

Metodi Quantitativi per la Finanza
Appunti del corso

Claudio Ortelli

Semestre Estivo
2011

Indice

1	Fonti e raccolta di dati finanziari	3
1.1	Fornitori di dati finanziari	3
1.2	Codici d'identificazione	5
1.3	Datastream, classi di strumenti finanziari e tipologie di dati	6
2	Portfolio Models - Introduction	11
2.1	Notazione	11
2.2	Rendimenti percentuali e logaritmici: definizioni e proprietà	13
2.3	Flashback: varianze, covarianze e coefficiente di correlazione	20
2.4	Valore atteso e varianza del rendimento di un portafoglio	23
2.5	Portafoglio efficiente: concetti generali e definizione	24
3	Introduzione al linguaggio di programmazione R	26
3.1	Premessa	26
3.2	Installazione	26
3.3	La linea di comando	28
3.4	Strutture semplici e strutture di dati	29
3.5	Funzioni e programmazione da Excel	49
4	Calculating the Variance-Covariance Matrix	50
4.1	In Excel	50
4.2	In R	50
4.3	The Single-Index Model	51
5	Calculating Efficient Portfolios When are No Short-Sale Restrictions	53
5.1	Notazione e definizioni preliminari	53
5.2	La matematica della frontiera efficiente	53
6	The Lognormal Distribution	66
6.1	Introduzione: il cammino dei prezzi azionari	66
6.2	Flashback: Funzione di ripartizione, di densità, quantile	67
6.3	La distribuzione lognormale	69
6.4	Simulazione di variabili aleatorie	70
6.5	The Geometric random walk with drift	71

7	Value at Risk - VaR	75
7.1	Introduzione	75
7.2	VaR con rendimenti lognormali	76
7.3	A Three-Asset Problem	77
7.4	Historical simulation	77

Capitolo 1

Fonti e raccolta di dati finanziari

1.1 Fornitori di dati finanziari

La raccolta e l'archiviazione di dati è un tema estremamente importante in finanza. Pensate semplicemente ad una banca che offre un servizio di gestione patrimoniale. Essa deve, tra le altre cose, essere in grado di

- calcolare il valore (attuale e passato) dei portafogli in deposito,
- valutare il rischio associato alle singole posizioni assunte sui vari mercati finanziari (azionario, dei cambi, reddito fisso, derivati, ecc.) ed in aggregato.

A tale scopo la banca (il discorso è tuttavia generalizzabile a qualsiasi operatore finanziario) deve conoscere la composizione attuale e passata di un portafoglio (le quantità) nonché i prezzi d'acquisto e/o vendita, i prezzi correnti e i prezzi passati di tutti gli strumenti in esso contenuti. Con il costante aumento della potenza di calcolo disponibile e grazie

- all'avvento di internet
- al diffondersi di database relazionali (vedi informatica I e II)

la domanda e l'offerta di dati finanziari sono entrambe cresciute enormemente nel corso dell'ultimo decennio. Oltre ai banali esempi sopracitati esiste tutta una serie di attività legate al settore finanziario che fanno largo uso di dati. Fra queste, senza pretesa di esaustività, citiamo

- Calcolo della performance
- Scomposizione della performance
- Calcolo del rischio
- Ottimizzazione del portafoglio
- Analisi tecnica

- Analisi di bilancio
- Pricing di derivati
- Formazione di aspettative

È quindi chiaro quanto sia importante saper reperire in maniera

1. affidabile: il flusso di dati non deve essere interrotto (applicazioni di borsa real time), la qualità deve essere impeccabile (i dati non devono e non possono contenere errori),
2. efficace: i dati devono arrivare in tempo utile e poter essere elaborati facilmente,
3. economica: i costi devono essere tali da non pregiudicare l'attività (la redditività del business)

tutte le informazioni desiderate. Esistono a tal scopo dei cosiddetti fornitori di dati (data providers), ovvero ditte specializzate nel fornire informazioni legate all'attività finanziaria o più in generale economica nazionale e mondiale. Fra le principali citiamo

- Reuters
- Bloomberg
- Datastream
- Telekurs
- Infotec
- VWD group
- Yahoo finance
- Swiss Exchange
- Swissquote
- Google finance
- Altri ...

Qui all'Università della Svizzera italiana abbiamo a disposizione due importanti fonti di dati finanziari: Datastream, che sarà oggetto di approfondimento nel corso di questo capitolo e l'applicazione Reuters 3000 Xtra (un grazie all'associazione Finance Floor USI senza la quale l'applicazione non sarebbe disponibile).

Osservazione 1. Per tutta la durata di questo corso il termine *dati* sarà da considerarsi quale sinonimo di *serie storiche* dei prezzi dei vari strumenti finanziari che analizzeremo (ci limiteremo a considerare azioni, tassi di cambio e alcuni indici). Tuttavia il concetto generale di dati in finanza è da interpretare in maniera assai più ampia. Citiamo ad esempio

- i bilanci pubblicati dalle aziende (essenziali per gli analisti al fine della loro valutazione)
- le serie macroeconomiche divulgate dai diversi uffici governativi (PIL, indice dei prezzi, massa monetaria, tasso di disoccupazione, ..., importazioni, esportazioni, dati settoriali, ...)
- le news
- ...

1.2 Codici d'identificazione

Quando cerchiamo i dati di un qualsiasi prodotto o strumento finanziario (un'azione, un'obbligazione, un fondo, un indice, ecc.) è necessario conoscere il suo identificativo. Per identificativo intendiamo una sequenza alfanumerica che identifica l'emissione di ogni strumento finanziario in maniera univoca. Purtroppo non esiste uno standard unico: ogni borsa ed ogni data provider ha la propria metodologia di classificazione. Noi studieremo in dettaglio Datastream. Tuttavia è utile osservare quanto segue:

1. Esistono almeno due standard molto utilizzati a livello internazionale:

- **ISIN**
ISIN = International Securities Identification Number. An ISO coding system that uniquely identifies a specific securities issue. The organization that allocates ISINs in any particular country is the National Numbering Agency (NNA).
- **SEDOL**
SEDOL = Stock Exchange Daily Official List. The stock code used to identify all securities issued in the UK or Eire. It is the basis of the ISIN code for UK securities, and is composed of a 7-digit number allocated by the master file service of the London Stock Exchange.

Conoscere uno di questi due identificativi è particolarmente utile poiché, soprattutto per quanto riguarda il codice ISIN, praticamente tutti i data provider lo riconoscono. La banca dati Datastream ad esempio, accetta entrambi questi codici.

2. In Svizzera è ancora molto utilizzato il numero di valore (in tedesco: Valorennummer). Tuttavia il codice ISIN sembra lentamente ma inesorabilmente avere il sopravvento.
3. Talvolta è necessario verificare attentamente che i dati raccolti corrispondano a quanto desiderato. Capita infatti che lo stesso prodotto finanziario venga trattato su più borse o in più valute. È quindi molto facile commettere errori nella scelta della serie da scaricare se non si presta la massima attenzione all'origine dei dati. In caso di dubbio e quando ciò è possibile andate a verificare e/o cercare direttamente nelle varie borse gli identificativi appropriati. Su Wikipedia trovate i collegamenti alle principali piazze finanziarie, raggruppate per area geografica. Fra queste scegliamo quattro esempi:

- *Mercato svizzero*

Seguite poi Shares ... Company -> List of Companies

-> ABB LTD -> Share details [VN: ... ISIN: CH0012221716 ... Trading currency: ...]

- *Mercato svedese*

Cercate ABB e controllate il codice ISIN e la moneta in cui è quotata l'azione. Come potrete notare la medesima azione (il codice ISIN è identico, cioè CH0012221716) è quotata in valuta diversa!

- *Mercato inglese*

- *Mercato italiano*

-> Quotazioni -> Cerca Strumento -> Azioni -> Roma -> ISIN: IT0001008876

- *Titoli azionari americani*

-> SYMBOL LOOKUP (The Coca-Cola Company)

Symbol: KO

Website Coca Cola: <http://www.coca-cola.com>

4. Datastream

Il codice alfanumerico ISIN può essere d'aiuto nella ricerca dell'azione desiderata. Datastream infatti contiene questo codice nel suo "mnemonic" (è questo il nome del codice interno a Datastream che identifica univocamente ogni singola serie).

Ad esempio la Roma ha quale codice alfanumerico ASR e Datastream ha quale mnemonic "I:ASR". La Coca-Cola ha quale codice alfanumerico "KO" e Datastream ha quale mnemonic "U:KO". La prima lettera denota quindi il paese (I=Italy, U=United States, C=Canada, ...), il carattere ":" serve quale delimitatore e poi segue il codice alfanumerico. È importante ricordarsi questa costruzione laddove non è stato possibile rintracciare il codice ISIN poiché Datastream permette una ricerca per "mnemonic" (è inoltre possibile utilizzare nella ricerca caratteri jolly quali "*").

1.3 Datastream, classi di strumenti finanziari e tipologie di dati

Al quarto piano dell'Istituto di Finanza è disponibile un pc con Datastream Advance. Il login deve essere effettuato col nome utente: "ospite_datastream", mentre la password è: "datastream". Per lanciare il programma andate come d'abitudine nel menu Start -> All Programs -> Datastream -> Datastream 5.1. Per scaricare dati da Datastream è necessario procedere come segue:

1. Aprire Datastream Advance e selezionare da "Data category" il tipo di strumento per cui si desidera ottenere la serie storica.
2. Nel riquadro inferiore sinistro "Analysis" selezionare "Single Series - Data". Selezionare *Time Series Data*.

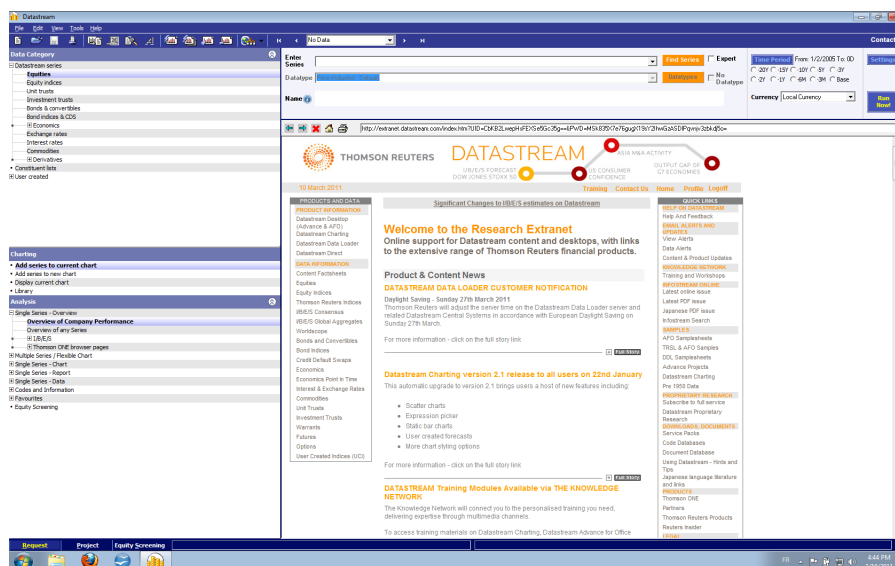


Figura 1.1: L'applicativo Datastream.

3. Nella parte superiore centrale della finestra di Datastream se conoscete il codice SEDOL o quello ISIN inseritelo direttamente nel campo *Series* dopo aver attivato il Checkbox denominato *Expert entry*.
4. Nel caso in cui non si conoscano né il codice ISIN, né il codice SEDOL e neppure il *mnemonic* Datastream, quale aiuto alla ricerca è possibile definire dei filtri. I criteri dipendono dalla categoria in cui si effettua la ricerca. Per esempio, i criteri disponibili per la categoria Equity sono diversi da quelli della categoria Economics. Per eseguire una ricerca cliccare il bottone "Find series".
5. Selezionate il periodo per il quale desiderate scaricare i dati (bottone "Time period" oppure cliccando sul checkbox corrispondente). Sempre nel medesimo riquadro ma più a destra premere il bottone *Settings...*. Scegliere la frequenza desiderata: daily, weekly, ecc. .
6. Selezionare la valuta. *Local currency* significa che verrà utilizzata la valuta propria del paese (stock exchange) in cui è scambiato il titolo.
7. Una volta che la serie desiderata è stata individuata, premere il bottone *Run Now*. I risultati della ricerca sono visualizzati nel riquadro inferiore centrale. Se la ricerca è stata fruttuosa, potete esportare i dati in un foglio excel premendo il bottone con il simbolo di excel nella "Toolbar" proprio sotto i menu a cascata.

Datastream associa a ciascuna classe di strumenti finanziari (azioni, indici azionari, bonds, ...) più di un tipo di serie storica scaricabile. I tipi di dati disponibili variano a seconda della classe di strumenti. Ad esempio, il tipo di default per la classe equity è il "Price (Adjusted)". Cliccando il bottone arancione con la scritta "*Datatype*" si ottiene la lista completa dei tipi disponibili. Ad ogni tipo

è associato un codice. Il "Price (Adjusted - Default)" ha codice uguale a P. Discutiamo ora i seguenti tipi:

1. Adjustment Factor (accumulated, Mnemonic = AF)
2. Adjustment Factor (not accumulated, Mnemonic = AX)
3. Dividends per share (Mnemonic = DPS)
4. Marketvalue (Mnemonic = MV)
5. Price (unadjusted) (Mnemonic = UP)
6. Ex-Dividend Date (Mnemonic = XDDE)
7. Date - Dividend Payment (Mnemonic = PYD)

Come potrete constatare voi stessi, cambiando la classe di strumenti finanziari e selezionando nuovamente il bottone *Datatype* la lista di tipi di dato fra cui scegliere cambierà.

Datastream prevede¹ inoltre la possibilità di scaricare le serie storiche direttamente in un file Excel, senza dover passare dall'applicazione Datastream Advance. Due sono i principali vantaggi di un simile approccio. Da un lato è possibile scaricare simultaneamente più di una serie storica e di un tipo di dato. Secondariamente l'aggiornamento di serie già scaricate richiede pochi passi ed un eventuale ulteriore elaborazione può essere automatizzata utilizzando la grande flessibilità di VBA (Visual Basic for Applications). Un esempio verrà presentato in classe. L'Help di Datastream Advance alla voce *Creating time series requests in Excel* è un buon punto di partenza.

Esercizio 1. Desideriamo analizzare lo Swiss Market Index (SMI).

1. Trovate su internet le azioni che attualmente sono incluse nello Swiss Market Index. Suggerimento: visitate il seguente sito
http://www.six-swiss-exchange.com/indices/overview_en.html
2. Di quanti titoli è composto lo SMI? Quali sono?
3. Qual è il criterio utilizzato per includere un titolo nell'indice?
4. Qual è il peso di ciascun titolo all'interno dell'indice?
5. Ogni quanto tempo vengono ricalcolati i pesi?
6. Ogni quanto tempo viene aggiornata la lista dei titoli da includere nello SMI?
7. Scaricate le serie storiche *a frequenza giornaliera* dello SMI degli ultimi 3 anni di osservazioni.
8. Scegliete 3 azioni attualmente rappresentate nell'indice e scaricate la serie storica *a frequenza giornaliera* degli ultimi 3 anni di osservazioni.

¹Si veda la Figura 1.4.

9. Eseguite il grafico di ogni serie separatamente. Eseguite poi un XY Scatter plot dell'indice SMI con ciascuna delle 3 serie scaricate. Cosa osservate?
10. Che differenza c'è con lo SMIC index?

□

The screenshot shows the Thomson Datastream Navigator search interface. The 'Data Category' is set to 'Equities'. The search criteria are as follows:

- Name: Contains UBS.S
- DS Mnemonic: Contains
- DS Code: Contains
- Market: Equals
- Base Date (dd/mm/yy): Before
- Currency:
- Primary Quote: ☒ All ☐ Yes ☐ No
- Major Security: ☒ All ☐ Yes ☐ No
- Adjusted Prices: ☒ All ☐ Yes ☐ No
- Status: ☐ All ☒ Active ☐ Dead ☐ Suspended
- Instrument Type: ☒ All ☐ American Depositary Receipt ☐ Closed-End Fund ☐ Equity ☐ Exchange-Traded Fund ☐ Genussschein (Profit Participation Certificate) ☐ Global Depositary Receipt ☐ Investment Trust ☐ Preference Share ☐ United States Warrant
- Exchange:

Buttons: Search, Reset All Criteria

Search Operator to use with multiple values for one field: And OR

Displayed Results: 10

Changes on Datastream: Significant Datastream Charting enhancements released - Jan 22nd

Figura 1.2: Criteri di ricerca di un titolo azionario.

The screenshot shows the Thomson Datastream Navigator Datatype Search interface. The 'Data Category' is set to 'Equities'. The search criteria are as follows:

- Find: Name Contains
- Filter / Search: Reset
- Displayed Results: 100

The search results are displayed in a list:

- ADR Parent Code
- Adjusted Price Indicator
- AIC Classification
- Base Date
- Classification Code Level 1
- Classification Name Level 1
- Classification Code Level 2
- Classification Name Level 2
- Classification Code Level 3
- Classification Name Level 3
- Classification Code Level 4
- Classification Name Level 4
- Company Name
- Date Of Last Name Change
- Code - Datastream
- Expanded Company Name
- Conversion Rate Of National Currency Non-Price Data To Euro
- Non-Price Currency Prior To The Event
- Date - Non-Price Currency Change
- Code - Datastream For Euro Data
- Date - Financial Year End Euro Accounts
- Instrument Type
- Expanded Name
- Conversion Rate Of National Currency Data To Euro
- Currency Prior To The Event
- Equity Status

Datatype Definition: To view the definition for a particular Datatype simply select the image and the definition will be displayed here

Items 1-100 of 116

Back Page: 1 2 Next

Figura 1.3: Selezione dei “datatypes” azionari.

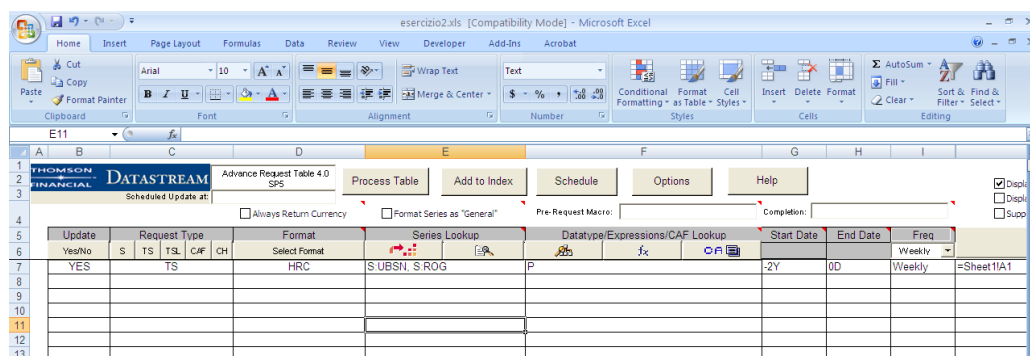


Figura 1.4: Request table in Excel.

Capitolo 2

Portfolio Models - Introduction

2.1 Notazione

In questo e nei prossimi capitoli faremo ampio uso del calcolo matriciale e delle nozioni di statistica multivariata apprese durante i corsi di Algebra Lineare e Introduzione all'Econometria. Prima di adoperare questi potenti strumenti presentiamo la notazione che verrà utilizzata. Purtroppo Benninga utilizza una notazione diversa da quella utilizzata nei corsi sopracitati. È questa una situazione alla quale dovrete abituarvi: in finanza ed in econometria non esiste nessuna convenzione per quanto riguarda la notazione. Ogni autore utilizzerà perciò i simboli a lui più congeniali. Per quanto riguarda questo corso indicheremo con

- $P_t :=$ Vettore aleatorio $N \times 1$ dei prezzi al tempo t .
- $r_t :=$ Vettore aleatorio $N \times 1$ dei rendimenti al tempo t (quando necessario verrà specificato il tipo - % o logaritmico - di rendimento).
- $V_t :=$ La matrice $N \times N$ delle varianze e covarianze di r_t .
- $e_t :=$ Il vettore $N \times 1$ dei rendimenti attesi, in altre parole

$$e_t := E[r_t].$$

Per indicare che il vettore aleatorio r_t ha valore atteso e_t e matrice delle varianze e covarianze V_t scriveremo semplicemente $r_t \sim (e_t, V_t)$. Quest'ultima espressione può essere semplificata ipotizzando che il valore atteso e la matrice delle varianze e covarianze dei rendimenti siano costanti nel tempo¹. Scriviamo allora semplicemente

$$r_t \sim (e, V) . \quad (2.1)$$

¹Abbiamo già utilizzato un'ipotesi simile nel modello di regressione classica

$$y_t = x_t' \beta + \epsilon_t$$

dove il valore atteso e la varianza dell'errore ϵ_t (di dimensione 1×1) sono costanti ed uguali rispettivamente a 0 e σ^2 . Le ipotesi già applicate su ϵ_t vengono ora applicate ed in un certo senso estese a r_t (di dimensione $N \times 1$). Infatti, l'ipotesi di stazionarietà dei primi due momenti è la stessa (si tratta di una semplice generalizzazione al caso multivariato).

	A	B	C	D	E	F	G
	PRICE AND RETURN DATA FOR WALMART (WMT) AND TARGET (TGT) Yahoo adjusts prices for dividends						
1							
2		Prices			Returns		
3	Date	WMT	TGT		WMT	TGT	
4	5-Jul-01	26.07	37.40				
5	1-Aug-01	22.00	33.53		-16.97%	-10.92%	<-- =LN(C5/C4)
6	4-Sep-01	20.07	30.73		-9.18%	-8.72%	<-- =LN(C6/C5)
7	1-Oct-01	20.02	30.15		-0.25%	-1.91%	<-- =LN(C7/C6)
8	1-Nov-01	23.35	36.38		15.39%	18.78%	
9	3-Dec-01	24.79	39.79		5.98%	8.96%	
10	2-Jan-02	23.03	43.04		-7.36%	7.85%	
11	4-Feb-02	18.09	40.66		-24.14%	-5.69%	
12	1-Mar-02	19.17	41.85		5.80%	2.88%	
13	1-Apr-02	20.25	42.36		5.48%	1.21%	
14	1-May-02	22.79	40.29		11.82%	-5.01%	
15	3-Jun-02	20.52	37.03		-10.49%	-8.44%	
55	3-Oct-05	50.93	55.37		7.95%	6.97%	
56	1-Nov-05	52.19	53.31		2.44%	-3.79%	
57	1-Dec-05	55.31	54.76		5.81%	2.68%	
58	3-Jan-06	55.63	54.54		0.58%	-0.40%	
59	1-Feb-06	58.46	54.29		4.96%	-0.46%	
60	1-Mar-06	60.23	51.90		2.98%	-4.50%	
61	3-Apr-06	65.46	52.99		8.33%	2.08%	
62	1-May-06	60.84	48.92		-7.32%	-7.99%	
63	1-Jun-06	67.51	48.87		10.40%	-0.10%	
64	3-Jul-06	67.65	49.17		0.21%	0.61%	

Figura 2.1: Esempio con $N = 2$ titoli, Benninga Capitolo 8.2 .

Esempio 1. Nell'esempio a pagina 240 di Benninga troviamo $N = 2$ titoli (azioni in questo caso) con i relativi prezzi. Le $n = 61$ osservazioni riportate sono a frequenza mensile: $t = \text{"2001-07-05", ...,"2006-07-03"}$. Generalmente si è soliti disporre le serie storiche in ordine cronologico partendo dall'osservazione più distante nel tempo. Come con la matrice X della regressione classica la t -esima riga conterrà i prezzi osservati di tutti gli N titoli alla data t mentre la prima (seconda ...) colonna conterrà tutte le osservazioni del primo (secondo ...) titolo. Nell'esempio di Benninga la prima colonna (colonna **Date**) contiene l'indice temporale² t ma questa colonna non fa parte della matrice X .

Riserveremo gli indici i e j per indicare l' i -esimo ed il j -esimo titolo: $V_{ij} = Cov(r_{it}, r_{jt})$. Avremo allora che $V_{ii} = Cov(r_{it}, r_{it}) = Var(r_{it})$ è la varianza del rendimento dell' i -esimo titolo. Denotiamo inoltre con

- q_t := Il vettore deterministico $N \times 1$ delle quantità (il numero) di titoli in portafoglio al tempo t . Quantità negative indicano vendite allo scoperto. Ad esempio $q_{it} = 5$ indica che, al tempo t , il portafoglio contiene 5 unità dell' i -esimo titolo. Qualora le quantità dovessero essere costanti (la struttura del portafoglio non viene modificata) eviteremo di utilizzare l'indice temporale e scriveremo semplicemente q .

²Notiamo che la frequenza con cui vengono osservati i prezzi è il mese. Per tale motivo in questo esempio l'unità di misura dell'indice temporale t è il mese.

- $P_{pt} :=$ Il valore (“prezzo”) del portafoglio al tempo t . Per definizione di q_t avremo $P_{pt} = q'_t P_t$. Come potete vedere si tratta del prodotto scalare tra il vettore dei prezzi e quello delle quantità.
- $w_t :=$ Il vettore $N \times 1$ dei pesi del portafoglio. Ogni componente w_{it} rappresenta la percentuale del valore totale del portafoglio investita nel titolo i al tempo t . Formalmente abbiamo

$$w_{it} = \frac{q_{it} P_{it}}{P_{pt}} .$$

Poiché sono permesse vendite allo scoperto (short selling), w_{it} può avere segno negativo. Tuttavia, ad ogni istante temporale t , la somma delle componenti di w_t deve forzatamente essere uguale ad 1. In altre parole deve valere $S'w_t = 1$, dove come al solito S rappresenta il vettore somma a N componenti.

- $r_{pt} :=$ Il rendimento del portafoglio al tempo t .

2.2 Rendimenti percentuali e logaritmici: definizioni e proprietà

Supponiamo, senza perdita di generalità, che l'unità di misura del tempo t sia il giorno.

Definizione 1. Il rendimento percentuale giornaliero alla data t dell'azione i è

$$r_{it} := \frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} = \frac{P_{it}}{P_{i(t-1)}} - 1 . \quad (2.2)$$

Osservazione 2. Dalla Definizione 1 di rendimento percentuale otteniamo l'identità

$$1 + r_{it} = \frac{P_{it}}{P_{i(t-1)}} . \quad (2.3)$$

Definizione 2. Il rendimento logaritmico giornaliero alla data t dell'azione i è

$$r_{it} := \ln \left(\frac{P_{it}}{P_{i(t-1)}} \right) . \quad (2.4)$$

Osservazione 3. Se la data t corrisponde alla “Ex-Dividend date”, allora il calcolo del rendimento percentuale e logaritmico vanno modificati come segue:

$$r_{it} = \frac{P_{it} + d_{i(t-1)} - P_{i(t-1)}}{P_{i(t-1)}} = \frac{P_{it} + d_{i(t-1)}}{P_{i(t-1)}} - 1 ,$$

rispettivamente

$$r_{it} = \ln \left(\frac{P_{it} + d_{i(t-1)}}{P_{i(t-1)}} \right) .$$

Il motivo per cui viene aggiunto il dividendo è il seguente. Al momento dell'acquisto dell'azione in data $t-1$ l'acquirente mantiene il diritto a percepire il

dividendo pur rivendendo l'azione il giorno t . Quindi a fronte di un costo iniziale di $P_{i(t-1)}$ il ricavo totale è dato dal prezzo di vendita P_{it} più il dividendo $d_{i(t-1)}$. Quando il rendimento è calcolato su un periodo superiore al giorno, ad esempio l'anno, $d_{i(t-1)}$ deve comprendere tutti i dividendi versati nel periodo (alcune azioni pagano dividendi a frequenza semestrale o addirittura trimestrale). Se il periodo è particolarmente lungo occorre tener conto del fatto che il dividendo è reinvestito. Questo comporta un aumento del numero di azioni che, a sua volta, va ad incidere sul calcolo del rendimento per il periodo successivo (si veda l'esempio numerico a pagina 256 di Benninga).

Esempio 2. Continuando con l'esempio precedente (vedi pag. 240 di Benninga) possiamo calcolare i 12 rendimenti logaritmici *mensili* delle due azioni A e B utilizzando la funzione LN (logaritmo naturale) di Excel. In questo esempio gli eventuali dividendi versati nel periodo non sono stati considerati. 11.1501cm - 5.7024cm Esempio calcolo rendimenti Figura esempio calcolo rendimenti 'KHMGB01.wmf' Poiché l'osservazione per il calcolo del rendimento alla data 0 non è disponibile (sarebbero necessari i prezzi alla data $t = -1$) la matrice dei rendimenti osservati avrà una riga in meno rispetto alla matrice dei prezzi osservati (vedi Figura ??).

□

2.2.1 Aggregazione cross-section dei rendimenti

Consideriamo in questo paragrafo la proprietà “*cross-section*” dei rendimenti di un portafoglio. La domanda a cui si desidera rispondere è la seguente: qual è la relazione tra il rendimento del portafoglio su un singolo periodo e i rendimenti dei singoli titoli in esso contenuto? La proprietà 1 dà una risposta a questa domanda.

Proprietà 1. Il rendimento percentuale del portafoglio soddisfa la seguente relazione

$$r_{pt} = w'_{t-1} r_t . \quad (2.5)$$

Quando i rendimenti sono logaritmici la proprietà 1 non è più mantenuta.

Proprietà 2. La Proprietà 1 non è soddisfatta dal rendimento logaritmico.

Le due proprietà appena discusse riguardano l'aggregazione *ad un medesimo istante temporale* di più titoli. Per tale motivo si parla della proprietà “*cross-section*” dei rendimenti. È questo un aspetto statico rispetto al tempo ma dinamico (aggrego) rispetto alle componenti del portafoglio. Vogliamo ora analizzare la situazione opposta, ovvero cosa capita quando l'aggregazione è temporale e non rispetto ai titoli in portafoglio. Questo implica analizzare il rendimento di *un singolo titolo* su un periodo di n giorni in funzione degli n rendimenti giornalieri.

2.2.2 Aggregazione temporale dei rendimenti

Consideriamo dapprima *un unico titolo*. In particolare, calcoliamo il suo rendimento su un orizzonte di n periodi utilizzando i rendimenti calcolati sui singoli periodi precedenti. Ad esempio, avendo a disposizione $n = 12$ rendimenti mensili come è possibile calcolare il rendimento annuo?

Definizione 3. Il rendimento percentuale dell' i -esimo titolo calcolato fra $t - n$ e t (su n giorni) è definito come

$$r_{i(t-n),t} := \frac{P_{it} - P_{i(t-n)}}{P_{i(t-n)}} = \frac{P_{it}}{P_{i(t-n)}} - 1 \quad (2.6)$$

Proprietà 3. Il rendimento percentuale su n periodi soddisfa la seguente relazione

$$r_{i(t-n),t} = \prod_{k=1}^n (1 + r_{i(t-k+1)}) - 1. \quad (2.7)$$

Osservazione 4. Al fine di snellire la notazione, il rendimento giornaliero o in generale il rendimento calcolato sul singolo periodo ($n = 1$) ora notato $r_{i(t-1),t}$ verrà semplicemente scritto come r_{it} .

Definizione 4. Il rendimento logaritmico dell' i -esima azione calcolato fra $t - n$ e t è definito come

$$r_{i(t-n),t} = \ln \left(\frac{P_{it}}{P_{i(t-n)}} \right) \quad (2.8)$$

Proprietà 4. Il rendimento logaritmico su n periodi soddisfa la seguente relazione

$$r_{i(t-n),t} = \sum_{k=1}^n r_{i(t-k+1)}. \quad (2.9)$$

Le Proprietà 1-4 sono molto importanti per comprendere come viene calcolato il rischio associato al movimento dei prezzi di mercato, il cosiddetto *market-risk* (si veda il Capitolo 7: VALUE AT RISK). Notiamo innanzi tutto che ci sono due dimensioni da considerare. La prima riguarda il fatto che il portafoglio è composto da più titoli (dimensione *cross-section*). Questo ci obbliga a considerare un vettore r_t di N rendimenti. Le Proprietà 1 e 2 riguardano proprio questa dimensione: la Proprietà 1 afferma che il rendimento percentuale di un portafoglio è dato dalla *combinazione lineare* dei rendimenti percentuali di ogni singolo titolo. I pesi di tale combinazione lineare corrispondono alle percentuali con cui i titoli erano rappresentati in portafoglio al momento dell'investimento, cioè $t - 1$. Il rendimento logaritmico non possiede questa interessante proprietà (Proprietà 2). Questo significa che il rendimento logaritmico di un portafoglio non è uguale alla somma pesata dei rendimenti logaritmici sui singoli titoli.

La seconda dimensione, che riguarda invece le Proprietà 3 e 4, concerne l'aspetto *temporale* dell'investimento: il rendimento percentuale di un singolo titolo aggrega in maniera geometrica rispetto al tempo (Proprietà 3) mentre il rendimento logaritmico aggrega in maniera lineare (Proprietà 4). È molto importante aver ben capito e sapere distinguere questi due aspetti al fine di non cadere in confusione e commettere errori. In particolare, come vedremo tra breve e nei prossimi capitoli, a dipendenza dalle ipotesi iniziali, dal tipo di rendimento e dalle sue proprietà d'aggregazione giungeremo a conclusioni diverse rispetto alle proprietà statistiche (distribuzione, momenti, ...) dei rendimenti.

Un'ultima osservazione: quando si discute dell'aspetto temporale di aggregazione dei rendimenti è necessario specificare la frequenza con cui si osservano i prezzi e l'orizzonte sul quale si desidera effettuare l'aggregazione.

Ad esempio, se disponiamo di rendimenti giornalieri e desideriamo calcolare i rendimenti settimanali avremo un orizzonte temporale di 5 periodi.

Consideriamo ora *un intero portafoglio* e chiediamoci se l'uguaglianza (2.7) è valida anche per un portafoglio composto da più titoli. La risposta è sì, ma con un'importante precisazione.

Definizione 5. Il rendimento percentuale calcolato fra $t-n$ e t di un portafoglio è definito come

$$r_{p(t-n),t} := \frac{P_{pt} - P_{p(t-n)}}{P_{p(t-n)}} = \frac{P_{pt}}{P_{p(t-n)}} - 1 . \quad (2.10)$$

Proprietà 5. Il rendimento percentuale su n periodi soddisfa la seguente relazione

$$r_{p(t-n),t} = \prod_{k=1}^n (1 + r_{p(t-k+1)}) - 1, \quad (2.11)$$

dove i singoli rendimenti $r_{p(t-k+1)}$ sono calcolati utilizzando la relazione (2.5) ma ovviamente con un vettore di pesi w_{t-k} diverso per ogni periodo.

Osservazione 5. La proprietà 5 evidenzia una caratteristica legata alla proprietà cross-section dei rendimenti percentuali di un portafoglio: i pesi w_{it} generalmente *non* sono costanti nel tempo (a meno di intervenire sulla composizione del portafoglio e ribilanciarlo continuamente). Questo perché i prezzi dei titoli che costituiscono un portafoglio non variano in ugual misura e di conseguenza il peso relativo di ciascun titolo rispetto al valore totale del portafoglio tenderà a variare nel tempo.

Esercizio 2. 1. Dimensione *cross-section* dei rendimenti

Il valore P_{pt} al tempo t di un portafoglio composto da solo due titoli è semplicemente

$$P_{pt} = q_1 P_{1,t} + q_2 P_{2,t}$$

dove q_1 e q_2 (per ipotesi costanti nel tempo) rappresentano la quantità della prima e rispettivamente della seconda azione in portafoglio.

(a) Dimostrate *algebricamente* che il rendimento percentuale del portafoglio

$$r_{p(t+1)} = (P_{p(t+1)} - P_{pt}) / P_{pt}$$

è uguale alla somma ponderata dei rendimenti percentuali

$$r_{i(t+1)} = (P_{i(t+1)} - P_{it}) / P_{it}$$

su ogni singola azione, ovvero

$$r_{p(t+1)} = w_{1t} r_{1(t+1)} + w_{2t} r_{2(t+1)} ,$$

verificando che ciascun peso w_{it} corrisponde alla percentuale iniziale dell' i -esima azione in portafoglio.

(b) Verificate quanto avete dimostrato nel punto precedente utilizzando i dati dell'Esempio 1 (file Chapter 07.xls). Assumete $q_1 = 1$ e $q_2 = 2$.

- (c) Mostrate numericamente con un semplice controesempio che ciò non è valido per i rendimenti logaritmici.
2. Variazione nel tempo della composizione percentuale del portafoglio
- Per risolvere quest'esercizio utilizzate le serie storiche contenute nel *Sheet1* del file *Esercizio2.xls*. Consideriamo un portafoglio costituito da un'azione di *Ubs* ed un'azione di *Roche*.
- (a) In base ai dati a vostra disposizione a quanto ammonta l'unità di misura del tempo t ?
 - (b) Eseguite il grafico delle due serie storiche.
 - (c) Calcolate il prezzo (o valore) del portafoglio a tutte le date disponibili.
 - (d) Utilizzando le serie dei prezzi calcolate i rendimenti percentuali di *Ubs*, *Roche*.
 - (e) Utilizzando la serie storica del valore (prezzo) del portafoglio calcolatene i rendimenti percentuali.
 - (f) Calcolate i pesi w_{it} di *Ubs* e *Roche*. Sono costanti? Sommano sempre a 1? Fate un grafico delle serie storiche dei due pesi.
 - (g) Calcolate i rendimenti percentuali del portafoglio utilizzando i rendimenti percentuali di *Ubs*, *Roche* ed i pesi w_t . Verificate che essi corrispondano ai rendimenti calcolati al punto (e).

3. Rendimenti su più periodi: aggregazione temporale

Consideriamo dapprima i rendimenti *logaritmici* settimanali di un'azione i , notati r_{it}^w ($w = \text{weekly}$) in funzione dei rendimenti giornalieri. Per definizione essi sono uguali a

$$r_{it}^w := r_{i(t-5),t} = \ln(P_{it}/P_{i(t-5)}) = \sum_{k=1}^5 r_{i(t-k+1)}.$$

- (a) Esplicitate la sommatoria che definisce r_{it}^w per $t = 5$. Disegnate l'asse temporale ed evidenziate con un colore l'intervallo temporale che comprende i rendimenti logaritmici giornalieri utilizzati per il calcolo di r_{i5}^w . Sul medesimo asse eseguite la stessa procedura ma per un valore di $t = 11$. Notate sovrapposizioni fra i colori? Interpretate il risultato?
- (b) Assumiamo che i rendimenti logaritmici giornalieri a date diverse non siano fra loro correlati e che $r_{i,t} \sim (\mu, \sigma^2)$. Cosa possiamo dedurre riguardo
 - i. al valore atteso del rendimento logaritmico settimanale?
 - ii. alla varianza del rendimento logaritmico settimanale?
 - iii. alla covarianza fra due rendimenti settimanali $\text{Cov}(r_{i(t-l)}^w, r_{it}^w)$?
Calcolate la covarianza per valori di l pari a 1, 2, 5 e 6.
Suggerimento: per ciascuna coppia $(r_{i(t-l)}^w, r_{it}^w)$ eseguite il disegno dell'asse temporale t ed indicate, con colori diversi, i due intervalli di tempo utilizzati per calcolare i rendimenti logaritmici settimanali corrispondenti. Calcolate i periodi in comune e confrontate con il valore della covarianza.

- (c) Se i rendimenti logaritmici giornalieri $r_{i,t}$ fossero *i.i.d.* $\sim (\mu, \sigma^2)$, cambierebbe qualcosa nei risultati del punto precedente? Argomentate la vostra risposta.
- (d) Se i rendimenti logaritmici giornalieri $r_{i,t}$ fossero *i.i.d.* $\sim N(\mu, \sigma^2)$, quale sarebbe la distribuzione di r_{it}^w e perché?
Consideriamo ora i rendimenti *percentuali* sempre di un'unica azione i
- (e) Qual è la formula che esprime i rendimenti percentuali settimanali di un'azione i in funzione dei suoi rendimenti percentuali giornalieri?
- (f) Se i rendimenti percentuali giornalieri $r_{i,t}$ fossero non correlati $\sim (\mu, \sigma^2)$, cosa potremmo concludere riguardo
- i. al valore atteso del rendimento percentuale settimanale?
 - ii. alla varianza del rendimento percentuale settimanale?
 - iii. alla covarianza fra due rendimenti settimanali $Cov(r_{i(t-l)}^w, r_{it}^w)$?
Suggerimento: non perdetevi troppo tempo su questo punto ...
- (g) Se i rendimenti percentuali giornalieri fossero *i.i.d.* $\sim (\mu, \sigma^2)$, cambierebbe qualcosa nei risultati del punto precedente? Argomentate la vostra risposta.
- (h) Se i rendimenti percentuali giornalieri fossero *i.i.d.* $\sim N(\mu, \sigma^2)$, la distribuzione di r_{it}^w sarebbe ancora normale? Argomentate la vostra risposta.

□

In genere un portafoglio titoli pur essendo valutato in un'unica moneta (la cosiddetta moneta di riferimento) è costituito da titoli denominati in valute diverse (chf, eur, usd, jpy, ...). Sorge quindi il problema di come scomporre il rendimento o il rischio di un titolo (o in aggregato di un portafoglio) nella parte dovuta al movimento del tasso di cambio e in quella dovuta al movimento del prezzo in valuta estera.

Per chiarirci le idee, consideriamo un'azione trattata in *EUR* alla borsa di Francoforte, il cui prezzo sarà notato P_{it}^{eur} . Ovviamente quali investitori svizzeri siamo interessati al suo prezzo in franchi, cioè $P_{it} = P_{eur,t} P_{it}^{eur}$. Il termine $P_{eur,t}$ denota quindi il tasso di cambio fra *EUR* e *CHF* alla data t : ad esempio $P_{eur,t} = 1.43 \text{ Fr/Eur}$.

Esempio 3. Un investitore svizzero ha acquistato 10 azioni della BASF. L'azione della BASF è quotata in euro. La tabella seguente riporta i prezzi a tre date successive.

Data	Prezzo in eur	Cambio 1 eur = x_{chf}	Prezzo in chf
7-Apr-09	23.40	1.516	
8-Apr-09	23.43	1.523	
9-Apr-09	24.96	1.522	

Evidentemente il prezzo (e quindi il rendimento) in *chf* della BASF è soggetto a due fonti di incertezza: i movimenti del prezzo della BASF in *eur* e i movimenti del tasso di cambio.

□

Quattro sono le domande che concernono i rendimenti di titoli quotati in moneta estera:

Domanda 1. Dato l' i -esimo titolo, che relazione sussiste fra il suo rendimento percentuale in *CHF* ed il suo rendimento percentuale in *EUR* su un singolo periodo?

Domanda 2. Dato l' i -esimo titolo, che relazione c'è fra il suo rendimento logaritmico in *CHF* ed il suo rendimento logaritmico in *EUR* su un singolo periodo?

Domanda 3. Come si calcola il rendimento percentuale in *CHF* fra $t-n$ e t (su n periodi) utilizzando i rendimenti percentuali calcolati su un singolo periodo (confronta la formula (2.7))?

Domanda 4. Come si calcola il rendimento logaritmico in *CHF* fra $t-n$ e t utilizzando i rendimenti logaritmici calcolati su un singolo periodo?

Risponderete a queste domande risolvendo l'esercizio seguente.

Esercizio 3. Utilizziamo quale unità di misura del tempo t il giorno. Consideriamo un'unica azione³, quotata in dollari al New York Stock Exchange e il cui prezzo in moneta locale (*usd*) è indicato⁴ con P_t^{usd} . Quali investitori svizzeri siamo interessati ai rendimenti in moneta di riferimento, ovvero in *chf*. Indicando con $P_{usd,t}$ il tasso di cambio al tempo t e con $r_{usd,t}$ il suo rendimento percentuale, abbiamo la semplice relazione $P_t = P_{usd,t} P_t^{usd}$.

D1. Trovate la formula che lega il rendimento giornaliero percentuale $r_t := (P_t - P_{t-1}) / P_{t-1}$ ai rendimenti percentuali giornalieri $r_{usd,t}$ e r_t^{usd} . Cosa notate?

Suggerimento:

$$r_t = \frac{P_{usd,t} P_t^{usd} - P_{usd,t} P_{t-1}^{usd} + P_{usd,t} P_{t-1}^{usd} - P_{usd,t-1} P_{t-1}^{usd}}{P_{usd,t-1} P_{t-1}^{usd}}$$

$$\frac{P_{usd,t} P_t^{usd} - P_{usd,t} P_{t-1}^{usd}}{P_{usd,t-1} P_{t-1}^{usd}} = (1 + r_{usd,t}) r_t^{usd}$$

D2. Mostrate che il rendimento giornaliero logaritmico in *chf* è uguale alla somma dei due rendimenti logaritmici giornalieri e cioè quello sul prezzo in valuta locale più quello sul tasso di cambio.

³Eviteremo quindi di utilizzare l'indice i a pedice del suo prezzo o del suo rendimento.

⁴Utilizziamo la convenzione che quando un prezzo (o un rendimento) è espresso in una valuta diversa dalla valuta di riferimento, tale valuta sarà posta ad apice del prezzo (rendimento).

- D3. Parte 1. Derivate la relazione che lega il rendimento percentuale in *CHF* fra $t-n$ e t , notato $r_{(t-n),t}$, ai rendimenti percentuali $r_{usd,(t-n),t}$ e $r_{(t-n),t}^{usd}$.
Suggerimento: riutilizzate, adattando l'indice temporale, il suggerimento del punto 1.
- D3. Parte 2. Ricavate la formula per il calcolo del rendimento $r_{(t-n),t}$ a partire dai rendimenti percentuali sui singoli periodi $r_{usd,t}$ e r_t^{usd} .
- D4. Derivate la relazione che lega il rendimento logaritmico in *CHF* fra $t-n$ e t , notato $r_{(t-n),t}$, ai rendimenti logaritmici $r_{usd,(t-n),t}$ e $r_{(t-n),t}^{usd}$. Utilizzando questo risultato derivate la formula per il calcolo $r_{(t-n),t}$ a partire dai rendimenti logaritmici sui singoli periodi $r_{usd,t}$ e r_t^{usd} (proprietà di aggregazione temporale).
- D5. Assumiamo che i vettori $(r_{usd,t}, r_t^{usd})'_{t=1,2,\dots}$ dei rendimenti logaritmici formino una successione *i.i.d.* $\sim N(\mu, \Sigma)$. Cosa possiamo concludere riguardo alla successione di rendimenti logaritmici in *CHF* r_t ?

□

2.3 Flashback: varianze, covarianze e coefficiente di correlazione

La varianza di una variabile aleatoria X (scalare) sappiamo essere definita da

$$Var(X) := E(X - E(X))^2 \quad (2.12)$$

mentre la covarianza con la variabile aleatoria Y (anch'essa scalare) è uguale

$$Cov(X, Y) := E(X - E(X))(Y - E(Y)) \quad (2.13)$$

Infine il coefficiente di correlazione fra X e Y è definito come

$$Cor(X, Y) := \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \quad (2.14)$$

Per quanto riguarda varianza e covarianza, esistono formule alternative che possono essere ricavate a partire dalla loro definizione utilizzando la proprietà di linearità del valore atteso. Tali formule sono

$$Var(X) = E(X^2) - (E(X))^2$$

e

$$Cov(X, Y) = E(XY) - E(X)E(Y) \quad .$$

Poiché nel caso scalare la moltiplicazione è commutativa, vale che

$$Cov(X, Y) = Cov(Y, X).$$

Lo stesso non vale nel caso in cui X e/o Y siano dei vettori aleatori. In tal caso occorre prestare attenzione all'ordine in cui X ed Y appaiono nella funzione di

Cov. Ricordiamo che la *Cov* fra due vettori aleatori $\underset{(r \times 1)}{X}$ e $\underset{(c \times 1)}{Y}$ è definita nel seguente modo (si noti il trasposto nella formula)

$$\underset{(r \times c)}{Cov}(X, Y) = E \left[(X - E(X)) (Y - E(Y))' \right] .$$

Come nel caso scalare potete dimostrare che

$$Cov(X, Y) = E(XY') - E(X) E(Y)'$$

e quando $X = Y$ la formula della covarianza si riduce alla varianza del vettore aleatorio X :

$$Cov(X, X) = Var(X) = E \left[(X - E(X)) (X - E(X))' \right] .$$

Ricordiamo che, per definizione, la varianza di un vettore aleatorio è una matrice quadrata, simmetrica, avente quale elemento nella posizione (i, j) la covarianza fra l' i -esima e la j -esima componente di X , $Cov(X_i, X_j)$. Allo stesso modo è possibile definire in maniera del tutto analoga la *matrice di correlazione* del vettore aleatorio X come la matrice quadrata avente nella posizione (i, j) la correlazione fra l' i -esima e la j -esima componente di X , $Cor(X_i, X_j)$. Poiché $Cor(X_i, X_j) = Cor(X_j, X_i)$, anche la matrice di correlazione di X è simmetrica! In particolare, la diagonale di qualsiasi matrice di correlazione è formata da 1, mentre fuori dalla diagonale devono esserci dei numeri compresi nell'intervallo $[-1, 1]$. La matrice di correlazione è positiva semidefinita. La matrice di correlazione del vettore aleatorio X sarà notata

$$Cor(X) .$$

Abbiamo studiato nella prima parte del corso la convergenza in probabilità dei momenti empirici verso i momenti teorici (si riveda la parte riguardante la *Convergenza dei momenti del campione*). La versione empirica della (2.12) è, dato un campione di numerosità n , semplicemente uguale⁵

$$\widehat{Var}(X) = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2 , \quad (2.15)$$

mentre la versione empirica della (2.13) altro non è che

$$\widehat{Cov}(X, Y) = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) (y_t - \bar{y}) . \quad (2.16)$$

Esercizio 4. Consideriamo il vettore aleatorio⁶ $Z := (X, Y)' \sim (\mu, V)$ di dimensione $N = 2$, dalla cui popolazione bivariata estraiamo un campione aleatorio di numerosità $n = 3$. Questo significa che avremo a che fare con le realizzazioni di 3 vettori aleatori $Z_1 = (X_1 \ Y_1)'$, $Z_2 = (X_2 \ Y_2)'$ e $Z_3 = (X_3 \ Y_3)'$. I tre vettori Z_1 , Z_2 e Z_3 sono fra loro indipendenti (ad esempio X_3 è indipendente da

⁵In MS Excel questa è esattamente la formula sottostante alla funzione *varp()*.

⁶Il vettore μ è semplicemente il vettore dei valori attesi di Z :

$$\mu = E(Z) = E \left[\begin{pmatrix} X \\ Y \end{pmatrix} \right] = \begin{pmatrix} E[X] \\ E[Y] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} .$$

Y_1 o X_1) mentre le due componenti X_t e Y_t avranno covarianza uguale a V_{12} . Possiamo impilare i 3 vettori Z_t disponendoli nelle righe di una matrice

$$\begin{bmatrix} Z'_1 \\ Z'_2 \\ Z'_3 \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ X_3 & Y_3 \end{bmatrix} = [x, y] .$$

Come nella regressione possiamo leggere ed interpretare il contenuto di questa matrice per riga o per colonna. Le colonne conterranno tutte le 3 (n in generale) osservazioni della medesima variabile, la X ad esempio (prima colonna). Le righe invece conterranno le osservazioni su tutte le variabili ad un determinato istante (di una determinata estrazione). x ed y sono due vettori aleatori contenenti tutte le n osservazioni della variabile X e rispettivamente della Y .

1. Quale sarà la dimensione della matrice $Cov(x, y)$? Calcolatela esplicitamente in funzione della matrice V .
2. Riscrivete la (2.16) come

$$\widehat{Cov}(X, Y) = \frac{1}{3} x' M y ,$$

e date esplicitamente il contenuto della matrice M . Di che matrice si tratta? Quali sono le sue caratteristiche?

3. Consideriamo il prodotto (S è il vettore somma a 3 componenti)

$$(y - \mu_2 S) (x - \mu_1 S)' .$$

Date dimensione e contenuto di ogni simbolo ed esplicitate il prodotto. Calcolatene poi il valore atteso.

4. Applicando le proprietà di linearità e commutatività della traccia e del valore atteso (già viste ed utilizzate per il calcolo del valore atteso di una forma quadratica) calcolate il valore atteso di

$$\frac{1}{3} (x - \mu_1 S)' M (y - \mu_2 S) . \quad (2.17)$$

e mostrate che è uguale a $\frac{2}{3} Cov(X, Y)$.

5. Dopo aver sciolto le parentesi della (2.17) calcolate il valore atteso di ciascun termine. Utilizzando il risultato del punto precedente date $E \left(\widehat{Cov}(X, Y) \right)$.
6. Generalizzate il risultato al caso con n osservazioni.

□

Osservazione 6. L'esercizio precedente ci mostra che, come nel caso della stima della varianza, anche per la stima della covarianza lo stimatore (2.16) non è corretto poiché sarebbe necessario dividere per $n - 1$. Tuttavia la distorsione (Bias) è dell'ordine $\frac{1}{n}$ e diventa trascurabile (tende a 0) al tendere di n all'infinito.

Osservazione 7. La dimensionalità cross-section del problema⁷ è espressa dalle dimensioni del vettore aleatorio Z dalla cui distribuzione sono campionate le osservazioni. La dimensione temporale è data dal numero⁸ $n = 3$ di campionamenti.

Tornando al coefficiente di correlazione: poiché entrambi gli stimatori convergono in probabilità alla varianza di X (Y) e alla covarianza fra X ed Y , utilizzando una delle proprietà del p lim (quale?) avremo che lo stimatore

$$\widehat{Cor}(X, Y) := \frac{\widehat{Cov}(X, Y)}{\sqrt{\widehat{Var}(X)}\sqrt{\widehat{Var}(Y)}}$$

è uno stimatore consistente della correlazione fra X ed Y .

Concludiamo questa sezione con una semplice applicazione di statistica multivariata. Sia V la matrice delle varianze e covarianze del vettore aleatorio N -dimensionale Z e Λ la matrice diagonale con elementi $\Lambda_{ii} = Var(Z_i)$, $i = 1, \dots, N$. Vale allora la relazione

$$Cor(Z) = \Lambda^{-1/2} V \Lambda^{-1/2} . \quad (2.18)$$

Esercizio 5. Sia $Z = (X \ Y)'$ un vettore aleatorio a $N = 2$ componenti di valore atteso nullo e matrice delle varianze e covarianze

$$V = \begin{bmatrix} 4 & -1 \\ -1 & 25 \end{bmatrix} .$$

1. Calcolate la matrice Λ e $\Lambda^{-1/2}$.
2. Verificate che *post-moltiplicare* V per la matrice *diagonale* $\Lambda^{-1/2}$ significa moltiplicare la i -esima *colonna* di V per il corrispettivo i -esimo elemento della diagonale di $\Lambda^{-1/2}$.
3. Verificate che *pre-moltiplicare* V per la matrice *diagonale* $\Lambda^{-1/2}$ significa moltiplicare la i -esima *riga* di V per il corrispettivo i -esimo elemento della diagonale di $\Lambda^{-1/2}$.
4. Calcolate $Cor(Z)$.

2.4 Valore atteso e varianza del rendimento di un portafoglio

Abbiamo dimostrato che il rendimento percentuale $r_{p(t-n),t}$ di un portafoglio è la somma ponderata dei rendimenti $r_{i(t-n),t}$ delle azioni che lo compongono. Il vettore di questa combinazione lineare è $w_{(t-n)}$. Esso contiene le posizioni percentuali ad inizio periodo. Per semplicità di notazione assumiamo $n = 1$, tralasciando l'indice temporale sia dei pesi che dei rendimenti. Poiché il vettore dei rendimenti $r \sim (e, V)$ e $r_p = w'r$, avremo per la famosissima formula $AV A'$, che $Var(r_p) = w'Vw$. Il valore atteso di r_p sarà invece uguale a $\mu_p = w'e$.

⁷Ci riferiamo al numero di titoli in portafoglio.

⁸Tale numero corrisponde al numero di date per le quali abbiamo le osservazioni degli N rendimenti azionari.

2.4.1 Caso con due soli titoli

Il caso con due soli titoli è triviale: $r_p = w_1 r_1 + w_2 r_2$ da cui segue semplicemente che

$$\sigma_p^2 := \text{Var}(r_p) = w_1^2 V(r_1) + w_2^2 V(r_2) + 2w_1 w_2 \text{Cov}(r_1, r_2) \quad (2.19)$$

Utilizzando il vincolo $w_1 + w_2 = 1$ è possibile riscrivere la (2.19) solo in funzione di w_1 (o w_2 ovviamente) ed eseguire un grafico della funzione $\sigma_p^2(w_1)$ ovvero della varianza di r_p in funzione del peso w_1 . Il tutto è lasciato come semplice esercizio.

2.4.2 Caso generale

Il caso con N titoli è praticamente già stato discusso nell'introduzione.

2.5 Portafoglio efficiente: concetti generali e definizione

Supponiamo che i rendimenti percentuali r_t dei titoli in portafoglio siano indipendenti ed identicamente distribuiti secondo la legge normale:

$$r_t \sim i.i.d. N(e, V).$$

Il rendimento r_{pt} del portafoglio, quale combinazione lineare dei rendimenti dei singoli titoli, sarà pertanto una variabile aleatoria distribuita secondo la legge normale, di valore atteso μ_p e varianza σ_p^2 .

Sappiamo inoltre che per caratterizzare una variabile aleatoria normale è necessario e sufficiente specificare il suo valore atteso e la sua varianza. Un investitore avrà dunque quali unici parametri di scelta il valore atteso e la varianza del portafoglio. È quindi ovvio che se i rendimenti sono variabili aleatorie normali, un individuo *razionale* ed *avverso al rischio* preferirà investire in un portafoglio A a bassa volatilità anziché in un portafoglio B di medesimo rendimento atteso ma volatilità superiore. È dunque legittimo chiedersi se, per un fisso rendimento atteso $\mu_p = c$, esiste un portafoglio a varianza minima fra tutti i portafogli di rendimento atteso superiore o uguale a c . Un simile portafoglio, se esiste, sarà chiamato portafoglio frontiera (“envelope portfolio”).

Tuttavia nulla ci impedisce di ragionare partendo dal rischio associato al portafoglio, identificato da $\sigma_p^2 = d$. In tal caso dovremo chiederci se esiste un portafoglio che massimizza il rendimento atteso fra tutti i portafogli con varianza minore o al massimo uguale a σ_p^2 . Un portafoglio del genere, se esiste, è chiamato portafoglio efficiente.

Grafico

Nel definire il portafoglio frontiera ed il portafoglio efficiente abbiamo utilizzato nel nostro ragionamento la varianza del portafoglio. Avremmo potuto fare lo stesso ragionamento in termini della volatilità σ_p del portafoglio e giungere esattamente alla stessa conclusione in quanto le due misure di dispersione sono legate fra loro da una trasformazione monotona crescente: dati due portafogli p_1 e p_2 :

$$\sigma_{p_1}^2 > \sigma_{p_2}^2 \iff \sigma_{p_1} > \sigma_{p_2} \ .$$

Capitolo 3

Introduzione al linguaggio di programmazione R

3.1 Premessa

R è un ambiente di sviluppo per l'analisi dei dati nato come estensione del linguaggio di programmazione S sviluppato negli anni 80 nei Laboratori Bell americani dell'AT&T. R è un ambiente statistico aperto e di pubblico dominio ("open source") che, nel rispetto dei diritti d'autore, può essere copiato, distribuito e modificato da chiunque. Inizialmente scritto da Ross Ihaka e Robert Gentleman del Dipartimento di Statistica dell'Università di Auckland, Nuova Zelanda, dal 1997 lo sviluppo di R è affidato ad un gruppo internazionale denominato "R Core Team". Il codice sorgente e gli eseguibili del programma di base e delle varie librerie aggiuntive possono essere scaricati accedendo all'indirizzo web

<http://www.r-project.org>

dove sono scaricabili le versioni del linguaggio per diversi sistemi operativi (Linux, Mac, Windows, ...).

Questo documento non è un manuale ma vuole essere semplicemente una traccia o un riassunto di quanto presentato in classe. Per ulteriori approfondimenti e precisazioni dei temi trattati, oltre ai manuali ed alla ricca documentazione disponibile all'indirizzo sopracitato, si vedano i riferimenti bibliografici indicati all'indirizzo

<http://www.r-project.org/doc/bib/R-publications.html> .

Buona documentazione in varie lingue fra cui l'italiano è pure ottenibile all'indirizzo

<http://stat.ethz.ch/CRAN/other-docs.html> .

3.2 Installazione

Il programma R è stato installato sul disco "Condivisione" nella cartella *programs/R/* (la versione attuale è la 2.10.1). Tuttavia i normali utenti non hanno

diritto di scrittura su tale disco e pertanto prima di lanciare il programma dovete creare una shortcut in una cartella sulla quale avete diritto di lettura e scrittura. In pratica dovete seguire i seguenti passi:

1. Creare una cartella che verrà utilizzata come home di R. Suggesto di creare la cartella “R” nella directory “My Documents”. In tal modo sarà possibile accedere ai dati da qualsiasi pc interno ed esterno all’USI.
2. Andate poi in “\\usi\dfs\Condivisione\programs\R\2.10.1\bin”.
3. Create una shortcut di “Rgui.exe” sul vostro desktop. L’icona “Rgui.exe” verrà utilizzata per lanciare R.
4. Selezionare l’icona della shortcut appena creata col tasto destro del mouse e scegliere dal menu a cascata la voce properties.
5. Nel campo “Start in” immettere il percorso della directory creata al punto 1. Nel mio caso il percorso è

“\\usi\dfs\istituti\Documents\ortellic\My Documents\R”.

Voi dovrete avere un percorso simile ma con “ortellic” rimpiazzato dal vostro identificativo d’accesso (login name).

6. Chiudere la finestra delle proprietà della shortcut.

Ora siete pronti ad avviare la vostra prima sessione di R!

È possibile configurare R andando ad aggiungere, eliminare o modificare il contenuto del file Rprofile.site che si trova nella sottocartella “etc” della cartella in cui è stato installato il programma. All’USI l’applicazione R è installata in

“\\usi\dfs\Condivisione\programs\R\2.10.1”

e il file Rprofile.site si troverà quindi in

“\\usi\dfs\Condivisione\programs\R\2.10.1\etc” .

Tuttavia, poiché non avete diritto di scrittura è necessario operare diversamente. Anziché modificare il file Rprofile.site ne facciamo una copia e la mettiamo nella cartella di home di R (da noi creata in precedenza al punto 1). All’avvio di R il programma cerca automaticamente nella cartella di home la presenza del file “.Rprofile” e ne legge il contenuto. Modifichiamo il contenuto di Rprofile.site e in seguito cambiamo il suo nome¹ in “.Rprofile”. In tal modo ci assicuriamo che dopo l’avvio di R il programma abbia caricato la configurazione da noi desiderata. Un esempio di .Rprofile è dato qui di seguito²:

¹Da explorer non è possibile rinominare un file con un nome che inizia con un punto. È necessario farlo da linea di comando (command prompt). Evito per ragioni di tempo di spiegare come aggirare il problema. La semplice alternativa è quella di copiare un file già esistente per poi modificarlo secondo le proprie necessità. Voi potete utilizzare il file .Rprofile che si trova sul disco Corsi.

²Attenzione: i file che iniziano con un punto sono normalmente nascosti. Per vederli è necessario aprire la cartella contenente il file che si desidera visualizzare, selezionare il Menu

Tools -> Folder Options ...

selezionare l’etichetta “View” e selezionare l’opzione “Show hidden files and folders”.

```
# Things you might want to change
options(papersize="a4")
# options(editor="notepad")
# options(pager="internal")
# to prefer Compiled HTML help
options(chmhelp=FALSE)
# to prefer HTML help
options(htmlhelp=TRUE)
```

Il carattere “#” serve per commentare il codice. Una linea che inizia con tale carattere non verrà eseguita!

Se desiderate installare R a casa dovete scaricare il programma:

1. Andate al sito *www.r-project.org*
2. Cliccate su CRAN sul lato sinistro della finestra. Dopodiché selezionate un “CRAN Mirrors”, io generalmente seleziono quello dell’ETH-Zurigo:

<http://cran.ch.r-project.org/>

In base alla piattaforma scaricate l’ultima versione di R, al momento la 2.10.1 ed installatela (a giorni dovrebbe essere rilasciata la versione 2.11.0).

Osservazione 8. Gli utilizzatori del sistema operativo Linux trovano R nei pacchetti della loro distribuzione preferita.

3.3 La linea di comando

Alla sua apertura il programma R mostra una sola finestra. Questa finestra è chiamata *finestra di comando* in quanto essa contiene la *linea di comando*, cioè il luogo dove scrivere i comandi che verranno poi eseguiti dal programma. A prima vista il tutto può sembrare molto scomodo, soprattutto per coloro che sono abituati a lavorare tramite mouse, menu a cascata, bottoni ecc. . Questo in parte è vero ... ma solo nella fase iniziale. Infatti, man mano che apprenderete i comandi di R vi accorgete della velocità e dei vantaggi di questo sistema.

Particolarmente importante, soprattutto durante la fase di apprendimento, è il comando *help(‘nome’)*. Con questo comando richiamate la pagina del manuale di R che descrive la sintassi ed il funzionamento del comando *nome*. Per esempio, digitando *help(help)* appare la pagina con la descrizione del comando *help*. Potete leggere in tale pagina che una scorciatoia al comando *help()* è data dal punto interrogativo ‘?’. Al posto di *help(help)* avremmo anche potuto scrivere *?help*. Ancora un’osservazione prima di iniziare. Se voi digitate *help* senza aggiungere le parentesi tonde verrà visualizzato il codice della funzione *help*. Questo comportamento vale per tutti i comandi di R. Infine, per ricerche avanzate digitare *help.start()*.

Un primo grande vantaggio della linea di comando è dato dal fatto di poter richiamare e quindi ripetere velocemente un comando o una serie di comandi impartiti in precedenza. Utilizzando le frecce “in alto” e “in basso” è possibile scorrere la storia dei comandi impartiti. Per ottenere una lista degli ultimi

comandi eseguiti digitare il comando *history()*. Il numero massimo di comandi tenuti a mente è 512 (questo parametro è tuttavia modificabile). Di default *history()* propone la lista degli ultimi 25 comandi. È possibile tramite l'argomento *max.show* aumentare o diminuire il numero di comandi mostrati: *history(max.show=3)*.

3.4 Strutture semplici e strutture di dati

Cominciamo utilizzando R come un'elementare macchina calcolatrice.

3.4.0.1 Operatori matematici

Per quanto riguarda gli operatori matematici semplici quali la somma, la sottrazione, la moltiplicazione, la divisione e l'elevamento a potenza, R non si differenzia dagli altri linguaggi.

Somma	+
Sottrazione	-
Moltiplicazione	*
Divisione	/
Divisione intera	%/%
Elevamento a potenza	**oppure ^

Ad esempio

```
> 5+7*3
[1] 26
```

Per modificare le priorità in una sequenza di operazioni si utilizzano le parentesi tonde:

```
> (5+7)*3
[1] 36
> 2*3^2
[1] 18
> (2*3)^2
[1] 36
```

3.4.0.2 Assegnamenti

In R si assegna alla variabile *x* il valore 2 tramite l'operatore '=' oppure '<-':

```
> x <- 2
```

Per visualizzare il contenuto di *x* basta eseguire

```
> x
```

oppure

```
> print(x)
```

Per rimuovere la variabile *x* si utilizza il comando remove: 'rm()'

```
> rm(x)
```

Per visualizzare la lista degli oggetti in memoria (variabili, funzioni, ecc.) si usa il comando `objects()`:

```
> objects()
```

Osservazione 9. Un'assegnazione può essere effettuata anche da sinistra verso destra. In altre parole

```
> 3 -> z
```

è un comando valido di R. Infatti

```
> z  
[1] 3
```

3.4.0.3 Successioni

Spesso è necessario creare delle successioni di numeri. In R esiste il comando *seq* (*limite inferiore*, *limite superiore*, *incremento*):

```
> seq(1,5,1)  
[1] 1 2 3 4 5  
> seq(1996, 2003,1)  
[1] 1996 1997 1998 1999 2000 2001 2002 2003  
> seq(1, 0, -0.2)  
[1] 1.0 0.8 0.6 0.4 0.2 0.0
```

Il limite inferiore così come l'incremento sono definiti per default uguali ad 1. Così avremo che

```
> a <- seq(5)  
> a  
[1] 1 2 3 4 5
```

Quest'ultima operazione può essere svolta anche con l'operatore "due punti":

```
> b <- 1:5  
> b  
[1] 1 2 3 4 5
```

Sono pure ammesse oltre che molto utili espressioni del tipo

```
> 0:5/10  
[1] 0.0 0.1 0.2 0.3 0.4 0.5
```

3.4.0.4 Concatenamento e copie multiple di elementi

L'operatore *c()* serve a concatenare più elementi fra loro.

```
> nome.cognome <- c("Claudio","Ortelli")  
> nome.cognome  
[1] "Claudio" "Ortelli"
```

L'oggetto creato è un vettore le cui componenti sono tutto del medesimo tipo (*character* in questo caso). Se si tenta di concatenare oggetti di formato diverso la funzione *c()* trasformerà tutti i dati nel tipo più generale, nel prossimo esempio in formato stringa:

```
> c(1,2.5,"nome")
[1] "1" "2.5" "nome"
```

Lavorare con vettori è utile in quanto è possibile sfruttare la capacità di R di effettuare determinate operazioni su ogni componente del vettore tramite un unico comando. Oltre ad un tempo d'esecuzione inferiore rispetto ad un ciclo *for* si ottiene anche una migliore leggibilità del codice.

```
> x <- c(2,3,-1)
> x^2
[1] 4 9 1
```

Tramite la funzione *rep* (*oggetto*, *numero ripetizioni*) è possibile copiare un oggetto un certo numero di volte. Per esempio

```
> rep(1,5)
[1] 1 1 1 1 1
> rep(x,2)
[1] 2 3 -1 2 3 -1
```

'numero ripetizioni' può essere un vettore contenente numeri interi non negativi. In tal caso anche 'oggetto' deve essere un vettore della medesima lunghezza. Il risultato sarà un vettore dove ogni componente del vettore 'oggetto' è ripetuta tante volte quanto è indicato alla posizione corrispondente del vettore 'numero ripetizioni'.

```
> x
[1] 2 3 -1
> rep(x,c(3,2,1))
[1] 2 2 2 3 3 -1
```

3.4.0.5 Operazioni matematiche

Abbiamo visto le operazioni matematiche su scalari al punto (3.4.0.1). Cosa accade se applichiamo tali operazioni fra due vettori numerici? La risposta è semplice: gli operatori aritmetici agiscono su vettori componente per componente. La prossima tabella mostra tale proprietà.

```
> b <- 1:4
> a <- 2*(0:3)
```

a	b	a+b	a-b	a*b	a/b	a^b
0	1	1	-1	0	0	0
2	2	4	0	4	1	4
4	3	7	1	12	1.33	64
6	4	10	2	24	1.5	1296

È quindi possibile scrivere formule anche relativamente complesse.

```
> z <- (a^2 - b^2)^2 / (a^2 + b^2)
> z
[1] 1.000000 0.000000 1.960000 7.692308
```

3.4.0.6 Tipi di parentesi e loro significato

R possiede tre tipi di parentesi: parentesi tonde, quadre e graffe. Le parentesi tonde servono a definire l'ordine nel quale le operazioni matematiche vanno eseguite. Le parentesi quadre invece servono a definire le componenti di un vettore o di una matrice sulle quali vogliamo operare (vedremo più tardi come definire una matrice in R). Infine, le parentesi graffe servono a delimitare blocchi di codice come ad esempio in caso di istruzioni `if { ... } else { ... }`. Definiamo ora un vettore qualsiasi di 3 componenti, per esempio

```
> x <- 1:3*10
> x
[1] 10 20 30
> x[1]          #accesso al primo elemento
[1] 10
> x[2]          #accesso al secondo elemento
[1] 20
> x[3]          #accesso al terzo elemento
[1] 30
> x[4]          #accesso al quarto elemento
[1] NA
```

Il termine **NA** significa “not available”. È possibile dare un vettore di interi quale argomento all'interno delle parentesi quadre:

```
> x[rep(1,5)]
[1] 10 10 10 10 10
> x[c(1,3)]
[1] 10 30
```

Indici negativi significano “non voglio”, cioè

```
> x[-1]         #non voglio il primo elemento
[1] 20 30
> x[-c(1,3)]    #non voglio il primo & il terzo elemento
[1] 20
```

Tuttavia non è possibile mischiare indici positivi e negativi: `x[c(-1,3)]` non funziona.

3.4.0.7 Operatori logici

Le variabili logiche sono variabili binarie, cioè assumono solo due valori vero (1 o TRUE) e falso (0 o FALSE). In R si definisce una variabile logica assegnandole il valore TRUE o FALSE, esattamente come si usa fare con una qualsiasi altra variabile numerica o di tipo stringa. Spesso le variabili logiche sono il risultato di un operatore logico quale ad esempio “>=”, “<”:

```
> 5 >=4
[1] TRUE
```

In R troviamo i seguenti operatori logici:

Operatore	Funzione
<	minore a
>	maggiore a
<=	minore o uguale a
>=	maggiore o uguale a
==	uguale a
!=	diverso da

```
> 5!=1
[1] TRUE
```

Così come gli operatori matematici, anche gli operatori logici operano componente per componente quando vengono applicati a vettori o matrici.

```
> x <- 1:3
> x < 3
[1] TRUE TRUE FALSE
> y <- 3:1
> y
[1] 3 2 1
> x<y
[1] TRUE FALSE FALSE
```

Una caratteristica molto utile degli operatori logici applicati a vettori consiste nel fatto che è possibile contare il numero di componenti che soddisfano una certa condizione. Per esempio, supponiamo di voler sapere quali componenti del vettore 1:10 sono maggiori o uguali a 7.

```
> 1:10 >= 7
[1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE
TRUE TRUE
> sum(1:10 >= 7)
[1] 4
```

La funzione *sum()* ha tramutato FALSE in 0 e TRUE in 1 e sommato tutte le componenti.

3.4.0.8 Estrazione di indici

I valori logici sono molto utilizzati per estrarre da strutture più complesse solo quelle parti che soddisfano determinate condizioni. Per esempio, supponiamo di avere osservazioni sull'altezza e sul peso di 9 individui,

```
> nome <- c("Claudio", "Maria", "Marco", "Matteo", "Francesca",
"Giovanni", "Marta", "Giorgio", "Alberto")
> altezza <- 140:148
> peso <- c(seq(50, 64, 2),59)
```

Quali sono gli individui che hanno un peso superiore a 60 Kg? Per rispondere a questa domanda è sufficiente eseguire il comando

```
> nome[peso>60]
[1] "Marta", "Giorgio"
```

E quanto sono alti?

```
> altezza[peso>60]
[1] 146 147
```

È possibile specificare più condizioni, legandole fra loro tramite gli operatori & ("logical and"), | ("logical or"). Ad esempio

```
> nome[peso>60 & peso != 64]
[1] "Marta"
```

3.4.1 Strutture di dati

Nel precedente paragrafo abbiamo visto come costruire una struttura "vettore" partendo da semplici variabili scalari (tramite la funzione di concatenamento `c()`). Come ben sapete dai vari corsi di statistica, matematica ed econometria, in molte circostanze è indispensabile generalizzare il concetto di vettore a strutture multidimensionali (chiamate generalmente arrays) di cui le matrici sono l'esempio più noto. Conosceremo ora le strutture multidimensionali contenute in R.

3.4.1.1 Data Frames

I data frames sono delle tabelle in cui è consentito inserire colonne di formato (tipo) diverso fra loro. Il comando per generare un data frame è il seguente:

data.frame (Elemento1, Elemento2, ...)

Tutti gli argomenti che vengono elencati in `data.frame` *devono* avere il medesimo numero di righe. Tornando all'esempio precedente dei nove individui:

```
> tabella <- data.frame(nome,altezza,peso)
> tabella
```

	nome	altezza	peso
1	Claudio	141	50
2	Maria	142	52
3	Marco	143	54
4	Matteo	144	56
5	Francesca	145	58
6	Giovanni	146	60
7	Marta	147	62
8	Giorgio	148	64
9	Alberto	149	59

Un esempio importante per la creazione di un `data.frame` è costituito dall'importazione di dati da una tabella in formato csv. Il comando da utilizzare è **read.table** (si veda la documentazione per la descrizione dettagliata digitando `"help(read.table)"`):

```
> dati <- read.table("//usi/dfs/istituti/Documents/ortellic/
  My Documents/R/daticorso.csv",header=TRUE,sep=";",
  colClasses=c("character",rep("numeric",6)))
```

Il primo argomento indica il nome del file da aprire. Si noti che mentre Windows utilizza il carattere “

textbf” per separare nomi di cartelle, R utilizza lo standard unix e cioè il carattere “/”. Quando il file da importare si trova nell’attuale cartella di lavoro (cioè la cartella indicata nel campo “start in” dell’icona con cui avviate il programma), è sufficiente indicare il nome del file. Avremmo dunque potuto anche scrivere

```
> dati <- read.table("daticorso.csv", header=TRUE, sep=";",
  colClasses=c("character",rep("numeric",6)))
```

Il secondo argomento indica se la prima riga di dati contiene le intestazioni delle colonne. Il parametro **sep** determina il carattere utilizzato per separare il contenuto delle colonne. L’ultimo argomento, **colClasses**, determina il tipo di dato contenuto in ciascuna colonna (logical, integer, real, complex e character sono i tipi di vettori (colonne) che si possono importare in un data frame).

Prima di eseguire l’importazione è bene dare un’occhiata al file contenente i dati. Aprite dunque il file con un semplice elaboratore di testo (Notepad) e verificate che il carattere di separazione sia la virgola, il punto e virgola o qualche altro carattere particolare. Modificate il comando di conseguenza (**sep**="*").

Per accedere a dati contenuti in un data frame si usano le parentesi quadre. Al loro interno vanno indicati i numeri delle righe e delle colonne che si vogliono estrarre. Se si desiderano tutte le righe o tutte le colonne di un data frame è sufficiente lasciare vuoto l’indice corrispondente. Per esempio

```
> dati[1,]
      date  SP100  MIB30   DAX   SMI   EUR   CHF
1  1/1/1999  604.03  35152  5002.39  7160.7  0.8516  1.413
```

Per vedere la prima, la terza e l’ultima riga delle prime 3 colonne del data frame “dati”

```
> dati[c(1,3,1141),1:3]
      date  SP100  MIB30
1  1/1/1999  604.03  35152
3  1/5/1999  615.38  37807
1141 5/16/2003  475.72  24740
```

Così come con i vettori è possibile utilizzare indici negativi sia per le righe che per le colonne di un data frame. Ricordo che non è possibile avere indici di segno opposto per indicare rispettivamente escludere righe o colonne.

```
> dati[c(1,3,1141),-(2:5)]
      date   EUR   CHF
1  1/1/1999  0.8516  1.4130
3  1/5/1999  0.8486  1.3697
1141 5/16/2003  0.8665  1.3250
```

Diamo alcuni comandi molto utili per lavorare con oggetti di tipo data frame.

1. Dimensioni di un data frame:

Il comando **dim** restituisce un vettore contenente il numero di righe e di colonne del data frame.

```
> dim(dati)
[1] 1141 7
```

2. Aggiungere una colonna ad un data frame

Per aggiungere una colonna utilizzare il comando **cbind()**:

```
> #genera 5 numeri aleatori normali di media 0 e deviazione
standard 1
> caso <- rnorm(5)
dati.test <- cbind(dati[1:5,1:3], caso)
```

Da notare che R assegna quale nome della nuova colonna il nome della variabile utilizzata e cioè **caso**.

```
> dati.test
      date   SP100   MIB30      caso
1 1/1/1999  604.03  35152  0.26866637
2 1/4/1999  604.93  37417 -0.03481095
3 1/5/1999  615.38  37807 -0.50321912
4 1/6/1999  631.00  37807 -0.37532667
5 1/7/1999  632.76  37797  0.07938619
```

3. Per modificare il nome delle colonne di un data frame utilizzare il comando **names()**:

```
> names(dati.test) <- c("Data","S&P100","MIB","Simulazione")
```

Il risultato del comando **names** è un vettore. Quindi se volessimo modificare solo alcuni dei nomi dovremmo utilizzare le parentesi quadre:

```
> names(dati.test) [c(2,4)] <- c("S&P","Sim")
```

4. Per modificare il nome delle righe di un data frame utilizzare il comando **row.names()**:

```
> row.names(dati.test) [1:3] <- c("uno","due","tre")
> dati.test
```

	Data	S&P	MIB	Sim
uno	1/1/1999	604.03	35152	0.26866637
due	1/4/1999	604.93	37417	-0.03481095
tre	1/5/1999	615.38	37807	-0.50321912
4	1/6/1999	631.00	37807	-0.37532667
5	1/7/1999	632.76	37797	0.07938619

Per accedere ai dati di un data frame è pure possibile scrivere i nomi delle colonne/righe

```
> dati.test["uno","MIB"]  
[1] 35152
```

mentre per accedere ai dati di un'intera colonna, oltre al comando **dati.test[, "MIB"]**, è consentita la sintassi

```
> dati.test$MIB  
[1] 35152 37417 37807 37807 37797
```

Il comando **attach()** permette di utilizzare direttamente i nomi delle colonne di un data frame senza doverne anteporre il nome. Ciò significa che dopo l'esecuzione del comando

```
> attach(dati.test)
```

anziché scrivere **dati.test\$MIB** sarà consentito scrivere solo

```
> MIB
```

Data frames diversi possono contenere colonne (variabili) con nomi identici. Nel caso che due o più data frames contenenti variabili col medesimo nome siano stati richiamati col comando **attach()**, R andrà a prendere i dati dall'ultimo data frame richiamato. Ad ogni modo esiste il comando **detach('nome')** che elimina il data frame 'nome' dalla lista.

3.4.1.2 Matrici ed arrays

Le matrici sono degli arrays a due dimensioni. Gli elementi di una matrice sono indicati tramite due indici, l'indice riga e l'indice colonna. In R un array può avere fino ad 8 dimensioni. Come i vettori anche le matrici (e più in generale gli arrays dunque) devono essere omogenee rispetto al formato dei dati. Per intenderci, non è possibile che una matrice contenga una colonna di stringhe ed una colonna di numeri interi. Per creare una matrice si usa il comando *matrix()*.

$$matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE)$$

Per default la matrice è riempita per colonne.

```
> matrix(1:6,nrow=3)  
      [,1] [,2]  
[1,]    1    4  
[2,]    2    5  
[3,]    3    6
```

Se viene indicato il numero delle righe ma non quello delle colonne, R aggiunge tante colonne quante sono necessarie ad esaurire i dati che gli sono stati passati. È bene prestare molta attenzione quando si costruisce una matrice. Quando il numero di dati non corrisponde al numero di elementi della matrice, R ci avverte con un messaggio. Quando il numero di elementi della matrice è inferiore al numero di dati, R utilizza solo una parte di essi. Nel caso contrario R riutilizza i dati più volte fino al completamento della matrice.

```
> matrix(1:2,nrow=3,ncol=4)
      [,1] [,2] [,3] [,4]
[1,]    1    2    1    2
[2,]    2    1    2    1
[3,]    1    2    1    2
```

Il comando **as.matrix()** tenta di convertire in matrice l'argomento che gli viene dato. Un'immediata applicazione è la conversione della parte numerica del data frame 'dati' in matrice³:

```
> serie <- as.matrix(dati[,2:7])
```

Per default **as.matrix** mantiene i nomi delle colonne e delle righe. Come per un data frame anche per una matrice è possibile assegnare o modificare i nomi delle righe/colonne. Il comando è tuttavia leggermente diverso. Ecco un esempio:

```
> dim(serie)
[1] 1141 6
> indice.riga <- paste("riga",1:1141,sep="")
> dimnames(serie)[[1]] <- indice.riga
> serie[1:3,1:2]
      SP100  MIB30
riga1  604.03  35152
riga2  604.93  37417
riga3  615.38  37807
```

Come per i data frames è possibile utilizzare i nomi al posto dei numeri di riga/colonna

```
> serie["riga14",]
      SP100      MIB30      DAX      SMI      EUR      CHF
626.0100  35382.0000  5143.0600  7300.2000  0.8642  1.3887
```

3.4.1.3 Calcolo matriciale

Il grosso vantaggio nell'aver trasformato il data frame 'dati' nella matrice 'serie' consiste nel fatto di poter utilizzare il calcolo matriciale. Gli operatori **+**, **-**, *****, **/** e **^** operano componente per componente quando utilizzati con matrici.

```
> A <- matrix(1:4, nrow=2,ncol=2)
> B <- matrix(4, nrow=2,ncol=2)
> A-B
      [,1] [,2]
[1,]   -3   -1
[2,]   -2    0
```

Quando applichiamo una funzione matematica ad una matrice, ad esempio la funzione esponenziale **exp()** o la sua inversa **log()**, è come se applicassimo la funzione ad ogni componente della matrice:

³Il comando **as.data.frame(serie)** converte la matrice 'serie' in un oggetto di classe data frame mantenendone le dimensioni ed il tipo.

```
> exp(matrix(c(1,0,0,1),nrow=2))
      [,1] [,2]
[1,] 2.718282 1.000000
[2,] 1.000000 2.718282
```

La moltiplicazione matriciale vera e propria è definita tramite l'operatore `%*%`, mentre per matrici quadrate non singolari l'inversione è data dalla funzione `solve()`:

```
> solve (A-B)
      [,1] [,2]
[1,] 0 -0.5
[2,] -1 1.5
> solve(A-B) %*% (A-B)
      [,1] [,2]
[1,] 1 0
[2,] 0 1
```

Osservazione 10. La funzione `solve()` permette di risolvere sistemi lineari del tipo $Ax = b$. Per motivi di carattere numerico si raccomanda di utilizzare la funzione `solve(A, b)` anziché `solve(A) %*% b` per il calcolo di espressioni quali $A^{-1}b$.

Tra le molte funzioni matriciali ecco alcune di particolare importanza:

- **t(A)**: ritorna la trasposta di A.
- **det(A)**: calcola il determinante di A.
- **eigen(A)**: ritorna una variabile di tipo 'list' contenente il vettore degli autovalori e la matrice degli autovettori di A. (Rimandiamo al punto 3.4.1.7 per la spiegazione delle variabili di classe 'list').
- **diag(A)**: se A è una matrice ritorna gli elementi della diagonale principale di A, se A è un vettore ritorna una matrice diagonale dove gli elementi della diagonale principale sono le componenti del vettore A.

Domanda 5. Come fareste a calcolare la traccia di una matrice?

3.4.1.4 Concatenamento di matrici

Capita spesso di dover aggiungere righe e/o colonne ad una matrice. I comandi `rbind()` e `cbind()` servono a tal fine.

```
> A
      [,1] [,2]
[1,] 1 3
[2,] 2 4
> col3 <- c(5,6)
> cbind(A,col3)
      [,1] [,2] col3
[1,] 1 3 5
[2,] 2 4 6
```


3.4.1.5 La funzione `apply()`

La funzione **`apply()`** permette di applicare una qualsiasi funzione (fra cui in particolare funzioni matematiche quali ad esempio **`mean()`**, **`max()`** o **`min()`**) rispetto ad una delle k dimensioni di un array k dimensionale. Noi per semplicità abbiamo trattato solo matrici ($k = 2$) ma l'uso di **`apply()`** si estende ad un qualsiasi array multidimensionale. La sintassi è la seguente

`apply(X, MARGIN, FUN)`

```
> A
      [,1] [,2]
[1,]    1    3
[2,]    2    4
> apply(A,1,sum) # applica la funzione sum alle righe di A
[1] 4 6
> apply(A,2,mean)
[1] 1.5 3.5
```

3.4.1.6 Calcolo di returns e di log returns

Ora che abbiamo i dati nella matrice serie possiamo calcolare agevolmente i returns,

```
> returns <- (serie[2:1141,]-serie[1:1140,])/serie[1:1140,]
```

e i log returns

```
> r <- log(serie[2:1141,]/serie[1:1140,])
```

oppure anche il comando (si veda **`help(diff)`**)

```
> r <- apply(log(serie[1:1141,]),2,diff)
```

Spiegate il significato del seguente comando:

```
> apply(r==0,2,sum)
      SP100  MIB30  DAX  SMI  EUR  CHF
      41      35   33   40   22   36
```

3.4.1.7 Liste

Una lista è un vettore le cui componenti possono essere un qualsiasi oggetto di R. Le liste sono particolarmente utili per raggruppare in maniera ordinata scalari, vettori, matrici, arrays di formato e dimensione diversa. Una lista può contenere altre liste. La sintassi per creare una variabile di tipo lista è la seguente

`list(elemento1, elemento2, ...)`

Esempio 4. Costruzione di una lista.

```
> mia.lista <- list(matrix(1:6,nrow=2),c("Nome","Cognome"), dati)
> is.list(mia.lista)
[1] TRUE
```

La variabile 'mia.lista' contiene ora tre oggetti: il primo è una matrice (2×3) di numeri interi, il secondo è un vettore di stringhe ed il terzo è un data frame. Per accedere ad uno o più elementi della lista è necessario utilizzare le **doppie** parentesi quadre:

```
> mia.lista [[1]]
      [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

Per accedere direttamente ad elementi della matrice **mia.lista [[1]]** è sufficiente aggiungere le parentesi quadre con i rispettivi indici desiderati

```
> mia.lista [[1]][2,]
[1] 2 4 6
```

Se viene utilizzato il comando

$$list(nome1 = elemento1, nome2 = elemento2, \dots)$$

è possibile riferirsi direttamente ai singoli elementi della lista utilizzando il loro nome (come nel caso dei data frames):

```
> mia.lista <- list(A=matrix(1:6,nrow=2),b=c("Nome","Cognome"),
dati=dati)
> mia.lista$b
[1] "Nome" "Cognome"
> mia.lista$A[2,]
[1] 2 4 6
```

Per sapere la lunghezza della lista (il numero di elementi) utilizzate il comando **length()**:

```
> length(mia.lista)
[1] 3
```

Per aggiungere un nuovo elemento alla lista:

```
mia.lista[[5]] <- c("Il","mio","nuovo","elemento")
```

Come avrete forse notato, non è obbligatorio utilizzare indici adiacenti. Ora però mia.lista ha un "buco".

```
> mia.lista[[4]]
NULL
```

L'elemento in posizione 5 della lista non ha un nome. Un metodo alternativo per ampliare una lista è quello di utilizzare direttamente il nome desiderato, cioè

```
> mia.lista$ultimo <- "Ultimo elemento"
> mia.lista$ultimo
[1] "Ultimo elemento"
```

la lista viene ampliata di un elemento (posizione 6) ed è possibile accedere a quest'ultimo col suo nome.

Per eliminare un elemento è sufficiente assegnare il valore **NULL**.

```
> mia.lista[[4]] <- NULL
```

Quando si elimina un elemento da una lista, tutti gli elementi con indice superiore a quello eliminato vengono scalati di uno. Ciò significa che ora la lista `mia.lista` nella posizione 4 contiene ...

```
> mia.lista[[4]]
[1] "II" "mio" "nuovo" "elemento"
```

Non è possibile utilizzare indici multipli con le liste. Un comando del tipo `mia.lista[[2:3]]` non è accettato.

Concludiamo questo paragrafo sulle liste dicendo che le liste sono utilizzate per raccogliere l'output di comandi complessi, si veda ad esempio il comando utilizzato per stimare modelli lineari (`?lm`).

3.4.2 Grafici

Come vedremo fra poco R è particolarmente adatto all'analisi grafica di dati. Inizialmente occorre un investimento per apprendere i principali comandi. Il comando principale è `plot()`. Ecco un esempio.

```
> #creazione di un vettore contenente i nomi delle serie
> nomi <- dimnames(r) [[2]]
> i <- 1
> #plot della prima serie di returns
> plot (r[,i],
       type="l",
       main= paste("Log returns di",nomi[i]),
       sub=paste("Media:",
       format(mean(r[,i]),digits=2)),
       xlab="Osservazione",
       ylab=nomi[i])
```

Per avere un'idea di quanto due serie storiche possano essere correlate fra loro è utile fare un plot di una serie sull'altra.

```
> i <- 1; j <- 2
> plot (r[,i],r[,j],
       main = paste (nomi[i], "vs", nomi[j]),
       xlab=nomi[i],
       ylab=nomi[j],
       col="blue")
```

`help(par)` indica tutte le opzioni disponibili.

3.4.2.1 Inserimento di elementi grafici

- Il comando `abline(a=0,b=1)` inserisce una linea retta di pendenza **b** ed intercetta **a** nella finestra grafica attiva

```
> abline(a=0,b=1)
```

- Per inserire una linea orizzontale lungo i punti 0.01, 0.02 e 0.03 utilizzare

```
> abline(h=c(0.01,0.02,0.03))
```

mentre per linee verticali

```
abline(v=c(0.01,0.02,0.03))
```

- Per inserire una freccia utilizzare `arrows(x1,y1,x2,y2)` dove (x1,y1) sono le coordinate della coda e (x2,y2) quelle della punta della freccia.

```
> arrows(-0.06,0,-0.04,0.05)
```

- Per inserire del testo utilizzare il comando `text(x,y,text)`

```
> arrows(-0.04,0.05,"Guarda che bel grafico")
```

3.4.2.2 Altri tipi di grafici

- Il comando `pairs()`: matrici di grafici

Il comando `pairs()` genera una matrice contenente i crossplot di tutte le coppie (x,y) di variabili (colonne) di una matrice di dati.

```
> pairs(r[,1:4])
```

- Istogramma

```
> i <- 1
```

```
> hist (r[,i],
        probability=TRUE,
        main=paste("Istogramma di",nomi[i]),
        xlab=nomi[i],
        breaks=17)
```

- Q-Q plot rispetto alla normale

```
> qqnorm(r[,i],main=paste("Normal Q-Q Plot per",nomi[i]),
          xlab="Quantili teorici",ylab="Quantili empirici")
> qqline(r[,i])
```

- Q-Q plot

```
> i <- 1; j <- 2
```

```
> qqplot(r[,i],r[,j],main=paste("Q-Q Plot fra",nomi[i],"e",nomi[j]),
          xlab=nomi[i],ylab=nomi[j])
```

- Il comando `acf()` stima la funzione di autocovarianza o autocorrelazione e per default ne esegue il grafico Esempio:

```
> acf(r[,i],lag.max=10,type="correlation",main=nomi[i])
```

- Il comando **ccf()** calcola ed esegue il grafico della cross-correlation, cross-covariance function. Esempio:

```
> i <- 2
> j <- 3
> ccf(r[,i],r[,j],lag.max=10, type="correlation", main=
  paste("Correlazione dei returns di ",nomi[i],"e",nomi[j],
  sep=" "))
```

Domanda 6. Come interpretate quest'ultimo grafico?

3.4.2.3 Alcuni utili comandi grafici

- Il comando **windows()** apre una nuova finestra grafica e la rende attiva. Ciò significa che il prossimo grafico sarà visualizzato nella finestra appena aperta. Quando ci sono più finestre grafiche aperte il comando **dev.set('#device')** permette di scegliere quale deve essere la finestra in cui ridirigere l'output. **'#device'** indica il numero della finestra o più precisamente del device. Per ottenere il vettore contenente tutti i numeri di device aperti al momento utilizzare il comando **dev.list()**. Il comando **dev.cur()** restituisce il numero del device attivo al momento (*'cur'* da *'current'*) mentre **dev.off('#device')** chiude il device specificato (per default chiude il device corrente).
- Il comando **par(mfcol=c(nr,nc))** e **par(mfrow=c(nr,nc))** apre una finestra il cui contenuto sarà una matrice di $nr \times nc$ grafici. Se è utilizzato il primo (secondo) comando, la matrice verrà riempita per colonne (righe).
- Il comando **postscript("miofile.ps")** apre come device un file testo di nome "miofile.ps". Esempio:

```
> postscript("foo.ps",paper="a4",width=5,height=4,bg="yellow")
> plot(r[,6],main=paste("Log-returns di ",nomi[6],sep=""),
  type="l",xlab="Osservazioni",ylab=nomi[6])
> dev.off()
```

Le dimensioni della larghezza e dell'altezza del grafico sono date in pollici (1 pollice = 2.54 centimetri). È importante ricordarsi di chiudere il device una volta terminato!

- I comandi **pdf()**, **bpm()**, **jpeg()** e **png()** sono simili al comando **postscript()** e servono a generare grafici nei rispettivi formati.

3.4.3 Comandi if, for, while, repeat

3.4.3.1 Il comando if

La sintassi del comando è la seguente

if (condizione) comando else comando

Esempio 5. Comando if semplice.

```
> if (rnorm(1) > 0) print ("Maggiore di 0") else print ("Minore o uguale a 0")
```

Se si desidera eseguire più comandi, è necessario racchiudere il codice fra parentesi graffe "{ ...}".

Esempio 6. Comando if.

```
> if (rnorm(1)>0) {  
  plot (rnorm(10),main="Normale")  
  print ("Terminato")  
} else {  
  plot (rt(10,df=6),main="t-6")  
  print ("Terminato")  
}
```

Osservazione 11. Abbiamo visto in precedenza che se x ed y sono due vettori il comando $x == y$ verifica l'uguaglianza componente per componente ed il risultato consiste in un vettore di TRUE/FALSE. Per tale motivo l'utilizzo di un'espressione del genere nel contesto dell'istruzione **if** può porre dei problemi se l'operatore binario $==$ non è utilizzato su scalari. R risolve il problema nella misura in cui considera solo la prima componente del vettore $x==y$. Per testare l'uguaglianza fra due oggetti (matrici, liste, vettori) si utilizza il comando **identical()**.

3.4.3.2 Il ciclo for

Capita sovente di eseguire più volte lo stesso comando o di dover applicare ripetutamente la stessa funzione ma con argomenti di valore diverso. La prima possibilità in situazioni del genere è quella di utilizzare il comando **for()**, la cui sintassi è

for (var in seq) comando

Esempio 7. Comando for.

```
> for (nome in nomi) print (nome)
```

Dall'esempio precedente vi sarete resi conto che è possibile eseguire un ciclo **for** utilizzando un vettore numerico o non numerico quale fonte dei dati con cui eseguire il ciclo. In maniera simile al comando **if**, se si desidera eseguire più comandi è necessario utilizzare le parentesi graffe.

Esempio 8. Comando for.

```
> for (indice in 1:10) {  
  x <- rnorm(1000)  
  if (exists("media")) {  
    media <- c(media,mean(x))  
  } else {  
    media <- mean(x)  
  }  
  if (exists("varianza")) {  
    varianza <- c(varianza,var(x))  
  }  
}
```

```

    } else {
      varianza <- var(x)
    }
  }
}

```

È possibile terminare anzitempo un ciclo **for** utilizzando il comando **break**:

```

> for (i in 1:10) {
  if (i > 5) break
  print (i)
}

```

3.4.3.3 Il ciclo while

Con un ciclo di tipo **for** il numero di ripetizioni è determinato anzitempo. Con il comando **while()** è invece possibile stabilire dinamicamente quante volte ripetere il ciclo: il ciclo viene ripetuto fin tanto che la condizione precisata fra parentesi tonde è vera.

```

> i <- 1
> while (i <= 5) {
  print (i)
  i <- i + 1
}

```

Osservazione 12. Se la condizione precisata nel comando **while** non è inizialmente vera, il ciclo non verrà mai eseguito. Come con il comando **for** è possibile interrompere l'esecuzione di un ciclo **while** utilizzando il comando **break**.

3.4.3.4 Il ciclo repeat

L'ultimo comando per generare dei cicli è il comando **repeat**. Repeat non necessita un condizione iniziale come il comando **while** e questo assicura che il ciclo venga sempre eseguito almeno una volta. Per terminare un ciclo **repeat** è necessario utilizzare il comando **if** associato al comando **break**.

Esempio 9. Comando break.

```

> i <- 1
> repeat {
  print (i)
  i <- i + 1
  if (i > 5) break
}

```

3.4.3.5 Il comando next

Il comando **next** può essere adoperato in un qualsiasi ciclo **for**, **while** o **repeat**. Il suo scopo è quello di interrompere l'esecuzione del codice all'interno del ciclo e saltare immediatamente all'inizio del ciclo. Ecco un esempio:

```

x <- rnorm(10)
for (i in 1:10) {
  if (x[i] < 0 ) next
  print (sqrt(x[i]))
}

```

3.4.3.6 Vettorizzazione di cicli

Per ragioni di efficienza e velocità d'esecuzione è opportuno evitare, quando possibile, i cicli **for**, **while** o **repeat** in quanto rallentano l'esecuzione del programma. Spesso possiamo riscrivere un ciclo in forma vettoriale. Prendendo ad esempio l'ultimo ciclo appena trattato (quello del comando **next**) esso può essere riscritto come segue:

```
> x <- rnorm(10)
> sqrt(x[x>0])
```

È questa una caratteristica molto importante di R (ed altri linguaggi di alto livello). Tuttavia ciò può avere anche degli svantaggi, come ad esempio un più largo uso della memoria.

3.4.4 Funzioni

Una funzione non è altro che un insieme ordinato di comandi che vengono raggruppati sotto un medesimo nome. Le funzioni possono ma non devono avere degli argomenti di input ed allo stesso modo possono ma non devono restituire un risultato. Vediamo un esempio:

```
> cubo <- function (x=10) { return(x^3) }
```

Con questo comando abbiamo definito la funzione **cubo**. Essa è ora contenuta nella lista degli oggetti con le altre variabili (**objects()**). Per cancellare la funzione **cubo()** si utilizza il comando **rm(cubo)**.

Alla sinistra del comando d'assegnazione troviamo sempre il nome della funzione che vogliamo creare mentre alla sua destra avremo la parola chiave **function** seguita in parentesi tonde dagli argomenti che vogliamo passare. È possibile assegnare ad ogni argomento un valore di default, nel nostro esempio 10, tramite l'operatore di uguaglianza. Ciò significa che se al momento della chiamata non viene specificato nessun valore per la variabile **x**, allora la funzione **cubo** assume un valore di **x** pari a 10. Infatti, provate a scrivere "**cubo()**"! Per finire viene specificato il codice fra parentesi graffe.

Valgono inoltre le seguenti osservazioni.

1. È ammesso utilizzare qualsiasi variabile per cui l'esecuzione del codice è ben definita. Nel nostro primo esempio **x** potrebbe essere un qualsiasi vettore, data frame o matrice di soli numeri.
2. Se vogliamo che la funzione ritorni il risultato ma senza che appaia sullo schermo occorre utilizzare il comando **return(invisible(x))**. In questo caso il comando

```
> cubo(4)
```

calcola il cubo di 4 ma senza mostrare il risultato.

```
> y <- cubo(4)
```

```
> y
```

```
[1] 64
```


3. Le variabili dichiarate all'interno di una funzione cessano di esistere una volta che la funzione è terminata. Tuttavia se l'assegnazione è fatta utilizzando l'operatore '«-', anziché creare una variabile locale alla funzione, R genera una variabile globale, cioè una variabile definita a livello dello spazio di lavoro interattivo.
4. Ci possono essere situazioni in cui il numero di argomenti che vengono passati alla funzione sono variabili: si pensi alla funzione di concatenamento `c()`. R permette quindi di creare funzioni alle quali è possibile passare un numero qualsiasi di argomenti. Il simbolo per indicare ciò è dato dai tre puntini '...'. Si veda il manuale di R per maggiori informazioni.
5. È possibile testare l'esistenza di un argomento utilizzando il comando `missing()`. Esempio:

```
> cubo <- function (x) { if (missing(x))
  print("Argomento non specificato") else return(x^3) }
> cubo()
[1] "Argomento non specificato"
```

6. Quando ci sono più argomenti è importante l'ordine con il quale essi vengono dati alla funzione. Esempio:

```
> potenza <- function (base,esponente) { return(base^esponente)}
> potenza (1,2)
[1] 1
> potenza (2,1)
[1] 2
```

Se si è incerti si può sempre assegnare esplicitamente i valori alle variabili al momento della chiamata.

```
> potenza(esponente=2,base=1)
[1] 1
```

R permette la chiamata alla funzione anche con nomi incompleti. Ad esempio

```
> potenza(esp=2,ba=1)
[1] 1
```

3.4.4.1 La distribuzione dei log-returns

Vogliamo analizzare più in dettaglio la distribuzione e le proprietà statistiche dei rendimenti logaritmici o log-returns. Ci interessiamo dapprima ai quantili. In R è possibile ottenere i quantili desiderati utilizzando la funzione `quantile()`.

```
> i <- 1
> quantile(r[i], seq(0,1,0.1))
```

La funzione `quantile` può essere utilizzata per fare il plot della funzione di ripartizione empirica dei nostri dati:

```
> plot(quantile(r[i], seq(0, 1, 0.01)), seq(0, 1, 0.01), type="l",  
main=nomi[i], xlab="Return",ylab="Probability")
```

Per stimare la funzione di densità esiste il comando `density()`:

```
> plot(density(r[i]), main=paste("Densità stimata dei log-returns di  
", nomi[i]))
```

3.5 Funzioni e programmazione da Excel

È possibile utilizzare R direttamente da Excel tramite Visual Basic. Per fare questo è necessario installare uno o più pacchetti reperibili al seguente indirizzo:

[http : //sunsite.univie.ac.at/rcom](http://sunsite.univie.ac.at/rcom).

La descrizione del processo d'installazione va oltre lo scopo di questa introduzione. All'utente inesperto si consiglia l'installazione del pacchetto *RAndFriendsLight* reperibile all'indirizzo (selezionate l'etichetta **Download** del precedente link)

[http : //sunsite.univie.ac.at/rcom/download.html](http://sunsite.univie.ac.at/rcom/download.html)

il quale installa e configura automaticamente, oltre che il pacchetto base di R, anche tutte le altre componenti necessarie al funzionamento e la comunicazione di R con Microsoft Excel.

Capitolo 4

Calculating the Variance-Covariance Matrix

4.1 In Excel

Si vedano le dispense tratte dal Capitolo 8 di Benninga.

4.2 In R

Supponiamo che i prezzi azionari siano in una matrice ($n \times N$) P dove ad ogni colonna corrispondono le osservazioni dei prezzi di una certa azione. I prezzi sono ordinati cronologicamente, partendo dalla data più distante. I rendimenti logaritmici possono essere calcolati nel seguente modo:

1. Calcola il logaritmo dei prezzi

```
> lnP <- log(P)  
> nrOss <- nrow(lnP)
```
2. Calcola i rendimenti storici

```
> lnRet <- lnP[-1,]-lnP[-nrOss,]
```
3. Calcola i rendimenti storici medi

```
> lnRetMedio <- mean(lnRet)
```
4. Calcola la matrice di covarianza

```
> V <- cov(lnRet,use="pair")
```
5. Calcola le volatilità storiche

```
> volStoriche <- sqrt(diag(V))  
> names(volStoriche) <- colnames(V)
```
6. Calcola la matrice di correlazione

```
> rho <- cor(lnRet,use="pair")
```

O in alternativa potremmo calcolare **rho** nel modo seguente

```
7. rho <- diag(1/volStoriche) %*% V %*% diag(1/volStoriche)
```

Domanda 7. Spiegate cosa avviene nel punto 7. Suggerimento: confrontate il codice con la formula (2.18).

Come potete vedere il calcolo di tutte queste quantità è estremamente semplice e rapido.

4.3 The Single-Index Model

Nel corso del semestre invernale avete incontrato un modello chiamato "The single-factor model" (Haugen, pp. 136). Questo modello lo ritroviamo ora in Benninga con il nome di "The Single-Index Model" (Capitolo 8.5, pp. 156). Evidenzieremo ora alcune proprietà del modello applicandolo a dati svizzeri.

Consideriamo dunque $N = 5$ azioni incluse nello SMI e i cui rendimenti logaritmici¹ sono delle variabili aleatorie notate r_i , $i = 1, \dots, 5$. Il rendimento dell'indice di mercato, lo SMI nel nostro caso, sarà indicato con r_x . E' ipotizzata una relazione lineare fra r_i e r_x del tipo

$$r_{it} = \alpha_i + \beta_i r_{xt} + \varepsilon_{it} \quad t = 1, \dots, n \quad (4.1)$$

Per fisso t possiamo raccogliere i rendimenti delle $N = 5$ azioni nel vettore aleatorio $r_t = (r_{1t}, r_{2t}, \dots, r_{Nt})'$. Le N equazioni possono essere scritte in forma compatta come

$$r_t = \alpha + \beta r_{xt} + \varepsilon_t \quad t = 1, \dots, n. \quad (4.2)$$

Osservazione 13. La (4.2) è dunque un sistema di $N = 5$ equazioni: benvenuti nel mondo multivariato!

Ipotesi sulla variabile esplicativa (stocastica):

Ipotesi 1. $E(r_{xt}) = \mu_x$

Ipotesi 2. $Var(r_{xt}) = \sigma_x^2 < \infty$.

Ipotesi sugli errori:

Ipotesi 3. $\varepsilon_t \sim i.i.d. (0, \Sigma)$.

Ipotesi 4. $Cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$ per $i \neq j$.

Ipotesi sulla relazione tra r_{xt} e gli errori:

Ipotesi 5. $Cov(r_{xt}, \varepsilon_t) = 0$.

Ipotesi 6. Rendimenti a istanti temporali diversi non sono fra loro correlati.

Osservazione 14. È questo un modo compatto di esprimere le N regressioni in un'unico grande modello di regressione. La sua struttura è molto simile a quella solita ma con l'importante differenza che la variabile spiegata è ora un vettore (dobbiamo spiegare N rendimenti!). Automaticamente il vettore di parametri α contiene le N costanti degli N modelli (4.1) di regressione semplice ed ε_t contiene gli N termini d'errore. Le componenti di ε_t non sono fra loro correlate in quanto

¹D'ora in poi in questo esercizio parleremo semplicemente di rendimenti intendendo però rendimenti logaritmici.

$cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$ per $i \neq j$. Tuttavia la loro varianza non è necessariamente uguale, infatti $Var(\varepsilon_{it}) = \sigma_i^2$. Come nella regressione classica gli errori sono *i.i.d.*, quindi *non correlati nel tempo*. Poiché la variabile esplicativa è aleatoria è necessario introdurre l'Ipotesi 6.

Esercizio 6. Single-Index Model

1. Si dia dimensione e contenuto di ogni simbolo della (4.2).
2. Si espliciti la struttura della matrice $\Sigma = Var(\varepsilon_t)$ (Ipotesi 3) in funzione dei σ_i^2 . Che particolarità presenta Σ ? Date infine la dimensione di $Cov(r_{xt}, \varepsilon_t)$.
3. Calcolate $e = E(r_t)$.
4. Calcolate, utilizzando il modello 4.2 e le sue ipotesi, $Var(r_{it})$ e $Cov(r_{it}, r_{jt})$ quando $i \neq j$.
5. Si dia $Var(r_t)$ (la matrice delle varianze e covarianze di r_t , notata V), in funzione di β , σ_x^2 e Σ .

Il file '**esercizio6.xls**' contiene 158×6 prezzi osservati a frequenza *settimanale*. Le sei serie storiche considerate si riferiscono a $N = 5$ azioni svizzere ed all'indice SMI.

6. Per tutte e sei le serie storiche di prezzi calcolate dapprima i rendimenti ed in seguito il rendimento medio. Utilizzando sempre i rendimenti calcolatene la matrice di covarianza, facendo in modo di lasciare quale ultima serie l'indice SMI. Della matrice di covarianza appena calcolata evidenziate in rosso la sottomatrice della covarianza stimata di r_t , denotata \hat{V}_{dati} .

Le 5 equazioni stimabili avranno tutte forma

$$r_{i,t} = \alpha_i + \beta_i r_{x,t} + \varepsilon_{i,t}, \quad t = 1, \dots, 157$$

7. Utilizzando esclusivamente i risultati del punto precedente (dunque senza eseguire la regressione) calcolate per ciascuna equazione le stime di α_i , β_i ed il vettore dei residui stimati.

Suggerimento: Nella parte comune quando abbiamo trattato la regressione stocastica abbiamo visto che i coefficienti stimati di una regressione semplice (costante più una sola variabile esplicativa) sono funzioni delle varianze e covarianze stimate. Adattate questo risultato al caso in esame.

8. Calcolate per ogni equazione la varianza stimata di ε_{it} , notata $\hat{\sigma}_i^2$.
9. Proponete $\hat{\Sigma}$, lo stimatore della matrice Σ (prendete spunto dal punto 2).
10. Utilizzando $\hat{\beta}$, $\hat{\sigma}_x^2$ e $\hat{\Sigma}$ calcolate \hat{V} (cf. punto 5). Confrontate poi \hat{V} con \hat{V}_{dati} . Notate delle differenze? Se sì, come le spiegate?

□

Capitolo 5

Calculating Efficient Portfolios When are No Short-Sale Restrictions

5.1 Notazione e definizioni preliminari

In questo capitolo utilizzeremo la notazione introdotta precedentemente. Ci troviamo al tempo t e dobbiamo costruire un portafoglio la cui struttura verrà mantenuta sino al periodo successivo, ovvero $t + 1$. Poiché il modello è statico, evitiamo di apporre l'indice temporale alle variabili: r_p indicherà quindi il rendimento $r_{p(t+1)}$ del portafoglio, r sarà il vettore dei rendimenti azionari e così via. Assumiamo che

- esistano $N \geq 2$ asset rischiosi il cui vettore dei rendimenti attesi è notato $e := E(r)$,
- il rendimento r_i di un qualsiasi asset non possa essere espresso quale combinazione lineare dei rendimenti dei restanti $N - 1$ asset, il che implica una matrice V delle varianze e covarianze dei rendimenti non singolare,
- ci siano almeno due asset di rendimento atteso diverso,
- sia ammesso vendere allo scoperto (short selling).

5.2 La matematica della frontiera efficiente

Sappiamo che un portafoglio p si trova sulla frontiera se ha varianza minima rispetto a tutti i portafogli con medesimo rendimento. Matematicamente avremo che un portafoglio p è sulla frontiera se e solo se w_p , il vettore degli N pesi (weights) di p , è soluzione del seguente problema di minimizzazione quadratica:

$$\min_{\{w\}} \frac{1}{2} w' V w$$

sotto i vincoli

$$\begin{aligned} w'e &= E[r_p] \\ w'S &= 1 \end{aligned}$$

Osservazione 15. Grazie all'ipotesi che gli N asset non hanno tutti il medesimo rendimento atteso ed alla possibilità di vendere allo scoperto è possibile costruire portafogli aventi un qualsiasi rendimento atteso. Da ciò segue che ci sarà sempre almeno un portafoglio che soddisfa il primo vincolo e quindi, per qualsiasi rendimento atteso desiderato, il problema ha sempre soluzione. Ovviamente se ulteriori vincoli dovessero essere aggiunti ai due precedenti l'esistenza di una soluzione non sarebbe più garantita.

5.2.1 La soluzione

Formando il lagrangiano $\mathcal{L}(w, \lambda, \gamma)$, w_p è soluzione di

$$\min_{\{w, \lambda, \gamma\}} \mathcal{L} = \frac{1}{2} w' V w + \lambda (E[r_p] - w'e) + \gamma (1 - w'S)$$

ovvero

$$\frac{\partial \mathcal{L}}{\partial w} = V w_p - \lambda e - \gamma S = \begin{matrix} 0 \\ (N \times 1) \end{matrix} \quad (5.1)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = E[r_p] - w'_p e = 0 \quad (5.2)$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = 1 - w'_p S = 0 \quad (5.3)$$

Poiché V è definita positiva, le condizioni di primo ordine (5.1), (5.2) e (5.3) sono necessarie e sufficienti per un minimo globale. Risolvendo la (5.1) rispetto a w_p si ottiene

$$w_p = \lambda V^{-1} e + \gamma V^{-1} S \quad (5.4)$$

e premoltiplicando (5.4) per e' ed utilizzando la (5.2):

$$E[r_p] = \lambda e' V^{-1} e + \gamma e' V^{-1} S \quad (5.5)$$

Premoltiplicando sempre la (5.4) per S' ed utilizzando (5.1):

$$1 = \lambda S' V^{-1} e + \gamma S' V^{-1} S \quad (5.6)$$

Grazie a (5.5) e (5.6) abbiamo 2 equazioni in 2 incognite, λ e γ . Risolvendo

$$\lambda = \frac{CE[r_p] - A}{D} \quad (5.7)$$

$$\gamma = \frac{B - AE[r_p]}{D} \quad (5.8)$$

dove

$$\begin{aligned} A &= S'V^{-1}e = e'V^{-1}S \\ B &= e'V^{-1}e \\ C &= S'V^{-1}S \\ D &= BC - A^2 \end{aligned}$$

Poiché l'inversa di una matrice definita positiva è definita positiva, B e C sono entrambi strettamente maggiori di 0. Essendo

$$0 \stackrel{!}{<} (Ae - BS)'V^{-1}(Ae - BS) = \underset{>0}{B}(BC - A^2)$$

segue che

$$(BC - A^2) > 0$$

e quindi $D > 0$. Sostituiamo ora λ e γ in (5.4) così da ottenere

$$w_p = g + hE[r_p] \tag{5.9}$$

dove

$$\begin{aligned} \underset{(N \times 1)}{g} &= \frac{1}{D} [BV^{-1}S - AV^{-1}e] \\ \underset{(N \times 1)}{h} &= \frac{1}{D} [CV^{-1}e - AV^{-1}S] \end{aligned}$$

5.2.2 Proprietà della frontiera

Le equazioni (5.1), (5.2) e (5.3) definiscono delle condizioni necessarie e sufficienti affinché la soluzione w_p sia sulla frontiera (tuttavia non necessariamente sulla parte efficiente). Quindi se w_p si trova sulla frontiera, esso avrà una rappresentazione in termini della (5.9), cioè in termini di g e h . Per contro, se w_p è rappresentabile tramite la (5.9), w_p soddisfa le (5.1), (5.2) e (5.3) e quindi esso è sulla frontiera.

L'interpretazione dei portafogli associati al vettore di pesi g e $g + h$ è la seguente:

- g corrisponde al vettore di pesi del portafoglio frontiera avente 0 "expected return". Infatti

$$w_p = g + h0 = g.$$

- $g + h$ corrisponde al vettore di pesi del portafoglio frontiera con rendimento atteso 1

$$w_p = g + h1 = g + h.$$

Teorema 1. Qualsiasi portafoglio situato sulla frontiera è generato tramite combinazione lineare dei due portafogli di pesi g e $g + h$.

Dimostrazione. Sia r^* un qualsiasi rendimento e w_{q^*} il portafoglio frontiera di rendimento atteso $E[r_{q^*}] = r^*$. La formula (5.9) afferma che tale portafoglio avrà pesi uguali a

$$w_{q^*} = g + hE[r_{q^*}].$$

Consideriamo la combinazione lineare dei portafogli g e $g + h$ con pesi $\{1 - E[r_{q^*}], E[r_{q^*}]\}$:

$$(1 - E[r_{q^*}])g + E[r_{q^*}](g + h) = g + hE[r_{q^*}] = w_{q^*}.$$

Tale combinazione lineare genera esattamente il portafoglio q^* ! Poiché r^* era un qualsiasi rendimento il teorema è dimostrato. \square

Osservazione 16. Notiamo che la precedente argomentazione fa indirettamente uso del fatto che i due portafogli frontiera g e $g + h$ hanno rendimento atteso diverso, 0 e 1. È possibile generalizzare ulteriormente il teorema precedente. È infatti possibile dimostrare che l'intera frontiera può essere generata a partire da due qualsiasi portafogli frontiera distinti¹, non necessariamente g e $g + h$.

Dimostrazione. Prendiamo due portafogli frontiera distinti, notati rispettivamente p_1 e p_2 , di rendimento atteso $E[r_{p_1}]$, $E[r_{p_2}]$ e pesi $w_{p_1} = g + hE[r_{p_1}]$, $w_{p_2} = g + hE[r_{p_2}]$. Sia ora p un qualsiasi portafoglio frontiera di rendimento atteso $E[r_p]$. Poiché $E[r_{p_1}] \neq E[r_{p_2}]$ l'equazione in α

$$E[r_p] = \alpha E[r_{p_1}] + (1 - \alpha)E[r_{p_2}] \quad (5.10)$$

ha quale unica soluzione

$$\alpha^* = \frac{E[r_p] - E[r_{p_2}]}{E[r_{p_1}] - E[r_{p_2}]}.$$

Consideriamo ora la combinazione lineare dei portafogli p_1 e p_2 di pesi $\{\alpha^*, 1 - \alpha^*\}$. Avremo

$$\begin{aligned} \alpha^* w_{p_1} + (1 - \alpha^*) w_{p_2} &= \alpha^* (g + hE[r_{p_1}]) + (1 - \alpha^*) (g + hE[r_{p_2}]) \\ &= g + h(\alpha^* E[r_{p_1}] + (1 - \alpha^*) E[r_{p_2}]) \\ &= g + hE[r_p] = w_p. \end{aligned}$$

Poiché il portafoglio p era un qualsiasi portafoglio frontiera, abbiamo dimostrato che anche con p_1 e p_2 è possibile costruire l'intera frontiera. \square

5.2.3 La covarianza fra portafogli frontiera

La covarianza fra due portafogli frontiera qualsiasi p e q è data da

$$Cov(r_p, r_q) = w_p' V w_q = \frac{C}{D} \left(E[r_p] - \frac{A}{C} \right) \left(E[r_q] - \frac{A}{C} \right) + \frac{1}{C} \quad (5.11)$$

La formula (5.11) ci permette di concludere che la covarianza fra due portafogli efficienti è sempre positiva.

Domanda 8. Spiegate perché la covarianza fra due portafogli efficienti è sempre positiva.

¹E quindi di valore atteso diverso fra loro!

Quando $p = q$ la precedente formula si riduce alla varianza di r_p che potrà essere scritta come

$$\begin{aligned}\sigma_p^2 &= \frac{C}{D} \left(E[r_p] - \frac{A}{C} \right)^2 + \frac{1}{C} \\ &= \frac{1}{D} \left(C(E[r_p])^2 - 2AE[r_p] + \frac{A^2 + D}{C} \right) \\ &= \frac{1}{D} (C(E[r_p])^2 - 2AE[r_p] + B)\end{aligned}$$

che è una parabola nello spazio $(\sigma^2, E[r])$.

Grafico

oppure un'iperbole di equazione

$$\frac{\sigma_p^2}{1/C} - \frac{(E[r_p] - A/C)^2}{D/C^2} = 1 \quad (5.12)$$

nello spazio $(\sigma, E[r])$.

Grafico

Calcoliamo ora il differenziale totale della funzione definita implicitamente dalla (5.12):

$$f(x, y) = 1$$

dove $x = \sigma_p$, $y = E[r_p]$

$$\begin{aligned} \frac{x^2}{1/C} - \frac{\left(y - \frac{A}{C}\right)^2}{D/C^2} &= 1 \\ 2Cx \, dx - \frac{C^2}{D} 2\left(y - \frac{A}{C}\right) dy &= 0 \end{aligned}$$

da cui

$$\frac{dy}{dx} = \frac{Dx}{Cy - A} \Rightarrow \frac{dE[r_p]}{d\sigma_p} = \frac{D\sigma_p}{CE[r_p] - A} \quad (5.13)$$

Sapendo che la retta passa per il punto p di coordinate $(\sigma_p, E[r_p])$ possiamo ora calcolare il coefficiente a della tangente $y = a + bx$ alla frontiera nel punto p :

$$\begin{aligned} a &= E[r_p] - b\sigma_p \stackrel{(5.13)}{=} E[r_p] - \frac{D\sigma_p}{CE[r_p] - A} \sigma_p \\ &= E[r_p] - \frac{D\sigma_p^2}{CE[r_p] - A} \end{aligned} \quad (5.14)$$

Risolvendo rispetto a σ_p^2 nella (5.12) e sostituendo nella (5.14) otteniamo:

$$\begin{aligned} a &= E[r_p] - \frac{D\sigma_p^2}{CE[r_p] - A} \\ &= E[r_p] - \frac{\left(E[r_p] - \frac{A}{C}\right)^2 D}{\frac{D}{C^2} C (CE[r_p] - A)} - \frac{D}{C} \frac{1}{(E[r_p]C - A)} \\ &= E[r_p] - \frac{\left(E[r_p] - \frac{A}{C}\right)^2 D}{C^2 \left(E[r_p] - \frac{A}{C}\right) \frac{D}{C^2}} - \frac{D}{C^2} \frac{1}{\left(E[r_p] - \frac{A}{C}\right)} \\ &= E[r_p] - \left(E[r_p] - \frac{A}{C}\right) - \frac{D}{C^2} \frac{1}{\left(E[r_p] - \frac{A}{C}\right)} \\ &= \frac{A}{C} - \frac{D}{C^2} \frac{1}{\left(E[r_p] - \frac{A}{C}\right)} \end{aligned}$$

La retta tangente alla frontiera nel punto $(\sigma_p, E[r_p])$ avrà dunque equazione

$$E[r] = \frac{A}{C} - \frac{D}{C^2} \frac{1}{\left(E[r_p] - \frac{A}{C}\right)} + \frac{D\sigma_p}{CE[r_p] - A} \sigma. \quad (5.15)$$

5.2.4 Il portafoglio a varianza minima

Il portafoglio a varianza minima, notato mvp , di valore atteso $\frac{A}{C}$ e varianza $\frac{1}{C}$ ha la seguente proprietà:

$$Cov(r_q, r_{mvp}) = Var(r_{mvp}) \quad (5.16)$$

per ogni portafoglio q , non necessariamente sulla frontiera. Vogliamo dimostrare questo risultato.

Esercizio 7. Proprietà del minimum-variance-portfolio (*mvp*)

Sia dunque q un qualsiasi portafoglio. Consideriamo un portafoglio p definito tramite una combinazione lineare di q e *mvp* con pesi a e $(1 - a)$.

1. Si calcoli la varianza di p in funzione del peso a , σ_q^2 , σ_{mvp}^2 e $Cov(r_q, r_{mvp})$.
2. Si imposti il problema di minimizzazione della varianza di p in funzione del parametro a . Si dia la condizione di primo ordine, notata (c1), e si verifichi che un'eventuale soluzione è un punto di minimo.
3. Ovviamente, essendo *mvp* il portafoglio a varianza minima, esso è anche soluzione del precedente problema di minimizzazione e quindi avremo che $p = mvp$. Quanto vale dunque il parametro a ?
4. Inserite il valore di a dedotto al punto precedente nell'equazione (c1) e dimostrate la (5.16).

□

5.2.5 La frontiera efficiente

I portafogli frontiera di rendimento atteso superiore ad $\frac{A}{C}$ (rendimento atteso del *mvp*) sono chiamati portafogli efficienti. Abbiamo visto in precedenza che la combinazione lineare di due portafogli frontiera è ancora un portafoglio frontiera. Questa proprietà è generalizzabile alla combinazione lineare di un numero qualsiasi di portafogli frontiera.

Proprietà 1. Qualsiasi combinazione lineare di portafogli frontiera è un portafoglio frontiera.

Proprietà 2. Combinazioni lineari convesse (cioè $\alpha_i \geq 0$ e $\sum \alpha_i = 1$) di portafogli efficienti sono portafogli efficienti. Da questa proprietà segue che l'insieme dei portafogli efficienti è un insieme convesso.

Esercizio 8. Dimostrazione della Proprietà 1

Siano w_i , $i = 1, 2, \dots, m$, m portafogli frontiera ed α_i , $i = 1, 2, \dots, m$, m numeri reali tali che $\sum \alpha_i = 1$. Si dia in funzione dei rendimenti attesi dei portafogli w_i , notati $E(\tilde{r}_i)$, $i = 1, 2, \dots, m$, il rendimento atteso del portafoglio p definito da $w_p = \sum \alpha_i w_i$ e si dimostri che p è un portafoglio frontiera.

Suggerimento: cercate di dimostrare che w_p è esprimibile nella solita forma $g + hE(\tilde{r}_p)$.

Esercizio 9. Dimostrazione della Proprietà 2

Siano ora w_i , $i = 1, 2, \dots, m$, m portafogli efficienti ed $\alpha_i \geq 0$, $i = 1, 2, \dots, m$, m numeri reali tali che la loro somma $\sum \alpha_i = 1$. Dimostrate che se $w_p = \sum \alpha_i w_i$, allora p è un portafoglio efficiente.

Suggerimento: dovete dimostrare che

1. p è un portafoglio frontiera,
2. $E(\tilde{r}_p) > \frac{A}{C}$.

5.2.6 Portafogli frontiera a 0 covarianza

Una proprietà importante della frontiera è che per qualsiasi portafoglio frontiera p diverso da mvp esiste un portafoglio frontiera, notato $zc(p)$, tale che $Cov(r_p, r_{zc(p)}) = 0$. Utilizzando l'equazione caratterizzante la covarianza fra due portafogli frontiera è possibile determinare il rendimento atteso del portafoglio $zc(p)$ semplicemente eguagliando la $Cov(r_p, r_{zc(p)})$ a 0.

Nel paragrafo 5.2.3 abbiamo calcolato la tangente alla frontiera nel punto $(\sigma_p, E[r_p])$:

$$E[r] = \underbrace{E[r_p] - \frac{D\sigma_p^2}{CE[r_p] - A}}_a + \underbrace{\frac{D\sigma_p}{CE[r_p] - A}}_b \sigma.$$

È possibile dimostrare che il coefficiente a è uguale al rendimento atteso del portafoglio $zc(p)$. Questo implica che il rendimento del portafoglio a covarianza 0 di un portafoglio p può essere trovato graficamente trovando il punto di intersezione fra l'asse delle ordinate e la retta tangente alla frontiera nel punto $(\sigma_p, E[r_p])$.

5.2.7 La security market line

Vogliamo ora caratterizzare la relazione fra il rendimento atteso di un qualsiasi portafoglio frontiera diverso da mvp ed il rendimento di un qualsiasi altro portafoglio (non necessariamente sulla frontiera quindi). Sia dunque $p \neq mvp$ un portafoglio frontiera. Utilizzando la (5.4) otteniamo

$$\begin{aligned} Cov(r_p, r_q) &= w_p' V w_q = \lambda e' V^{-1} V w_q + \gamma S' V^{-1} V w_q \\ &= \lambda E[r_q] + \gamma \end{aligned}$$

Sostituendo la (5.7) e (5.8) a λ e γ :

$$E[r_q] = \frac{AE[r_p] - B}{CE[r_p] - A} + Cov(r_p, r_q) \frac{D}{CE[r_p] - A}$$

e tramite l'equazione (5.12)

$$\begin{aligned} E[r_q] &= \frac{A}{C} - \frac{D/C^2}{E[r_p] - A/C} + \underbrace{\frac{Cov(r_p, r_q)}{\sigma_p^2}}_{\beta_{q,p}} \underbrace{\left[\frac{1}{C} + \frac{(E[r_p] - A/C)^2}{D/C} \right]}_{\text{tramite (5.12) } \sigma_p^2} \frac{D}{CE[r_p] - A} \\ &= E[r_{zc(p)}] + \beta_{q,p} \left(E[r_p] - A/C + \frac{D/C^2}{E[r_p] - A/C} \right) \\ &= E[r_{zc(p)}] + \beta_{q,p} (E[r_p] - E[r_{zc(p)}]) \quad (5.17) \\ &= \beta_{q,p} E[r_p] + (1 - \beta_{q,p}) E[r_{zc(p)}] \quad (5.18) \end{aligned}$$

Il rendimento atteso di un qualsiasi portafoglio q può quindi scriversi come una combinazione lineare dei rendimenti attesi del portafoglio p e del suo "zero covarianza" $zc(p)$ portafoglio con pesi uguali a $\beta_{q,p}$ e rispettivamente $(1 - \beta_{q,p})$.

Notiamo che $zc(zc(p)) = p$ (il portafoglio "zero covarianza" del portafoglio $zc(p)$ è p) per qualsiasi portafoglio frontiera $p \neq mvp$. Questo ci permette di ripetere² l'argomentazione invertendo i ruoli fra p e $zc(p)$, cioè prendendo quale

²Tuttavia deve valere $\sigma_{zc(p)}^2 > 0$.

portafoglio frontiera iniziale $zc(p)$ ed ottenendo per analogia³

$$E[r_q] = \beta_{q,zc(p)} E[r_{zc(p)}] + (1 - \beta_{q,zc(p)}) E[r_p] \quad (5.19)$$

Poiché $E[r_{zc(p)}] \neq E[r_p]$, l'equazione

$$E[r_q] = \alpha E[r_{zc(p)}] + (1 - \alpha) E[r_p]$$

possiede un'unica soluzione che notiamo α^* . Confrontando i pesi di $E[r_{zc(p)}]$ nell'equazione (5.18) e (5.19) otteniamo la seguente identità:

$$(1 - \beta_{q,p}) = \alpha^* = \beta_{q,zc(p)}.$$

Quindi, riprendendo l'equazione (5.18) essa può essere riscritta come

$$E[r_q] = \beta_{q,p} E[r_p] + \beta_{q,zc(p)} E[r_{zc(p)}] \quad (5.20)$$

Le relazioni (5.18), (5.19) e (5.20) sono tutte equivalenti fra loro.

5.2.8 Asset non rischioso

Assumiamo ora che esista un asset non rischioso il cui rendimento è certo. Notiamo con p un portafoglio frontiera degli $N + 1$ asset e con w_p il vettore dei pesi di p negli investimenti a rischio. w_p sarà dunque la soluzione al seguente problema di minimizzazione vincolato:

$$\min \frac{1}{2} w' V w$$

sotto i vincoli

$$w' e + (1 - w' S) r_f = E[r_p] \quad (5.21)$$

Come in precedenza formiamo il lagrangiano \mathcal{L} : w_p sarà soluzione del seguente problema

$$\min_{\{w, \lambda\}} \frac{1}{2} w' V w + \lambda (E[r_p] - w' e - (1 - w' S) r_f)$$

Le condizioni di primo ordine saranno

$$V w_p = \lambda (e - r_f S) \quad (5.22)$$

$$r_f + w_p' (e - r_f S) = E[r_p] \quad (5.23)$$

$$\text{da cui} \quad w_p = \lambda V^{-1} (e - r_f S) \quad (5.24)$$

Premoltiplicando (5.24) per $(e - r_f S)'$ e risolvendo su λ

$$\lambda = \left[\underbrace{(e - r_f S)' V^{-1} (e - r_f S)}_H \right]^{-1} (E[r_p] - r_f)$$

da cui segue che

$$w_p = V^{-1} (e - r_f S) \frac{E[r_p] - r_f}{H} \quad (5.25)$$

³Per un qualsiasi portafoglio frontiera k vale: $E[\tilde{r}_q] = \beta_{q,k} E[\tilde{r}_k] + (1 - \beta_{q,k}) E[\tilde{r}_{zc(k)}]$.

La varianza di r_p :

$$\sigma_p^2 = w_p' V w_p = \frac{(E[r_p] - r_f)^2}{H} \quad (5.26)$$

o equivalentemente

$$\sigma_p = \begin{cases} \frac{E[r_p] - r_f}{\sqrt{H}} & \text{se } E[r_p] \geq r_f \\ -\frac{E[r_p] - r_f}{\sqrt{H}} & \text{se } E[r_p] < r_f \end{cases}$$

Grafico

La frontiera di tutti gli asset è composta da due semirette originate in $(0, r_f)$ con pendenza \sqrt{H} e $-\sqrt{H}$. Tre i casi possibili:

1. $r_f < \frac{A}{C}$

T è il punto di tangenza fra la semiretta $r_f + \sqrt{H}\sigma$ e la frontiera.

- Portafogli situati sul segmento $\overline{r_f T}$ consistono in una combinazione lineare convessa tra r_f ed il portafoglio T .
- Portafogli situati sulla parte $\overline{r_f T}^c$ utilizzano una posizione negativa (short selling) di risk free.
- Posizioni sulla semiretta $r_f - \sqrt{H}\sigma$ sono costituite tramite posizioni negative (short) in T .

2. $r_f > \frac{A}{C}$

Situazione "simmetrica" rispetto alla precedente.

3. $r_f = \frac{A}{C}$

In questo caso

$$\begin{aligned} H &= (e - S r_f)' V^{-1} (e - S r_f) \\ &= e' V^{-1} e - r_f e' V^{-1} S - r_f S' V^{-1} e + r_f^2 S' V^{-1} S \\ &= B - 2 r_f A + r_f^2 C \\ &= \frac{BC - A^2}{C} = \frac{D}{C} > 0 \end{aligned}$$

dove A, B e C sono definiti come in precedenza. Otteniamo dunque che $H = \frac{D}{C}$ e

$$E[r_p] = \frac{A}{C} \pm \sqrt{D/C} \sigma.$$

Ma questi sono esattamente gli asintoti della parabola definita nel piano $(\sigma, E[r])$ già incontrati nei paragrafi precedenti. Abbiamo quindi la seguente situazione:

Grafico

La frontiera di tutti gli asset non è generata da una combinazione tra il riskless asset ed un portafoglio di asset rischiosi. Infatti, sostituendo nella (5.25) $r_f = \frac{A}{C}$ e premoltiplicando w_p per S'

$$\begin{aligned} S'w_p &= S'V^{-1}(e - \frac{A}{C}S) \frac{E[r_p] - r_f}{H} \\ &= (A - \frac{A}{C}C) \frac{E[r_p] - r_f}{H} \\ &= 0 \end{aligned}$$

Quindi un qualsiasi portafoglio sulla frontiera di tutti gli asset è costruito investendo tutto nel risk free asset e tenendo un cosiddetto "arbitrage portfolio": un portafoglio i cui pesi sommano a zero.

5.2.9 La Security market line rivista

Sia q un qualsiasi portafoglio e w_q i rispettivi pesi delle componenti a rischio. Sia p un portafoglio frontiera. Assumiamo che $E[r_p] \neq r_f$. Allora⁴

$$Cov(r_q, r_p) = w_q' V w_p = \frac{(E[r_q] - r_f)(E[r_p] - r_f)}{H}$$

⁴In quanto per la 5.25 $w_p = V^{-1}(e - r_f S) \frac{E[r_p] - r_f}{H}$ e quindi

$$Cov(r_q, r_p) = w_q' V V^{-1}(e - r_f S) \frac{E[r_p] - r_f}{H}.$$

Terminate utilizzando la 5.21.

Utilizzando ora la (5.26), cioè $\sigma_p^2 = \frac{(E[r_p] - r_f)^2}{H}$ otteniamo

$$\begin{aligned} Cov(r_q, r_p) &= \frac{(E[r_q] - r_f)(E[r_p] - r_f)}{H} \quad | : \sigma_p^2 \\ \frac{Cov(r_q, r_p)}{\sigma_p^2} &= \frac{(E[r_q] - r_f)}{(E[r_p] - r_f)} \end{aligned}$$

da cui è immediato ottenere

$$E[r_q] = r_f + \beta_{qp} (E[r_p] - r_f) \quad (5.27)$$

Questa relazione è indipendente dal segno di $\frac{A}{C} - r_f$. Inoltre vale la seguente relazione

$$r_q = r_f + \beta_{qp} (r_p - r_f) + \tilde{\epsilon}_{qp}$$

dove $\tilde{\epsilon}_{qp} := r_q - r_f - \beta_{qp} (r_p - r_f)$. È immediato verificare che per qualsiasi portafoglio frontiera p diverso r_f vale che $Cov(\tilde{\epsilon}_{qp}, r_p) = E(\tilde{\epsilon}_{qp}) = 0$.

5.2.10 CAPM

Vogliamo derivare le conclusioni del Capital Asset Pricing Model che postula una relazione d'equilibrio di tipo lineare fra i rendimenti attesi di mercato. Definiamo il portafoglio di mercato, ovvero l'insieme degli asset trattati in un'economia. Sia W_i il patrimonio dell' i -esimo individuo, $i = 1, \dots, I$ mentre w_{ij} la percentuale del suo patrimonio investita nel j -esimo asset, $j = 1, \dots, N$. La sostanza totale nell'economia è dunque

$$W := \sum_{i=1}^I W_i$$

Definiamo ora w_j^m come la percentuale del valore dell'asset j sul totale della sostanza W . Esemplicando, supponiamo che gli unici assets siano quelli trattati nell'indice Swiss Market Index⁵. W sarà quindi la capitalizzazione totale dell'indice mentre w_{ubs}^m sarà il valore totale delle azioni *ubs* divise per la capitalizzazione dell'indice. È chiaro che $\sum_j w_j^m = 1$. I pesi w_j^m sono chiamati i pesi del portafoglio di mercato. Affinché il mercato sia in equilibrio dovrà essere

$$\underbrace{\sum_{i=1}^I W_i w_{ij}}_{\text{domanda aggregata } j\text{-esimo asset}} = \underbrace{w_j^m W}_{\text{capitaliz. } j\text{-esimo asset}} \quad (5.28)$$

Dividendo entrambi i lati per W otteniamo

$$\sum_{i=1}^I \underbrace{\frac{W_i}{W}}_{c_i = \% \text{ sostanza } i\text{-esimo individuo}} w_{ij} = w_j^m \quad (5.29)$$

Notiamo che $c_i = \frac{W_i}{W} \geq 0$ e $\sum_{i=1}^I c_i = 1$. Inoltre, poiché questa relazione vale per tutti gli asset j , essa può essere riformulata in termini vettoriali. Definendo

⁵Si veda a tal proposito il file `historical_basket_nssmi_20090415.csv` e la discussione in classe sul suo contenuto.

w_i e w^m rispettivamente come il vettore dei pesi del portafoglio dell' i -esimo individuo ed il vettore dei pesi del portafoglio di mercato, la condizione d'equilibrio 5.29 può essere riscritta nel modo seguente:

$$\sum_{i=1}^I \underset{(1 \times 1)}{c_i} \underset{(N \times 1)}{w_i} = \underset{(N \times 1)}{w_m} . \quad (5.30)$$

Quest'ultima equazione esprime la condizione d'equilibrio fra domanda ed offerta di titoli ed afferma che, in equilibrio, il portafoglio di mercato (lato destro della 5.30) è una combinazione lineare convessa dei portafogli dei singoli individui. Ora, se ogni individuo ha un portafoglio efficiente, per le proprietà viste negli esercizi 8 e 9 anche il portafoglio di mercato sarà sulla frontiera nonché efficiente. In particolare, esso sarà il portafoglio di tangenza T visto nel paragrafo precedente.

Grafico

Capitolo 6

The Lognormal Distribution

6.1 Introduzione: il cammino dei prezzi azionari

In questo capitolo ci proponiamo di modellare la dinamica nel tempo del prezzo di un'azione. È evidente che il prezzo futuro di un'azione (fra un giorno, un mese, un anno, o anche solo fra un minuto) è aleatorio, esso non è determinabile con assoluta certezza. Il modello quindi non potrà essere deterministico.

Quali sono le proprietà desiderabili per un modello aleatorio del prezzo di un'azione? Se utilizzassimo un modello di probabilità, quali caratteristiche dovrebbe avere la distribuzione del prezzo al tempo t ? Inoltre, come evolve il prezzo nel tempo e quindi come cambia la sua distribuzione?

Analizzando i prezzi osservabili sui vari mercati finanziari possiamo identificare alcune caratteristiche che li contraddistinguono e che il modello probabilistico dovrebbe riuscire a descrivere.

1. I prezzi cambiano con continuità: tanto più piccolo è l'arco di tempo fra un'osservazione e l'altra e tanto minore sarà la variazione nel prezzo dell'azione.
2. Il prezzo non è mai negativo (può essere 0 in caso di fallimento della ditta ma in tal caso non è più osservato).
3. Il rendimento atteso di un investimento azionario tende a crescere tanto più il titolo è tenuto in portafoglio.
4. L'incertezza associata al rendimento di un investimento azionario tende a crescere all'aumentare della durata dell'investimento stesso. La varianza del prezzo aumenta nel tempo!

Possiamo disegnare il grafico a frequenza giornaliera dei prezzi di ADECCO. Questo grafico è anche chiamato “the stock price path”. Esso rappresenta il cammino effettivamente realizzato fra gli infiniti cammini che, ex-ante, avrebbero potuto realizzarsi.

È bene osservare quanto segue: in realtà quando costruiamo il grafico dei prezzi di una qualsiasi azione noi stiamo rappresentando le realizzazioni di n variabili aleatorie, dove n rappresenta il numero di giorni osservati. Infatti, ad ogni data t noi osserviamo la realizzazione di P_t , la cui distribuzione non

necessariamente sarà uguale alla distribuzione di P_{t-1} o P_{t+1} . A dipendenza del modello probabilistico utilizzato per descrivere la dinamica temporale dei prezzi le distribuzioni di P_t e P_{t+1} potranno variare. Inoltre, talvolta è necessario calcolare la distribuzione congiunta di due o più prezzi a date differenti, come ad esempio P_t e P_{t+20} .

Supponiamo di avere a disposizione un modello probabilistico ed essere in grado di simulare i cammini del processo dei prezzi $\{P_t\}_{t=1}^n$. Potremmo allora, quale esercizio, simulare $M = 1000$ cammini $\{P_t^s\}_{t=1}^n$, $s = 1, \dots, M$ indipendenti l'uno dall'altro ed in seguito, per una qualsiasi data \tilde{t} , stimare valore atteso e varianza di $P_{\tilde{t}}$ utilizzando le M realizzazioni simulate disponibili per quella data. Dalla teoria asintotica sappiamo che al crescere di M la media $\frac{1}{M} \sum_{s=1}^M P_{\tilde{t}}^s$ convergerà ad $E(P_{\tilde{t}})$ e la varianza stimata convergerà alla varianza $Var(P_{\tilde{t}})$.

In realtà l'economia genera un solo ed unico cammino, cioè $M = 1$. La vita non permette di tornare indietro nel tempo e ripetere l'esperimento. Per tale motivo non è possibile stimare il valore atteso del prezzo o la volatilità del rendimento di una qualsiasi data \tilde{t} facendo una media di osservazioni: alla data \tilde{t} abbiamo un'unica osservazione! Sarà necessario appoggiarsi alle proprietà del modello di probabilità che noi supponiamo generi le osservazioni. In un primo passo, utilizzando i dati a disposizione, si stimeranno i parametri del modello dai quali, in una seconda fase, si dedurranno le informazioni cercate.

6.2 Flashback: Funzione di ripartizione, di densità, quantile

A una variabile aleatoria X (continua o discreta) è sempre associata una distribuzione (o legge) di probabilità P , ovvero una funzione che assegna la probabilità che X assuma singoli valori (nel caso discreto) oppure che cada in singoli intervalli di valori (nel caso continuo).

Esempio 10. La legge di probabilità di Poisson si indica con $P(\lambda)$ e per qualsiasi evento $A \subset \{0, 1, 2, \dots\}$ è data dall'espressione:

$$P_X(A) = \sum_{k \in A} \frac{\exp(-\lambda) \lambda^k}{k!} .$$

□

Tramite la legge di probabilità P possiamo definire la cosiddetta funzione di ripartizione, notata F_X .

Definizione 6. La funzione di ripartizione della variabile aleatoria X è la funzione

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

tale che per ogni $x \in \mathbb{R}$

$$F_X(x) = P(X \in (-\infty, x]) .$$

Osservazione 17. Quando è chiaro a quale variabile aleatoria ci si riferisce si può semplicemente scrivere F anziché F_X .

Definizione 7. Una funzione di ripartizione F è detta assolutamente continua se esiste una funzione non negativa f tale che

$$F(x) = \int_{-\infty}^x f(y)dy .$$

La funzione f è chiamata funzione di densità di F .

Esempio 11. Sia $X \sim N(\mu, \sigma^2)$. La funzione di densità è uguale a

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) .$$

□

Variabili aleatorie aventi una funzione di ripartizione assolutamente continua presentano alcuni vantaggi pratici per quanto riguarda il calcolo delle probabilità. Infatti, dato un qualsiasi intervallo $A = [a, b]$

$$P(X \in A) = \int_a^b f(y)dy .$$

Notate che non sempre è possibile ottenere una soluzione in forma analitica dell'integrale precedente. In taluni casi, come ad esempio con la distribuzione normale, il valore dell'integrale deve essere approssimato tramite procedure numeriche.

Definizione 8. Per semplicità consideriamo variabili aleatorie le cui funzioni di ripartizione F sono assolutamente continue e *strettamente* crescenti¹. Per tali V.A. la funzione quantile è definita per $\alpha \in (0, 1)$ come

$$q_\alpha = F^{-1}(\alpha) .$$

Data la variabile aleatoria X varrà quindi che q_α soddisfa la seguente identità:

$$P(X \leq q_\alpha) = \alpha$$

Esempio 12. La funzione di ripartizione di una variabile aleatoria X distribuita secondo la legge esponenziale è uguale a

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-\lambda x) & x \geq 0 \end{cases}$$

dove $\lambda > 0$ è un parametro detto intensità. La funzione quantile è uguale a

$$q_\alpha = F^{-1}(\alpha) = -\frac{\ln(1-\alpha)}{\lambda} .$$

□

¹Se F possiede dei punti di discontinuità o non è strettamente crescente occorre modificare la definizione di quantile nel modo seguente:

$$q_\alpha = \inf \{x \in \mathbb{R} : \alpha \leq F(x)\} .$$

Inserire Grafico!

Teorema 2. Densità di funzioni di V.A.. Sia X una variabile aleatoria assolutamente continua con funzione di densità f_X . Definiamo la nuova variabile aleatoria

$$Y = g(X)$$

dove g è una funzione biiettiva, derivabile e con funzione inversa g^{-1} . Allora Y è variabile aleatoria assolutamente continua con funzione di densità

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|.$$

6.3 La distribuzione lognormale

Definizione: Una variabile casuale X ha distribuzione lognormale, con parametri a e b^2 , se $\ln(X)$ ha distribuzione normale con media a e varianza b^2 . Equivalentemente $X = \exp(Y)$ dove $Y \sim N(a, b^2)$.

- Funzione di densità:

La distribuzione lognormale ha densità uguale a

$$f(x) = \exp\left\{-\frac{(\ln(x) - a)^2}{2b^2}\right\} / (x\sqrt{2\pi}b) \quad \text{per } x > 0.$$

- Funzione di ripartizione:

E' possibile dimostrare (vedi serie esercizi) che la sua funzione di ripartizione è data da

$$F(x) = G\{[-a + \ln(x)]/b\} \quad \text{per } x > 0,$$

dove G rappresenta la funzione di ripartizione della distribuzione normale standard, cioè $N(0, 1)$.

- Funzione quantile:

La funzione quantile di una variabile aleatoria lognormale è data da

$$F^{-1}(\alpha) = \exp[a + bG^{-1}(\alpha)] \quad \text{per } 0 < \alpha < 1.$$

- Valore atteso e varianza in funzione dei parametri a e b^2

– Il valore atteso di una variabile aleatoria lognormale è uguale a

$$E(X) = \exp(a + b^2/2).$$

– La varianza è uguale a

$$V(X) = \exp[2(a + b^2)] - \exp(2a + b^2).$$

Esercizio 10. Distribuzione lognormale. Utilizzando il Teorema 2:

1. Derivate la funzione di densità di $Y = \exp(X)$ dove $X \sim N(\mu, \sigma^2)$.

2. Mostrate che la funzione di ripartizione di Y può essere scritta come

$$F(y) = G\{[-\mu + \ln(y)]/\sigma\} \text{ per } y > 0,$$

dove G rappresenta la funzione di ripartizione della distribuzione normale *standard*.

Suggerimento: Sappiamo che $F(y) = P(Y \leq y) = P(\exp(X) \leq y) \dots$

3. Ricavate la funzione quantile di Y

$$q_\alpha = F^{-1}(\alpha) = \exp[\mu + \sigma G^{-1}(\alpha)] \text{ per } 0 < \alpha < 1.$$

□

6.4 Simulazione di variabili aleatorie

La funzione *rand()* di MS Excel permette di generare un numero aleatorio estratto da una distribuzione uniforme $[0, 1]$. Per la precisione, leggendo la pagina d'aiuto della funzione stessa si scopre che i numeri generati saranno compresi nell'intervallo $[0, 1)$. Per generare N variabili uniformi indipendenti sarà dunque sufficiente applicare la funzione *rand()* ad N celle.

Una seconda possibilità per generare numeri aleatori in Excel è quella di utilizzare il comando del menu "**Tools | Data Analysis...**" e poi "**Random Number Generation**". Questo comando permette di simulare matrici di variabili aleatorie campionando non solo dalla distribuzione uniforme $[0, 1)$ ma da tutta una serie di distribuzioni discrete e non. Di particolare interesse è la casella "**Random seed**". Il valore in essa contenuto serve ad inizializzare il generatore di numeri aleatori (si tratta infatti una funzione deterministica ...). In pratica, due simulazioni aventi il medesimo Random seed saranno identiche. Se la casella Random seed viene lasciata vuota si otterranno simulazioni diverse.

6.4.0.1 Una procedura generale

E' possibile generare numeri aleatori campionati dalla distribuzione desiderata utilizzando la funzione quantile corrispondente abbinata ad un generatore $U[0, 1]$. L'idea, molto semplice, è la seguente:

- Dapprima simuliamo un'osservazione da $U[0, 1]$: $X \sim U[0, 1]$
- In un secondo tempo calcoliamo $Y = F^{-1}(X)$, dove F^{-1} è la funzione quantile della distribuzione desiderata.

La variabile aleatoria $Y = F^{-1}(X)$ ha dunque quale funzione di distribuzione F . Per dimostrare questo basta notare che

$$P(Y \leq c) \underset{(1)}{=} P(F^{-1}(X) \leq c) \underset{(2)}{=} P(X \leq F(c)) \underset{(3)}{=} F(c)$$

Il primo uguale si ottiene inserendo la definizione di Y , il secondo applicando F ad entrambi i lati della disuguaglianza e l'ultimo è una semplice conseguenza del fatto che X è una variabile aleatoria uniforme $[0, 1]$.

6.5 The Geometric random walk with drift

Indichiamo come d'abitudine il prezzo al tempo t di un'azione con P_t (poiché consideriamo una sola azione evitiamo l'uso dell'indice i). Sia invece Δt un intervallo di tempo, ad esempio un giorno, espresso nell'unità di misura di t . Supponiamo che il prezzo P_t segua un processo del tipo

$$P_{t+\Delta t} = P_t \exp(\mu \Delta t + \sigma \sqrt{\Delta t} \varepsilon_{t+\Delta t}), \quad \varepsilon_t \sim i.i.d. N(0, 1), \quad \sigma > 0. \quad (6.1)$$

I parametri μ e σ sono chiamati rispettivamente termine di *drift* e termine di *diffusione* (o di volatilità). Come osservato in Benninga, se ponessimo il termine di diffusione uguale a 0 il modello si ridurrebbe ad un'equazione deterministica:

$$P_{t+\Delta t} = P_t \exp(\mu \Delta t),$$

il prezzo dell'azione crescerebbe in modo esponenziale ad un tasso composto continuamente μ . Il rendimento logaritmico fra t e $t + \Delta t$ è appunto pari a $\ln(P_{t+\Delta t}/P_t) = \mu \Delta t$. Quando $\sigma > 0$ il rendimento è aleatorio: $P_{t+\Delta t}/P_t$ ha distribuzione lognormale e la sua varianza dipende da Δt e da σ .

Esercizio 11. Dimostrate che se i prezzi sono generati dall'equazione 6.1, condizionatamente a P_t il prezzo $P_{t+\Delta t}$ è distribuito secondo la legge lognormale.

L'intervallo temporale Δt può essere interpretato come la distanza di tempo fra un'osservazione e l'altra (intervallo di campionamento) e (6.1) è l'equazione che descrive la dinamica dei prezzi osservati a tali istanti temporali. Gli indici temporali dei prezzi osservati saranno nell'insieme $(\dots, t - \Delta t, t, t + \Delta t, t + 2\Delta t, \dots)$. E' questa una notazione un po' insolita in quanto generalmente $\Delta t = 1$. In realtà si tratta solo di un modo diverso e più generale di indicizzare le osservazioni.

6.5.1 Caratteristiche del modello

Il modello (6.1) presenta alcune interessanti proprietà che vogliamo analizzare. Innanzi tutto osserviamo che, applicando il logaritmo naturale ad entrambi i lati della (6.1), il modello può essere scritto come

$$\ln(P_{t+\Delta t}) = \ln(P_t) + \mu \Delta t + \sigma \sqrt{\Delta t} \varepsilon_{t+\Delta t}$$

e ridefinendo appropriatamente i termini otteniamo

$$X_{t+\Delta t} = X_t + \tilde{\mu} + \tilde{\varepsilon}_{t+\Delta t}. \quad (6.2)$$

Si tratta di un caso particolare del famoso modello $AR(1)$ già incontrato nella parte comune di questo corso. In questo caso il coefficiente della variabile ritardata è uguale a 1. In letteratura questo modello prende il nome di *random walk con drift*. Il termine random walk (cammino aleatorio) deriva dal fatto che la dinamica di questo modello è puramente casuale. La costante $\tilde{\mu}$, detta appunto drift, genera in X una tendenza a crescere (quando $\tilde{\mu} > 0$ ovviamente).

Utilizzando la (6.2), sviluppiamo in modo ricorsivo

$$\begin{aligned} X_{t+\Delta t} &= X_t + \tilde{\mu} + \tilde{\varepsilon}_{t+\Delta t}, \\ X_{t+2\Delta t} &= X_{t+\Delta t} + \tilde{\mu} + \tilde{\varepsilon}_{t+2\Delta t} = X_t + 2\tilde{\mu} + \tilde{\varepsilon}_{t+2\Delta t} + \tilde{\varepsilon}_{t+\Delta t}, \\ X_{t+3\Delta t} &= X_{t+2\Delta t} + \tilde{\mu} + \tilde{\varepsilon}_{t+3\Delta t} = X_t + 3\tilde{\mu} + \tilde{\varepsilon}_{t+3\Delta t} + \tilde{\varepsilon}_{t+2\Delta t} + \tilde{\varepsilon}_{t+\Delta t}, \end{aligned}$$

e in questo modo dopo n passi avremo che

$$X_{t+n\Delta t} = X_t + n\tilde{\mu} + \sum_{i=1}^n \tilde{\varepsilon}_{t+i\Delta t} \quad (6.3)$$

$X_{t+n\Delta t}$ è dunque composto da tre termini:

- X_t : il valore del processo al tempo t che sarà aleatorio se non osservato o deterministico (conosciuto) se osservato
- $n\tilde{\mu}$: la costante deterministica $n\tilde{\mu}$
- $\sum_{i=1}^n \tilde{\varepsilon}_{t+i\Delta t}$: un ultimo termine aleatorio composto dalla somma degli errori degli ultimi n periodi.

Supponiamo che X_t sia conosciuto e quindi deterministico. Sarà allora semplice calcolare il valore atteso di $X_{t+n\Delta t}$ e vedere che

- esso non è costante nel tempo ma aumenta linearmente in n ,
- la varianza di $X_{t+n\Delta t}$ non è costante nel tempo. Essa è uguale a $n(\sigma^2\Delta t)$ e in maniera simile al valore atteso aumenta linearmente in n .

Se invece X_t fosse anch'esso aleatorio potremmo calcolare media e varianza di

$$\underbrace{X_{t+n\Delta t} - X_t}_{\ln(P_{t+n\Delta t}/P_t)} = n\tilde{\mu} + \sum_{i=1}^n \tilde{\varepsilon}_{t+i\Delta t}$$

Il rendimento logaritmico quale funzione di n possiede quindi le medesime proprietà di $X_{t+n\Delta t}$ quando X_t è osservato. Infine, il termine "geometric" random walk, deriva dal fatto che sono i logaritmi del prezzo a seguire un random walk e non il prezzo stesso (come abbiamo visto ne deriva una crescita esponenziale).

Tornando al modello iniziale è immediato vedere che il modello presuppone rendimenti logaritmici normali:

$$r_{t,t+\Delta t} = \ln(P_{t+\Delta t}/P_t) = \mu\Delta t + \sigma\sqrt{\Delta t}\varepsilon_{t+\Delta t},$$

e quindi

$$r_{t,t+\Delta t} \sim N(\mu\Delta t, \sigma^2\Delta t).$$

Considerando invece il rendimento su un orizzonte di n periodi, sulla base dei calcoli svolti in precedenza abbiamo che

$$r_{t,t+n\Delta t} = \ln(P_{t+n\Delta t}/P_t) = n\mu\Delta t + \sigma\sqrt{\Delta t} \sum_{i=1}^n \varepsilon_{t+i\Delta t} \quad (6.4)$$

e quindi

$$r_{t,t+n\Delta t} \sim N(n\mu\Delta t, n\sigma^2\Delta t).$$

Quando $n\Delta t = 1$ dalla (6.4) segue semplicemente che $r_{t,t+1} \sim N(\mu, \sigma^2)$. μ e σ sono quindi il valore atteso e la volatilità del rendimento logaritmico su un periodo pari all'unità di misura prescelta, l'anno ad esempio.

Esercizio 12. Rispetto alle proprietà d'aggregazione temporale dei rendimenti logaritmici viste nei precedenti capitoli, quali sono le nuove conclusioni a cui il modello “geometric random walk” permette di giungere?

Esercizio 13. Benninga, pagina 288

Supponiamo che t indichi il tempo in anni. Noi desideriamo simulare a frequenza giornaliera. Poiché in un anno ci sono (circa) 250 giorni lavorativi, dobbiamo scegliere $\Delta t = 1/250$. Supponiamo che valore atteso e volatilità annui del rendimento logaritmico siano rispettivamente uguali al 10% e 20% e che il prezzo iniziale dell'azione sia di \$25. Simulate un anno di osservazioni giornaliere.

Inserire grafico

Capitolo 7

Value at Risk - VaR

7.1 Introduzione

Nell'ambito della gestione del rischio finanziario possiamo identificare i seguenti tipi di rischi

1. Market risk

Rischio causato da movimenti inattesi ed avversi di prezzi, tassi d'interesse e tassi di cambio.

2. Credit risk

Rischio generato da un possibile fallimento della controparte relativamente ad un contratto finanziario (per esempio obbligazioni Parmalat).

3. Liquidity risk

Rischio dovuto alla poca liquidità dell'asset ed alla conseguente impossibilità d'acquisto/vendita sul mercato in caso di necessità.

4. Operational risk

Rischio causato da errori nell'esecuzioni di ordini d'acquisto/vendita, frodi, ecc.

Il Value at Risk (VaR) è attualmente la più utilizzata misura di rischio di mercato. Il VaR di un portafoglio di assets (non necessariamente di sole azioni) è una stima dell' $(1 - \alpha)$ -quantile della distribuzione della perdita di valore del portafoglio (in termini assoluti o percentuali) su di un orizzonte temporale di T periodi.

Nel calcolo del VaR abbiamo quindi due "gradi di libertà", ovvero possiamo decidere il livello di significatività α e l'orizzonte temporale T per cui stimare la distribuzione del valore (rendimento) del portafoglio ed in seguito l' $(1 - \alpha)$ -quantile. La scelta di α e di T dipendono dall'uso che si desidera fare del VaR. Ad esempio, il comitato¹ di Basilea II situato presso la Bank for

¹"The Committee does not possess any formal supranational supervisory authority, and its conclusions do not, and were never intended to, have legal force. Rather, it formulates broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements ...".

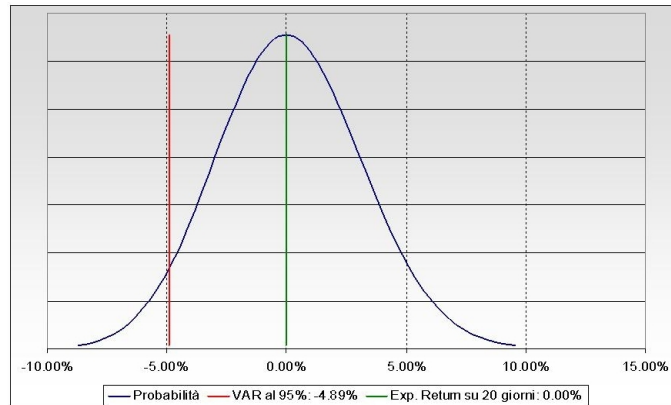


Figura 7.1:

international settlements (<http://www.bis.org/bcbs/aboutbcbs.htm>) suggerisce $T = 5 - 10$ giorni ed $\alpha = 99\%$ al fine di definire i margini di capitale necessari alle banche quale garanzia sui rischi derivanti dalle loro attività di trading. Per un gestore patrimoniale o una cassa pensione invece, T potrebbe essere pari ad un anno ed $\alpha = 95\%$.

Ci sono almeno tre modi per stimare il VaR:

1. Assumere un modello parametrico dei rendimenti, stimarne i parametri ed infine calcolare il VaR.
2. Utilizzare la simulazione di Monte Carlo.
3. Utilizzare tecniche non parametriche.

Abbiamo visto il modello lognormale dei rendimenti azionari, ed è questo il modello che vogliamo ora analizzare in relazione al VaR.

7.2 VaR con rendimenti lognormali

Partendo dalla definizione di una variabile aleatoria lognormale X è evidente che per caratterizzare la sua distribuzione è necessario specificare unicamente valore atteso μ e varianza σ^2 di $\ln(X)$ (che appunto sappiamo essere normale). Di conseguenza, poiché μ e σ^2 determinano in maniera univoca e completa la distribuzione di X essi ne determineranno anche il rischio, e di conseguenza il VaR.

Supponiamo quindi che l'azione S_T abbia una distribuzione lognormale con parametri annuali ($T = 1$) pari a (μ, σ^2) . Questo significa che, indicando con S_0 il valore iniziale², $\ln(S_T) = \ln(S_0) + T\mu + \sigma\sqrt{T}U$, con $U \sim N(0, 1)$. $\ln(S_T)$ ha quindi valore atteso $\ln(S_0) + T\mu$ e varianza $\sigma^2 T$. Utilizzando la formula per il calcolo del valore atteso di una v.a. lognormale otteniamo che

$$E(S_T) = \exp\left(\ln(S_0) + T\mu + \frac{\sigma^2 T}{2}\right) = S_0 \exp\left(\left(\mu + \frac{\sigma^2}{2}\right)T\right).$$

²Utilizzate la notazione dell'equazione (6.4) per dedurre che $T = n\Delta t$.

Notate che $E(S_T) \neq S_0 \exp(\mu T) = \exp(E[\ln(S_T)])$!

Per quanto riguarda il calcolo del VaR di quest'azione con un orizzonte di T anni ad un livello di confidenza del 99%:

$$q_{1\%} = \exp\left(\ln(S_0) + T\mu + \sigma\sqrt{T}G^{-1}(1\%)\right) = S_0 \exp(T\mu + \sigma\sqrt{T}G^{-1}(1\%)).$$

Per un'azione $S_0 = 80$ con $\mu = 10\%$, $\sigma = 30\%$ ed un orizzonte temporale pari a $T = 1$ avremo un $q_{1\%} = 44$ e quindi un $VaR(T; 99\%) = -(44 - 80) = 36$. In percentuale $VaR(T; 99\%) = 45\%$. Notate che il segno è stato cambiato in quanto siamo interessati alla coda sinistra (negativa) della distribuzione. L'interpretazione è che 1 anno su 100 la *perdita* sarà superiore a 36 fr (superiore al 45%).

Avremmo pure potuto calcolare il quantile $q_{1\%}$ del rendimento e non del valore dell'azione. Infatti, sapendo che

$$\ln(S_T/S_0) = T\mu + \sigma\sqrt{T}U,$$

sempre con $\mu = 10\%$, $\sigma = 30\%$ e $T = 1$:

$$q_{1\%} = T\mu + \sigma\sqrt{T}G^{-1}(1\%) = 10\% + 30\% \cdot (-2.326) = -0.5979.$$

Con una probabilità del 1% il rendimento lognormale sarà inferiore o uguale a -0.5979 . Ora, essendo $S_T = S_0 \exp(r_T)$ otteniamo $S_T = 80 \exp(-0.5979) = 44$, che corrisponde ad un VaR del 45% calcolato come $-100 \frac{44-80}{80}$.

7.3 A Three-Asset Problem

Nel caso che il portafoglio consista in due o più azioni abbiamo visto nell'esercizio 2 della serie 4 che il rendimento logaritmico di un portafoglio, $\ln(V_T/V_0)$, è diverso dalla somma pesata dei rendimenti logaritmici sulle singole posizioni. Tuttavia, sempre nella serie 4, avete dimostrato che

$$(V_T - V_0)/V_0 = \sum_{i=1}^N w_{i,0} \frac{P_{i,T} - P_{i,0}}{P_{i,0}} \quad (7.1)$$

Facendo uso della (7.1) e del fatto che³

$$\frac{P_{i,T} - P_{i,0}}{P_{i,0}} \simeq \ln\left(1 + \frac{P_{i,T} - P_{i,0}}{P_{i,0}}\right) = \ln(P_{i,T}/P_{i,0}) \sim N(T\mu_i, \sigma_i^2 T)$$

approssimiamo la distribuzione multivariata dei rendimenti percentuali $r_{0,T}$ con la distribuzione normale multivariata dei rendimenti logaritmici $\ln(P_T/P_0)$.

7.4 Historical simulation

Nei paragrafi precedenti abbiamo visto come calcolare il VaR utilizzando il modello parametrico lognormale. Tuttavia quando è necessario calcolare il VaR di

³Per x prossimi allo 0 vale infatti che $\ln(1+x) \simeq x$.

portafogli composti da numerosi strumenti finanziari quali ad esempio bonds, derivati, swaps, futures ecc. è necessario rivedere l'approccio in quanto, ad esempio, la relazione fra i rendimenti su un call e rendimenti sul relativo sottostante non è più lineare. Semplificazioni ed ipotesi aggiuntive sono quindi necessarie.

Abbiamo precedentemente visto come simulare variabili aleatorie indipendenti aventi la distribuzione desiderata. Una prima possibilità sarebbe quindi quella di utilizzare il modello parametrico per simulare la distribuzione futura dei prezzi alla data desiderata. In pratica si procederebbe a tappe:

1. Utilizzando le osservazioni a disposizione, stimate i parametri del modello parametrico
2. $i = 1$;
3. Utilizzando il modello parametrico e le stime ottenute simula un *rendimento* all'orizzonte desiderato.
4. Tramite il rendimento simulato calcola il prezzo corrispondente.
5. Calcola il valore totale del portafoglio.
6. Se $i = 1000$ termina, altrimenti $i = i + 1$ e torna la punto 3.

Utilizzando la distribuzione empirica delle 1000 simulazioni calcola il quantile desiderato.

Questa metodologia chiamata **simulazione di Monte Carlo** dipende dalla validità del modello parametrico utilizzato.

La simulazione storica (historical simulation) tende a staccarsi da un modello parametrico ed utilizza esclusivamente la distribuzione empirica dei rendimenti (logaritmici). E' consuetudine utilizzare rendimenti giornalieri per eseguire la seguente variante dell'algoritmo descritto in precedenza. Supponiamo di avere $N + 1$ osservazioni giornaliere di prezzi, tassi d'interesse, tassi di cambio... . Allora:

1. $i = 1$;
2. Estraiete a caso una delle N date per le quale potete calcolare i rendimenti giornalieri.
3. Prendete tutti i rendimenti corrispondenti alla data estratta e moltiplicateli per \sqrt{T} , dove T corrisponde alla lunghezza in giorni dell'orizzonte temporale per il quale desiderate calcolare il *VaR*.
4. Con i rendimenti scalati calcolate i prezzi degli strumenti in portafoglio.
5. Calcola il valore totale del portafoglio.
6. Se $i = 1000$ termina, altrimenti $i = i + 1$ e torna la punto 2.

Utilizzando la distribuzione empirica delle 1000 simulazioni calcolate il quantile desiderato.

Osservazione:

1. Ad ogni ripetizione i è *fondamentale* che tutti i rendimenti siano della medesima data. Questo per mantenere la struttura di correlazione *cross-section* fra i rendimenti stessi. Quanto è spiegato nel capitolo 12.5 di Benninga è sbagliato!
2. Questa metodologia è chiamata col nome di historical simulation.