\boldsymbol{\theta})\}_{\boldsymbol{\theta} \in \Theta}$), sotto ulteriori condizioni di regolarità⁶, vale l'uguaglianza classica della matrice d'informazione di Fisher

$$\mathbf{A}(\boldsymbol{\theta}^*) = \mathbf{B}(\boldsymbol{\theta}^*) = \mathcal{I}(\boldsymbol{\theta}^*),$$

con \mathcal{I} matrice d'informazione di Fisher, cosicchè

$$\mathbf{C}(\boldsymbol{\theta}^*) = \mathcal{I}^{-1}(\boldsymbol{\theta}^*)$$

ed il teorema 14 contiene il noto risultato degli stimatori di massima verosimiglianza come caso particolare.

18.1.3. Stime di quasi massima verosimiglianza per modelli GARCH. Per semplificare la notazione ci si limiterà a considerare la stima di un modello GARCH(1,1). L'estensione a modelli di ordine superiore è immediata. Sia $f_{\sigma_t^2}(x)$ una generica funzione di densità con media nulla e varianza σ_t^2 , e si supponga di volere stimare i parametri del seguente modello GARCH(1,1)

$$\begin{aligned}\varepsilon_t | \mathcal{I}_{t-1} &\sim f(\varepsilon_t | \sigma_t^2), \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2,\end{aligned}$$

per $t = 1, 2, \dots, n$, con $\mathcal{I}_t = (\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1)$. Sia $\boldsymbol{\theta}$ il vettore dei parametri da stimare, contenente (ω, α, β) oltre ad altri possibili parametri necessari a definire $f_{\boldsymbol{\theta}}(x|\sigma_t^2)$ (per esempio gradi di libertà o parametri di forma).

⁶In particolare il supporto di $F(\mathbf{x}|\boldsymbol{\theta})$ non deve dipendere da $\boldsymbol{\theta}$.

La funzione di verosimiglianza di tale modello è

$$L(\boldsymbol{\theta}) \propto f_{\boldsymbol{\theta}}(\varepsilon_n|\sigma_n^2) \cdot f_{\boldsymbol{\theta}}(\varepsilon_{n-1}|\sigma_{n-1}^2) \cdot \dots \cdot f_{\boldsymbol{\theta}}(\varepsilon_2|\sigma_2^2) \cdot f_{\boldsymbol{\theta}}(\varepsilon_1|\sigma_1^2),$$

dove sorge il solito problema della determinazione della distribuzione iniziale $f_{\boldsymbol{\theta}}(\varepsilon_1|\sigma_1^2)$. Dato che la densità non condizionata di ε_t è di ardua determinazione, si può agire in tre modi alternativi: i) considerare $\mathcal{I}_0 = \{\emptyset\}$ e utilizzare per il calcolo di σ_1^2 i valori attesi non condizionati di ε_t^2 e σ_t^2 , ii) considerare $\mathcal{I}_0 = \{\emptyset\}$ ed inserire σ_1^2 nel vettore $\boldsymbol{\theta}$ dei parametri da stimare, iii) considerare $\mathcal{I}_0 = \{\varepsilon_0\}$, dove ε_0 è una osservazione precampionaria da utilizzare per il calcolo di σ_1^2 , mentre σ_0^2 può essere trattato come parametro aggiuntivo in $\boldsymbol{\theta}$ o sostituito con il suo valore non condizionato.

Le densità più utilizzate nella pratica sono la normale a media nulla

$$f(\varepsilon_t|\sigma_t^2) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2}\right\},$$

la t di Student generalizzata in modo tale da avere gradi di libertà $\nu \in \mathbb{R}^+$ e varianza σ_t^2

$$f_{\nu}(\varepsilon_t|\sigma_t^2) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)\sigma_t^2}} \left[1 + \frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right]^{-\frac{\nu+1}{2}},$$

e la *Generalized Error Distribution*, GED(r),

$$f_r(\varepsilon_t|\sigma_t^2) = \frac{r \exp\left[-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda}\right|^r\right]}{\lambda 2^{\frac{r+1}{r}} \Gamma(1/r)},$$

con $r > 0$ e

$$\lambda = \frac{1}{2^{1/2}} \left[\frac{\Gamma(1/r)}{\Gamma(3/r)} \right]^{1/2},$$

che ha media nulla, varianza unitaria⁷ e curtosi governata dal parametro r :

$$r \in (0, 2) \quad \Leftrightarrow \quad \kappa > 3$$

$$r = 2 \quad \Leftrightarrow \quad \text{GED} = \mathcal{N}(0, 1)$$

$$r \in (2, +\infty) \quad \Leftrightarrow \quad \kappa < 3,$$

dove κ è la curtosi.

La massimizzazione della funzione di log-verosimiglianza può essere condotta per mezzo di metodi numerici, mentre gradiente ed Hessiana possono essere calcolati anche analiticamente.

La funzione di verosimiglianza può essere facilmente generalizzata per un modello regressivo eteroschedastico

$$y_t = \mathbf{x}_t' \boldsymbol{\gamma} + \varepsilon_t,$$

dove \mathbf{x}_t , $t = 1, 2, \dots, n$ sono vettori (colonna) di variabili predeterminate, $\boldsymbol{\gamma}$ è un vettore (colonna) di parametri e ε_t è un processo GARCH. Per fare ciò è sufficiente sostituire nelle funzioni di densità $(y_t - \mathbf{x}_t' \boldsymbol{\gamma})$ a ε_t ed aggiungere i parametri $\boldsymbol{\gamma}$ nel vettore $\boldsymbol{\theta}$ dei parametri rispetto a cui massimizzare la funzione di verosimiglianza.

⁷Per ricavare la densità di $\varepsilon_t|\mathcal{I}_{t-1}$, si ricordi che, se la variabile casuale X ha densità $f(x)$, allora $Y = X \cdot \sigma$ ha densità

$$Y \sim \frac{1}{\sigma} f\left(\frac{y}{\sigma}\right).$$

18.2. Test di restrizioni sui parametri

18.2.1. Test di Wald e test LM. Si supponga di volere verificare una ipotesi parametrica del tipo

$$H_0 : s(\boldsymbol{\theta}^*) = \mathbf{0}$$

dove $s : \Theta \mapsto \mathbb{R}^r$ è una funzione vettoriale continua di $\boldsymbol{\theta}$, con gradiente in $\boldsymbol{\theta}^*$, $\nabla s(\boldsymbol{\theta}^*)$ finito, contro l'alternativa $H_1 : s(\boldsymbol{\theta}^*) \neq \mathbf{0}$.

Tale ipotesi può essere testata per mezzo della *statistica di Wald*, che sotto H_0 ha distribuzione asintotica χ_r^2

$$W_n = n \cdot s(\hat{\boldsymbol{\theta}}_n)' [\nabla s(\hat{\boldsymbol{\theta}}_n) \mathbf{C}_n(\hat{\boldsymbol{\theta}}_n) \nabla s(\hat{\boldsymbol{\theta}}_n)']^{-1} s(\hat{\boldsymbol{\theta}}_n) \stackrel{d}{\rightarrow} \chi_r^2,$$

dove $\hat{\boldsymbol{\theta}}_n$ è lo stimatore di quasi massima verosimiglianza del modello senza le restrizioni e $\mathbf{C}_n(\hat{\boldsymbol{\theta}}_n)$ è lo stimatore della matrice di covarianza di $\hat{\boldsymbol{\theta}}_n$ definito nella precedente sezione. Quando il modello è specificato correttamente $\mathbf{C}_n(\hat{\boldsymbol{\theta}}_n)$ può essere sostituito da $\mathbf{A}_n(\hat{\boldsymbol{\theta}}_n)$ o da $\mathbf{B}_n(\hat{\boldsymbol{\theta}}_n)$, assumendo la forma usuale.

Un test asintoticamente equivalente è il *test dei moltiplicatori di Lagrange*, detto anche *score test*. Sia $\tilde{\boldsymbol{\theta}}_n$ lo stimatore di quasi massima verosimiglianza vincolato

$$\begin{cases} \tilde{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta}} l_n(\boldsymbol{\theta} | \mathbf{X}) \\ s(\boldsymbol{\theta}^*) = \mathbf{0} \end{cases},$$

la statistica LM è data da

$$\begin{aligned} LM_n &= \nabla l_n(\tilde{\boldsymbol{\theta}}_n | \mathbf{X})' \mathbf{A}_n(\tilde{\boldsymbol{\theta}}_n)^{-1} \nabla s(\tilde{\boldsymbol{\theta}}_n)' \cdot \\ &\quad \cdot [\nabla s(\tilde{\boldsymbol{\theta}}_n) \mathbf{C}_n(\tilde{\boldsymbol{\theta}}_n) \nabla s(\tilde{\boldsymbol{\theta}}_n)']^{-1} \nabla s(\tilde{\boldsymbol{\theta}}_n) \mathbf{A}_n(\tilde{\boldsymbol{\theta}}_n)^{-1} \nabla l_n(\tilde{\boldsymbol{\theta}}_n | \mathbf{X}) \stackrel{d}{\rightarrow} \chi_r^2, \end{aligned}$$

dove la convergenza in distribuzione vale sotto l'ipotesi nulla. Si può dimostrare che $(W_n - LM_n) \xrightarrow{p} 0$ al divergere di n . Ancora, quando il modello è correttamente specificato $\mathbf{C}_n(\hat{\boldsymbol{\theta}}_n)$ è sostituito da $\mathbf{A}_n(\hat{\boldsymbol{\theta}}_n)$, e la statistica LM assume la sua forma usuale.

Quando l'ipotesi da testare è che uno o più parametri di $\boldsymbol{\theta}^*$ siano nulli, il test LM è conveniente in quanto solo la verosimiglianza del modello più semplice, cioè quello privo dei parametri sotto test, è da stimare.

18.2.2. Test ARCH-LM. Sia ε_t una serie storica osservabile, oppure il residuo rispetto ad un modello di regressione stimato consistentemente e si voglia testare l'ipotesi nulla di assenza di eteroschedasticità condizionata autoregressiva:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0,$$

con α_i parametri di un modello ARCH(q).

Si può dimostrare che, se il modello vincolato è quello “vero”, il test LM è asintoticamente equivalente al seguente test, di pratica implementazione.

- (1) Si regredisca ε_t^2 su ε_{t-i}^2 , $i = 1, 2, \dots, q$ e una costante.
- (2) Si calcoli $n \cdot R^2$, dove R^2 è il coefficiente di determinazione della regressione ausiliaria calcolata al punto 1. e n la numerosità campionaria.
- (3) Si confronti la statistica $n \cdot R^2$ con i valori critici di una $\chi^2(r)$, a cui $n \cdot R^2$ converge in distribuzione.

Tale test prende il nome di ARCH-LM e può essere applicato anche ai residui standardizzati di un modello GARCH, per verificare la presenza di ulteriore eteroschedasticità condizionata autoregressiva.

Estensioni della modellistica GARCH

19.1. Modelli GARCH asimmetrici

Un fatto empirico delle serie storiche finanziarie notato già negli anni '70 è che notizie negative, cioè shocks che influiscono negativamente sul prezzo di un titolo azionario, hanno un impatto sulla volatilità maggiore rispetto a notizie positive. Dal punto di vista della teoria economica questa proprietà delle serie dei rendimenti dei titoli azionari può essere attribuita alla considerazione che al diminuire del valore dell'azione, cresce il rapporto (debito dell'azienda)/(valore azione), che è un parametro utilizzato spesso dagli investitori per valutare il rischio di detenere azioni di quell'azienda. Tale evidenza empirica, nota in letteratura con il nome di *leverage effect*, tuttavia non è colta dai modelli GARCH fino ad ora esaminati, infatti l'impatto di ε_{t-1} su σ_t^2 è governato da $\alpha_1 \varepsilon_{t-1}^2$, che non dipende in nessun modo dal segno di ε_t . In questa sezione si mostreranno alcuni modelli a eteroschedasticità condizionata autoregressiva, che consentono di modellare l'effetto asimmetrico che shock positivi e negativi hanno sulla volatilità.

Uno strumento assai utile nel valutare il *leverage effect* consentito dai modelli GARCH asimmetrici, che saranno fra poco trattati, è la *News Impact Curve* (NIC), definita come la funzione $NIC : \mathbb{R} \mapsto \mathbb{R}^+$, di ε_{t-1} che produce un determinato valore di σ_t^2 , fissate le altre variabili del modello ai loro valori stazionari. Per il modello GARCH(1,1) è facile vedere che

$$NIC_{GARCH}(\varepsilon_{t-1}) : \sigma_t^2 = A + \alpha_1 \varepsilon_{t-1}^2 \text{ con } A = \omega + \beta_1 \sigma^2$$

dove $\sigma^2 = \omega/(1 - \alpha_1 - \beta_1)$ è la varianza non condizionata (o stazionaria, o di lungo periodo) del processo GARCH. In figura 1 è rappresentata la NIC di un processo GARCH(1,1), che è evidentemente simmetrica rispetto al punto $\varepsilon_{t-1} = 0$.

19.1.1. TGARCH. Un modello GARCH asimmetrico molto semplice è il *Threshold GARCH* (detto anche TGARCH o TARCH). Il modello TGARCH è definito da

$$\begin{aligned} \varepsilon_t | \mathcal{I}_{t-1} &= \sigma_t \xi_t \quad \xi_t \sim \text{i.i.d.}(0, 1) \\ \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \gamma \varepsilon_{t-1}^2 \mathbb{I}_{(-\infty, 0)}(\varepsilon_{t-1}) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \end{aligned}$$

dove $\mathbb{I}_A(x)$ è una funzione-indicatore, così definita

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{altrimenti} \end{cases},$$

con A sottoinsieme dello spazio di cui x è elemento. Il modello è asimmetrico rispetto alle informazioni, perché quando ε_{t-1} è positivo (informazione positiva), l'impatto sulla volatilità σ_t^2 è determinato dal parametro α_1 , mentre quando ε_{t-1} è negativo (informazione negativa), l'impatto su σ_t^2 è determinato dal valore $\alpha_1 + \gamma$. La

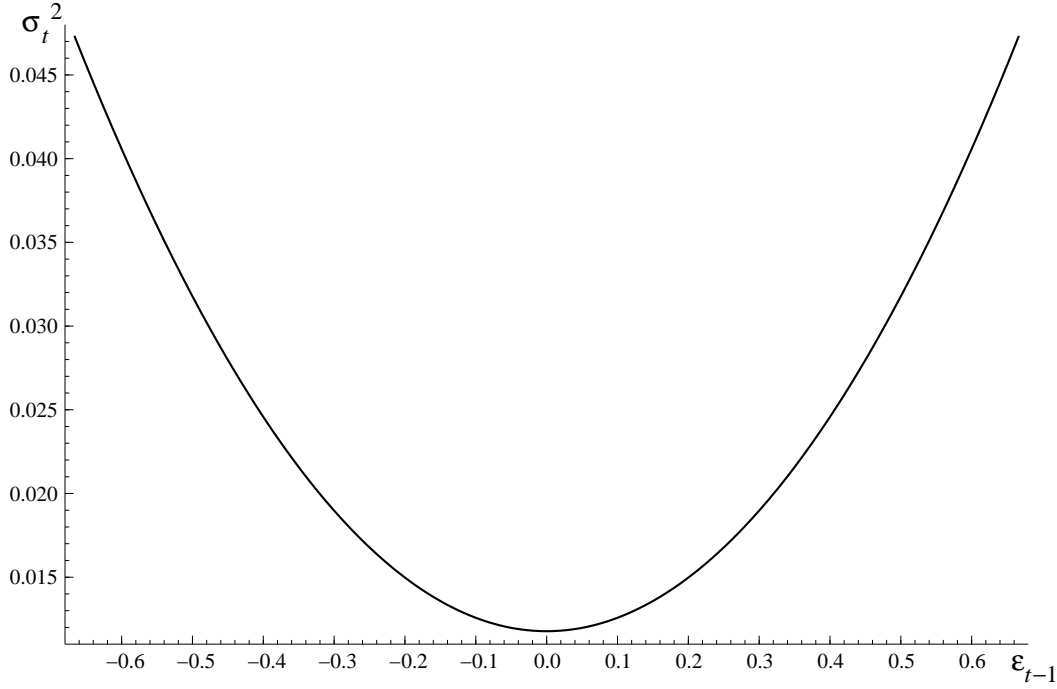


FIGURA 1. News Impact Curve di un processo GARCH(1,1) con $\alpha_1 = 0.080$, $\beta_1 = 0.910$, $\omega = 0.001$.

NIC di un modello TGARCH(1,1) è

$$NIC_{TGARCH}(\varepsilon_{t-1}) : \sigma_t^2 = A + \begin{cases} \alpha_1 \varepsilon_{t-1}^2 & \text{per } \varepsilon_{t-1} \geq 0 \\ (\alpha_1 + \gamma) \varepsilon_{t-1}^2 & \text{per } \varepsilon_{t-1} < 0 \end{cases},$$

con $A = \omega + \beta_1 \sigma^2$.

È possibile testare il *leverage effect* implementato nel modello TGARCH su di una serie storica per mezzo di un test t sul coefficiente γ .

19.1.2. EGARCH. Un processo GARCH asimmetrico con proprietà piuttosto interessanti è l'*exponential GARCH* (EGARCH)

$$\begin{aligned} \varepsilon_t | \mathcal{I}_{t-1} &= \sigma_t \xi_t & \xi_t &\sim \text{i.i.d.}(0, 1), \\ \ln(\sigma_t^2) &= \omega + \sum_{i=1}^q \alpha_i g(\xi_{t-i}) + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2), \\ g(\xi_t) &= [\phi \xi_t + \psi(|\xi_t| - \mathbb{E}|\xi_t|)], \end{aligned}$$

con $\alpha_1 = 1$. È facile vedere che la variabile casuale $g(\xi_t)$ ha media nulla, e si può dimostrare che se $\Pr(\xi_t > 0) = \Pr(\xi_t < 0) = 0.5$, allora le varie $g(\xi_t)$ ad istanti di tempo t diversi sono tra loro incorrelate, sebbene non indipendenti, e quindi $g(\xi_t)$ è un white noise in senso debole. Pertanto il processo $\{\ln(\sigma_t^2)\}$ è quindi di tipo ARMA(q, p).

Un aspetto interessante della formulazione EGARCH è che non vi è bisogno di imporre vincoli sui parametri affinché la varianza σ_t^2 sia sempre positiva, infatti l'esponenziale di un qualsiasi valore reale è funzione sempre positiva. Ciò permette al modello EGARCH di cogliere anche reazioni oscillatorie della volatilità agli shocks $g(\xi_t)$.

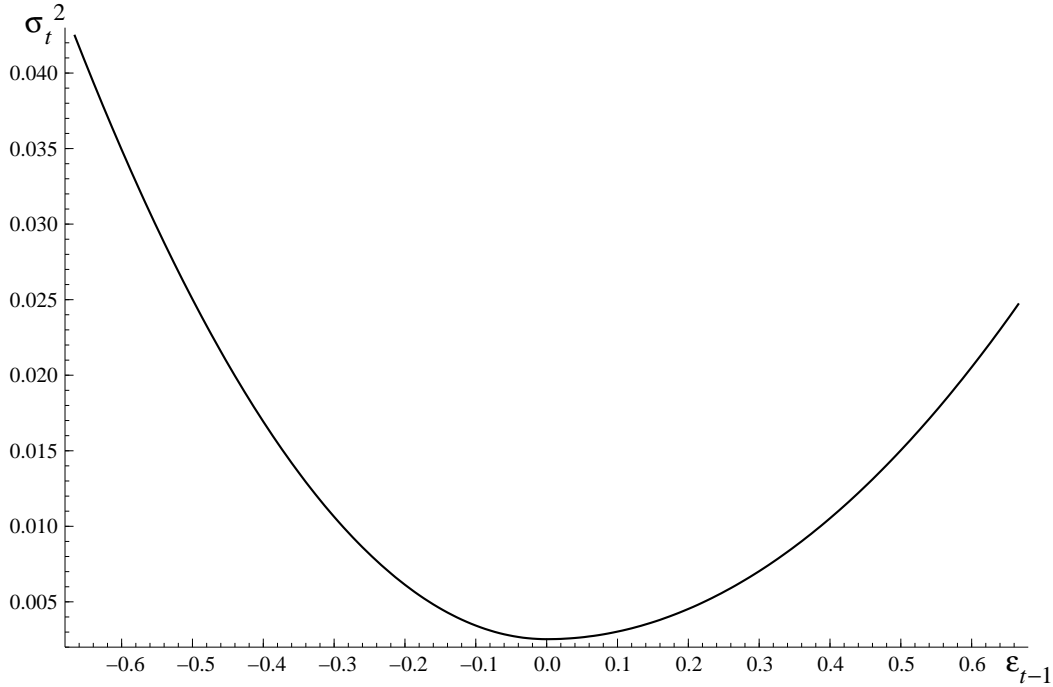


FIGURA 2. News Impact Curve di un processo TGARCH(1,1) con $\alpha_1 = 0.050, \gamma = 0.040, \beta_1 = 0.900, \omega = 0.001$.

Per valori negativi di ξ_t , $g(\xi_t)$ è funzione lineare di ξ_t con coefficiente angolare $(\phi - \psi)$, mentre per valori positivi $g(\xi_t)$ è funzione lineare di ξ_t con coefficiente angolare $(\phi + \psi)$, permettendo alla volatilità di reagire in modo asimmetrico a notizie positive e negative. In presenza di *leverage effect* ϕ è, quindi, negativo.

Quando $\phi = 0$ e $\psi > 0$, la reazione di $\ln(\sigma_t^2)$ a ξ_{t-1} è simmetrica e positiva (risp. negativa) se il modulo di ξ_t è maggiore (risp. minore) del suo valore medio. Un modello EGARCH che riproduce alcune delle evidenze empiriche delle serie storiche finanziarie ha $\phi < 0, \psi > 0$ con la distribuzione di ξ_t leptocurtica, come ad esempio la GED con parametro $r < 2$. Per applicazioni pratiche, il valore atteso di $|\xi_t|$ con ξ_t distribuita come una GED(r) è

$$\mathbb{E}|\xi_t| = \frac{\Gamma(2/r)}{\sqrt{\Gamma(1/r)\Gamma(3/r)}},$$

che nel caso di normalità ($r = 2$) si riduce a $\sqrt{2/\pi}$.

La NIC del modello EGARCH(1,1) è data da (si ricordi che $\alpha_1 = 1$)

$$NIC_{EGARCH}(\varepsilon_{t-1}) : \sigma_t^2 = \begin{cases} A \exp \left[\frac{\phi - \psi}{\sigma} \varepsilon_{t-1} \right] & \text{se } \varepsilon_{t-1} < 0 \\ A \exp \left[\frac{\phi + \psi}{\sigma} \varepsilon_{t-1} \right] & \text{se } \varepsilon_{t-1} \geq 0 \end{cases},$$

con

$$A = \sigma^{2\beta_1} \exp[\omega - \psi \mathbb{E}|\xi|].$$

19.1.3. Test di asimmetria. Una volta stimato un modello asimmetrico è naturalmente possibile testare la significatività del parametro responsabile dell'asimmetricità del modello per mezzo di un test t . Altrimenti si può costruire un test LM nel seguente modo.

- (1) Si stimi un processo GARCH(1,1) e si calcolino le serie $\hat{\sigma}_t^2$ e $\hat{\xi}_t = \varepsilon_t / \sigma_t^2$.

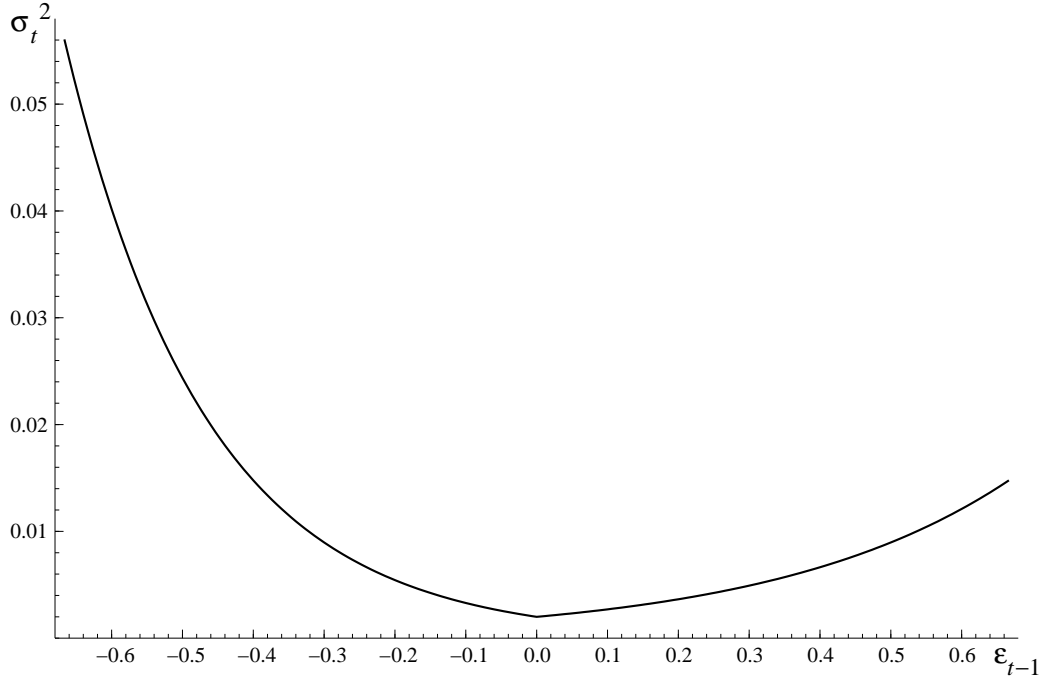


FIGURA 3. News Impact Curve di un processo EGARCH(1,1).

- (2) Si regredisca $\hat{\xi}_t^2$ su $\mathbf{z}_t = (1, \varepsilon_{t-1}^2, \sigma_{t-1}^2)'$ e uno o più indicatori di errata specificazione (indicatori di asimmetria), per esempio

$$(19.1.1) \quad \hat{\xi}_t^2 = \mathbf{z}'_t \boldsymbol{\gamma} + \mathbb{I}_{-\infty,0}(\varepsilon_{t-1})\lambda + \epsilon_t$$

oppure

$$(19.1.2) \quad \hat{\xi}_t^2 = \mathbf{z}'_t \boldsymbol{\gamma} + \mathbb{I}_{-\infty,0}(\varepsilon_{t-1})\lambda_1 + \varepsilon_{t-1} \mathbb{I}_{-\infty,0}(\varepsilon_{t-1})\lambda_2 + \varepsilon_{t-1} \mathbb{I}_{(0,+\infty)}(\varepsilon_{t-1})\lambda_3,$$

per il modello (19.1.1) l'ipotesi alternativa alla simmetria è che vi siano due valori di $\mathbb{E}(\xi_t^2)$, uno in caso di notizie positive, l'altro in caso di notizie negative; il modello (19.1.2) ingloba ulteriori due variabili adatte a modellare l'ipatto diverso che il modulo di ε_{t-1} può avere su ξ_t^2 , a seconda del segno degli shocks.

- (3) Si calcoli l' R^2 della regressione ausiliaria e lo si moltiplichi per la numerosità campionaria n . Sotto l'ipotesi nulla che il modello GARCH sia stato correttamente specificato, nR^2 della regressione (19.1.1) ha distribuzione asintotica $\chi^2(1)$ e nR^2 della regressione (19.1.2) ha distribuzione asintotica $\chi^2(3)$.

19.2. Il premio a rischio ed il modello GARCH-M

Per la teoria economica, in un mercato di investitori avversi al rischio, un'attività finanziaria più rischiosa deve fornire rendimenti attesi più alti. Questo incremento del rendimento al crescere della variabilità del titolo è detto premio a rischio. Un modello in grado di cogliere questo variare del rendimento al variare del

rischio può essere colto per mezzo del modello *GARCH in mean* o GARCH-M:

$$Y_t | \mathcal{I}_{t-1} = \nu + \delta g(\sigma_t^2) + \sigma_t \xi_t \quad \text{con } \xi_t \sim \text{i.i.d.}(0, 1)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i} + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2,$$

dove $\varepsilon_t = [Y_t - \nu - \delta g(\sigma_t^2)]$ e $g(\sigma_t^2)$ è una funzione monotona in σ_t^2 , spesso σ_t , σ_t^2 , $\ln(\sigma_t)$ o $\ln(\sigma_t^2)$. Il valore atteso di $Y_t | \mathcal{I}_{t-1}$ è

$$\mathbb{E}(Y_t) = \nu + \delta g(\sigma^2)$$

che dipende anche dalla volatilità σ_t^2 . Ovviamente il modello è generalizzabile aggiungendo altri regressori nell'equazione del livello.

Il processo $\{Y_t\}$, è debolmente stazionario quando $\{\sigma_t^2\}$ è debolmente stazionario, ed essendo $\{\sigma_t^2\}$ positivamente autocorrelato anche Y_t (se $\delta g(\sigma_t^2)$ è funzione strettamente crescente) è positivamente autocorrelato.

Modelli GARCH multivariati

20.1. Processi autoregressivi vettoriali (VAR)

Un processo autoregressivo vettoriale (VAR) è la naturale estensione multivariata di un processo AR. Sia $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{Kt})'$, con $t \in \mathbb{Z}$, una successione di vettori casuali K -dimensionali, definita su un opportuno spazio di probabilità. Il processo VAR di ordine p è definito da

$$(20.1.1) \quad \mathbf{Y}_t = \boldsymbol{\nu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t$$

dove $\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p$ sono matrici $(K \times K)$ di coefficienti di regressione, $\boldsymbol{\nu}$ è un vettore K -dimensionale di costanti e $\boldsymbol{\varepsilon}_t$ è un *white noise vettoriale*, cioè un processo con i seguenti momenti per ogni $t \in \mathbb{Z}$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma} \text{ positiva definita}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0} \quad \forall s \neq t$$

Un processo vettoriale \mathbf{Y}_t è stazionario in senso debole se per ogni $t \in \mathbb{Z}$

$$\mathbb{E}(\mathbf{Y}_t) = \boldsymbol{\mu}$$

$$\mathbb{E}[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-h} - \boldsymbol{\mu})'] = \boldsymbol{\Gamma}_Y(h) \text{ positiva definita, } \forall h \in \mathbb{Z},$$

dove le matrici $\boldsymbol{\Gamma}_Y(h)$ sono dette di autocovarianza.

Il processo VAR(p) definito in equazione 20.1.1 è stazionario in senso debole se e solo se

$$\det(\mathbf{I}_K - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_p z^p) \neq 0 \quad \text{per } |z| \leq 1,$$

che per il processo VAR(1) equivale alla condizione che tutti gli autovalori della matrice $\boldsymbol{\Phi}_1$ siano in modulo strettamente minori di 1.

Un processo VAR(p) ha sempre una rappresentazione VAR(1), infatti ponendo

$$\tilde{\mathbf{Y}}_t = \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p+1} \end{pmatrix}, \quad \tilde{\boldsymbol{\nu}} = \begin{pmatrix} \boldsymbol{\nu} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\boldsymbol{\Phi}} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_K & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_K & \mathbf{0} \end{pmatrix}$$

$$\tilde{\boldsymbol{\varepsilon}}_t = \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix},$$

si ottiene la rappresentazione VAR(1) del processo (20.1.1)

$$(20.1.2) \quad \tilde{\mathbf{Y}}_t = \tilde{\boldsymbol{\nu}} + \tilde{\boldsymbol{\Phi}} \tilde{\mathbf{Y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t$$

La condizione di stazionarietà del processo $\text{VAR}(p)$ diventa quindi: il processo $\text{VAR}(p)$ è stazionario se e solo se tutti gli autovalori della matrice $\tilde{\Phi}$ sono in modulo strettamente minori dell'unità. Il caso di autovalori in modulo maggiori di 1 genera processi esplosivi, di difficile interpretazione economica. Quando uno o più autovalori di $\tilde{\Phi}$ sono in modulo esattamente uguali a uno vi sono due possibilità: i) uno o più processi univariati $Y_{i,t}$, componenti il vettore \mathbf{Y}_t , sono integrati¹ e vanno quindi opportunamente differenziati, ii) due o più processi scalari del vettore \mathbf{Y}_t sono *cointegrati*, cioè, pur non essendo i singoli processi stazionari, esiste una combinazione lineare di tali processi, che è stazionaria. Due o più processi sono pertanto cointegrati quando esiste un trend, o andamento di lungo periodo, comune ai diversi processi. Data l'alta integrazione dei mercati finanziari, dovuta alla facilità di movimento dei capitali è lecito aspettarsi che le serie degli indici delle diverse borse valori mondiali, ma anche la gran parte dei titoli in esse quotati siano cointegrati. Esiste un'ampia letteratura sui processi cointegrati e sulla relativa inferenza a cui si rimanda per eventuali approfondimenti.

20.2. GARCH multivariati

20.2.1. Nozioni preliminari.

TEOREMA 15 (fattorizzazione di Choleski). *Per qualsiasi matrice quadrata \mathbf{A} , positiva definita, esiste un'unica matrice triangolare inferiore \mathbf{P} con diagonale principale positiva, tale che*

$$\mathbf{A} = \mathbf{P}\mathbf{P}'.$$

La maggior parte dei packages e degli ambienti di sviluppo statistico-matematici implementano routines per il calcolo di \mathbf{P} .

DEFINIZIONE 13 (operatore *vech*). *L'operatore vech trasforma una matrice $m \times m$ in un vettore di dimensione $m(m+1)/2$, incolonnando gli m vettori colonna che formano la matrice stessa, escludendo gli elementi al di sopra della diagonale principale.*

Per esempio,

$$\text{vech} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{pmatrix}.$$

20.2.2. Rappresentazione vech del GARCH multivariato. Sia $\boldsymbol{\xi}_t$, $t \in \mathbb{Z}$, un processo white noise multivariato m -dimensionale con matrice di covarianza \mathbf{I}_m , e siano $\mathbf{P}_t\mathbf{P}_t'$ le fattorizzazioni di Choleski delle matrici definite positive \mathbf{H}_t . Il processo GARCH multivariato è definito da

$$\boldsymbol{\varepsilon}_t = \mathbf{P}_t\boldsymbol{\xi}_t,$$

dove ogni elemento unico di \mathbf{H}_t dipende linearmente da ogni elemento unico delle matrici $(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}', i = 1, \dots, q$ e $\mathbf{H}_{t-j}, j = 1, \dots, p$. I momenti condizionati di $\boldsymbol{\varepsilon}_t$ sono dati da

$$\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{P}_t\mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{0}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{P}_t\mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')\mathbf{P}_t' = \mathbf{P}_t\mathbf{I}_m\mathbf{P}_t' = \mathbf{H}_t$$

¹Si ricordi che un processo integrato di ordine d è un processo non stazionario, la cui differenza d -esima è stazionaria.

Utilizzando l'operatore vech , si può scrivere

$$\text{vech}(\mathbf{H}_t) = \boldsymbol{\omega} + \sum_{i=1}^q \mathbf{A}_i \text{vech}(\boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i}') + \sum_{j=1}^p \mathbf{B}_j \text{vech}(\mathbf{H}_{t-j}),$$

con $\boldsymbol{\omega}$ vettore $[m(m+1)/2]$ -dimensionale e $\mathbf{A}_i, \mathbf{B}_j$ matrici $[m(m+1)/2 \times m(m+1)/2]$ di coefficienti.

Per meglio comprendere le formule appena viste, si consideri un modello con $p = q = 1$ e $m = 2$ nella forma vech :

$$\begin{pmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

Il modello così come è stato definito presenta alcuni problemi. Il numero dei parametri da stimare è $m(m+1)/2 + (p+q)[m(m+1)/2]^2$, che già per $p = q = 1$ e $m = 5$ risulta pari a 465 (!!!). I vincoli sui parametri da imporre affinché \mathbf{H}_t sia sempre definita positiva sono tutt'altro che banali e di complicata implementazione pratica in fase di stima. È quindi necessario porre dei vincoli sul modello, in modo tale da ridurre il numero di parametri da stimare e semplificare le condizioni di definitezza positiva di \mathbf{H}_t .

20.2.2.1. Modello vech diagonale. Se si impone la diagonalità delle matrici di coefficienti \mathbf{A} e \mathbf{B} , si ottiene il modello vech diagonale in cui la varianza i -esima, $h_{ii,t}$, dipende solo dai rispettivi $\varepsilon_{i,t-h}^2$, $h = 1, \dots, q$, e dal proprio passato e la covarianza (i, j) -esima dipende solo da $(\varepsilon_{i,t-h} \varepsilon_{j,t-h})$, $h = 1, \dots, q$, e dal proprio passato. Per $p = q = 1$ e $m = 2$:

$$\begin{pmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} + \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

Il modello contiene ora $[m(m+1)/2](1+p+q)$ parametri da stimare. Nel caso $p = q = 1$ e $m = 5$, i parametri sono 45.

20.2.2.2. Modello diagonale a correlazione condizionata costante. Una formulazione ancora più restrittiva del modello GARCH multivariato è quella che presuppone che la correlazione condizionata tra i vari elementi di $\boldsymbol{\varepsilon}_t$ rimanga sempre costante, ossia per $i = 1, 2, \dots, m$

$$\sigma_{i,t}^2 = \omega_i + \sum_{s=1}^q \alpha_{i,s} \varepsilon_{i,t-s}^2 + \sum_{r=1}^p \beta_{i,r} \sigma_{i,t-r}^2$$

e per $i \leq j$

$$\sigma_{ij,t} = \rho_{ij} \sigma_{i,t} \sigma_{j,t}.$$

L'inferenza di questo modello è piuttosto semplice (i parametri diventano $m(q+p) + m(m+1)/2$, e nel caso $p = q = 1$, $m = 5$ sono 25), ma il vincolo di correlazione costante tra i rendimenti di tutti i titoli è un grosso fardello. La definitezza positiva delle matrici \mathbf{H}_t è garantita dalle $(m^2 - m)/2$ disuguaglianze

$$-1 < \rho_{ij} < 1 \quad \text{per } i \leq j.$$

20.3. Rappresentazione BEKK del GARCH multivariato

Una rappresentazione del GARCH multivariato che permette una ricca struttura di dipendenza, rispetto ai modelli ristretti visti sopra, assicurando al contempo la definitezza positiva delle matrici \mathbf{H}_t è la forma BEKK, che prende il nome dalle iniziali dei quattro studiosi che l'hanno proposta (Baba, Engle, Kraft e Kroner):

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{i=1}^q \mathbb{A}_i \boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i}' \mathbb{A}_i' + \sum_{j=1}^p \mathbb{B}_j \mathbf{H}_{t-j} \mathbb{B}_j',$$

con \mathbf{C} matrice $(m \times m)$ triangolare inferiore con elementi sulla diagonale principale positivi, e \mathbb{A}_i e \mathbb{B}_j matrici $(m \times m)$ di coefficienti. Per garantire che non esistano diversi valori numerici dei coefficienti a cui corrisponde il medesimo processo (per esempio il processo è il medesimo per i valori $\mathbb{A}_i = \hat{\mathbb{A}}_i$ e $\mathbb{A}_i = -\hat{\mathbb{A}}_i$), o in altre parole che il modello sia identificato, è sufficiente imporre la positività degli elementi ($i = 1, j = 1$) delle matrici \mathbb{A}_i e \mathbb{B}_j .

Dato che nel modello BEKK \mathbf{H}_t è rappresentata come somma di prodotti di matrici quadrate per se stesse trasposte, se anche le matrici iniziali $\mathbf{H}_0, \mathbf{H}_{-1}, \dots, \mathbf{H}_{-p+1}$, usate per inizializzare la stima, sono positive definite, allora il modello produrrà matrici \mathbf{H}_t sempre positive definite².

Nel caso bivariato con $p = q = 1$ il modello BEKK diventa

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} c_{11} & c_{21} \\ 0 & c_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix},$$

dove gli elementi sopra (o sotto) la diagonale di \mathbf{H}_t sono ridondanti per la simmetria delle matrici di covarianza. Per facilitare la comprensione della dinamica permessa dal modello e dei vincoli parametrici imposti, la prima equazione della varianza e quella della covarianza possono essere scritte per esteso come

$$\begin{aligned} h_{11,t} &= c_{11}^2 + a_{11}^2 \varepsilon_{1,t-1}^2 + 2a_{11}a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{12}^2 \varepsilon_{2,t-1}^2 \\ &\quad + b_{11}^2 h_{11,t-1} + 2b_{11}b_{12} h_{21,t-1} + b_{12}^2 h_{22,t-1} \\ h_{21,t} &= c_{21}c_{11} + a_{11}a_{22} \varepsilon_{1,t-1}^2 + (a_{12}a_{21} + a_{11}a_{22}) \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{12}a_{22} \varepsilon_{2,t-1}^2 \\ &\quad + b_{11}b_{22} h_{11,t-1} + (b_{12}b_{21} + b_{11}b_{22}) h_{21,t-1} + b_{12}b_{22} h_{22,t-1} \\ h_{22,t} &= c_{22}^2 + a_{21}^2 \varepsilon_{1,t-1}^2 + 2a_{21}a_{22} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{22}^2 \varepsilon_{2,t-1}^2 \\ &\quad + b_{21}^2 h_{11,t-1} + 2b_{21}b_{22} h_{21,t-1} + b_{22}^2 h_{22,t-1}, \end{aligned}$$

con $h_{12,t} = h_{21,t}$. I parametri totali da stimare sono, per il modello BEKK, $m(m+1)/2 + (p+q)m^2$, che per $p = q = 1$ e $m = 5$ diventano 65.

20.4. Modelli GARCH multivariati basati sulla stima di modelli GARCH univariati

I modelli GARCH multivariati visti sopra, richiedendo la stima simultanea di un numero di parametri che esplode al crescere delle dimensioni del vettore $\boldsymbol{\varepsilon}_t$, hanno utilizzi limitati come, ad esempio, la previsione

²Si ricordi che, se \mathbf{G} è una matrice quadrata a rango pieno, allora $\mathbf{G}\mathbf{G}'$ è definita positiva.

della matrice di covarianza di piccoli portafogli (solitamente di fondi o valute piuttosto che di titoli), o l'utilizzo congiunto delle serie dei rendimenti e dei volumi scambiati per prevedere la volatilità di un titolo.

Data la relativa semplicità e le buone proprietà di convergenza della stima del GARCH univariato, molti studiosi hanno cercato di risolvere il problema della stima di un GARCH multivariato scomponendolo nella stima di diversi GARCH univariati.

20.4.1. Constant Correlation GARCH. Questo primo modello, introdotto da Bollerslev (il padre del GARCH) sfrutta la seguente rappresentazione della matrice di covarianza condizionata \mathbf{H}_t

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t,$$

dove \mathbf{D}_t è una matrice diagonale contenente le deviazioni standard condizionate

$$\mathbf{D}_t = \begin{pmatrix} \sigma_{1,t} & 0 & \dots & 0 \\ 0 & \sigma_{2,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{m,t} \end{pmatrix}$$

e \mathbf{R} è la matrice di correlazione di ε_t supposta costante per ogni t

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1m} \\ \rho_{21} & 1 & \dots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \dots & 1 \end{pmatrix}.$$

Il modello è equivalente al vech a correlazione costante, da cui eredita anche i limiti; tuttavia la stima può essere condotta in due passi:

- (1) si stimano m modelli GARCH univariati per ognuna delle serie considerate e si costruiscono le stime $\hat{\mathbf{D}}_t$ delle matrici \mathbf{D}_t riempiendo la diagonale di $\hat{\mathbf{D}}_t$ con le varianze stimate con i GARCH univariati $\hat{\sigma}_{i,t}^2$.
- (2) si stimano le correlazioni della matrice \mathbf{R} per mezzo delle correlazioni campionarie condizionate alla stima delle varianze condizionali

$$(20.4.1) \quad \bar{\rho}_{ij} = \frac{\sum_{t=1}^n (\varepsilon_{i,t}/\hat{\sigma}_{i,t})(\varepsilon_{j,t}/\hat{\sigma}_{j,t})}{\sqrt{\sum_{t=1}^n (\varepsilon_{i,t}/\hat{\sigma}_{i,t})^2 \sum_{t=1}^n (\varepsilon_{j,t}/\hat{\sigma}_{j,t})^2}}.$$

20.4.2. Dynamic Conditional Correlation GARCH. Engle (il padre dell'ARCH), di recente, ha proposto il Dynamic Conditional Correlation (DCC) GARCH, partendo dalla scomposizione della matrice di covarianza

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t,$$

dove \mathbf{D}_t è una matrice diagonale contenente le deviazioni standard condizionate e gli elementi della matrice di correlazione condizionata $\mathbf{R}_t = [\rho_{ij,t}]$ evolvono secondo una semplice trasformazione di un GARCH(1,1):

$$(20.4.2) \quad \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{i,t}q_{j,t}}}$$

con

$$(20.4.3) \quad q_{ij,t} = \nu_{ij} + \alpha \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta q_{ij,t-1}.$$

I parametri α e β vengono supposti uguali per ogni equazione, e per ridurre ulteriormente il numero di parametri da stimare si può sfruttare la relazione tra le costanti ν_{ij} e le correlazioni marginali (o non condizionali) ρ_{ij}

$$\nu_{ij} = \rho_{ij}(1 - \alpha - \beta),$$

in modo tale da poter stimare ρ_{ij} separatamente per mezzo delle correlazioni campionarie $\bar{\rho}_{ij}$.

La stima del modello viene condotta per mezzo dei seguenti tre passi:

- (1) si stimano m modelli GARCH univariati per ognuna delle serie considerate e si costruiscono le matrici \mathbf{D}_t , calcolando anche i residui standardizzati $\hat{\boldsymbol{\xi}}_t = \hat{\mathbf{D}}_t^{-1} \boldsymbol{\varepsilon}_t$,
- (2) si stimano le correlazioni marginali ρ_{ij} per mezzo delle correlazioni campionarie $\bar{\rho}_{ij}$ come da equazione 20.4.1,
- (3) condizionalmente ai risultati ottenuti nei punti precedenti, si massimizza la funzione di log-verosimiglianza gaussiana condizionata alle stime precedenti

$$L(\alpha, \beta) = -\frac{1}{2} \sum_{t=1}^n (\log |\mathbf{R}_t| + \boldsymbol{\xi}_t' \mathbf{R}^{-1} \boldsymbol{\xi}_t)$$

rispetto ad α e β , dove la relazione che intercorre tra i parametri α e β , i dati ε_t e gli elementi di \mathbf{R}_t è data dalle formule (20.4.2) e (20.4.3).

Si può dimostrare che le stime così ottenute sono consistenti.

20.4.3. Orthogonal GARCH. Un altro modo per trasformare un problema multivariato in una serie di problemi univariati è l'*ortogonalizzazione* delle variabili originali. Un modo per ortogonalizzare (rendere incorrelate) le variabili originali e al contempo ridurre la dimensionalità del problema è l'*analisi delle componenti principali*.

L'analisi delle componenti principali consiste nell'individuazione di m combinazioni lineari, dette componenti principali, delle m variabili originali tali che

- (1) le componenti principali siano tra loro incorrelate,
- (2) la prima componente spieghi il massimo della varianza delle m variabili originali, la seconda spieghi il massimo della varianza residua delle variabili originali, ecc.

Sia $\mathbf{X} = [x_{t,i}]$, con $t = 1, \dots, n$ e $i = 1, \dots, m$, la matrice contenente nelle colonne le n osservazioni standardizzate di ciascuna variabile. La standardizzazione delle variabili è necessaria affinché l'unità di misura in cui è rappresentata ciascuna variabile non influenzi il risultato. Sia

$$\mathbf{V} = \mathbf{X}'\mathbf{X}/n$$

la matrice di correlazione campionaria delle variabili originali.

La matrice delle componenti principali ordinate è data da

$$(20.4.4) \quad \mathbf{P} = \mathbf{X}\mathbf{W}$$

dove \mathbf{W} è una matrice $m \times m$ contenente gli autovettori (normalizzati) di \mathbf{V} ordinati in maniera decrescente rispetto al corrispondente autovalore.

La varianza della componente i -esima è data dal corrispondente autovalore λ_i e la somma delle varianze delle m componenti è pari a m . La proporzione di varianza spiegata dalla i -esima componente è quindi λ_i/m .

Spesso poche componenti, per es. le prime k , sono sufficienti a spiegare gran parte della varianza (90% e più) e quindi ci si limita a considerare esse nelle successive analisi.

Per ricostruire le variabili originali standardizzate, partendo dalle m componenti si sfrutta la relazione

$$\mathbf{X} = \mathbf{P}\mathbf{W}'$$

che si ottiene dalla (20.4.4) e dall'ortogonalità della matrice dagli autovettori, che implica $\mathbf{W}^{-1} = \mathbf{W}'$. Per ottenere le variabili originali non standardizzate \mathbf{Y} bisogna moltiplicare ciascuna variabile in \mathbf{X} per la relativa deviazione standard e sommare la corrispondente media

$$y_{t,i} = \mu_i + w_{i,1}\sigma_i p_{t,1} + \dots + w_{i,m}\sigma_i p_{t,m} \quad t = 1, \dots, n.$$

Quando si utilizzano solamente le prima k componenti principali si può riscrivere la precedente formula come

$$y_{t,i} = \mu_i + w_{i,1}\sigma_i p_{t,1} + \dots + w_{i,k}\sigma_i p_{t,k} + \eta_{t,i} \quad t = 1, \dots, n,$$

dove $\eta_{t,i}$ è l'errore che commetto nel ricostruire la i -esima variabile trascurando le ultime $m - k$ componenti.

Utilizzando quest'ultima rappresentazione, la matrice di covarianza $\mathbf{\Sigma}$ delle variabili originali \mathbf{y}_i è quindi data da

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}' + \mathbf{\Sigma}_\eta,$$

dove $\mathbf{A} = [w_{ij}\sigma_i]$, $\mathbf{\Lambda}$ è la matrice diagonale delle varianze delle componenti principali utilizzate e \mathbf{V}_η è la matrice di varianza degli errori, che sarà in genere trascurabile.

L'orthogonal GARCH (O-GARCH) consiste nel calcolo delle prime k componenti principali di $\boldsymbol{\varepsilon}_t$ (che in genere sono rendimenti) e nell'applicazione di k modelli GARCH univariati alle componenti. Per mezzo di tali GARCH si costruiscono le matrici diagonali delle varianze condizionate $\mathbf{\Lambda}_t$ delle componenti e si stima la matrice di covarianza condizionata delle variabili originali \mathbf{H}_t per mezzo di

$$\mathbf{H}_t = \mathbf{A}\mathbf{\Lambda}_t\mathbf{A}',$$

dove $\mathbf{A} = [w_{ij}\sigma_i]$ contiene gli autovettori riscalati con le deviazioni standard σ_i delle variabili originali.

Parte 5

Strumenti derivati sul rischio di credito

Derivativi sul Rischio di Credito.

21.1. Introduzione

In molti contratti finanziari, un aspetto primario riguarda il cosiddetto rischio di *default* o *rischio di credito*, ossia il rischio che il debitore possa non ripagare a scadenza in tutto o in parte quanto dovuto. Un'esemplificazione grafica si ha nella Figura 1, nella quale è rappresentato il *cash-flow* di un'obbligazione su cui grava il rischio di insolvenza. In anni recenti un crescente numero di strumenti finanziari apposti è stato introdotto con il preciso intento di prezzare o di ripartire il rischio di *default*. Semplificando un po' le cose vi sono essenzialmente due classi di prodotti derivati sul rischio di credito:

- (1) **Prodotti basati sul rischio di *default*.** Si tratta di attività finanziarie il cui *payoff* dipende dall'occorrenza del *default* o da altri eventi relativi alla solvibilità quali il declassamento del titolo. Ai fini di questi contratti, è ovviamente essenziale che l'evento creditizio in oggetto possa essere verificato in modo incontrovertibile da entrambe le parti. L'aspetto cruciale di questi contratti è che essi dipendono da un *evento* finanziario e non da un prezzo sottostante e sono pertanto maggiormente simili a normali polizze assicurative che non ad altri derivati. A questa classe di prodotti appartengono, tra gli altri, i *credit default swaps* e le *credit default options*. Un aspetto determinante ed assai delicato in questi contratti è la determinazione del valore di recupero o *recovery value*, che potrebbe in principio essere controverso.

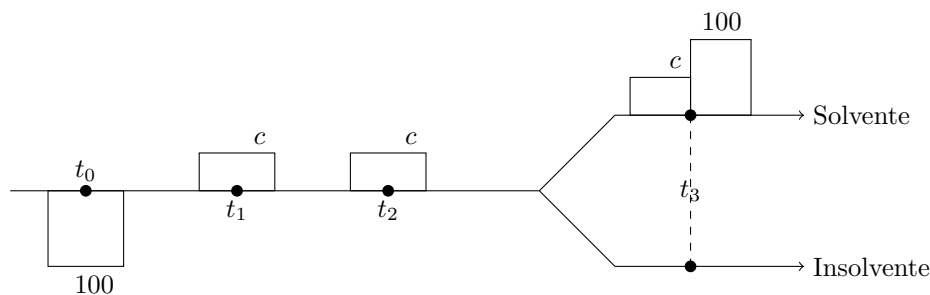


Figure 10: *Bond* con rischio di default.

- (2) **Prodotti basati su Credit Spread.** Per questo tipo di titoli, il *payoff* dipende dalle variazioni dello *spread* delle attività sottostanti rispetto ad un'attività di riferimento o *benchmark*. A questa classe appartengono, tra gli altri, le *credit spread options* – il cui *payoff* è l'eccesso del *credit spread* rispetto al prezzo di esercizio – e i *credit spread swaps* nel quale le due parti scambiano tra loro i *credit spreads* relativi a due distinti emittenti. Un altro importante derivato che appartiene a questo gruppo è il *total return swaps*.

A titolo di illustrazione di quanto detto in precedenza si consideri il seguente esempio.

ESEMPIO 9. Una banca concede un prestito di K euro per la durata di 1 anno e fissa il tasso di interesse a R . Lo scopo della banca è di ottenere un rendimento finale pari a 5% nonostante il fatto che vi sia una probabilità pari a p che l'impresa finanziata fallisca e che, in tal caso, la banca possa al massimo sperare di ricevere una frazione $0 < \delta < 1$ del fido erogato, inclusi gli interessi. In altre parole, la banca deve fissare il tasso R in modo tale che:

$$K(1 + 5\%) \leq E(K(1 * R)) = P\delta K(1 + R) + (1 - P)K(1 + R)$$

Ne segue che R deve necessariamente soddisfare la disuguaglianza

$$R \geq \frac{1 + 5\%}{1 - P(1 - \delta)} - 1 = \frac{5\% + P(1 - \delta)}{1 - P(1 - \delta)}$$

Il termine di destra di questa espressione è chiaramente più grande del 5% ed è altresì crescente rispetto a p e $1 - \delta$, il parametro che misura la cosiddetta *loss given default*.

Questo esempio chiarisce, nella sua semplicità, che il tasso di interesse richiesto dalla banca dipende inversamente dall'affidabilità dell'impresa creditrice, come certificata, ad esempio, dal *rating*. Quanto più elevata è la probabilità di *default* e/o le perdite da insolvenza, quanto maggiore sarà il tasso di interesse prevalente sul mercato dei prestiti.

La posizione lunga in un contratto derivato sul credito viene spesso indicata come *protection buyer*.

21.2. Prodotti Basati sul Rischio di *Default*.

21.2.1. Credit Default SWAP (CDS). Si consideri un'obbligazione con rischio di *default* il cui *coupon* sia pari a c rappresentata come nella Figura 1 nella quale abbiamo fatto implicitamente l'ipotesi che il tasso di recupero sia 0 e che il momento del *default* possa essere previsto correttamente. Naturalmente, come in tutti i contratti *SWAP*, i pagamenti intermedi possono essere isolati, come nella Figura 21.2.1

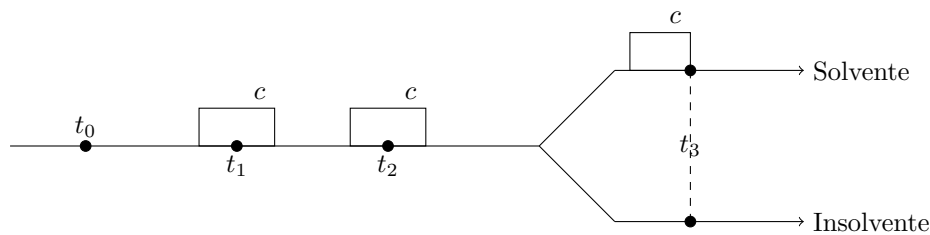
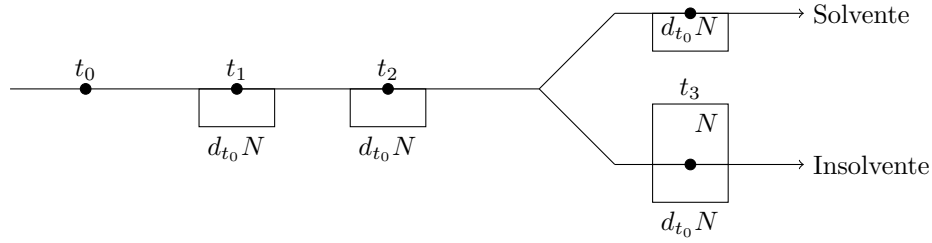
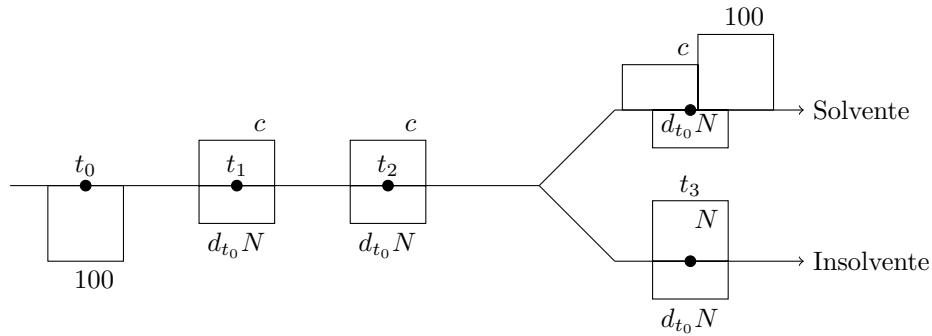


Figure 11: *CDS*.

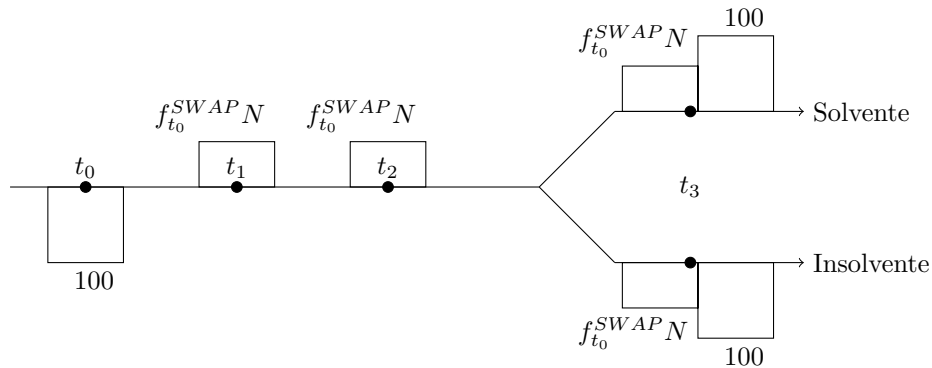
Un *CDS* è un contratto in cui il *protection buyer* si impegna ad effettuare una serie di pagamenti fissati preliminarmente, $d_{t_0}N$, proporzionali al principale N . In cambio, egli riceve in caso di *default* del titolo sottostante un pagamento monetario pari a N a fronte del quale consegna al *protection seller* gli strumenti di debito scelti da un paniere fissato di titoli ammissibili per la consegna e di valore facciale pari a N . L'idea di fondo è che il *protection buyer*, che potrebbe ad esempio avere una posizione lunga in un'obbligazione con rischio di *default*, trasferisca al venditore il bond stesso, ricevendone in cambio il suo valore facciale. Il diagramma relativo al *CDS* è rappresentato nella Figura 21.2.1

Figure 12: Il cash-flow di un *CDS*.

In un *CDS* il pagamento fisso viene in genere stabilito pari a $d_{t_0}N = c - f_{t_0}^{SWAP}N$ ossia la differenza tra il *coupon* del titolo con rischio di *default* e il tasso fisso di un classico *SWAP* su tasso di interesse moltiplicato per il valore facciale del bond sottostante. Se il detentore del bond acquista protezione, allora il suo *payoff* finale sarà come rappresentato nella Figura 21.2.1

Figure 13: *CDS*.

ossia nella Figura 21.2.1

Figure 14: *CDS*.

nella quale abbiamo posto

$$N = 100 + c$$

Come si intuisce facilmente, il *payoff* finale corrisponde a quello di un titolo privo di rischio con tasso fisso pari a $f_{t_0}^{SWAP}$ e valore facciale pari a 100. In altre parole, combinando l'obbligazione con rischio di *default* e un *CDS* abbiamo ottenuto lo stesso risultato che si avrebbe acquistando un deposito a tasso variabile più uno *SWAP* sul tasso di interesse. Abbiamo cioè dimostrato l'equazione contrattuale

$$\text{Obbligazione rischiosa} + \text{CDS} = \text{IRS} + \text{Deposito privo di rischio}$$

21.3. Total Return SWAPS (TRS).

Un *TRS* è un normale *SWAP* come ad esempio l'*equity SWAP*. Una delle due parti contraenti riceve un tasso variabile quale il *LIBOR* più uno spread e, in cambio, paga il rendimento complessivo scaturito dal titolo sottostante, vuoi sotto forma di pagamenti in conto reddito vuoi sotto forma di *capital gain*. Il fatto che i *capital gains* siano anch'essi pagati dall'acquirente introduce una differenza essenziale con i più tradizionali *IRS* nei quali solo i *coupons* vengono scambiati.

La parte corta di questo contratto viene chiamata *total return payer*; quella lunga *total return buyer*. Il contratto risulta in genere interessante per le banche che lo usano per gestire il rischio derivante dai fidi erogati. Il contratto si estingue se giunge a scadenza oppure se il credito sottostante è interessato da un evento creditizio pre specificato. Il pagamento finale può avvenire in termini monetari o fisici, tramite il trasferimento della proprietà di un titolo. Nel primo caso, il *total return payer* paga l'incremento percentuale di valore del titolo moltiplicato per il principale. Naturalmente, se questo pagamento fosse negativo, ciò corrisponde ad un pagamento che il *payer* riceve dalla controparte. Nel caso di consegna fisica, il *total returns payer* deve consegnare il titolo sottostante includendo ogni forma di liquidità ed ogni titolo debitorio secondario in cambio di NS_{t_0} , il capitale principale moltiplicato per il valore iniziale dell'attività. Quando la regolazione del pagamento è monetaria, in genere il valore finale dell'asset viene determinata da una terza parte (spesso un *financial advisor*).

A titolo di esempio si consideri una banca che ha concesso un prestito ad un'impresa. Naturalmente tale fido implica un rischio di *default* per coprirsi dal quale la banca potrebbe acquistare un *TRS* che la impegna a pagare il tasso *LIBOR* + 120 bpts ricevendone in cambio il *Libor*+40pbts. Lo spread tra i due tassi nasce dal rischio di *default* insito nel fido. Infatti, se l'impresa va in fallimento prima di aver ripagato il finanziamento la banca non sarà ora più tenuta a fare pagamenti alla controparte nello *SWAP*.

ESEMPIO 10. Si consideri la posizione di un investitore che ha investito 1 milione di euro in un'obbligazione che paga una cedola pari al 4% ed il cui valore iniziale era $P = 100$. L'investitore partecipa ad un *TRS* in cui il *total return receiver* deve pagare il *LIBOR* + 1%. La seguente tabella illustra i dettagli del contratto. La prima colonna indica le variazioni del prezzo dell'obbligazione.

<i>Data</i>	<i>Prezzo dell'obbligazione</i>	<i>LIBOR</i>	<i>Coupon</i>	<i>TRP</i>	<i>TRR</i>	<i>Pagamento netto</i>
t	100	2.5%	4%	40,000	35,000	5,000
$t+1$	101	3%	4%	50,000	40,000	10,000
$t+2$	30	2.5%	4%	70,000	1,035,000	-965,000

21.4. Credit Spread Options, Futures e SWAPS.

Uno *spread* sui tassi di interesse indica il rendimento aggiuntivo richiesto dal mercato per sostenere i rischi impliciti in una determinata attività finanziaria. I prodotti basati sul credit spread, quali i *futures* o le opzioni sul credit spread implicano un tasso di interesse di riferimento ed un titolo specificato dall'inizio.

Lo spread viene quindi calcolato come la differenza tra il rendimento di tale titolo e il tasso di riferimento. Quest'ultimo tasso non deve necessariamente essere privo di rischio, anche se è così in molti casi concreti.

In un *credit spread SWAP* (*CSS*) vi sono tipicamente due strutture tra loro alternative: il *CSS* assoluto o relativo. Nel primo caso una delle due parti paga uno spread fisso concordato tra i contraenti all'inizio del contratto e l'altra paga la differenza tra un titolo sottostante ed un'attività priva di rischio specificata anch'essa nel contratto. Nel secondo tipo di contratti, una delle due controparti paga lo spread tra il titolo 1 e il titolo privo di rischio e riceve in cambio lo spread tra il titolo 2 e quella stessa attività priva di rischio. In molti contratti il tasso privo di rischio prescelto è il *LIBOR*.

Contratti forward e opzioni sullo spread sono più o meno simili ai forward e alle opzioni standard ad eccezione del fatto che il prezzo di esercizio (o di consegna) viene espresso in termini di uno spread. Per tradurre numericamente lo spread in un prezzo, si applica un particolare moltiplicatore, detto il fattore *duration*. Ne segue che il *payoff* finale di un, diciamo, contratto forward è $DN(S_T^{spr} - F(t, T))$ dove D è la *duration*, N il principale, T la scadenza, S^{spr} il tasso di mercato di riferimento and $F(t, T)$ il prezzo (spread) di consegna, fissato alla data t .

21.5. Credit Linked Notes.

Le Credit Linked Notes sono attività finanziarie (generalmente obbligazioni) che incorporano un qualche derivato sul credito allo scopo di trasferire in tutto o in parte il rischio di credito implicito. In genere si distingue tra:

- (1) titoli strutturati;
- (2) titoli ;
- (3) obbligazioni sintetiche;
- (4) cartolarizzazione di portafogli di obbligazioni

21.6. Obbligazioni Strutturate

Si tratta di titoli a reddito fisso combinati con qualche derivato di credito come il *CDS* o *CSO*. Questi prodotti offrono all'investitore la garanzia vuoi della cedola vuoi del valore di mercato del titolo. In particolare, le *Credit Default Linked Notes* consentono di ridurre in modo significativo l'esposizione al rischio di *default*. In generale questi strumenti generano la possibilità di partecipare al *cash-flow* del titolo, quale il rendimento di un prestito bancario, senza tuttavia dover necessariamente prendere posizione nel titolo stesso. Inoltre, la protezione verso il rischio di credito offerta dai derivati sul credito inclusi in questi prodotti consente la possibilità di ottenere dei *rating* anche per mercati che tradizionalmente ne sono privi.

L'investitore riceve una cedola quale il *LIBOR* più uno *spread* e, in assenza di eventi creditizi, il valore facciale del titolo a scadenza. Tuttavia se vi è un evento di *default*, come l'incapacità di pagare la cedola o di rimborsare il valore facciale, la bancarotta o una ristrutturazione debitoria, l'investitore riceve il valore facciale al netto del valore di recupero o della consegna del titolo.

21.7. Repackaged Notes

Questi prodotti finanziari sono stati introdotti in tempi relativamente recenti ma hanno subito guadagnato una certa popolarità. Lo scopo principale di chi investe in questi titoli è quello di accrescere la propria liquidità il che li rende particolarmente interessanti per quelle istituzioni, quali le banche, titolari di grossi portafogli di passività. Questi prodotti si sono rivelati essere piuttosto utili nell'attrarre gli investitori verso

investimenti che tradizionalmente erano considerati piuttosto illiquidi. Inoltre essi hanno reso più facile soddisfare i requisiti patrimoniali.

La cartolarizzazione delle attività finanziarie è un tecnica di ingegneria finanziaria che si è sviluppata con l'introduzione degli *SWAP*, il cui scopo principale è quello di assemblare caratteristiche di titoli già esistenti per formare nuovi titoli.

ESEMPIO 11. *Nel 1985 Merrill Lynch ha intuito il forte interesse del mercato del reddito fisso in UK per obbligazioni a tasso fisso del Tesoro denominati in dollari che a quel tempo non esistevano sul mercato. Per completare il mercato con il lancio di una nuova attività finanziaria, Merrill Lynch acquistò \$ 100 milioni di titoli a tasso variabile denominati in dollari, emessi dal governo britannico. Contemporaneamente venne creata una società indipendente (special purpose vehicle SPV) allo scopo di trattare una nuova attività finanziaria, denominata MES. Merrill Lynch vendette quindi a MECS i titoli a tasso variabile e organizzò per conto di questa uno SWAP con la Prudential Global Funding Corporation. L'accordo implicava l'obbligo per MECS di pagare ogni trimestre il tasso LIBID a Prudential ricevendone in cambio un tasso annuo fisso pari a 9.75%. Questa operazione consentì di tradurre il rendimento variabile denominato in dollari in una cedola fissa, anch'essa in dollari, avendo il titolo originale come collaterale. Molte operazioni simili seguirono negli anni successivi.*

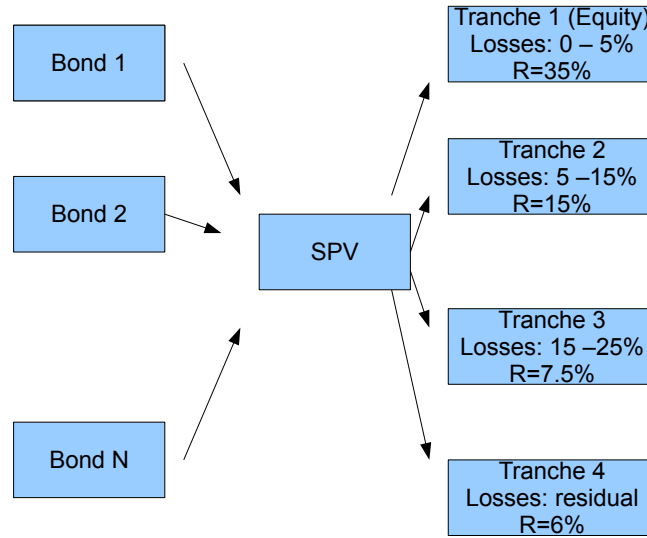
Molte operazioni di *repackaging* hanno preso da questa prima esperienza la stessa struttura di fondo: l'acquisto delle attività primarie sul mercato secondario, una serie di operazioni con derivati allo scopo di trasformare i *cash-flow* nel modo opportuno, l'emissione di titoli che rappresentino la nuova struttura del *cash-flow*. A seconda dei titoli originari, questo tipo di attività viene in genere denominato *Collateralized Debt Obligations (CDO)*, o *mortgage backed securities (MBS)* o più generalmente *Asset Backed Securities (ABS)*. Il primo esempio di *MBS* si ebbe nel 1977 con un'emissione di *Bank of America* su di un sottostante portafoglio di mutui. Un aspetto importante riguarda il ruolo svolto dalla società che eredita dall'*originator* l'insieme dei titoli di riferimento, obbligazioni o prestiti. Questa società può alternativamente essere uno *Special Purpose Vehicle*, come nel caso di Merrill Lynch, che emette veri e propri titoli, oppure un *Trust* che emette certificati. Nel primo caso la proprietà dei titoli sottostanti rimane nelle mani del *SPV* e i titoli emessi da quest'ultima vengono spesso chiamati titoli *pay-through*. Nell'esempio di Bank of America, il Trust si incarica solo della gestione dei titoli originari per conto degli investitori finali i quali sono a tutti gli effetti i proprietari di questi ultimi. I titoli emessi vengono in tal caso denominati titoli *pass-through*.

Uno schema base è quello illustrato nella Figura 1.

In questo esempio grafico, ogni *tranche* corrisponde a una percentuale del valore di mercato del portafoglio sottostante, 5%, 10%, 10% e 75%. Le tranches differiscono tra loro per il grado di *seniority*: l'ultima è quella più senior ed infatti è toccata da perdite solo se il portafoglio sottostante subisce una perdita superiore al 25%. La prima tranche è quella meno privilegiata ed in genere viene denominata *equity* in quanto essa perde l'intero valore se solo il portafoglio subisce una perdita del 5% del suo valore nominale. Nella maggioranza dei casi il creditore originario non vende l'*equity* sul mercato a causa dell'elevato che sarebbe necessario promettere per rendere un tale investimento appetibile per i risparmiatori.

Un semplice modello consente di chiarire alcuni punti. Normalizziamo a 1 il valore facciale di un certo titolo. Con $0 < \delta < 1$ indichiamo il tasso di recupero: in caso di *default* la banca recupera δ altrimenti 1. La probabilità di *default* è di $1 - \pi$:

π	$1 - \pi$
1	δ

FIGURA 1. *Collateralized Debt Obligation*

Assumendo neutralità al rischio, il valore dell'asset per la banca ammonta a

$$V = \pi + \delta(1 - \pi)$$

Gli investitori privati assegnano la medesima probabilità all'evento di *default* ma sono più pessimisti sul tasso di recupero, ceh reputano essere $0 < \delta_0 < \delta$. Sempre sotto l'ipotesi di neutralità al rischio, il valore del titolo per l'investitore ammonta a

$$\begin{aligned} V_I &= \pi + \delta_0(1 - \pi) \\ &= \pi + \delta(1 - \pi) - (\delta - \delta_0)(1 - \pi) \\ &= V - (\delta - \delta_0)(1 - \pi) \end{aligned}$$

Dunque il costo L di liquidazione è

$$L = V - V_I = (\delta - \delta_0)(1 - \pi)$$

In sostanza il costo di liquidazione è tanto più alto quanto maggiore è la probabilità di *default* e quanto minore è il tasso di recupero stimato dall'investitore.

Supponiamo ora che il creditore originario metta in piedi un'operazione strutturata nella quale il titolo originario viene suddiviso in due tranches, una *senior* e l'altra *junior*. La diversa seniority, dovuta ai diversi strumenti di garanzia impliciti, determina due diversi parametri di recupero per ciascuna tranche: $\delta_s > \delta > \delta_j = 0$. Di conseguenza il valore di mercato per le due tranches, la cui quota è pari a α e $1 - \alpha$ rispettivamente

$$V_s^\alpha = \alpha[\pi + (1 - \pi)\delta_s]$$

dove α è la quota della tranche senior. Il valore V_j dell'altra tranche è dunque

$$V_j^\alpha = (1 - \alpha)\pi$$

Il valore totale dell'emissione così strutturata ammonta a

$$\begin{aligned} V^\alpha &= V_s^\alpha + V_j^\alpha \\ (21.7.1) \qquad &= \pi + (1 - \pi)\alpha\delta_s \end{aligned}$$

Il costo di liquidazione $L^\alpha = V_C - V^\alpha$.

Supponiamo che l'operazione si possa realizzare con

$$\alpha = \frac{\delta}{\delta_s}$$

Si ha allora

$$\begin{aligned} V^\alpha &= \pi + (1 - \pi)\alpha\delta_s \\ &= \pi + (1 - \pi)\frac{\delta}{\delta_s}\delta_s \\ &= \pi + (1 - \pi)\delta \\ &= V \end{aligned}$$

In altre parole, il metodo di emissione consente al creditore originario di recuperare il valore intero del titolo senza incorrere in alcun costo di liquidazione.

Parte 6

Modelli binomiali di prezzo.

Il modello

22.1. Alberi

Fissiamo la terminologia di base e la notazione che utilizzeremo nei capitoli successivi. L'idea di fondo è estremamente semplice: si tratta di modellare l'evoluzione graduale dell'incertezza di modo che ad ogni stadio non vi è che un numero finito di possibili scenari da prendere in esame. In particolare siamo interessati allo schema dei cosiddetti *alberi binomiali* nei quali in ogni possibile stato del mondo vi sono solamente due possibili esiti. In particolare, poiché l'incertezza alla quale siamo interessati è quella relativa all'andamento del valore di mercato di un qualche titolo finanziario, immaginiamo che ad ogni stadio possa esservi solo un rialzo od un ribasso, contrassegnati rispettivamente dall'indice u e d . Una illustrazione grafica assai facile ed intuitiva è quella fornita dalla Figura sottostante

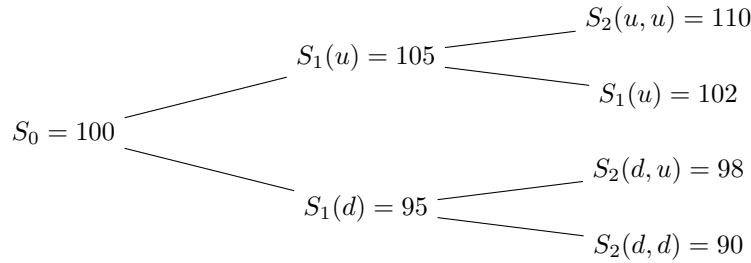


Figure 1.: Albero binomiale non ricombinante.

In questo semplice esempio ci sono quattro possibili esiti: $\Omega = \{(u, u), (u, d), (d, u), (d, d)\}$ ed Ω è per l'appunto lo spazio degli stati. Possiamo dunque scrivere in modo più compatto $S_t(\omega)$ per indicare il prezzo alla data t in quell'unico nodo compatibile con l'esito finale ω . Ad esempio, scriviamo $S_1(u, u) = S_1(u, d) = 105$.

Un formalismo assai conveniente è poi quello degli alberi ricombinanti, in termini dei quali $S_2(u, d) = S_2(d, u)$ ovvero l'esito finale è il medesimo che vi sia stato un rialzo seguito da un ribasso o viceversa ossia indipendentemente dall'ordine con cui gli eventi *rialzo*, *ribasso* si alternano. Abbiamo pertanto

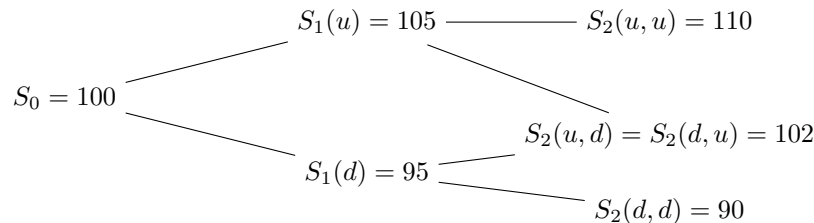


Figure 2.: Albero binomiale ricombinante.

Inoltre è utile supporre che la percentuale di aumento o di ribasso sia la medesima per tutte le date. Le indichiamo in tal caso con u e d anch'esse. Da ora in poi ci interesserà esclusivamente il caso di alberi binomiali ricombinanti ed omogenei. Naturalmente, possiamo costruire alberi binomiali con un qualunque numero di stadi. Poiché ad ogni stadio il numero di nodi aumenta di 1, con due stadi si avranno 3 nodi finali, con T stadi se ne hanno $T + 1$. Dunque, quanto maggiore è il numero di stadi tanto più ricca sarà la descrizione dei possibili scenari futuri. In tal modo, qualunque variabile aleatoria che assuma un numero finito di possibili valori può essere descritta adeguatamente da un corrispondente albero, mentre variabili aleatorie arbitrarie possono essere *approssimate* a piacere. Il fatto importante è che tale capacità di approssimazione dipende in realtà da un solo parametro, dato che $ud = 1$.

22.2. I mercati

Relativamente ai mercati introduciamo una serie di ipotesi, alcune delle quali piuttosto forti, alcune generali, altre specifiche.

22.2.1. Ipotesi generali. Ci atterremo alle seguenti ipotesi:

Mercati perfetti. Ciò significa che ciascun investitore può a proprio piacere comprare o vendere qualunque ammontare di una data attività. Non ci sono dunque vincoli di alcun tipo nell'espletare le transazioni.

Assenza di frizioni. L'acquisto e la vendita di un titolo non implica per l'investitore altro costo o guadagno che il prezzo del titolo stesso. In particolare non ci sono costi di transazione né tasse od altre frizioni quali il *bid-ask spread*.

Naturalmente nel modo reale questi aspetti non solo sono presenti ma spesso hanno anche una certa importanza. La ragione per la quale si introducono queste ipotesi è che risulta estremamente difficile ottenere dei risultati facilmente trattabili se si volesse tener conto di tali possibili imperfezioni.

Assenza di dividendi. In generale considereremo solo titoli che non pagano alcun dividendo.

Il titolo finanziario sarà indicato con il processo stocastico S che ne descrive il prezzo. Si tratterà in generale di un titolo rischioso nel senso che S_t dipenderà dallo stato del mondo $\omega \in \Omega$.

Titolo privo di rischio. Assumiamo che esista un'obbligazione priva di rischio, il cui prezzo alla data t sarà indicato dal simbolo B_t . Nella realtà possiamo identificare un tale titolo con un qualche *bond* con scadenza breve, quale un buono del Tesoro a tre mesi. Il fatto che sia privo di rischio si traduce nella circostanza che B_t è indipendente da ω . Il tasso di sconto è importante e pertanto scriveremo

$$(22.2.1) \quad B_{t,T} = \frac{B_T}{B_t} \quad T > t$$

Sappiamo dalla prima parte di queste note che sussiste una relazione diretta tra il prezzo di un'obbligazione ed il tasso di interesse. In particolare, $B_{t,T} = (1 + r)^{T-t} = \exp(r^c(T - t))$ a seconda che si utilizzi la capitalizzazione discreta o quella continua.

Default. Non ammetteremo che una delle due parti implicate in una transazione finanziaria possa non coprire i propri debiti. Dunque l'alea implicita in una transazione finanziaria è puramente oggettiva nel senso che non ha a che fare con il comportamento della controparte.

Completezza. I mercati finanziari saranno completi.

Ciò significa, semplicemente, che ad ogni nodo i rendimenti del titolo rischioso e del titolo privo di rischio, considerati come un vettore con indici u e d a seconda che vi sia un aumento o una riduzione, sono

linearmente indipendenti e dunque sufficienti a riprodurre il valore di ogni altra variabile aleatoria. In effetti abbiamo:

$$\begin{bmatrix} S_t(\omega) \\ B_t(\omega) \end{bmatrix} = \begin{bmatrix} u & 1+r \\ d & 1+r \end{bmatrix} \begin{bmatrix} S_{t-1}(\omega) \\ B_{t-1}(\omega) \end{bmatrix} = \mathbf{R} \begin{bmatrix} S_{t-1}(\omega) \\ B_{t-1}(\omega) \end{bmatrix}$$

dove la matrice

$$\mathbf{R} = \begin{bmatrix} u & 1+r \\ d & 1+r \end{bmatrix}$$

dei rendimenti uniperiodali ha rango pieno se e solo se i mercati sono completi.

Quello che segue è l'esempio di un mercato *non* completo

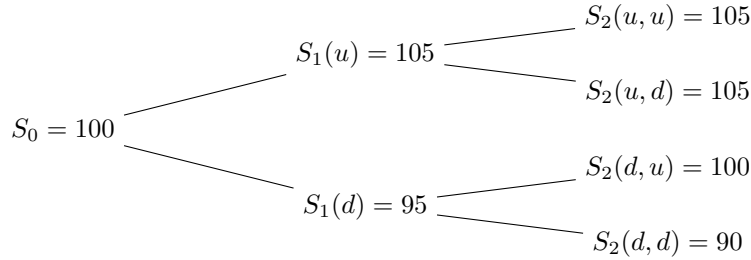


Figure 3.: Mercati incompleti.

Arbitraggio. Non vi è la possibilità di compiere profitti di arbitraggio.

Questo semplice principio sarà l'elemento determinante di tutto quel che segue. Il suo significato è semplicemente che se vi sono due titoli che hanno esattamente le stesse caratteristiche economiche allora dovranno anche avere il medesimo prezzo, poiché altrimenti, vendendo il titolo con prezzo maggiore e comprando quello con prezzo minore si potrebbe conseguire un guadagno certo, ossia privo di ogni sorta di rischio: si realizzerebbe in altre parole un'opportunità di arbitraggio.

Nel nostro semplice schema binomiale questo principio si traduce nella doppia disuguaglianza

$$(22.2.2) \quad u \geq B_{t,t+1} \geq d$$

per ogni t , dove le disuguaglianze sono vuoi entrambe strette oppure entrambe due uguaglianze. Un mercato finanziario è dunque completo e privo di possibilità di arbitraggio se e solo se:

$$(22.2.3) \quad u > B_{t,t+1} > d$$

22.2.2. Portafogli. Un portafoglio è una funzione θ tale che per ogni $\omega \in \Omega$ e t

$$\theta(\omega, t) = \begin{bmatrix} \delta(\omega, t) \\ \beta(\omega, t) \end{bmatrix}$$

dove $\delta(\omega, t)$ è l'ammontare acquistato (venduto, se una quantità negativa) del titolo rischioso e $\beta(\omega, t)$ l'ammontare del titolo privo di rischio. Il valore di mercato del portafoglio lo indichiamo con V_t^θ ed è semplicemente

$$(22.2.4) \quad V_t^\theta = [S_t, B_t] \cdot \theta_t = S_t \delta_t + B_t \beta_t$$

L'acquisto del portafoglio θ alla data $t-1$ con lo scopo di rivenderlo alla data successiva produce un guadagno pari a

$$(22.2.5) \quad G_t^\theta = [S_t - S_{t-1}, B_t - B_{t-1}] \cdot \theta_{t-1} = (S_t - S_{t-1})\delta_{t-1} + (B_t - B_{t-1})\beta_{t-1}$$

Quest'ultimo processo G^θ viene in genere chiamato processo dei profitti di scambio associato al portafoglio θ . Ovviamente

$$G_t^\theta = [S_t, B_t] \cdot \theta_{t-1} - [S_{t-1}, B_{t-1}] \cdot \theta_{t-1} = [S_t, B_t] \cdot \theta_{t-1} - V_{t-1}(\theta)$$

Analogamente possiamo definire la quantità

$$(22.2.6) \quad F_t^\theta = [S_t, B_t] \cdot (\theta_t - \theta_{t-1})$$

che descrivi i costi dovuti al ribilanciamento del portafoglio effettuato alla data t , ossia i finanziamenti aggiuntivi necessari per finanziare la strategia θ .

$$\begin{aligned} V_t^\theta &= [S_{t-1}, B_{t-1}] \cdot \theta_{t-1} + [S_t - S_{t-1}, B_t - B_{t-1}] \cdot \theta_{t-1} + [S_t, B_t] \cdot (\theta_t - \theta_{t-1}) \\ (22.2.7) \quad &= V_{t-1}^\theta + G_t^\theta + F_t^\theta \\ &= V_0^\theta + \sum_{i=1}^t G_i^\theta + \sum_{i=1}^t F_i^\theta \end{aligned}$$

Tra tutte le strategie di investimento, un interesse particolare l'hanno le cosiddette strategie *auto-finanziate*, ossia quelle strategie θ per le quali $F_t^\theta = 0$ per ogni t . Ogni strategia cosiddetta *buy-and-hold* è di questa natura: per essere auto-finanziato un deve essere tale da poter essere aggiustato senza ulteriori costi.

Talvolta conviene esprimere i prezzi in termini scontati: indicheremo i prezzi scontati con un segno sopra

$$\bar{S}_t = \frac{S_t}{B_t}, \quad \bar{V}_t^\theta = \bar{S}_t \delta_t + \beta_t, \quad \bar{G}_t^\theta = (\bar{S}_t - \bar{S}_{t-1}) \delta_{t-1}, \quad \bar{F}_t^\theta = [\bar{S}_t, 1] \cdot (\theta_t - \theta_{t-1})$$

Vale dunque l'analogo della (22.2.7) ma in termini scontati, cioè:

$$(22.2.8) \quad \bar{V}_t^\theta = V_0^\theta + \sum_{i=1}^t \bar{G}_i^\theta + \sum_{i=1}^t \bar{F}_i^\theta$$

Il principio di valutazione

Lo scopo di questo capitolo è quello di descrivere due modi equivalenti di affrontare il problem del *pricing* ossia di come stabilire il prezzo di un'attività finanziaria e di calcolare in modo esplicito tale prezzo in alcuni casi determinati. Faremo sempre l'ipotesi che i mercati siano completi e privi di opportunità di arbitraggio.

23.1. Il prezzo di replica

Come abbiamo già visto nel contesto degli alberi binomiali la completezza dei mercati e la mancanza di opportunità di arbitraggio sono verificate se e solo se

$$d < B_{t-1,t} < u$$

Grazie alla proprietà di completezza qualunque *contingent claim* X ammette una strategia $\theta(X)$ tale che: $V_T^\theta(X) = X$ o, equivalentemente, $\bar{V}_T^\theta(X) = \bar{X}$. Inoltre, la strategia $\theta(X)$ può essere del tipo autofinanziato. Di conseguenza, acquistare il *contingent claim* X e pagare il prezzo di mercato $\pi(X)$ fornisce all'acquirente esattamente gli stessi diritti che acquistare $\theta(X)$, al prezzo $V_0^\theta(X)$. Pertanto, se non vi sono opportunità di arbitraggio, deve necessariamente valere $\pi(X) = V_0^\theta(X)$. Detto in termini più formali

PROPOSIZIONE 1 (Pricing Principle). *Il prezzo iniziale di qualunque contingent claim non può che coincidere col costo iniziale sostenuto per finanziare quella strategia che consente di replicarne i pagamenti.*

Nella sua immediatezza un tale principio non richiede particolari commenti. Sembra invece più interessante considerarne l'applicazione ad alcuni esempi espliciti. A tal fine concentriamo l'attenzione sull'esempio più facile nella famiglia degli alberi binomiali.

ESEMPIO 12. *Consideriamo il seguente albero*

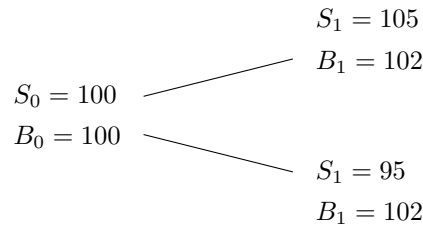


Figure 1.: Albero a uno stadio.

e poniamo che X sia un *contingent claim* con possibili valori X^u e X^d . Costruiamo il portafoglio $\theta = (\delta, \beta)$ di modo che $V_1^\theta(\omega) = X^\omega$ per ogni $\omega \in \{u, d\}$ e dunque tale che

$$\begin{cases} \delta_0 S_1^u + \beta_0 B_1 = X^u \\ \delta_0 S_1^d + \beta_0 B_1 = X^d \end{cases}$$

Ne segue che

$$(23.1.1) \quad \delta_0 = \frac{X^u - X^d}{S_1^u - S_1^d}$$

$$(23.1.2) \quad \beta_0 = \frac{1}{B_1} \frac{S_1^u X^d - X^u S_1^d}{S_1^u - S_1^d}$$

e quindi

$$(23.1.3) \quad V_0^\theta(X) = \delta_0 S_0 + \beta_0 B_0 = \frac{1}{B_{0,1}} \left\{ X^u \frac{B_{0,1} - d}{u - d} + X^d \frac{u - B_{0,1}}{u - d} \right\}$$

Nel caso in esame abbiamo implicitamente $u = 1,05$, $d = 0,95$ e $B_{0,1} = 1,02$ e dunque adattando la formula precedente al caso di opzioni CALL o PUT di tipo europeo e con prezzo di esercizio $k = 100$ otteniamo:

$$(23.1.4) \quad c_{1,100}(0) = 1,02^{-1} \left(5 \frac{1,02 - 0,96}{1,05 - 0,96} \right) = 3,268$$

$$(23.1.5) \quad p_{1,100}(0) = 1,02^{-1} \left(5 \frac{1,05 - 1,02}{1,05 - 0,96} \right) = 1,634$$

Benché l'esercizio sia del tutto triviale, esso illustra una proprietà interessante e tutt'altro che scontata. Infatti, nonostante i due contratti di opzione abbiano due payoff del tutto simmetrici (ognuno paga 5 in uno stato del mondo e 0 nell'altro) i loro prezzi differiscono in modo significativo. Naturalmente gli eventi finali potrebbero avere diverse probabilità ma i prezzi che abbiamo derivato sono del tutto indipendenti da ogni valutazione probabilistica.

L'esempio precedente solleva il seguente problema:

Che cosa spiega la differenza nel prezzo di una CALL e di una PUT quando i payoff finali sono del tutto simmetrici?

Naturalmente il nostro albero binomiale è piuttosto semplice e potremmo fare il medesimo esperimento con un albero più elaborato. Consideriamo a tal fine l'esempio seguente:

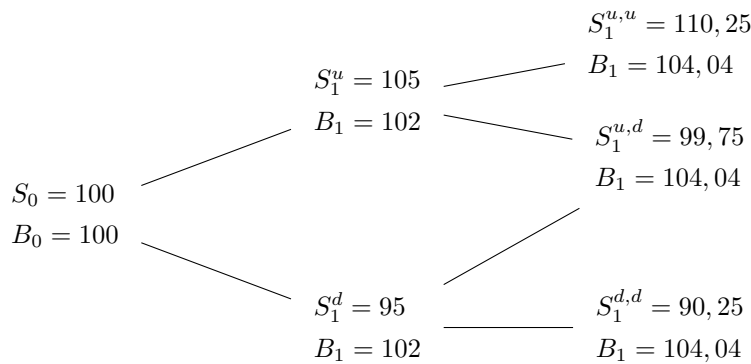
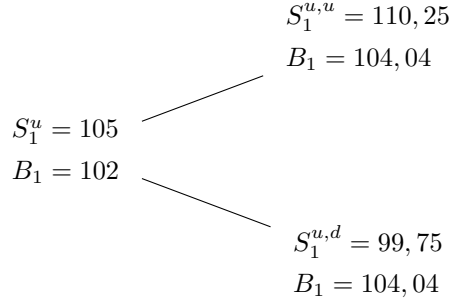


Figure 2.: Albero binomiale non ricombinante.

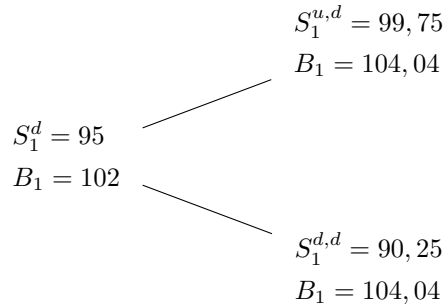
Alla data $t = 1$ vi sono ora due nodi distinti e la trading strategy θ^X potrebbe in linea di principio differire tra questi due stati. Poniamoci dal punto di vista del nodo superiore ossia poniamo attenzione al sotto albero

Figure 3.: Sotto albero relativo al nodo u .

Il modello che abbiamo isolato è del tutto analogo all'albero binomiale a due stadi che abbiamo considerato precedentemente. La soluzione ci è pertanto già nota

$$\begin{aligned}\delta_1^u &= \frac{X^{u,u} - X^{d,u}}{S_2^{u,u} - S_2^{d,u}} \\ \beta_1^u &= \frac{1}{B_2} \frac{S_2^{u,u} X^{u,d} - X^{u,u} S_2^{d,u}}{S_2^{u,u} - S_2^{d,u}} \\ V_1^u(\theta^X) &= B_{1,2}^{-1} \left\{ X^{u,u} \frac{B_{1,2} - d}{u - d} + X^{d,u} \frac{u - B_{1,2}}{u - d} \right\}\end{aligned}$$

la sola differenza essendo gli indici. Lo stesso ragionamento si applica al sottoalbero inferiore

Figure 4.: Sotto albero relativo al nodo d .

relativamente al quale si ottiene

$$\begin{aligned}\delta_1^d &= \frac{X^{d,u} - X^{d,d}}{S_2^{d,u} - S_2^{d,d}} \\ \beta_1^d &= \frac{1}{B_2} \frac{S_2^{d,u} X^{d,d} - X^{d,u} S_2^{d,d}}{S_2^{d,u} - S_2^{d,d}} \\ V_1^d(\theta^X) &= B_{1,2}^{-1} \left\{ X^{d,u} \frac{B_{1,2} - d}{u - d} + X^{d,d} \frac{u - B_{1,2}}{u - d} \right\}\end{aligned}$$

Abbiamo quindi derivato una regola che ci consente di determinare il portafoglio alla data $t = 1$ a seconda dello stato del mondo prevalente e tale da consentirci di replicare perfettamente il *contingent claim* X che scade la data successiva. Questa, tuttavia, non è ancora la soluzione che cercavamo poiché il nostro scopo è quello di capire come formare il nostro portafoglio di replica alla data iniziale, ossia $t = 0$. Tuttavia non è difficile comprendere che possiamo facilmente scegliere θ_0^X in modo tale che alla data successiva esso ci

consenta di finanziare il portafoglio contingente θ_1^X . In altre parole stiamo cercando un portafoglio di replica per il *contingent claim* $V_1(\theta^X)$ sull'albero

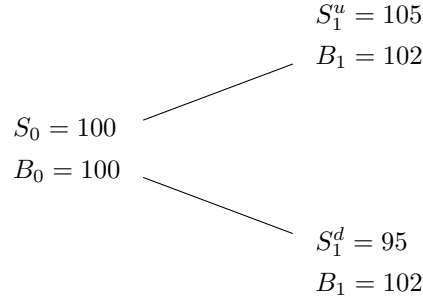


Figure 5.: Sotto albero iniziale.

Pertanto abbiamo la soluzione

$$(23.1.6) \quad \delta_0 = \frac{V_1^u(\theta^X) - V_1^d(\theta^X)}{S_1^u - S_1^d}$$

$$(23.1.7) \quad \beta_0 = \frac{1}{B_1} \frac{S_1^u V_1^d(\theta^X) - V_1^u(\theta^X) S_1^d}{S_1^u - S_1^d}$$

e – con l'assunzione di comodo $B_{0,1} = B_{1,2} = R$ tale per cui $B_{0,2} = R^2$

$$(23.1.8) \quad \begin{aligned} V_0(\theta^X) &= B_{0,1}^{-1} \left\{ V_1^u(\theta^X) \frac{B_{0,1} - d}{u - d} + V_1^d(\theta^X) \frac{u - B_{0,1}}{u - d} \right\} \\ &= R^{-2} \left\{ X^{u,u} \left(\frac{R - d}{u - d} \right)^2 + 2X^{u,d} \frac{u - R}{u - d} \frac{R - d}{u - d} + X^{d,d} \left(\frac{u - R}{u - d} \right)^2 \right\} \\ &= \pi(X) \end{aligned}$$

che rappresenta la nostra formula di prezzo.

La (23.1.8) è sufficientemente esplicita per dedurre una regola generale di fissazione del prezzo. Si noti che i coefficienti che compaiono nella seconda riga non dipendono che dal numero degli incrementi (e dunque delle riduzioni) del prezzo tra la data iniziale e la data finale. Procedendo per induzione otteniamo pertanto:

$$(23.1.9) \quad V_0(\theta^X) = R^{-T} \left\{ \sum_{j=0}^T X^{u^j, d^{T-j}} \binom{T}{j} \left(\frac{R - d}{u - d} \right)^j \left(\frac{u - R}{u - d} \right)^{T-j} \right\}$$

in cui u^j, d^{T-j} corrisponde allo stato del mondo ω nel quale si sono verificati j incrementi di prezzo su un totale di T variazioni.

ESEMPIO 13. Uno degli usi più immediati della formula (23.1.9) è quello di prezzare opzioni CALL e PUT. Si ottiene facilmente

$$(23.1.10) \quad c_{\tau,k}(t) = R^{-(\tau-t)} \left\{ \sum_{j=0}^{\tau-t} (S_t u^j d^{\tau-t-j} - k)^+ \binom{\tau-t}{j} \left(\frac{R - d}{u - d} \right)^j \left(\frac{u - R}{u - d} \right)^{\tau-t-j} \right\}$$

e

$$(23.1.11) \quad p_{\tau,k}(t) = R^{-(\tau-t)} \left\{ \sum_{j=0}^{\tau-t} (k - S_t u^j d^{\tau-t-j})^+ \binom{\tau-t}{j} \left(\frac{R - d}{u - d} \right)^j \left(\frac{u - R}{u - d} \right)^{\tau-t-j} \right\}$$

Nel caso rappresentato in questo esempio tali formule si traducono in numeri nel modo seguente (supponendo un valore iniziale $S_0 = 100$ e un prezzo d'esercizio $k = 100$)

(23.1.12)

$$c_{2,100}(t) = 1,02^{-2} \sum_{j=0}^2 100(1,05^j \times 0,95^{2-j} - 1)^+ \binom{2}{j} \left(\frac{1,02 - 0,95}{1,05 - 0,95} \right)^j \left(\frac{1,05 - 1,02}{1,05 - 0,95} \right)^{2-j} = 4,8275$$

(23.1.13)

$$p_{2,100}(t) = 1,02^{-2} \sum_{j=0}^2 100(1 - 1,05^j \times 0,95^{2-j})^+ \binom{2}{j} \left(\frac{1,02 - 0,95}{1,05 - 0,95} \right)^j \left(\frac{1,05 - 1,02}{1,05 - 0,95} \right)^{2-j} = 0,94435$$

e i seguenti portafogli

$$\theta^c = \begin{cases} \theta_0^c = \begin{bmatrix} \delta_0^c \\ \beta_0^c \end{bmatrix} = \begin{bmatrix} 105 & 102 \\ 95 & 102 \end{bmatrix}^{-1} \begin{bmatrix} 105 & 102 & 0 & 0 \\ 0 & 0 & 95 & 102 \end{bmatrix} \begin{bmatrix} \delta_1^{c,u} \\ \beta_1^{c,u} \\ \delta_1^{c,d} \\ \beta_1^{c,d} \end{bmatrix} = \begin{bmatrix} 0,70341 \\ -0,65513 \end{bmatrix} \\ \theta_1^{c,u} = \begin{bmatrix} \delta_1^{c,u} \\ \beta_1^{c,u} \end{bmatrix} = \begin{bmatrix} 110,25 & 104,04 \\ 99,75 & 104,04 \end{bmatrix}^{-1} \begin{bmatrix} 10,25 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,97619 \\ -0,93594 \end{bmatrix} \\ \theta_1^{c,d} = \begin{bmatrix} \delta_1^{c,d} \\ \beta_1^{c,d} \end{bmatrix} = \begin{bmatrix} 99,75 & 104,04 \\ 90,25 & 104,04 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{cases}$$

$$\theta^p = \begin{cases} \theta_0^p = \begin{bmatrix} \delta_0^p \\ \beta_0^p \end{bmatrix} = \begin{bmatrix} 105 & 102 \\ 95 & 102 \end{bmatrix}^{-1} \begin{bmatrix} 105 & 102 & 0 & 0 \\ 0 & 0 & 95 & 102 \end{bmatrix} \begin{bmatrix} \delta_1^{p,u} \\ \beta_1^{p,u} \\ \delta_1^{p,d} \\ \beta_1^{p,d} \end{bmatrix} = \begin{bmatrix} -0,29658 \\ 0,30603 \end{bmatrix} \\ \theta_1^{p,u} = \begin{bmatrix} \delta_1^{p,u} \\ \beta_1^{p,u} \end{bmatrix} = \begin{bmatrix} 110,25 & 104,04 \\ 99,75 & 104,04 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0,25 \end{bmatrix} = \begin{bmatrix} -0,02381 \\ 0,025231 \end{bmatrix} \\ \theta_1^{p,d} = \begin{bmatrix} \delta_1^{p,d} \\ \beta_1^{p,d} \end{bmatrix} = \begin{bmatrix} 99,75 & 104,04 \\ 90,25 & 104,04 \end{bmatrix}^{-1} \begin{bmatrix} 0,25 \\ 9,75 \end{bmatrix} = \begin{bmatrix} -1,0 \\ 0,96117 \end{bmatrix} \end{cases}$$

Da tutto ciò segue infine

$$\pi_0(\theta^c) = \begin{bmatrix} S_0 \\ B_0 \end{bmatrix} \begin{bmatrix} \delta_0^c \\ \beta_0^c \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \end{bmatrix} \begin{bmatrix} 0,70341 \\ -0,65513 \end{bmatrix} = 4,8275$$

e

$$\pi_0(\theta^p) = \begin{bmatrix} S_0 \\ B_0 \end{bmatrix} \begin{bmatrix} \delta_0^p \\ \beta_0^p \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \end{bmatrix} \begin{bmatrix} -0,29658 \\ 0,30603 \end{bmatrix} = 0,94435$$

il che conclude il nostro esempio.

Raffrontando (23.1.3) con (23.1.8) si giunge al seguente interrogativo

Per quale ragione nell'esempio precedente il prezzo dell'opzione CALL cresce con la durata della vita residua mentre accade l'opposto con l'opzione PUT?

Ancora una volta osserviamo che una tale domanda non è affatto triviale. Ad un primo sguardo ai *payoff* finali sembrerebbe infatti che debba valere l'opposto poiché l'opzione *PUT* è quella che presenta i *payoff* più elevati (naturalmente distribuiti in modo diversi da quelli della *CALL*). Nuovamente potrebbe venire in mente che la valutazione di probabilità degli eventi potrebbe giocare un qualche ruolo, sebbene non se ne sia fatto alcun uso nelle nostre formule.

A prescindere dalla due domande aperte formulate, vi è un certo numero di aspetti che merita un qualche commento.

OSSERVAZIONE 1. *Le formule dalla (23.1.1) alla (23.1.9) hanno un senso matematico compiuto se e solo se $u > d$, ossia nel caso di mercati completi. Cosa si deve concludere nei restanti casi?*

Chiaramente, in assenza di completezza il ruolo del principio di replica viene rimesso in discussione. SAI potrebbe pensare che per quei titoli che ammettono di essere replicati, i risultati precedenti debbano essere ancora validi. Ma qual è l'alternativa alla replica per quei titoli che non ammettono tale possibilità?

OSSERVAZIONE 2. *L'equazione (23.1.9) ha senso economico compiuto se e solo se $u > B_{t,t+1} > d$, cioè quando non vi sono opportunità di arbitraggio.*

Questo problema è molto chiaro nel problema ad uno stadio (23.1.3) poiché nel caso in cui una delle due disuguaglianze dovesse non valere, allora è assai facile costruire un *contingent claim* X con *payoff* strettamente positivo (ossia positivo in almeno uno stato del mondo e non negativo in quelli restanti) e prezzo non positivo. Nell'ambito di un albero a più stadi vale la stessa conclusione.

OSSERVAZIONE 3. *Il segno di δ^X ha una chiara interpretazione. Di nuovo si consideri un albero ad un solo stadio: δ^X è positivo per quei contingent claims X che sono positivamente correlati con il titolo sottostante, ossia che hanno un *payoff* maggiore nello stato u anziché in d .*

Questa stessa conclusione si estende a situazioni più generali, con $T > 1$, tramite induzione all'indietro. Se X e S sono positivamente correlati, allora ponendoci alla data $t = T - 1$, concludiamo che $\delta_{T-1}^X > 0$. Ma se questo è vero allora V_{T-1}^X è positivamente correlato con S_{T-1} (perché?) così che per replicarlo alla data $t = T - 2$ dovrà valere $\delta_{T-2}^X > 0$. Questo stesso modo di ragionare funziona solo per il nostro semplice caso in cui vi sono solo due attività, l'una delle quali priva di rischio, ed un numero finito di stati. Ci si potrebbe chiedere se questo sia vero in tutta generalità. La risposta è positiva, come vedremo in seguito.

OSSERVAZIONE 4. *Dalla (23.1.2) apprendiamo che $\beta > 0$ se e solo se $\frac{X^u}{X^d} < \frac{S^u}{S^d}$, una condizione che si può interpretare col dire che io contingent claim è meno volatile dell'attivo sottostante.*

Vedendo le cose da questo punto di vista la conclusione è davvero ovvia. Poiché il portafoglio di replica deve avere per definizione la medesima volatilità del titolo che deve replicare, quando le precedente disuguaglianza vale allora una quantità positiva del titolo privo di rischio è necessaria per mitigare l'eccesso di volatilità che caratterizza il sottostante. Nell'esempio di una *CALL*

$$\frac{X^u}{X^d} = \frac{(S^u - k)^+}{(S^d - k)^+} = \begin{cases} 0 & \text{if } S^u \leq k \\ \infty & \text{if } S^u > k \geq S^d \\ \frac{(S^u - k)^+}{(S^d - k)^+} & \text{if } S^d > k \end{cases}$$

e dunque $\beta^c \leq 0$.

OSSERVAZIONE 5. *Naturalmente il sottostante S coincide con una *CALL* con prezzo di esercizio nullo (se una tale opzione effettivamente esistesse).*

Allora

$$\begin{aligned}
c_{1,k}(0) &= B_{0,1}^{-1} \left\{ (S^u - k)^+ \frac{B_{0,1} - d}{u - d} + (S^d - k)^+ \frac{u - B_{0,1}}{u - d} \right\} \\
&\geq B_{0,1}^{-1} \left((S^u - k) \frac{B_{0,1} - d}{u - d} + (S^d - k) \frac{u - B_{0,1}}{u - d} \right) \\
&= B_{0,1}^{-1} \left(S^u \frac{B_{0,1} - d}{u - d} + S^d \frac{u - B_{0,1}}{u - d} \right) - B_{0,1}^{-1} k \\
&= c_{1,0}(0) - B_{0,1}^{-1} k \\
&= S_0 - B_{0,1}^{-1} k
\end{aligned}$$

Questa relazione è nota come *limite inferiore* del prezzo della *CALL*.

Poiché abbiamo fin qui completamente ignorato ogni aspetto probabilistico proviamo a calcolare alcune statistiche dell'albero binomiale (??).

$$\begin{aligned}
Var(\Delta S_1) &= E[(\Delta S_1 - E(\Delta S_1))^2] \\
&= p(\Delta S_1^u)^2 + (1-p)(\Delta S_1^d)^2 \\
&\quad - \left[p^2(\Delta S_1^u)^2 + (1-p)^2(\Delta S_1^d)^2 + 2p(1-p)\Delta S_1^u\Delta S_1^d \right] \\
(23.1.14) \quad &= \Delta S_1^u p(1-p)(\Delta S_1^u - \Delta S_1^d) - \Delta S_1^d p(1-p)(\Delta S_1^u - \Delta S_1^d) \\
&= p(1-p)(\Delta S_1^u - \Delta S_1^d)^2 \\
&= p(1-p)(S_1^u - S_1^d)^2
\end{aligned}$$

e covarianza

$$\begin{aligned}
COV(\Delta S_1, X_1) &= E(\Delta S_1 X_1) - E(\Delta S_1)E(X_1) \\
&= p(\Delta S_1^u X_1^u) + (1-p)(\Delta S_1^d X_1^d) \\
&\quad - (p^2 \Delta S_1^u X_1^u + (1-p)^2 \Delta S_1^d X_1^d + p(1-p)(\Delta S_1^u X_1^d + \Delta S_1^d X_1^u)) \\
(23.1.15) \quad &= \Delta S_1^u p(1-p)(X_1^u - X_1^d) - \Delta S_1^d p(1-p)(X_1^u - X_1^d) \\
&= p(1-p)(X_1^u - X_1^d)(\Delta S_1^u - \Delta S_1^d) \\
&= p(1-p)(X_1^u - X_1^d)(S_1^u - S_1^d)
\end{aligned}$$

Lavorando coi prezzi normalizzati \bar{S} si ottiene

$$\begin{aligned}
Var(\bar{\Delta S}_1) &= p(1-p)(\bar{S}_1^u - \bar{S}_1^d)^2 \\
COV(\bar{\Delta S}_1, X_1) &= p(1-p)(\bar{X}_1^u - \bar{X}_1^d)(\bar{S}_1^u - \bar{S}_1^d)
\end{aligned}$$

Queste formule piuttosto noiose hanno tuttavia un interessante sottoprodotto

$$\delta^X = \frac{COV(\bar{\Delta S}_1, \bar{X})}{Var(\bar{\Delta S}_1)}$$

Questa formula è la (ben nota) risposta al problema dei minimi quadrati seguente

$$(23.1.16) \quad \min_{\delta} E[(\bar{X} - \delta \bar{\Delta S}_1)^2]$$

Il nostro coefficiente di portafoglio δ^X è dunque il migliore nel senso dell'errore quadratico, ossia nel senso che riduce al valore minimo la varianza tra il *contingent claim* ed il portafoglio replicato (entrambi espressi in termini standardizzati). Ma è anche il migliore in ogni altro senso poiché esso consente di ridurre la varianza

a 0 (grazie alla completezza dei mercati). Il criterio dei minimi quadrati sarà in genere piuttosto differente – e dunque darà luogo a soluzioni diverse – a seconda della legge di probabilità sottostante

$$\min_{\delta} E[(\bar{X} - \delta \overline{\Delta S_1})^2] = E[\min_{\delta} (\bar{X} - \delta \overline{\Delta S_1})^2] = 0$$

Il portafoglio che genera l'errore quadratico minimo può però essere costruito in situazioni del tutto generali e non richiede la completezza dei mercati. Sia $T > 0$ il nostro orizzonte temporale (finito) e definiamo

$$(23.1.17) \quad R(\theta; X) = X - V_t^\theta$$

e

$$(23.1.18) \quad \bar{R}(\theta; X) = \bar{X} - \bar{V}_t^\theta$$

Se X è un *contingent claim*, e dunque non dà luogo a pagamenti intermedi, possiamo restringere l'attenzione all'insieme delle strategie autofinanziate e per $\theta \in \Theta_0$

$$\bar{R}(\theta; X) = \bar{X} - [V_0(\theta) + \underbrace{\sum_{t=0}^{T-1} \overline{\Delta S_{t+1}} \delta_t}_{\bar{G}(\theta)_{t+1}}]$$

di modo che otteniamo come soluzione alla (23.1.16)

$$(23.1.19) \quad \delta_t^X = \frac{COV(\bar{X}, \overline{\Delta S_{t+1}} | \mathcal{F}_t)}{Var(\overline{\Delta S_{t+1}} | \mathcal{F}_t)}$$

Calcolando il minimo si ottiene

$$E[\bar{R}(\theta^X; X)^2] = Var(\bar{X}) - \sum_{t=0}^{T-1} Var(\overline{\Delta S_{t+1}})(\delta_t^X)^2 = Var(\bar{X})(1 - \rho_X^2)$$

dove $\rho_X \in [-1, 1]$ indica il coefficiente di regressione lineare. Abbiamo quindi una formula – la (23.1.19) – che consente di estendere il nostro semplice risultato ottenuto con gli alberi binomiali al caso di uno spazio degli stati generale con un numero finito ma arbitrario di attività finanziarie. Tutto ciò di cui vi è bisogno è la condizione $\rho_X = \pm 1$ che è equivalente al dire che il *claim* X può essere replicato (perché?). I mercati sono allora completi se $\rho_X = \pm 1$ per ogni *contingent claim* X ¹.

23.2. Risk-Neutral Pricing

Nel paragrafo precedente non abbiamo fatto alcun uso della probabilità; in questa non useremo che questa. Grazie ad un'osservazione assai semplice saremo in grado di generalizzare il nostro approccio – e con esso le nostre formule – in modo molto importante. Alcune delle domande sollevate in precedenza troveranno qui la loro risposta.

L'irrelevanza della probabilità, messa ripetutamente in evidenza nel corso della trattazione dell'approccio di replica² implica che la scelta della probabilità di un aumento o una riduzione del prezzo del sottostante potrebbe essere sostituita da un qualunque numero (strettamente) compreso tra zero e 1. Per il principio di

¹Ad essere precisi la definizione di *contingent claim* in uno spazio di probabilità generale richiede maggior cautela.

²Non è del tutto vero che la probabilità non conti. In un trattamento più completo dell'argomento, una *legge di probabilità* P deve intendersi come una funzione che associa a certi insiemi un numero reale positivo $P(A)$. Ogni insieme C il cui complemento ha probabilità nulla - $P(C^c) = 0$ - viene chiamato il supporto di P . In effetti lo stesso albero - ossia l'insieme di nodi e diramazioni - deve considerarsi il supporto della legge di probabilità ed anche se non abbiamo fatto alcun riferimento ad una funzione data di probabilità abbiamo però utilizzato di continuo la descrizione del suo supporto, rappresentata dall'albero stesso. Questa osservazione un po' pedante consente di correggere la nostra affermazione iniziale col dire che, per un dato supporto, la distribuzione probabilistica è irrilevante.

non arbitraggio sappiamo che $u > B_{t,t+1} > d$ e ciò equivale a dire che esiste un numero reale $q_t \in (0, 1)$ tale che,

$$(23.2.1) \quad q_t u + (1 - q_t) d = B_{t,t+1}$$

o, equivalentemente,

$$(23.2.2) \quad q_t \bar{S}_{t+1}^u + (1 - q_t) \bar{S}_{t+1}^d = \bar{S}_t$$

Possiamo sempre interpretare qualsiasi grandezza compresa tra 0 e 1 alla stregua di una probabilità di un qualche evento. Dunque possiamo considerare q_t come la probabilità associata, alla data t all'evento di un rialzo del prezzo del sottostante. In termini di una tale interpretazione possiamo riscrivere la (23.2.2) come

$$(23.2.3) \quad E_Q(\bar{S}_{t+1} | \mathcal{F}_t) = \bar{S}_t$$

dove $E_Q(\cdot)$ sta ad indicare l'aspettativa calcolata rispetto alla misura Q secondo la quale $E_Q(\{u\} | \mathcal{F}_t) = q_t$. SI osservi che in un modello ad un solo stadio ciò prende la forma seguente

$$\bar{S}_0 = E_Q(\bar{S}_1)$$

La condizione (23.2.3) rende il processo del prezzo scontato, \bar{S} , una martingala (rispetto a Q) ossia un processo con media costante. In modo equivalente si può affermare che il rendimento atteso generato dal processo di prezzo uguaglia $B_{t,t+1}$ per ogni t , come si comprende facilmente dalla (23.2.1).

Si osservi che il passaggio alla misura Q non modifica la natura rischiosa del titolo sottostante S né del titolo privo di rischio. Ciononostante, S viene calcolato come un semplice valore atteso senza che si consideri in alcun modo il fattore del rischio. Quello così calcolato è esattamente il prezzo che prevarrebbe in un mondo popolato da investitori neutrali rispetto al rischio, un assunto del tutto lontano dal mondo reale e che non intendiamo certo introdurre qui. In effetti una tale ipotesi non è necessaria. Se il principio dell'assenza di opportunità di arbitraggio vale, allora investitori avversi al rischio si comportano *come se* fossero neutrali al rischio una volta adottata la misura Q . Per questa ragione una tale misura viene sovente denominata misura neutrale al rischio. Il principio che tentiamo di chiarire può allora essere formulato nel modo seguente

TEOREMA 16 (FTAP). *I mercati finanziari non ammettono alcuna opportunità di arbitraggio se e solo se esiste una misura di probabilità Q equivalente alla P e tale che i prezzi scontati delle attività finanziarie seguono un processo di martingala rispetto alla Q*

Questo teorema è noto come *teorema fondamentale dell' Asset Pricing* e, come abbiamo informalmente visto, è perfettamente vero nel nostro modello semplificato benché non sia lo stesso nel caso generale. Il requisito che Q sia equivalente a P significa semplicemente che col modificare la misura di probabilità non stiamo al contempo modificando la natura del problema e ci troviamo ancora a lavorare sul medesimo albero di partenza.

Tornando coi piedi sul terreno dei nostri semplici esempi, osserviamo che il valore di q_t può essere determinato facilmente sulla base della (23.2.1)

$$q_t = \frac{B_{t,t+1} - d}{u - d}$$

In termini di questa notazione la (23.1.3) diviene semplicemente

$$(23.2.4) \quad V_0(\theta^X) = B_{0,1}^{-1} \{X^u q + X^d (1 - q)\}$$

che nel caso di un'opzione *CALL* non è altro che

$$(23.2.5) \quad c_{1,k}(0) = B_{0,1}^{-1}[(S_0 u - k)^+ q + (S_0 d - k)^+ (1 - q)]$$

In effetti la (23.2.2) – o equivalentemente la (23.2.3) – non è altro che la nostra solita formula applicata ad una opzione *CALL* con prezzo di esercizio nullo. In realtà il principio di valutazione appena introdotto ha un'applicazione molto più generale, come vedremo. Dalla (??) si deduce che la probabilità di un incremento nel valore del sottostante si deduce che la probabilità di un incremento alla data t secondo la misura neutrale al rischio è funzione crescente di d e $B_{t,t+1}$ e decrescente di u . Questo ci aiuta a comprendere meglio – o almeno dovrebbe.

Rispetto alla misura Q il rendimento atteso generato da S_t – ossia $\frac{S_{t+1}}{S_t}$ – è pari a $B_{t,t+1}$. Affinché tale vincolo sia soddisfatto, la misura Q deve assegnare una probabilità inferiore a quegli eventi in cui il rendimento rischioso si allontana maggiormente dal valore prescritto. Se ad esempio $u - B_{t,t+1} > B_{t,t+1} - d$ allora la deviazione maggiore dal rendimento $B_{t,t+1}$ si registra con un incremento del prezzo ossia nello stato u e pertanto tale nodo deve risultare meno verosimile che il nodo d : in fatti $q_t < 0,5$. Un incremento nel valore di u deve essere compensato da una minor probabilità assegnata a quello specifico stato del mondo. La misura Q si dice neutrale al rischio nel senso che essa esprime implicitamente un'attitudine di cautela col considerare poco probabili ampi scostamenti dal rendimento del titolo privo di rischio. Conseguentemente, un aumento di $B_{t,t+1}$, rende l'evento di un rendimento più elevato più probabile (secondo la Q) per la semplice ragione che il rendimento normalizzato è ora inferiore.

Nell'esempio di pagina l'opzione CALL ha un prezzo superiore in quanto, nonostante l'apparente simmetria dei payoff, $q > 0,5$ cioè che la CALL sia in-the-money è più probabile (rispetto alla Q) che lo sia la PUT.

In effetti sotto la misura neutrale al rischio l'unica differenza nel prezzo di contratti finanziari con *payoff* perfettamente simmetrici sta nella probabilità assegnata ai diversi eventi.

Si può riscrivere la (23.2.5) come

$$\begin{aligned} c_{1,k}(0) &= R^{-1}[(S_1^u - k)^+ q + (S_1^d - k)^+ (1 - q)] \\ &= (\bar{S}_1^u - \bar{k})^+ q + (\bar{S}_1^d - \bar{k})^+ (1 - q) \end{aligned}$$

dove $\bar{k} = kR^{-1}$. Dalla (??) sappiamo che $\bar{S}_1^\omega = E_Q(\bar{S}_2 | \mathcal{F}_1)_\omega$ dove $\omega \in \{u, d\}$. Sostituendo

$$\begin{aligned} c_{1,k}(0) &= (\bar{S}_1^u - \bar{k})^+ q_0 + (\bar{S}_1^d - \bar{k})^+ (1 - q_0) \\ &= (E_Q(\bar{S}_2 | \mathcal{F}_1)_u - \bar{k})^+ q_0 + (E_Q(\bar{S}_2 | \mathcal{F}_1)_d - \bar{k})^+ (1 - q_0) \\ &\leq E_Q((\bar{S}_2 - \bar{k})^+ | \mathcal{F}_1)_u q_0 + E_Q((\bar{S}_2 - \bar{k})^+ | \mathcal{F}_1)_d (1 - q_0) \\ &= Q(E_Q((\bar{S}_2 - \bar{k})^+ | \mathcal{F}_1)) \\ &= Q((\bar{S}_2 - \bar{k})^+) \\ &= c_{2,k}(0) \end{aligned}$$

ci si assicuri di aver ben compreso le righe 4 e 6,

Si osservi che non abbiamo fatto alcun uso della struttura binomiale e dunque concludiamo che

TEOREMA 17. *Qualunque contingent claim il cui payoff finale è una funzione convessa (risp. concava) del valore del sottostante ha un prezzo tanto più elevato (risp. contenuto) quanto più lunga è la vita residua del titolo.*

Abbiamo ricavato la misura Q ragionando sul prezzo S . La sua applicazione, come abbiamo visto nella (23.2.4), è tuttavia ben più generale ed in effetti si estende a tutti i modelli finanziari finiti. Possiamo quindi riassumere la precedente discussione nel modo seguente

PROPOSIZIONE 2 (Risk Neutral Pricing). *Se esiste una misura neutrale al rischio Q allora il prezzo corrente di un qualunque titolo finanziario è uguale al valore atteso scontato (rispetto alla Q) del suo valore futuro.*

Questa conclusione così generale ci fornisce uno strumento estremamente utile nella valutazione dei titoli finanziari. Si osservi che, avendo ormai perso ogni interesse per il concetto della replicazione, non abbiamo neppure menzionato la completezza dei mercati. Ma allora come si forma il prezzo se i mercati sono incompleti?

ESEMPIO 14. *Consideriamo il mercato seguente*

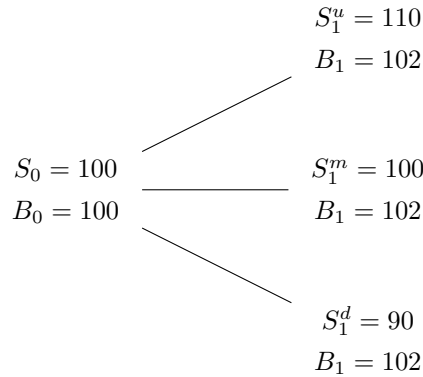


Figure 6.: Albero trinomiale a uno stadio.

Per entrambi i titoli, quello rischioso e quello privo di rischio, il valore nello stato m è la media aritmetica di quello assunto negli stati u e d : la stessa conclusione vale dunque per i portafogli composti da questi due titoli. Pertanto non possiamo replicare che quei titoli che hanno questa medesima proprietà. La probabilità neutrale al rischio si calcola come segue:

$$Q(\omega) = \begin{cases} q & \omega = u \\ 1, 2 - 2q & \omega = m \\ q - 0, 2 & \omega = d \end{cases}$$

Si osservi che q può essere un numero qualunque nell'intervallo (aperto) $(0, 2 - 0, 6)$ e vi è pertanto un'infinità di tali misure. Calcolare l'aspettativa da luogo a

$$E_Q(X) = 1, 2X^m - 0, 2X^d + 2q \left(\frac{X^u + X^d}{2} - X^m \right)$$

Questa espressione non dipende da q se e solo se $X^u + X^d = 2X^m$, dunque solo per quei titoli che possono essere replicati. Dunque l'incompletezza dei mercati si associa alla molteplicità delle misure neutrali al rischio.

Appendici

Alcune semplici regole su varianza e covarianza

Nel testo si utilizzano frequentemente alcune semplici proprietà della covarianza che, per comodità, sviluppiamo in questa appendice.

Anzitutto abbiamo la formula

$$(23.2.6) \quad \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

la quale è una diretta conseguenza della definizione di varianza: $\text{Var}(X) = E\{(X - E(X))^2\}$.

Un caso particolare della (23.2.6) si ha laddove $b = 1$. In tale circostanza possiamo guardare a $\text{Var}(aX + Y)$ come funzione reale del parametro a . Calcoliamo ad esempio il minimo, utilizzando la condizione del primo ordine:

$$0 = 2a \text{Var}(X) + 2 \text{Cov}(X, Y)$$

Dunque

$$(23.2.7) \quad a_{\min} = -\frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad \text{e} \quad \text{Var}(a_{\min}X + Y) = \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)}$$

e siccome la varianza è una quantità necessariamente non negativa, si conclude

$$(23.2.8) \quad \text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$$

la quale è nota come disuguaglianza di *Cauchy-Schwartz*. In particolare, nel caso in cui la (23.2.8) valga con segno di uguaglianza si avrebbe $\text{Var}(a_{\min}X + Y) = 0$ ossia,

$$(23.2.9) \quad Y = E(Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)}E(X) + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}X$$

ossia, Y è una funzione lineare di X e necessariamente $a_0 = E(Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)}E(X)$

Bibliografia

- [1] P. Artzner, F. Delbaen, M. Eber, D. Heath, (1999), *Coherent Measures of Risk*, Math. Fin. **9**, 203-228.
- [2] P. Christoffersen, *Principles of Financial Risk-Management*, 2001, Academic Press New York.
- [3] J. H. Cochrane, *Asset Pricing*, 2005, Princeton University Press, Princeton
- [4] D. M. Kreps, *Corso di Microeconomia*, 1994, Bologna Il Mulino.
- [5] J. Lintner (1965), s, *Prices, Risk and Maximal Gains from Diversification*, J. Fin. **20**, 587-615.
- [6] H. Markovitz (1937), s, *Portfolio Selection*, J. Fin. **7**, 77-91.
- [7] J. C. Rochet (1999), *Solvency Regulations and the Management of Banking Crisis*, Eur. Econ. Review **43**, 981-990.
- [8] H. Sharpe (1964), s, *Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk*, J. Fin. **19**, 425-442.