

# Regressione

Leo Pasquazzi

ORCID iD <https://orcid.org/0000-0002-2467-2667>

leo.pasquazzi@unimib.it

*Dipartimento di Statistica e Metodi Quantitativi  
Università degli Studi di Milano - Bicocca*

2022-02-28

# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
<b>2</b>	<b>Svolgimento di un'analisi di regressione</b>	<b>7</b>
2.1	I dati . . . . .	8
2.2	Fase 1: Scelta di una famiglia di possibili funzioni interpolanti . . . . .	12
2.3	Fase 2: Individuazione della funzione interpolante . . . . .	19
2.4	Fase 3: Valutazione della bontà d'adattamento . . . . .	21
2.5	Fase 4: Interpretazione . . . . .	31
<b>3</b>	<b>La retta ai minimi quadrati</b>	<b>33</b>
3.1	Proprietà fondamentali della media aritmetica, della devianza e della codevianza* . . . . .	33
3.2	Determinazione dei parametri . . . . .	42
3.3	Proprietà dei residui della retta ai minimi quadrati . . . . .	59
3.4	Scomposizione della devianza . . . . .	63
3.5	Indice di determinazione . . . . .	66
3.6	Formula indiretta per la devianza spiegata . . . . .	70
3.7	Coefficiente di correlazione lineare . . . . .	73
3.8	Applicazione in ambito finanziario . . . . .	82
3.8.1	L'alpha di Jensen . . . . .	82
3.8.2	Confronto tra fondi d'investimento . . . . .	89
<b>4</b>	<b>Interpolazione con la famiglia delle funzioni potenza</b>	<b>101</b>
4.1	Proprietà delle funzioni potenza . . . . .	101

4.2	Variazioni logaritmiche e variazioni relative . . . . .	106
4.3	Il parametro di elasticità . . . . .	110
4.4	Determinazione dei parametri di una funzione interpolante potenza . . . . .	111
4.5	Applicazione: la regressione di Cobb e Douglas . . . . .	116
<b>5</b>	<b>Il piano ai minimi quadrati</b>	<b>129</b>
5.1	Determinazione dei parametri del piano ai minimi quadrati . . . . .	130
5.2	Formule alternative per calcolare i coefficienti angolari del piano ai minimi quadrati . . . . .	148
5.3	Le proprietà dei residui del piano ai minimi quadrati	167
5.4	La scomposizione della devianza . . . . .	170
5.5	Indice di determinazione . . . . .	173
5.6	Miglioramento della bontà d'adattamento nel passaggio da una retta ai minimi quadrati ad un piano ai minimi quadrati . . . . .	176
5.7	Formula indiretta per il calcolo della devianza spiegata	182
5.8	Formula per calcolare l'indice di determinazione a partire dai coefficienti di correlazione lineare . . . . .	186
5.9	Il coefficiente di correlazione multiplo . . . . .	188
5.10	Coefficienti di correlazione parziale . . . . .	190
<b>6</b>	<b>Interpolazione con la famiglia delle funzioni di produzione di Cobb-Douglas generalizzata</b>	<b>201</b>
6.1	La famiglia delle funzioni di produzione di Cobb-Douglas generalizzata: definizione e proprietà . . . . .	201

6.2	Determinazione dei parametri di una funzione interpolante di Cobb-Douglas. . . . .	205
6.3	Applicazione . . . . .	208
<b>7</b>	<b>L'iperpiano ai minimi quadrati</b>	<b>216</b>
7.1	Definizione di iperpiano . . . . .	216
7.2	Determinazione dei parametri di un iperpiano ai minimi quadrati . . . . .	217
7.3	Formule alternative per calcolare i coefficienti angolari di un iperpiano ai minimi quadrati . . . . .	223
7.4	Proprietà dei residui di iperpiani ai minimi quadrati	226
7.5	Scomposizione della devianza e indice di determinazione . . . . .	227
7.6	Miglioramento della bontà d'adattamento nel caso di aggiunta di variabili esplicative . . . . .	228
7.7	Formula indiretta per il calcolo della devianza spiegata	229
7.8	Formula per calcolare l'indice di determinazione a partire dai coefficienti di correlazione lineare . . . . .	231
7.9	Coefficiente di correlazione multiplo . . . . .	232
7.10	Coefficienti di correlazione parziale . . . . .	233
7.11	Esempio . . . . .	236

# 1 Introduzione

Nel linguaggio statistico moderno il termine **regressione** viene utilizzato per indicare un tipo di analisi che ha lo scopo di ricavare una **funzione interpolante**. Un'analisi di regressione parte da un certo numero di osservazioni di due o più variabili e produce una funzione interpolante con andamento regolare che **descrive la relazione di fondo** tra le osservazioni della cosiddetta **variabile dipendente** e le osservazioni delle variabili rimanenti che vengono chiamate **indipendenti** o **esplicative**. Analisi di regressione vengono effettuate per

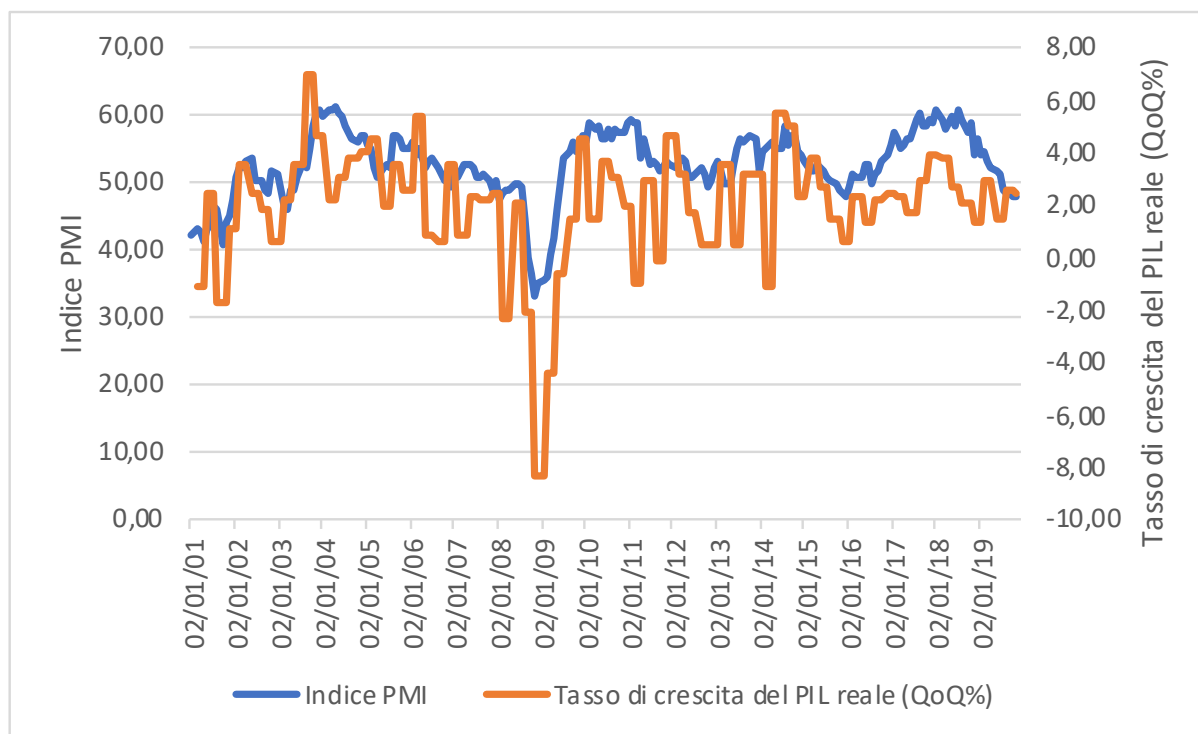
- i) indagare sulla relazione di fondo tra le osservazioni di diverse variabili,
- ii) per indagare sul processo che ha generato le osservazioni delle variabili
- iii) e per ottenere delle **previsioni** per future osservazioni della variabile dipendente.

Il termine "regressione" fu coniato verso la fine del diciannovesimo secolo da Francis Galton (un cugino di Charles Darwin). In un documento pubblicato nel 1885 Galton utilizzò il termine "regressione" per descrivere il fenomeno biologico secondo il quale la progenie di individui eccezionali (per statura) tende ad essere meno eccezionale dei propri genitori e più simile ai loro avi più distanti. Per Galton il termine "regressione" aveva solo questo significato biologico. Estendendo il lavoro di Galton, Udney Yule e Karl Pearson (due statistici) diedero al termine "regressione" il suo significato moderno.

Analisi di regressione vengono effettuate in ogni ambito scientifico. In ambito economico e finanziario, per esempio, si effettuano delle analisi di regressione per analizzare la relazione tra cosid-

detti *leading indicators* (ossia variabili che dovrebbero anticipare cambiamenti nel ciclo economico) e altre variabili che descrivono l'andamento dell'economia, come per esempio il prodotto interno lordo (PIL). Classici esempi di *leading indicators* sono i vari *Purchasing Managers' indexes* (*PMI indexes*) che riflettono le opinioni dei dirigenti di grandi aziende sull'andamento dell'economia. Siccome gli indici PMI vengono pubblicati con cadenza mensile mentre i dati ufficiali sulla crescita del PIL vengono pubblicati solo al termine di ogni trimestre, gli operatori economici e finanziari si affidano spesso agli indici PMI per prendere decisioni in modo più tempestivo. Per illustrare come un indice PMI può anticipare variazioni del tasso di crescita del PIL, il seguente grafico mostra l'andamento storico dell'indice PMI calcolato dall'*Institute for Supply Management* sovrapposto all'andamento storico del tasso di crescita del PIL reale degli Stati Uniti:

Figura 1.1: Dati Factset

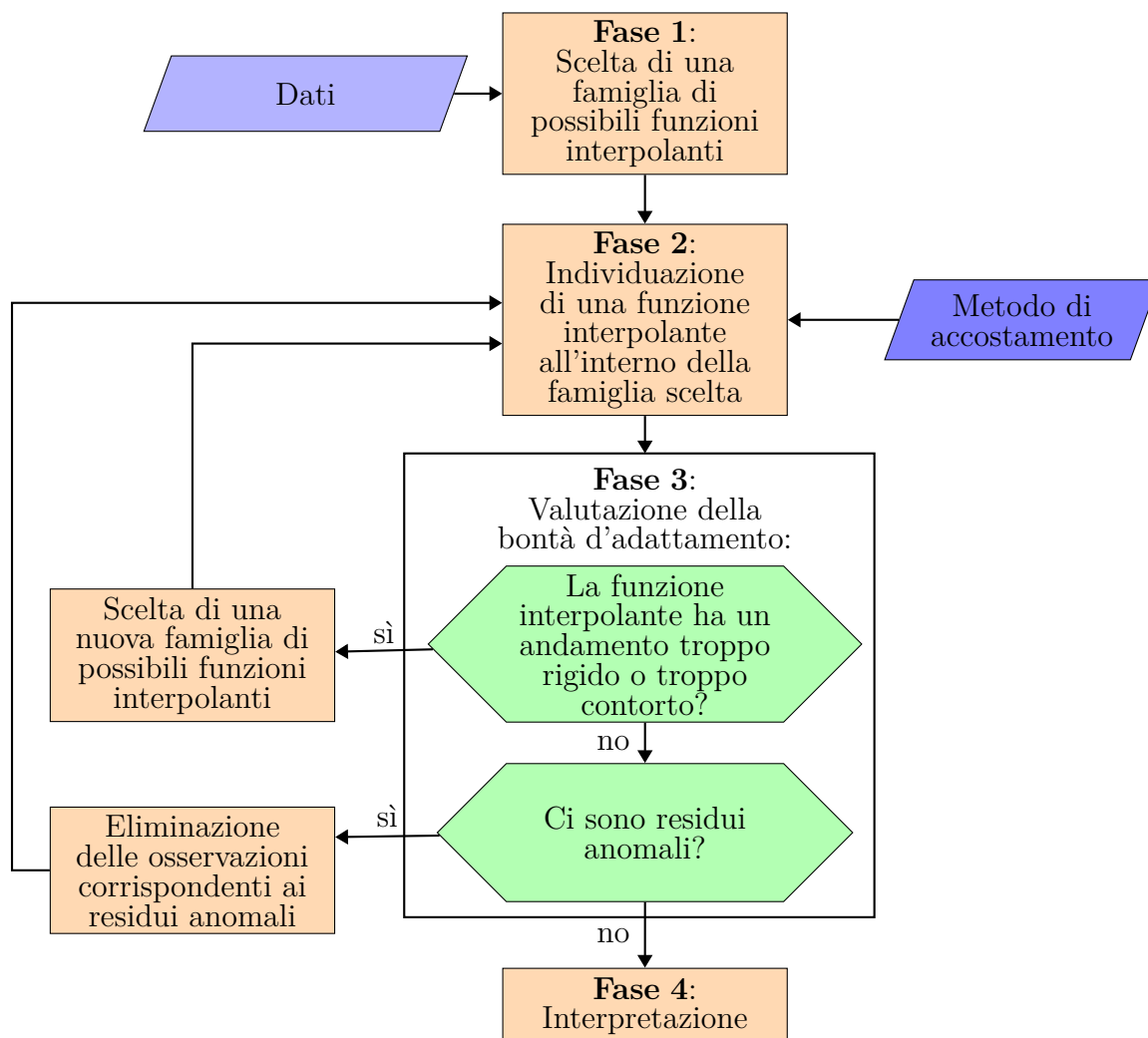


Esaminando con attenzione il grafico notiamo che l'andamento dell'indice PMI non anticipa sempre correttamente l'andamento del tasso di crescita del PIL. Per questo motivo gli operatori economici e finanziari effettuano regolarmente delle analisi di regressione per monitorare la relazione tra le due variabili.

## 2 Svolgimento di un'analisi di regressione

In questa sezione descriveremo lo svolgimento di un'analisi di regressione. Il diagramma di flusso in Figura 2.1 ne anticipa lo schema generale:

Figura 2.1: Svolgimento di un'analisi di regressione



## 2.1 I dati

Come si vede nel diagramma, un'analisi di regressione parte sempre dai *dati*, ovvero dalle osservazioni di un certo numero di variabili. D'ora in poi indicheremo dunque con  $X_1, X_2, \dots, X_k$  le variabili



oggetto di analisi e con

$$\begin{aligned} & (x_{11}, x_{21}, \dots, x_{k1}), \\ & (x_{12}, x_{22}, \dots, x_{k2}), \\ & \vdots \\ & (x_{1n}, x_{2n}, \dots, x_{kn}) \end{aligned}$$

$n$   $k$ -uple di valori osservati per tali variabili. Distingueremo due tipi di dati:

- *Dati longitudinali (logitudinal data)* dove i valori osservati per le variabili si riferiscono a diversi periodi di tempo. In questo caso ciascuna  $k$ -upla  $(x_{1i}, x_{2i}, \dots, x_{ki})$  si riferisce ad un dato periodo di tempo e il numero totale di periodi considerati è uguale a  $n$ . Per esempio, usando le serie storiche dell'indice PMI e del tasso di crescita del PIL in Figura 1.1 si può ottenere un *dataset longitudinale* dedicando ciascuna riga di un foglio di calcolo ad un singolo trimestre e riportando in ciascuna riga i tre valori mensili dell'indice PMI e il valore del tasso di crescita del PIL del trimestre corrispondente. La tabella in Figura 2.2 mostra la parte superiore di un foglio di calcolo che contiene questo *dataset longitudinale*:

Figura 2.2: Foglio "Dati riordinati" della cartella di lavoro "Indice PMI e tasso di crescita del PIL reale" (fonte dati: FactSet).

	A	B	C	D	E	F
1						
2	Anno	Trimestre	Indice PMI del primo mese del trimestre	Indice PMI del secondo mese del trimestre	Indice PMI del terzo mese del trimestre	Tasso di crescita del PIL reale (QoQ%)
3						
4	2001	GEN-MAR	42,3	42,1	43,1	-1,1
5		APR-GIU	42,7	41,3	43,2	2,4
6		LUG-SET	43,5	46,3	46,2	-1,7
7		OTT-DIC	40,8	44,1	45,3	1,1
8	2002	GEN-MAR	47,5	50,7	52,4	3,5
9		APR-GIU	52,4	53,1	53,6	2,4
10		LUG-SET	50,2	50,3	50,5	1,8
11		OTT-DIC	49	48,5	51,6	0,6
12	2003	GEN-MAR	51,3	48,8	46,3	2,2
13		APR-GIU	46,1	49	49	3,5
14		LUG-SET	54	50,2	50,4	7

- *Dati sezionali (cross-sectional data)* dove i valori osservati per le variabili si riferiscono a diverse unità statistiche appartenenti ad una popolazione oggetto di studio. In questo caso ciascuna  $k$ -upla  $(x_{1i}, x_{2i}, \dots, x_{ki})$  si riferisce ad una singola unità statistica e il numero totale di unità statistiche appartenenti alla popolazione è uguale a  $n$ . Un esempio di *dataset sezionale* si ottiene considerando la popolazione delle  $n = 10$  più grandi aziende del settore GICS delle comunicazioni incluse nell'indice di borsa S&P500 e rilevando per ciascuna di tali aziende la capitalizzazione di mercato, il totale dei compensi percepiti dai dirigenti nell'anno 2019, l'utile netto dell'anno 2019 e l'utile netto previsto per l'anno 2020. Figura 2.3 mostra questo *dataset sezionale* all'interno di un foglio di calcolo.

Figura 2.3: Foglio "Dati" della cartella di lavoro "Settore GICS delle comunicazioni" (fonte dati: FactSet).

	A	B	C	D	E	F	G
1	TICKER	NOME SOCIETÀ	SETTORE GICS	CAPITALIZZAZIONE DI MERCATO (in miliardi di USD)	COMPENSO DIRIGENTI ANNO 2019 (in milioni di USD)	UTILE NETTO (in miliardi di USD)	
2						ANNO 2019	ANNO 2020 (stima)
3	GOOG	Alphabet Inc. Class C	Communication Services	1096,723	281,946	34,343	4,175
4	FB	Facebook, Inc. Class A	Communication Services	648,350	112,844	18,485	0,847
5	DIS	Walt Disney Company	Communication Services	251,246	92,795	11,054	0,548
6	VZ	Verizon Communications Inc.	Communication Services	248,451	51,095	19,913	1,189
7	CMCSA	Comcast Corporation Class A	Communication Services	224,504	142,706	13,057	0,872
8	NFLX	Netflix, Inc.	Communication Services	215,702	126,379	1,867	0,820
9	T	AT&T Inc.	Communication Services	201,808	80,583	14,975	1,826
10	TMUS	T-Mobile US, Inc.	Communication Services	160,957	85,494	3,468	1,493
11	CHTR	Charter Communications, Inc. Class A	Communication Services	131,000	21,511	1,668	-0,276
12	ATVI	Activision Blizzard, Inc.	Communication Services	59,363	54,417	1,736	0,233

Di solito le analisi di regressione vengono effettuate su dei *dataset* molto più numerosi (ovvero con  $n$  molto più elevato) di quello rappresentato in Figura 2.3, che in questa dispensa useremo solo a fini didattici.

Di fatto, il *numero minimo di osservazioni* da utilizzare in un'analisi di regressione dipende dagli obiettivi dell'analisi che di solito ricadono in una o entrambe le seguenti categorie:

- raggiungere un determinato livello di precisione nella stima di uno o più parametri che descrivono il processo che ha generato i dati;
- raggiungere un determinato livello di precisione nelle previsioni basate sulla funzione interpolante.

Non sempre è possibile aumentare il numero di osservazioni onde raggiungere un obiettivo prestabilito. Infatti, negli *studi osservazionali* i dati provengono spesso dall'osservazione di un fenomeno che non può essere replicato sotto condizioni abbastanza simili e in

questi casi il numero  $n$  di osservazioni è necessariamente limitato. Anche se il problema della determinazione del numero minimo di osservazioni è molto rilevante, in questa dispensa non lo tratteremo (di solito viene trattato testi più avanzati).

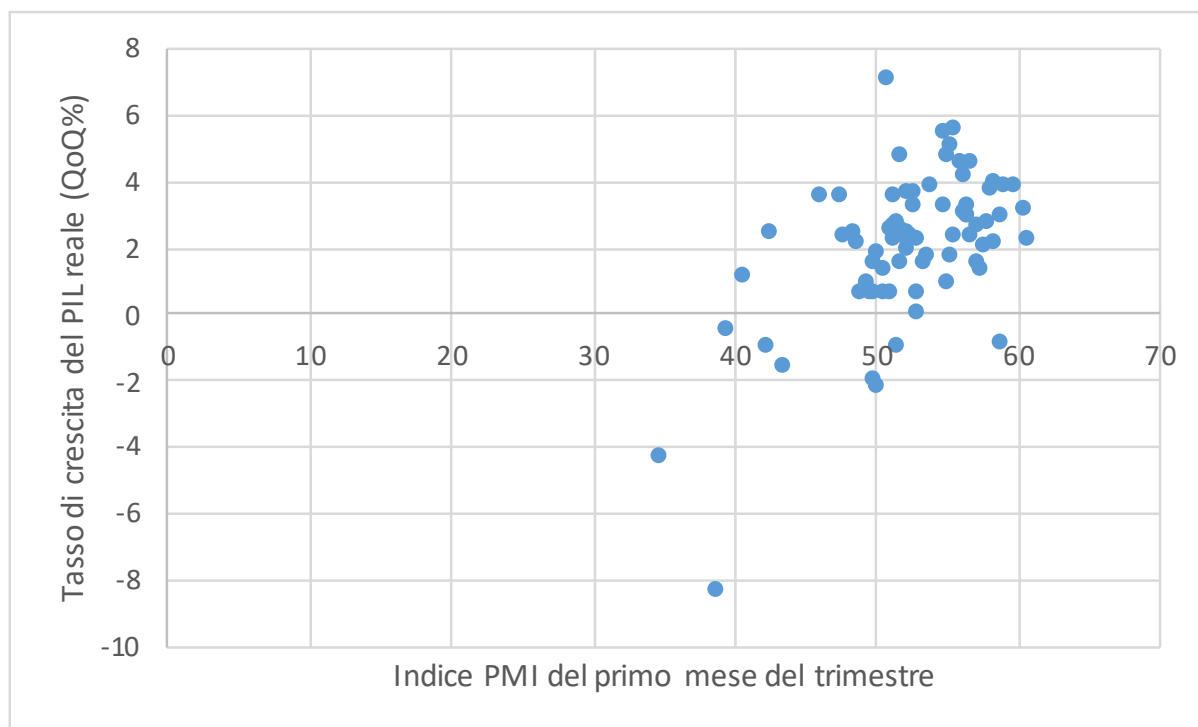
Prima di procedere ad un'analisi di regressione conviene sempre organizzare i dati in forma matriciale come illustrato nelle Figure 2.2 e 2.3. In questo modo si possono infatti reperire facilmente le osservazioni che si riferiscono ad un dato periodo di tempo (nel caso di dati longitudinali) oppure ad una data unità statistica (nel caso di dati sezionali) e individuare eventuali valori mancanti e/o valori palesemente anomali. Dopo questo passo preliminare si può procedere all'analisi di regressione vera a propria. Se non diversamente specificato, in questa dispensa indicheremo sempre con  $X_1$  la variabile dipendente e con  $X_2, X_3, \dots, X_k$  le variabili esplicative.

## 2.2 Fase 1: Scelta di una famiglia di possibili funzioni interpolanti

In questa fase iniziale dobbiamo scegliere una famiglia di funzioni della quale ci aspettiamo che contenga una funzione in grado di descrivere la relazione di fondo tra le osservazioni della variabile dipendente e le osservazioni della o delle variabili esplicative.

Nel caso di una singola variabile esplicativa possiamo aiutarci visualizzando i dati come punti in un piano cartesiano. Il grafico sottostante, per esempio, è stato ottenuto a partire dai dati contenuti nel [foglio di calcolo](#) in Figura 2.2:

Figura 2.4:



Nel grafico appare abbastanza evidente che la relazione di fondo tra le osservazioni delle due variabili sia monotona crescente e per semplicità sceglieremo quindi come famiglia di possibili funzioni interpolanti la famiglia delle rette, ovvero la famiglia di tutte le funzioni del tipo

$$x_1 = f(x_2; \alpha_1, \alpha_2) = \alpha_1 + \alpha_2 x_2, \quad x_2 \in \mathbb{R},$$

dove  $\alpha_1$  e  $\alpha_2$  sono due parametri reali. In alternativa, a seconda di come si sarebbe presentata la nuvola di punti, avremmo potuto anche scegliere una famiglia di funzioni interpolanti diversa come per esempio

- la *famiglia funzioni quadratiche* che sono definite come

$$x_1 = f(x_2; \alpha_1, \alpha_2, \alpha_3) = \alpha_1 + \alpha_2 x_2 + \alpha_3 x_2^2, \quad x_2 \in \mathbb{R},$$

dove  $\alpha_1, \alpha_2, \alpha_3$  sono tre parametri reali;

- la *famiglia delle funzioni polinomiali di grado  $g$*  che sono definite come

$$x_1 = f(x_2; \alpha_1, \dots, \alpha_k) = \alpha_1 + \alpha_2 x_2 + \dots + \alpha_g x_2^g, \quad x_2 \in \mathbb{R},$$

dove  $\alpha_1, \alpha_2, \dots, \alpha_g$  sono  $g$  parametri reali;

- la *famiglia delle funzioni potenza* che sono definite come

$$x_1 = f(x_2; \alpha_1, \alpha_2) = \alpha_1 \times x_2^{\alpha_2}, \quad x_2 > 0,$$

dove  $\alpha_1$  e  $\alpha_2$  sono due parametri reali;

- la *famiglia delle funzioni esponenziali* che sono definite come

$$x_1 = f(x; \alpha_1, \alpha_2) = \alpha_1 \times \alpha_2^{x^2}, \quad x_2 \in \mathbb{R}$$

dove  $\alpha_1$  è un parametro reale e  $\alpha_2$  è un parametro reale strettamente positivo;

- ...altre famiglie di funzioni...

Anche nel caso di due variabili esplicative ( $k = 3$ ) è ancora possibile rappresentare le osservazioni in uno spazio a tre dimensioni (la terza dimensione è riservata alla variabile dipendente). Tuttavia, a partire da una nuvola di punti tridimensionale è molto difficile inferire l'andamento della relazione di fondo tra le osservazioni delle variabili e per questo motivo nel caso di due o più variabili esplicative di solito si avvia un'analisi di regressione facendo riferimento ad una famiglia di funzioni interpolanti molto semplice come per esempio

- la famiglia dei piani (nel caso di due variabili esplicative), ovvero delle funzione di due variabili che possono essere definite come

$$x_1 = f(x_2, x_3; \alpha_1, \alpha_2, \alpha_3) = \alpha_1 + \alpha_2 x_2 + \alpha_3 x_3,$$

$$x_2, x_3 \in \mathbb{R},$$

con  $\alpha_1$ ,  $\alpha_2$  e  $\alpha_3$  tre parametri reali,

- o quella degli iperpiani (nel caso di tre o più variabili esplicative), ovvero delle funzioni di tre o più variabili che possono essere definite come

$$x_1 = f(x_2, x_3, \dots, x_k; \alpha_1, \alpha_2, \dots, \alpha_k) = \alpha_1 + \alpha_2 x_2 + \dots + \alpha_k x_k,$$

$$x_2, x_3, x_4 \in \mathbb{R},$$

con  $\alpha_1, \alpha_2, \dots, \alpha_k$  parametri reali.

Successivamente, alla luce dei risultati della fase di valutazione della bontà d'adattamento si procederà eventualmente ad affinare la scelta della famiglia di possibili funzioni interpolanti.

Non di rado, nella scelta della famiglia di possibili funzioni interpolanti ci si lascia anche guidare una qualche teoria che prevede già *a priori* un certo tipo di relazione di fondo tra le variabili oggetto di analisi. Cobb e Douglas, per esempio, in un celebre studio sull'economia statunitense pubblicato nel 1928<sup>1</sup> hanno cercato di descrivere la relazione di fondo tra "osservazioni" della quantità

---

<sup>1</sup>Cobb, C. W.; Douglas, P. H. (1928). "A Theory of Production". American Economic Review. 18 (Supplement): 139–165.

di lavoro e capitale e corrispondenti "osservazioni" della quantità prodotta.<sup>2</sup> A tal scopo hanno condotto un'analisi di regressione usando come famiglia di possibili funzioni interpolanti una particolare famiglia di funzioni omogenee di primo grado che successivamente divenne nota come famiglia di funzioni di produzione di Cobb-Douglas. In realtà, come i due autori spiegano nell'articolo scientifico che espone i risultati della loro ricerca, la scelta della famiglia di possibili funzioni interpolanti non era basata esclusivamente sulle osservazioni delle variabili oggetto di studio, ma è stata anche indotta dalla teoria marginalista e da un teorema dimostrato da Wicksteed.<sup>3</sup>

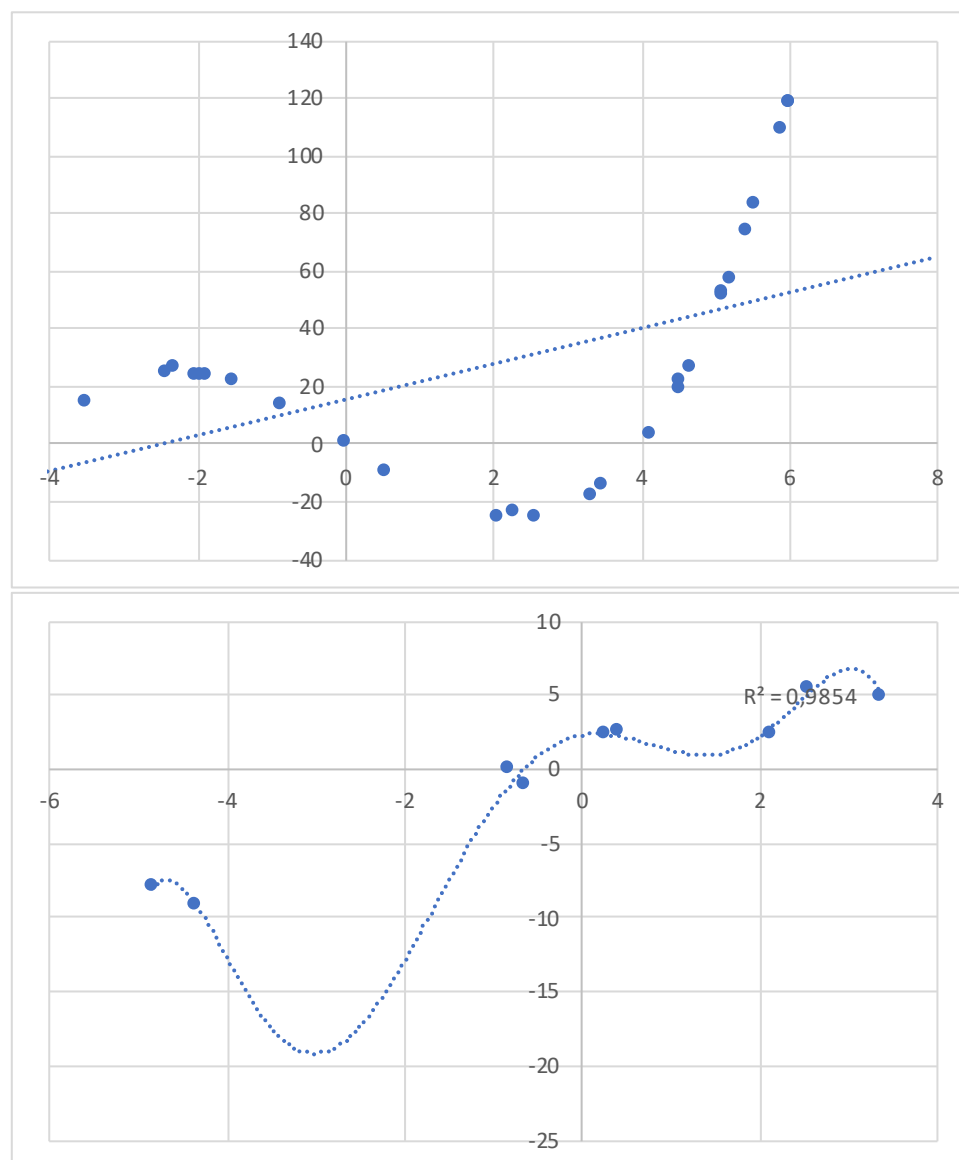
Due problemi importanti da tenere in considerazione quando si sceglie una famiglia di possibili funzioni interpolanti sono quelli dell'*underfitting* e dell'*overfitting*. Questi problemi possono insorgere quando si sceglie una famiglia di funzioni troppo ristretta (*underfitting*) o troppo ampia (*overfitting*). I due grafici sottostanti illustrano questi problemi:

---

<sup>2</sup>La parola "osservazioni" è scritta tra virgolette perché in realtà le quantità di lavoro e capitale usate da Cobb e Douglas sono state faticosamente dedotte da altre fonti statistiche disponibili all'epoca in cui hanno condotto le loro ricerche.

<sup>3</sup>Philip H. Wicksteed (1894), "An Essay on the Co-ordination of the Laws of Distribution", 1932 edition, Reprint No. 12, London: London School of Economics



Figura 2.5: *Underfitting* (in alto) e *overfitting* (in basso)

Nel primo grafico è stata usata la famiglia delle rette quando invece sarebbe stata più adeguata la famiglia dei polinomi di terzo grado, mentre nel secondo grafico è stata usata la famiglia dei polinomi di sesto grado quando invece sarebbe stata più adeguata la famiglia delle rette. Sia l'*underfitting* che l'*overfitting* danno luogo ad una rappresentazione sbagliata della relazione di fondo tra le osservazioni delle variabili oggetto di analisi. Inoltre, facen-

do delle previsioni sulla base di una funzione interpolante affetta da *underfitting* o *overfitting* si può incorrere in errori di previsione sistematici. Nella descrizione della fase di valutazione della bontà d'adattamento vedremo come l'*underfitting* può essere individuato mediante l'ausilio del grafico dei residui. L'*overfitting* è invece più difficile da individuare. Nella descrizione della fase di valutazione della bontà d'adattamento vedremo come può essere prevenuto partendo da una famiglia di funzioni interpolanti ristretta e ampliandola poco alla volta qualora il grafico dei residui segnalasse dell'*underfitting*.

Prima di passare alla fase successiva conviene ancora osservare che tutte le famiglie di funzioni che abbiamo precedentemente proposto sono *parametriche* nel senso che sono indicizzate da un numero finito di parametri: ad ogni combinazione di valori dei parametri  $\alpha_1, \alpha_2, \dots, \alpha_g$  corrisponde un'unica funzione all'interno della famiglia e viceversa. Tuttavia, nelle analisi di regressione si ricorre spesso anche a famiglie *non parametriche*, ovvero a famiglie di funzioni che non possono essere indicizzate (in modo ragionevole) da un numero finito di parametri. Esempi di famiglie non parametriche sono la famiglia di tutte le funzioni continue oppure la famiglia di tutte le funzioni con un certo numero di derivate continue, la famiglia di tutte le funzioni infinitamente derivabili ecc. Per non mettere troppa carne al fuoco, in questa dispensa ci limiteremo a considerare solo alcune famiglie parametriche.

## 2.3 Fase 2: Individuazione della funzione interpolante

Una volta scelta una famiglia di funzioni, si dovrà individuare al suo interno una funzione interpolante

$$\hat{x}_1 = f(x_2, x_3, \dots, x_k; \hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_g)$$

che riproduca la relazione di fondo tra le osservazioni delle variabili. A tal fine bisogna fare riferimento ad un qualche **metodo di accostamento** come per esempio il **metodo dei minimi quadrati**. Secondo questo metodo i valori dei parametri che identificano la funzione interpolante dovrebbero essere determinati in modo tale che risulti minimizzata la funzione

$$D(\alpha_1, \alpha_2, \dots, \alpha_g) = \sum_{i=1}^n [x_{1i} - f(x_{2i}, x_{3i}, \dots, x_{ki}; \alpha_1, \dots, \alpha_g)]^2$$

che in seguito chiameremo **distanza quadratica**.

Consideriamo un esempio. Se applichiamo il metodo dei minimi quadrati per trovare l'intercetta  $\hat{\alpha}_1$  e il coefficiente angolare  $\hat{\alpha}_2$  di una retta che interpola i punti in Figura 2.4, otteniamo (questo risultato verrà verificato nell'Esempio 3.2)

$$\hat{\alpha}_1 = -10,535 \quad \text{e} \quad \hat{\alpha}_2 = 0,239.$$

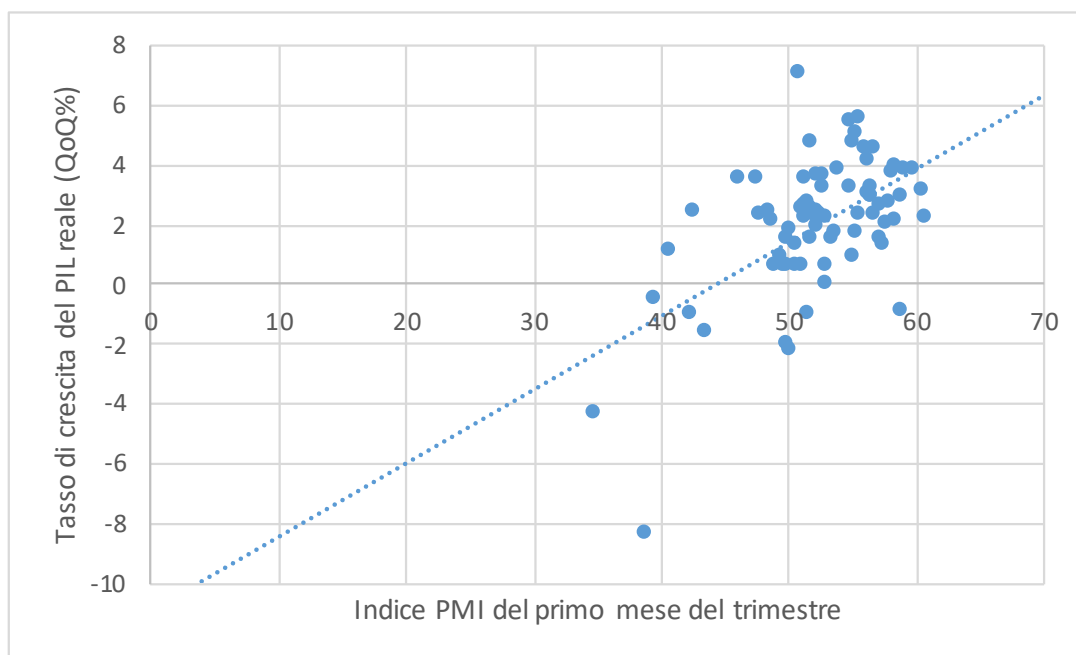
La retta ai minimi quadrati che interpola la nuvola di punti in Figura 2.4 è dunque data da

$$\hat{x}_1 = f(x_2; \hat{\alpha}_1, \hat{\alpha}_2) = -10,535 + 0,239x_2$$

e sovrapponendola alla nuvola di punti vediamo che il metodo dei minimi quadrati ha effettivamente prodotto una retta che descrive

la relazione di fondo tra le osservazioni dell'indice PMI e quelle del tasso di crescita del PIL reale:

Figura 2.6:



Ovviamente esistono tanti metodi alternativi a quello dei minimi quadrati che in alcune applicazioni potrebbero essere preferibili. Ma siccome in molte applicazioni importanti il metodo dei minimi quadrati produce delle funzioni interpolanti che hanno delle proprietà molto vantaggiose, e siccome il metodo dei minimi quadrati ha inoltre spesso il pregio di essere semplice e veloce da applicare, in questa dispensa ci limiteremo a trattare solo questo metodo.

## 2.4 Fase 3: Valutazione della bontà d'adattamento

Una volta determinata una funzione interpolante bisogna valutare la sua bontà d'adattamento rispetto ai dati. Questa valutazione si articola in tre momenti:

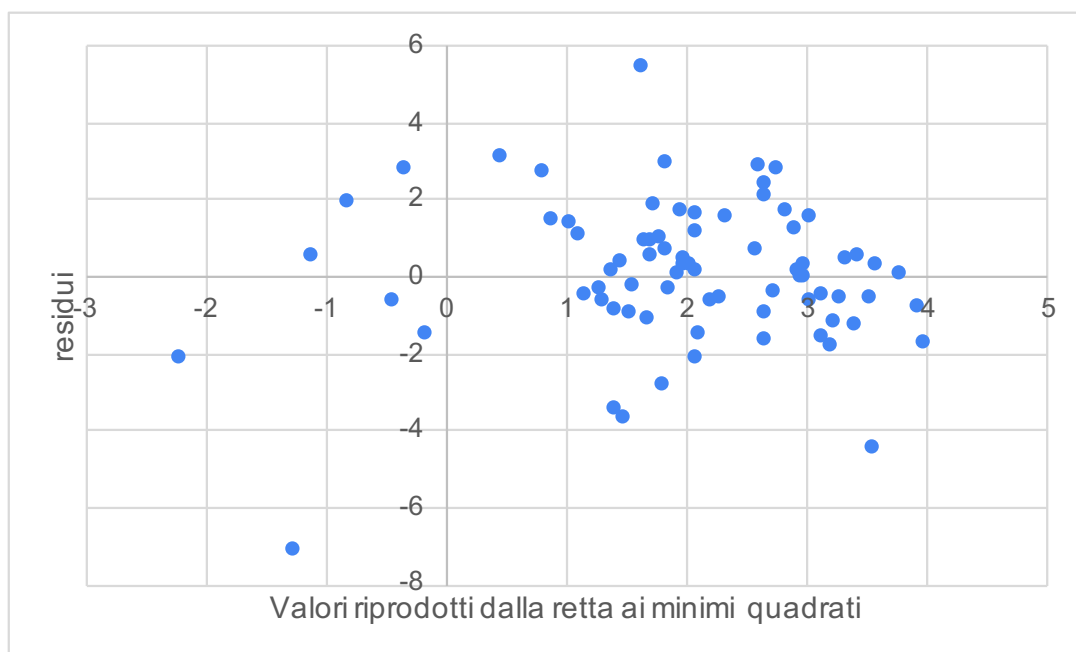
- 1) valutare se l'andamento della funzione interpolante è troppo rigido (*underfitting*) o troppo contorto (*overfitting*);
- 2) valutare se i dati contengono delle osservazioni anomale che potrebbero attirare verso di sé l'andamento della funzione interpolante;
- 3) calcolare degli indici sintetici che misurano la bontà d'adattamento.

Come abbiamo già anticipato nella descrizione della fase di scelta della famiglia di possibili funzioni interpolanti (vedi Sezione 2.2), per individuare casi di eccessiva rigidità o *underfitting* conviene costruire il cosiddetto **grafico dei residui**. Vediamo dunque che cos'è il grafico dei residui e come viene utilizzato per individuare casi di *underfitting*. Come suggerisce il suo nome, il grafico dei residui è basato sui residui di interpolazione, ovvero sugli scarti  $z_i$  tra i valori osservati per la variabile dipendente e i corrispondenti valori riprodotti dalla funzione interpolante:

$$\begin{aligned} z_i &= x_{1i} - f(x_{2i}, x_{3i}, \dots, x_{ki}; \hat{\alpha}_1, \dots, \hat{\alpha}_g) \\ &= x_{1i} - \hat{x}_{1i}, \quad i = 1, 2, \dots, n. \end{aligned}$$

Infatti, il grafico dei residui è un grafico a dispersione che riporta i valori riprodotti  $\hat{x}_{1i}$  in ascissa e i residui  $z_i$  in ordinata. Figura 2.7 mostra, per esempio, il grafico dei residui che si ottiene a partire dalla retta ai minimi quadrati in Figura 2.6:

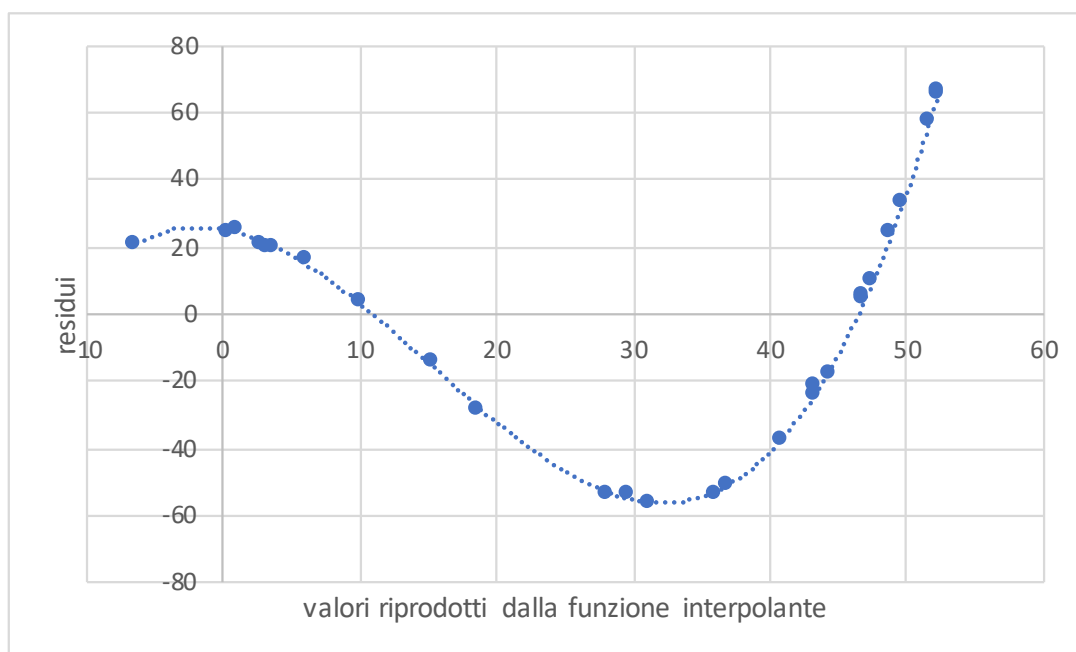
Figura 2.7:



Ora, per verificare se l'andamento di una funzione interpolante sia troppo rigido bisogna analizzare la **successione dei segni dei residui** nell'ordine in cui si presentano sul grafico dei residui (ovvero al crescere dei valori riprodotti  $\hat{x}_{1i}$ ). Se la funzione interpolante ha un andamento sufficientemente flessibile per cogliere la relazione di fondo tra le osservazioni delle variabili oggetto di analisi, la successione dei segni dei residui dovrebbe essere casuale come per esempio in Figura 2.7. Se invece osservassimo lunghe successioni di residui dello stesso segno, allora questo sarebbe un chiaro segnale che l'andamento della funzione interpolante sia troppo rigido (ov-

vero di *underfitting*). Il grafico dei residui che si ricava dalla prima funzione interpolante in Figura 2.5 (quella corrispondente alla situazione di *underfitting*) illustra che cosa accade in una situazione di *underfitting*:

Figura 2.8:



In questo grafico notiamo tre lunghe sequenze di residui dello stesso segno: da sinistra verso destra (ovvero all'aumentare dei valori riprodotti sull'asse delle ascisse) notiamo prima una lunga sequenza di residui positivi, poi una lunga sequenza di residui negativi e infine un'altra lunga sequenza di residui positivi. Come dicevamo, questo è un chiaro segnale che la funzione interpolante lineare nel primo grafico in Figura 2.5 abbia un andamento troppo rigido e che quindi si dovrebbe cercare un'altra funzione interpolante all'interno di una famiglia di possibili funzioni interpolanti più ampia della famiglia delle rette. Per capire come ampliare la

famiglia delle rette ci conviene studiare la relazione di fondo tra i residui e i valori interpolati. Come evidenziato dalla curva tratteggiata nel grafico dei residui in Figura 2.8, i valori  $z_i$  dei residui sono concentrati in prossimità di una funzione polinomiale di terzo grado che indicheremo con  $g(\cdot)$ . Questo significa che

$$z_i = x_{1i} - \hat{x}_{1i} \simeq g(\hat{x}_{1i}) \quad \text{per ogni } i = 1, 2, \dots, n,$$

ovvero che

$$x_{1i} \simeq \hat{x}_{1i} + g(\hat{x}_{1i}) \quad \text{per ogni } i = 1, 2, \dots, n.$$

Ora siccome

$$\hat{x}_{1i} = f(x_{2i}; \hat{\alpha}_1, \hat{\alpha}_2) \quad \text{per } i = 1, 2, \dots, n,$$

il grafico dei residui ci sta in pratica suggerendo che la funzione

$$\hat{\hat{x}}_1 = f^*(x_2) = f(x_2; \hat{\alpha}_1, \hat{\alpha}_2) + g(f(x_2; \hat{\alpha}_1, \hat{\alpha}_2))$$

sia una funzione interpolante più adatta per descrivere la relazione di fondo tra le osservazioni delle variabili  $X_1$  e  $X_2$ . E siccome nel caso in questione la funzione interpolante

$$f(x_2; \hat{\alpha}_1, \hat{\alpha}_2) = \hat{\alpha}_1 + \hat{\alpha}_2 \times x_2$$

è una retta (vedi il primo grafico in Figura 2.5), la nuova funzione interpolante  $\hat{\hat{x}}_1 = f^*(x_2)$  deve essere un polinomio di terzo grado (infatti,  $f(x_2; \hat{\alpha}_1, \hat{\alpha}_2)$  è una retta e dunque un polinomio di primo grado e siccome la funzione  $g(\cdot)$  è una funzione polinomiale di terzo grado, anche  $g(f(x_2; \hat{\alpha}_1, \hat{\alpha}_2))$  dovrà essere una funzione



polinomiale di terzo grado; ne consegue che la somma tra la funzione  $f(x_2; \hat{\alpha}_1, \hat{\alpha}_2)$  e la funzione  $g(f(x_2; \hat{\alpha}_1, \hat{\alpha}_2))$  sia anch'essa un polinomio di terzo grado). Questo ragionamento ci mostra dunque come il grafico dei residui ci ha indicato la strada per scegliere una famiglia di possibili funzioni interpolanti più adatta della famiglia delle rette (ovvero la famiglia dei polinomi di terzo grado)!!!! Non è difficile rendersi conto che questo ragionamento può essere replicato anche nel caso di più di una variabile esplicativa: se il grafico dei residui associato ad una data funzione interpolante  $\hat{x}_1 = f(x_2, x_3, \dots, x_k; \hat{\alpha}_1, \dots, \hat{\alpha}_g)$  ci suggerisce che

$$z_i = x_{1i} - \hat{x}_{1i} \simeq g(\hat{x}_{1i}) \quad \text{per ogni } i = 1, 2, \dots, n,$$

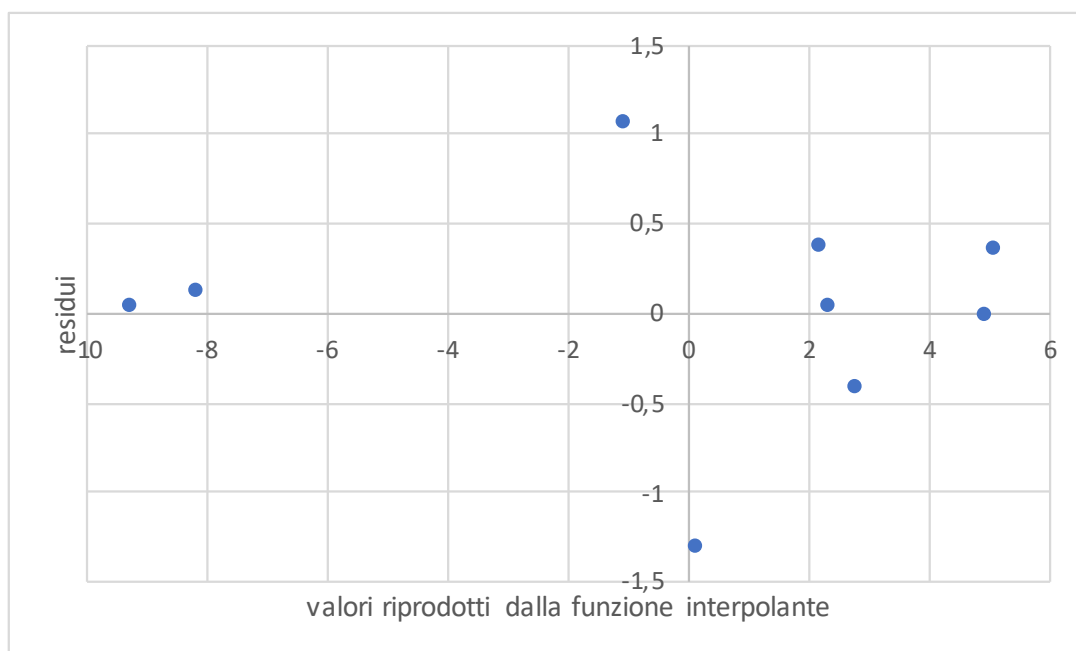
allora

$$\begin{aligned} \hat{\hat{x}}_1 &= f^*(x_2, x_3, \dots, x_k) \\ &= f(x_2, x_3, \dots, x_k; \hat{\alpha}_1, \dots, \hat{\alpha}_g) + \\ &\quad + g(f(x_2, x_3, \dots, x_k; \hat{\alpha}_1, \dots, \hat{\alpha}_g)) \end{aligned}$$

sarà una nuova funzione interpolante con andamento più flessibile che si adatta meglio ai dati. Anche se in molte applicazioni questo metodo per trovare una nuova funzione interpolante è molto efficace, bisogna tenere presente che in alcuni casi può anche causare dei problemi. Infatti, a volte l'andamento più flessibile della nuova funzione interpolante  $f^*(x_2, x_3, \dots, x_k)$  potrebbe essere difficile da interpretare e se la relazione di fondo tra i residui  $z_i$  e il valori interpolati  $\hat{x}_{1i}$  ha un andamento molto contorto, la nuova funzione interpolante  $f^*(x_2, x_3, \dots, x_k)$  erediterà questo andamento contorto e da una situazione di *underfitting* si potrebbe quindi passare ad una situazione di *overfitting*.

Pur essendo uno strumento molto potente per individuare e correggere situazioni di *underfitting*, il grafico dei residui purtroppo non fornisce nessun segnale in casi di *overfitting*. Per fare un esempio che illustra questa carenza, costruiamo il grafico dei residui della seconda funzione interpolante in Figura 2.5 (quella corrispondente alla situazione di *overfitting*):

Figura 2.9:



Notiamo che in questo grafico la successione dei segni dei residui sembra perfettamente casuale e inoltre che tra i residui e i valori riprodotti non sembra esserci nessuna relazione. Questo grafico dei residui suggerisce dunque un buon adattamento della funzione interpolante e non svela che essa in realtà ha un andamento molto contorto (vedi il secondo grafico in Figura 2.5) e che si potrebbe ottenere un adattamento (quasi) altrettanto buono con una funzione interpolante che ha un andamento molto più rego-

lare (per esempio con una retta). Come anticipato nella Sezione 2.2, per evitare di ottenere una funzione interpolante con andamento troppo contorto (soprattutto in situazioni dove si hanno più variabili esplicative e quindi è difficile o impossibile rappresentare graficamente la funzione interpolante sovrapposta alla nuvola di punti) conviene partire da una famiglia di possibili funzioni interpolanti molto ristretta (come per esempio la famiglia delle rette, dei piani o degli iperpiani) e ampliarla poco alla volta nel caso in cui il grafico dei residui evidenziasse un problema di *underfitting*. In questo modo si possono sfruttare le indicazioni fornite dal grafico dei residui per ampliare la famiglia di funzioni interpolanti in modo opportuno.

Ovviamente, nel caso di una singola variabile esplicativa conviene sempre costruire il grafico che sovrappone la funzione interpolante alla nuvola di punti per valutare se l'andamento della funzione interpolante sia troppo rigido o troppo contorto (vedi per esempio Figura 2.5). Tuttavia, nel caso di due o più variabili esplicative può essere complicato (due variabili esplicative) se non impossibile (più di due variabili esplicative) visualizzare la funzione interpolante sovrapposta alla nuvola di punti. In questi casi il grafico dei residui si rivela uno strumento assai utile.

Vediamo ora invece come con l'ausilio del grafico dei residui si possono anche individuare dei dati anomali. A tal fine bisogna innanzitutto calcolare un indice che misura l'ordine di grandezza medio dei residui. Di solito, soprattutto quando la funzione interpolante è stata individuata con il metodo dei minimi quadrati, si

utilizza la media quadratica dei residui, ovvero l'indice

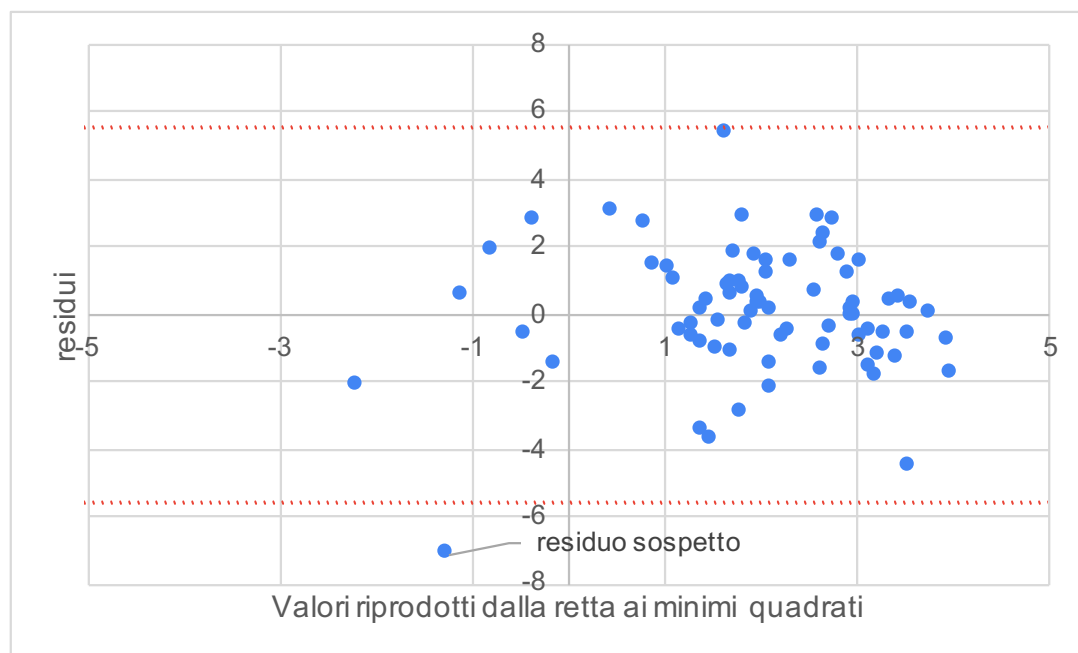
$$A_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2},$$

ma con altri metodi di accostamento può essere preferibile utilizzare l'indice

$$A_1 = \frac{1}{n} \sum_{i=1}^n |z_i|.$$

Infatti, entrambi questi indici misurano l'ordine di grandezza medio dei residui o, se si preferisce, la dispersione delle osservazioni attorno alla funzione interpolante. Una volta calcolato uno di questi indici si confrontano i singoli residui  $z_i$  con un opportuno multiplo dell'indice. Di solito si fa riferimento al triplo del valore dell'indice  $A_2$ : residui  $z_i$  i cui moduli  $|z_i|$  superano il *triplo del valore di  $A_2$*  dovrebbero essere considerati sospetti. Per facilitare l'individuazione di residui sospetti conviene tracciare nel grafico dei residui due linee orizzontali che intersecano l'asse delle ordinate in corrispondenza di  $\pm 3 \times A_2$  così come mostrato nel grafico sottostante:

Figura 2.10:



In questo modo i residui sospetti si troveranno al di fuori della fascia delimitata dalle due linee orizzontali. In presenza di residui  $z_i$  i cui moduli  $|z_i|$  superano di gran lunga il triplo del valore di  $A_2$  si procede all'eliminazione delle corrispondenti  $k$ -uple  $(x_{1i}, x_{2i}, \dots, x_{ki})$  e alla rideterminazione della funzione interpolante senza usare le  $k$ -uple che hanno generato i residui sospetti. Se la nuova funzione interpolante presenta un buon adattamento ai dati rimanenti, si potrà ritenere che le  $k$ -uple eliminate siano effettivamente il frutto di una qualche "anomalia" e si dovrà indagare sulle cause dell'anomalia.

Infine, per completare la fase di valutazione della bontà d'adattamento, di solito si calcolano anche degli indici di adattamento *relativi*, ovvero degli indici basati sul confronto tra la dispersione

delle osservazioni della variabile dipendente attorno alla funzione interpolante e la loro dispersione rispetto ad una ipotetica funzione interpolante che è costante e pari ad un dato valore rappresentativo (che di solito è la media aritmetica delle osservazioni della variabile dipendente). Due candidati "naturali" per questo tipo di confronto sono gli indici

$$A_1^* = \frac{A_1}{M_1(|X_1 - \bar{x}_1|)} = \frac{\frac{1}{n} \sum_{i=1}^n |z_i|}{\frac{1}{n} \sum_{i=1}^n |x_{1i} - \bar{x}_1|} = \frac{\sum_{i=1}^n |z_i|}{\sum_{i=1}^n |x_{1i} - \bar{x}_1|}$$

e

$$A_2^* = \frac{A_2}{\sigma_1} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}} = \sqrt{\frac{\sum_{i=1}^n z_i^2}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}}, \quad (1)$$

con

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} \quad \text{e} \quad \sigma_1 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2},$$

ma ovviamente si potrebbe pensare anche ad altri indici. Si noti che gli indici  $A_1^*$  e  $A_2^*$  restituiscono dei numeri puri, ovvero dei numeri privi di unità di misura e quindi possono essere utilizzati per confrontare l'adattamento di funzioni interpolanti che si riferiscono a variabili dipendenti di natura completamente diversa. Per esempio,  $A_2^* = 0,16$  significa che la dispersione media delle osservazioni della variabile dipendente attorno alla funzione interpolante è pari al 16% della loro dispersione media attorno alla loro media aritmetica. Per esprimere questo concetto in modo più sintetico conviene dire che l'ordine di grandezza medio dei residui (misurato dall'indice  $A_2$ ) è pari al 16% dello scarto quadratico medio delle

osservazioni della variabile dipendente. Come vedremo più avanti, nel caso di rette, piani o iperpiani ai minimi quadrati l'indice  $A_2^*$  è sempre un numero compreso tra 0 (dispersione nulla attorno alla funzione interpolante) e 1 (dispersione massima attorno alla funzione interpolante) ed è strettamente legato ad un altro indice di adattamento relativo il cui utilizzo (nel caso di rette, piani e iperpiani ai minimi quadrati) è molto più popolare: l'indice di determinazione.

## 2.5 Fase 4: Interpretazione

Nella fase interpretativa bisogna cercare di rispondere alle domande che ci hanno indotto ad intraprendere l'analisi di regressione. Nel caso in cui lo scopo fosse quello di trovare una semplice descrizione della relazione di fondo tra le osservazioni delle variabili oppure di indagare sul processo che ha generato i dati oggetto di analisi dovremmo dunque rispondere a domande del tipo:

- Come si comporta la funzione interpolante al variare delle variabili esplicative?
- Esistono delle teorie che per le variabili oggetto di analisi prevedono una relazione di fondo diversa da quella descritta dalla funzione interpolante? In caso affermativo, come si spiegano le discrepanze rispetto a quanto previsto dalla teoria?
- La funzione interpolante presenta un buon adattamento ai dati?

- La dispersione delle osservazioni della variabile dipendente attorno alla funzione interpolante è costante al variare della o delle variabili esplicative? (Si noti che il grafico dei residui può fornire indicazioni utili a riguardo, ma in ogni caso, per ottenere chiarezza su questo aspetto, si può sempre procedere ad una ulteriore analisi di regressione per indagare sulla relazione di fondo tra i moduli o i quadrati dei residui e le osservazioni delle variabili esplicative.)

Se inoltre avessimo l'intenzione di usare la funzione interpolante per effettuare delle previsioni sui valori che assumerà la variabile dipendente in corrispondenza di nuove osservazioni delle variabili esplicative, allora dovremmo anche chiederci:

- La funzione interpolante può essere utilizzata per fare delle previsioni? In caso affermativo, possiamo aspettarci che le previsioni siano precise?
- Quali ulteriori o altre variabili esplicative potrebbero condurre ad una funzione interpolante dalla quale si potrebbero ottenere previsioni più precise?
- Quali variabili esplicative potrei ignorare in una nuova analisi di regressione senza pregiudicare (o addirittura migliorando) la precisione di previsioni ottenute da una nuova funzione interpolante?

Rispondere a queste domande non è sempre facile e richiede non solo buone conoscenze statistiche e probabilistiche, ma anche



un'ottima conoscenza del processo che genera i dati. Negli esempi che vedremo in questa dispensa la fase interpretativa sarà limitata ad alcuni commenti sull'andamento della funzione interpolante.

### 3 La retta ai minimi quadrati

In questa sezione ricaveremo delle formule per calcolare i parametri di una retta ai minimi quadrati e dimostreremo le principali proprietà di questo tipo di funzione interpolante. Inoltre, definiremo e dimostreremo le principali proprietà di due importanti indici statistici legati alla retta ai minimi quadrati: l'indice di determinazione e il coefficiente di correlazione lineare. Per procedere con questo programma ci serviranno alcune proprietà fondamentali della media aritmetica, della devianza e della codevianza che verranno ricordate nella prossima sottosezione.

#### 3.1 Proprietà fondamentali della media aritmetica, della devianza e della codevianza\*

La media aritmetica, la devianza e la codevianza sono degli indici della statistica descrittiva che si incontrano in molte dimostrazioni e applicazioni. In questa dispensa indicheremo questi tre indici rispettivamente con  $M_1(X)$ ,  $dev(X)$  e  $codev(X_1, X_2)$ , dove  $X$ ,  $X_1$  e  $X_2$  sono le variabili di riferimento. Spesso indicheremo la media aritmetica anche con una lettera minuscola soprasssegnata con una linea orizzontale come per esempio  $\bar{x}$ ,  $\bar{x}_1$  e  $\bar{x}_2$ . Ricordiamo che la media aritmetica, la devianza e la codevianza sono definite come

$$M_1(X) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$dev(X) = \sum_{i=1}^n (x_i - \bar{x})^2$$

e

$$\text{cov}(X_1, X_2) = \sum_{i=1}^n (x_{1i} - \bar{x}_1) \times (x_{2i} - \bar{x}_2).$$

In quanto segue useremo le lettere  $a$  e  $b$  (eventualmente accompagnate da un pedice e/o un apice) per indicare delle costanti reali e nelle proprietà che coinvolgono le  $n$  trasformazioni lineari affini  $y_i = a + bx_i$  useremo l'espressione  $Y = a + b \times X$  per indicare che  $Y$  è la variabile che rappresenta i valori  $y_i$ . Più in generale, useremo l'espressione

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

per indicare che  $Y$  è la variabile che rappresenta le  $n$  trasformazioni lineari affini

$$y_i = a + b_1x_{1i} + b_2x_{2i} + \cdots + b_kx_{ki}.$$

### Proprietà della media aritmetica:

- a) La somma degli scarti tra  $n$  valori  $x_i$  e la loro media aritmetica è nulla:  
 $\sum_{i=1}^n (x_i - \bar{x}) = 0.$
- b) Proprietà di minimo: la somma degli scarti quadratici  $\sum_{i=1}^n (x_i - a)^2$  come funzione di  $a \in \mathbb{R}$  raggiunge il suo valore minimo se e solo se  $a = \bar{x}$ .
- c) Trasformazioni lineari affini di una unica variabile: se  $Y = a + bX$ , allora  
 $M_1(Y) = M_1(a + bX) = a + bM_1(X).$
- d) Somme: se  $Y = X_1 + X_2$ , allora

$$M_1(Y) = M_1(X_1 + X_2) = M_1(X_1) + M_1(X_2).$$

- e) Trasformazioni lineari affini di più variabili: se

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k,$$

allora

$$\begin{aligned} M_1(Y) &= M_1(a + b_1X_1 + b_2X_2 + \cdots + b_kX_k) \\ &= a + \sum_{j=1}^k b_j M_1(X_j). \end{aligned}$$

**Dimostrazione.** a) La dimostrazione è immediata:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n \times \bar{x} = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0.$$

b) La somma degli scarti quadratici può essere riscritta come

$$\begin{aligned}
 \sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - a)]^2 \\
 &= \sum_{i=1}^n [(x_i - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_i - \bar{x})(\bar{x} - a)] \\
 &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\bar{x} - a)^2 + 2 \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - a) \\
 &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2 + 2(\bar{x} - a) \sum_{i=1}^n (x_i - \bar{x})
 \end{aligned}$$

Ricordando che  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  (vedi la proprietà a)), possiamo dunque concludere che

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2.$$

La proprietà di minimo segue dunque dal fatto che  $n(\bar{x} - a)^2$  raggiunge il suo valore minimo se e solo se  $a = \bar{x}$ .

c) La media delle  $n$  trasformazioni lineari affini  $y_i = a + bx_i$  è data da

$$\begin{aligned}
 \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (a + bx_i) \\
 &= \frac{1}{n} \left( \sum_{i=1}^n a + \sum_{i=1}^n bx_i \right) \\
 &= \frac{1}{n} \sum_{i=1}^n a + \frac{b}{n} \sum_{i=1}^n x_i \\
 &= a + b\bar{x}.
 \end{aligned}$$

d) La media delle  $n$  somme  $y_i = x_{1i} + x_{2i}$  è data da

$$\begin{aligned}
 \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (x_{1i} + x_{2i}) \\
 &= \frac{1}{n} \left( \sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n x_{1i} + \frac{1}{n} \sum_{i=1}^n x_{2i} \\
 &= \bar{x}_1 + \bar{x}_2.
 \end{aligned}$$

e) La proprietà e) è un corollario delle proprietà c) e d). Omettiamo la dimostrazione.

□

**Proprietà della devianza (varianza):**

- a) Formula indiretta: la devianza può essere calcolata secondo la formula indiretta

$$dev(X) = \sum_{i=1}^n x_i^2 - n\bar{x}^2.$$

- b) Formula alternativa: La devianza può essere anche calcolata come

$$dev(X) = \sum_{i=1}^n (x_i - \bar{x})x_i.$$

- c) Trasformazioni lineari affini di una singola variabile: se  $Y = a + bX$ , allora

$$dev(Y) = dev(a + bX) = b^2 dev(X).$$

- d) Somme: se  $Y = X_1 + X_2$ , allora

$$dev(Y) = dev(X_1 + X_2) = dev(X_1) + dev(X_2) + 2codev(X_1, X_2).$$

- e) Trasformazioni lineari affini di più variabili: se

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k,$$

allora

$$\begin{aligned} dev(Y) &= dev(a + b_1X_1 + b_2X_2 + \dots + b_kX_k) \\ &= \sum_{j=1}^k b_j^2 dev(X_j) + 2 \sum_{j=2}^k \sum_{j'=1}^{j-1} b_{j'} b_j codev(X_{j'}, X_j). \end{aligned}$$

Dividendo per  $n$  si ottengono le analoghe proprietà della varianza.

**Dimostrazione.** a) La formula indiretta della devianza può essere ottenuta come segue:

$$\begin{aligned} dev(X) &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2. \end{aligned}$$

b) Partendo dalla formula definitoria della devianza si ottiene

$$\begin{aligned}
 dev(X) &= \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) \\
 &= \sum_{i=1}^n [(x_i - \bar{x})x_i - (x_i - \bar{x})\bar{x}] \\
 &= \sum_{i=1}^n (x_i - \bar{x})x_i - \sum_{i=1}^n (x_i - \bar{x})\bar{x} \\
 &= \sum_{i=1}^n (x_i - \bar{x})x_i - \bar{x} \sum_{i=1}^n (x_i - \bar{x}).
 \end{aligned}$$

Ricordando che  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  (proprietà a) della media aritmetica) si vede dunque che

$$dev(X) = \sum_{i=1}^n (x_i - \bar{x})x_i$$

come volevasi dimostrare.

c) Siccome  $\bar{y} = a + b\bar{x}$  (vedi le proprietà della media aritmetica), possiamo scrivere

$$\begin{aligned}
 dev(Y) &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n [(a + bx_i) - (a + b\bar{x})]^2 \\
 &= \sum_{i=1}^n [b(x_i - \bar{x})]^2 \\
 &= b^2 \sum_{i=1}^n [x_i - \bar{x}]^2 \\
 &= b^2 dev(X).
 \end{aligned}$$

d) Siccome  $\bar{y} = \bar{x}_1 + \bar{x}_2$  (vedi le proprietà della media aritmetica), possiamo scrivere

$$\begin{aligned}
 dev(Y) &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n [(x_{1i} + x_{2i}) - (\bar{x}_2 + \bar{x}_1)]^2 \\
 &= \sum_{i=1}^n [(x_{1i} - \bar{x}_1) + (x_{2i} - \bar{x}_2)]^2 \\
 &= \sum_{i=1}^n [(x_{1i} - \bar{x}_1)^2 + (x_{2i} - \bar{x}_2)^2 + 2(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)] \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 \\
 &\quad + 2 \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\
 &= dev(X_1) + dev(X_2) + 2codev(X_1, X_2).
 \end{aligned}$$

- e) Questa proprietà è un corollario delle proprietà c) e d). Omettiamo la dimostrazione.

□

**Proprietà della codevianza (covarianza):**

a)  $codev(X, X) = dev(X)$

b) Simmetria:  $codev(X_1, X_2) = codev(X_2, X_1)$ .

c) Formula indiretta: la codevianza può essere calcolata secondo la formula indiretta

$$codev(X_1, X_2) = \sum_{i=1}^n x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2.$$

d) Disuguaglianza di Cauchy-Schwarz:

$$[codev(X_1, X_2)]^2 \leq dev(X_1)dev(X_2)$$

con uguaglianza se e solo se tutti i punti  $(x_{2i}, x_{1i})$  giacciono su una retta.

e) Formula alternativa: la codevianza può essere anche calcolata come

$$codev(X_1, X_2) = \sum_{i=1}^n (x_{1i} - \bar{x}_1)x_{2i}.$$

f) Trasformazioni lineari affini di una singola variabile: se  $Y = a + bX_2$ , allora

$$codev(X_1, Y) = codev(X_1, a + bX_2) = bcodev(X_1, X_2).$$

g) Somme: se  $Y = X_2 + X_3$ , allora

$$codev(X_1, Y) = codev(X_1, X_2 + X_3) = codev(X_1, X_2) + codev(X_1, X_3).$$

h) Trasformazioni lineari affini di più variabili: se

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

e

$$Y' = a' + b'_1X'_1 + b'_2X'_2 + \dots + b'_kX'_k,$$

allora

$$\begin{aligned} codev(Y, Y') &= \\ &= codev(a + b_1X_1 + b_2X_2 + \dots + b_kX_k, a' + b'_1X'_1 + b'_2X'_2 + \dots + b'_kX'_k) \\ &= \sum_{j=1}^k \sum_{j'=1}^k b_j b'_{j'} codev(X_j, X'_{j'}). \end{aligned}$$

Dividendo per  $n$  si ottengono le analoghe proprietà della covarianza.

**Dimostrazione.** a) La dimostrazione dell'uguaglianza  $codev(X, X) = dev(X)$  segue immediatamente dalle definizioni della devianza e della covarianza:

$$codev(X, X) = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2 = dev(X).$$

b) La simmetria segue direttamente dalla commutatività del prodotto:

$$\begin{aligned} codev(X_1, X_2) &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\ &= \sum_{i=1}^n (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1) = codev(X_2, X_1). \end{aligned}$$

c) La formula indiretta della covarianza può essere dimostrata come segue:

$$\begin{aligned} codev(X_1, X_2) &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\ &= \sum_{i=1}^n (x_{1i}x_{2i} - \bar{y}_1x_{2i} - x_{1i}\bar{x}_2 + \bar{x}_1\bar{x}_2) \\ &= \sum_{i=1}^n x_{1i}x_{2i} - \bar{x}_1 \sum_{i=1}^n x_{2i} - \sum_{i=1}^n x_{1i}\bar{x}_2 + \sum_{i=1}^n \bar{x}_1\bar{x}_2 \\ &= \sum_{i=1}^n x_{1i}x_{2i} - \bar{x}_1 n\bar{x}_2 - n\bar{x}_1\bar{x}_2 + n\bar{x}_1\bar{x}_2 \\ &= \sum_{i=1}^n x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2. \end{aligned}$$

d) Nel caso in cui almeno una delle due devianze è nulla, la dimostrazione è immediata. Supponiamo, per esempio, che  $dev(X_1) = \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 = 0$ . Ovviamente questo caso si verifica se e solo se tutti i punti  $x_{1i}$  coincidono con la loro media aritmetica  $\bar{x}_1$ , ovvero se e solo se tutti i punti  $(x_{2i}, x_{1i})$  si trovano sulla retta orizzontale con ordinata  $x_1 = \bar{x}_1$ . Sostituendo  $x_{1i} = \bar{x}_1$  nella formula definitoria della covarianza vediamo che in questo caso si ha

$$codev(X_1, X_2) = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) = \sum_{i=1}^n 0(x_{2i} - \bar{x}_2) = 0.$$

Questo ragionamento dimostra quindi che nel caso  $dev(X_1) = 0$ , tutti i punti  $(x_{2i}, x_{1i})$  devono essere allineati lungo una retta e che in questo caso ambo i membri della disuguaglianza di Cauchy-Schwarz devono essere nulli. Ovviamente si può fare un ragionamento analogo anche nel caso in cui  $dev(X_2) = 0$  e nel caso in cui  $dev(X_1) = dev(X_2) = 0$  (in quest'ultimo caso tutti i punti  $(x_{2i}, x_{1i})$  coinciderebbero con il punto  $(\bar{x}_2, \bar{x}_1)$ ).

La dimostrazione della disuguaglianza di Cauchy-Schwarz per il caso in cui entrambe le devianze sono positive può essere ricavata dalle proprietà del coefficiente di correlazione lineare (vedi Sezione 3.7).



e) Partendo dalla formula definitoria della covarianza si ottiene

$$\begin{aligned}
 \text{cov}(X_1, X_2) &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\
 &= \sum_{i=1}^n [(x_{1i} - \bar{x}_1)x_{2i} - (x_{1i} - \bar{x}_1)\bar{x}_2] \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)x_{2i} - \sum_{i=1}^n (x_{1i} - \bar{x}_1)\bar{x}_2 \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)x_{2i} - \bar{x}_2 \sum_{i=1}^n (x_{1i} - \bar{x}_1)
 \end{aligned}$$

Per la proprietà a) della media aritmetica si ha  $\sum_{i=1}^n (x_{1i} - \bar{x}_1) = 0$ , e quindi possiamo concludere che

$$\text{cov}(X_1, X_2) = \sum_{i=1}^n (x_{1i} - \bar{x}_1)x_{2i}$$

come volevasi dimostrare.

f) Siccome  $\bar{y} = a + b\bar{x}_2$  (vedi le proprietà delle medie), possiamo concludere che

$$\begin{aligned}
 \text{cov}(X_1, Y) &= \text{cov}(X_1, a + bX_2) \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y}) \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)[(a + bx_{2i}) - (a + b\bar{x}_2)] \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)[b(x_{2i} - \bar{x}_2)] \\
 &= b \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\
 &= b\text{cov}(X_1, X_2).
 \end{aligned}$$

g) Siccome  $\bar{y} = \bar{x}_2 + \bar{x}_3$  (vedi le proprietà delle medie), possiamo concludere che

$$\begin{aligned}
 \text{codev}(X_1, Y) &= \text{codev}(X_1, X_2 + X_3) \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y}) \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)[(x_{2i} + x_{3i}) - (\bar{x}_2 + \bar{x}_3)] \\
 &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)[(x_{2i} - \bar{x}_2) + (x_{3i} - \bar{x}_3)] \\
 &= \sum_{i=1}^n [(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) + (x_{1i} - \bar{x}_1)(x_{3i} - \bar{x}_3)] \\
 &= \sum_{i=1}^n [(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)] + \sum_{i=1}^n [(x_{1i} - \bar{x}_1)(x_{3i} - \bar{x}_3)] \\
 &= \text{codev}(X_1, X_2) + \text{codev}(X_1, X_3).
 \end{aligned}$$

h) La proprietà h) è un corollario delle proprietà f) e g). Omettiamo la dimostrazione.  $\square$

## 3.2 Determinazione dei parametri

Consideriamo ora il [metodo dei minimi quadrati](#) per determinare all'interno della famiglia delle funzioni lineari una retta ai minimi quadrati. Come già accennato, i parametri di una retta ai minimi quadrati sono due valori

$$\alpha_1 = \hat{\alpha}_1 \quad \text{e} \quad \alpha_2 = \hat{\alpha}_2$$

che minimizzano la distanza quadratica

$$D(\alpha_1, \alpha_2) = \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i})^2.$$

Siccome  $D(\alpha_1, \alpha_2)$  è una funzione differenziabile, ogni suo punto di minimo si trova in corrispondenza di una coppia  $(\alpha_1, \alpha_2)$  che

annulla le due derivate parziali

$$\begin{aligned}\frac{\partial D}{\partial \alpha_1} &= 2 \times \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i}) \times (-1), \\ \frac{\partial D}{\partial \alpha_2} &= 2 \times \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i}) \times (-x_{2i}),\end{aligned}\tag{2}$$

e siccome  $D(\alpha_1, \alpha_2)$  è una funzione convessa (omettiamo la dimostrazione), qualunque coppia di parametri  $(\hat{\alpha}_1, \hat{\alpha}_2)$  che annulla le due derivate parziali deve essere un punto di minimi della funzione  $D(\alpha_1, \alpha_2)$ . Risolvendo il sistema di equazioni che si ottiene annullando le due derivate parziali, ovvero il sistema

$$\begin{cases} 2 \times \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) \times (-1) = 0 \\ 2 \times \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) \times (-x_{2i}) = 0. \end{cases}$$

troviamo quindi tutte le coppie di parametri  $(\hat{\alpha}_1, \hat{\alpha}_2)$  che identificano una retta ai minimi quadrati.

Per trovare le soluzioni del suddetto sistema di equazioni, conviene innanzitutto dividere per (-2) entrambe le equazioni. In questo modo otteniamo il sistema

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) \times x_{2i} = 0. \end{cases}\tag{3}$$

e spezzando la sommatoria nella prima equazione, vediamo che questa equazione può essere riscritta come

$$\begin{aligned}\hat{\alpha}_1 \times n + \hat{\alpha}_2 \times \sum_{i=1}^n x_{2i} &= \sum_{i=1}^n x_{1i} \quad \Rightarrow \\ \Rightarrow \hat{\alpha}_1 + \hat{\alpha}_2 \times \frac{1}{n} \sum_{i=1}^n x_{2i} &= \frac{1}{n} \sum_{i=1}^n x_{1i} \quad \Rightarrow\end{aligned}$$

$$\Rightarrow \hat{\alpha}_1 + \hat{\alpha}_2 \times \bar{x}_2 = \bar{x}_1. \quad (4)$$

Da quest'ultima equazione deduciamo che in corrispondenza di  $x_2 = \bar{x}_2$  una retta ai minimi quadrati deve necessariamente assumere il valore  $\bar{x}_1$ .

Ogni retta ai minimi quadrati passa per il punto con ascissa  $x_2 = \bar{x}_2$  e ordinata  $x_1 = \bar{x}_1$ .

Isolando il parametro  $\hat{\alpha}_1$  al primo membro della (4) vediamo ora che l'intercetta di una retta ai minimi quadrati deve essere data da:

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2. \quad (5)$$

A questo punto possiamo eliminare l'incognita  $\hat{\alpha}_1$  dalla seconda equazione del sistema (3): sostituendo l'espressione al secondo membro della (5) nella seconda equazione del sistema (3) otteniamo l'equazione

$$\begin{aligned} & \sum_{i=1}^n [x_{1i} - (\bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2) - \hat{\alpha}_2 x_{2i}] \times x_{2i} = 0 \quad \Rightarrow \\ \Rightarrow & \sum_{i=1}^n [(x_{1i} - \bar{x}_1) - \hat{\alpha}_2 \times (x_{2i} - \bar{x}_2)] \times x_{2i} = 0 \quad \Rightarrow \\ \Rightarrow & \sum_{i=1}^n [(x_{1i} - \bar{x}_1) \times x_{2i} - \hat{\alpha}_2 \times (x_{2i} - \bar{x}_2) \times x_{2i}] = 0 \quad \Rightarrow \\ \Rightarrow & \hat{\alpha}_2 \times \sum_{i=1}^n (x_{2i} - \bar{x}_2) \times x_{2i} = \sum_{i=1}^n (x_{1i} - \bar{x}_1) \times x_{2i}. \end{aligned}$$

Ricordando che la sommatoria al primo membro dell'ultima equazione è uguale a  $dev(X_2)$  (Sezione 3.1, formula alternativa per

la devianza) e che la sommatoria al secondo membro è uguale a  $codev(X_1, X_2)$  (Sezione 3.1, formula alternativa per la codevianza), vediamo che quest'ultima equazione può essere riscritta come

$$\hat{\alpha}_2 \times dev(X_2) = codev(X_1, X_2). \quad (6)$$

A questo punto distinguiamo due casi:

- 1) Il caso in cui  $dev(X_2) > 0$ . In questo caso l'equazione (6) ha un'unica soluzione che è data da

$$\hat{\alpha}_2 = \frac{codev(X_1, X_2)}{dev(X_2)} \quad (7)$$

e quindi ci sarà quindi un'unica retta ai minimi quadrati i cui parametri sono dati dalla (7) e dalla (5).

Se  $dev(X_2) > 0$ , allora esiste un'unica retta ai minimi quadrati i cui parametri sono dati da

$$\hat{\alpha}_2 = \frac{codev(X_1, X_2)}{dev(X_2)}$$

e da

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2.$$

- 2) Il caso in cui  $dev(X_2) = 0$ . In questo caso la variabile esplicativa  $X_2$  non presenta variabilità: i valori  $x_{2i}$  sono tutti uguali tra di loro e quindi uguali alla loro media  $\bar{x}_2$ . Ne consegue che sarà nulla anche la codevianza  $codev(X_1, X_2)$

perché

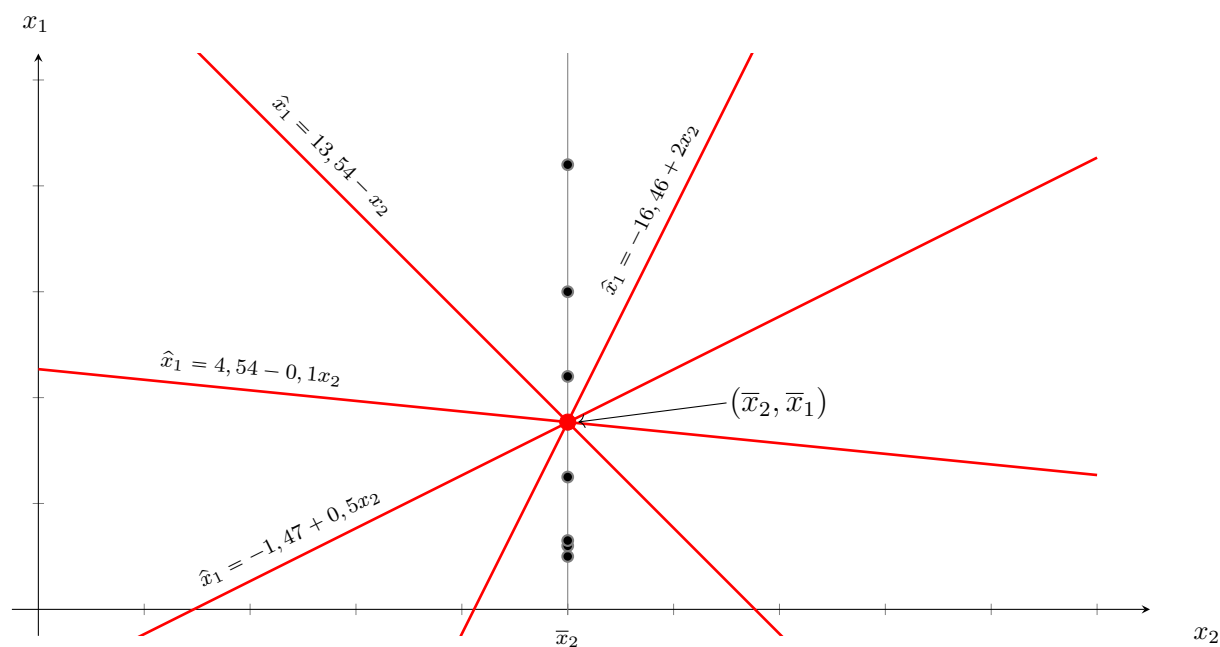
$$\begin{aligned} \text{codev}(X_1, X_2) &= \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\ &= \sum_{i=1}^n (x_{1i} - \bar{x}_1) \times 0 \\ &= 0. \end{aligned}$$

L'equazione (6) sarà dunque soddisfatta per qualsiasi valore di  $\alpha_2$  e quindi esisteranno infinite rette ai minimi quadrati. Per ottenere i parametri di una di queste rette basta scegliere in modo arbitrario un valore di  $\hat{\alpha}_2$  e sostituirlo nella (5) onde ottenere il corrispondente valore di  $\hat{\alpha}_1$ .

Se  $\text{dev}(X_2) = 0$ , allora qualsiasi retta che passa per il punto  $(\bar{x}_2, \bar{x}_1)$  sarà una retta ai minimi quadrati e in questo caso ci saranno dunque infinite rette ai minimi quadrati.

Per comprendere meglio il caso in cui  $\text{dev}(X_2) = 0$  conviene analizzare la situazione anche dal punto di vista grafico:

Figura 3.1:



Come si vede nella Figura 3.1, se  $dev(X_2) = 0$ , allora i punti  $x_{2i}$  sono tutti uguali tra di loro e quindi anche uguali alla loro media  $\bar{x}_2$ . Questo significa che tutti i punti osservati  $(x_{2i}, x_{1i})$  (si ricordi che la variabile dipendente  $X_1$  viene riportata in ordinata) giacciono sulla retta verticale che interseca l'asse delle ascisse in corrispondenza di  $x_2 = \bar{x}_2$ . In questo caso la distanza quadratica  $D(\alpha_1, \alpha_2)$  tra una generica retta  $x_1 = \alpha_1 + \alpha_2 \times x_2$  e la nuvola di punti dipende solo dal punto in cui la retta  $x_1 = \alpha_1 + \alpha_2 \times x_2$  interseca la retta verticale che contiene la nuvola di punti e la funzione  $D(\alpha_1, \alpha_2)$  raggiunge il suo minimo se e solo se il punto di intersezione è dato dal punto  $(\bar{x}_2, \bar{x}_1)$ . Ne consegue che tutte le rette (tranne quella verticale che non può essere espressa attraverso un'equazione del tipo  $x_1 = \alpha_1 + \alpha_2 \times x_2$ ) che passano per il

punto  $(\bar{x}_2, \bar{x}_1)$  minimizzano la distanza quadratica  $D(\alpha_1, \alpha_2)$  e per questo motivo esistono infinite rette ai minimi quadrati.

Ovviamente, nel caso in cui tutte le osservazioni di una variabile sono identiche si potrà dire ben poco sulla loro relazione di fondo con le osservazioni di un'altra variabile (almeno dal punto di vista statistico). Di solito nelle applicazioni osservazioni identiche si verificano per uno dei seguenti motivi:

- a) motivi accidentali (problemi di rilevazione, problemi di impostazione di un esperimento, ecc.)
- b) il processo che genera i dati genera *quasi sempre* lo stesso valore,
- c) il processo che genera i dati genera *sempre* lo stesso valore.

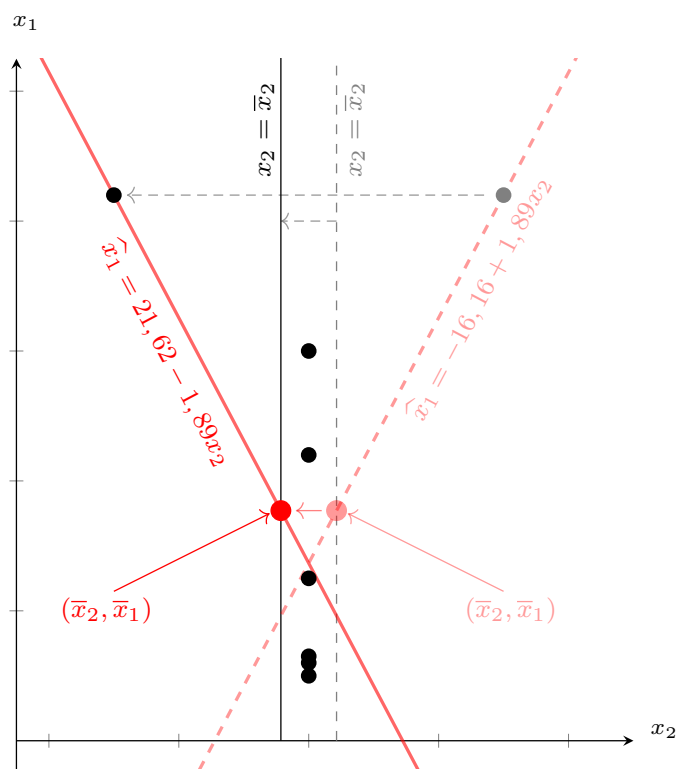
Nel primo caso il problema dei valori identici può essere risolto rilevando nuove osservazioni. Anche nel caso b) il problema può essere risolto facendo nuove rilevazioni, magari in modo più mirato per massimizzare la probabilità di ottenere valori diversi. Nel caso c), ovviamente, non c'è nulla da fare. Ovviamente, esistono anche delle applicazioni dove non è semplice distinguere tra i casi b) e c).

In ogni caso, siccome dal punto di vista statistico non ha senso studiare la relazione di fondo tra le osservazioni di due variabili quando le osservazioni di almeno una di esse sono tutte identiche, **d'ora in poi, salvo esplicita indicazione contraria, assumeremo sempre che le devianze delle variabili  $X_1$  e  $X_2$  siano entrambe strettamente positive.**



Tuttavia, nelle applicazioni può succedere che il valore di  $dev(X_2)$  sia molto prossimo a zero e in questo caso ci troveremo di fronte ad una nuvola di punti molto simile a quella rappresentata in Figura 3.1 anche se la retta ai minimi quadrati sarà unica. Come si vede nel grafico in Figura 3.2, in una tale situazione piccoli errori di rilevazione (o approssimazioni troppo grossolane) nei valori  $x_{2i}$  potrebbero causare grandi variazioni nei valori dei parametri che identificano una retta ai minimi quadrati.

Figura 3.2:



Osservando il grafico notiamo che il problema si verifica quando a causa di un errore di rilevazione (o di un'approssimazione troppo grossolana) un punto salta da un lato all'altro della retta verticale corrispondente all'ascissa  $x_2 = \bar{x}_2$ , trascinando con se la retta  $x_2 = \bar{x}_2$  e facendo in questo modo passare dall'altra parte della retta molti altri punti. Per valutare se è verosimile che gli errori di rilevazione (o le approssimazioni) possano avere causato dei salti di questo tipo bisogna assicurarsi che l'entità degli errori di rilevazione sia trascurabile rispetto allo scarto quadratico medio delle osservazioni della variabile  $X_2$ . In caso contrario l'andamento della retta ai minimi quadrati potrebbe non rappresentare fedelmente la relazione di fondo tra i veri valori delle variabili.

**Esempio 3.1** (Compensi e utili). Il foglio di calcolo in Figura 2.3 contiene l'utile netto dell'anno 2019 (variabile  $X_1$ ) e i compensi

dei dirigenti (variabile  $X_2$ ) delle  $n = 10$  più grandi società del settore delle comunicazioni incluse nell'indice di borsa S&P500. Per calcolare i parametri della retta ai minimi quadrati che descrive l'utile netto in funzione dei compensi dei dirigenti conviene copiare i dati relativi alle due variabili  $X_1$  e  $X_2$  in un nuovo foglio di calcolo e costruire la seguente tabella:

Figura 3.3: Foglio "Retta ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni".

	A	B	C	D	E	F	G
1							
2	TICKER	NOME SOCIETÀ	UTILE NETTO DELL'ANNO 2019 (in miliardi di USD)	COMPENSO DIRIGENTI ANNO 2019 (in milioni di USD)			
3			Variabile X1	Variabile X2	X1^2	X2^2	X1 * X2
4	GOOG	Alphabet Inc. Class C	34,343	281,946	1179,442	79493,810	9682,887
5	FB	Facebook, Inc. Class A	18,485	112,844	341,695	12733,741	2085,919
6	DIS	Walt Disney Company	11,054	92,795	122,191	8610,996	1025,761
7	VZ	Verizon Communications Inc.	19,913	51,095	396,528	2610,696	1017,475
8	CMCSA	Comcast Corporation Class A	13,057	142,706	170,485	20364,993	1863,312
9	NFLX	Netflix, Inc.	1,867	126,379	3,486	15971,586	235,938
10	T	AT&T Inc.	14,975	80,583	224,251	6493,561	1206,725
11	TMUS	T-Mobile US, Inc.	3,468	85,494	12,027	7309,200	296,493
12	CHTR	Charter Communications, Inc. Class A	1,668	21,511	2,782	462,732	35,881
13	ATVI	Activision Blizzard, Inc.	1,736	54,417	3,014	2961,196	94,468
14							
15		<b>SOMME</b>	<b>120,566</b>	<b>1049,77</b>	<b>2455,900</b>	<b>157012,511</b>	<b>17544,859</b>

Nell'ultima riga della tabella troviamo i valori delle somme

$$\sum_{i=1}^n x_{1i} = 120,566, \quad \sum_{i=1}^n x_{2i} = 1049,77,$$

$$\sum_{i=1}^n x_{2i}^2 = 157012,511, \quad \sum_{i=1}^n x_{1i} \times x_{2i} = 17544,859.$$

Usando queste somme otteniamo

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} = \frac{120,566}{10} = 12,0566$$

$$\bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2i} = \frac{1049,77}{10} = 104,977$$

$$\begin{aligned} dev(X_2) &= [[ \text{formula indiretta, vedi Sezione 3.1} ]] \\ &= \sum_{i=1}^n x_{2i}^2 - n \times \bar{x}_2^2 \\ &= 157012,511 - 10 \times 104,977^2 \\ &= 46810,806 \end{aligned}$$

$$\begin{aligned} codev(X_1, X_2) &= [[ \text{formula indiretta, vedi Sezione 3.1} ]] \\ &= \sum_{i=1}^n x_{1i} \times x_{2i} - n \times \bar{x}_1 \times \bar{x}_2 \\ &= 17544,859 - 10 \times 12,0566 \times 104,977 \\ &= 4888,202 \end{aligned}$$

e sostituendo queste quantità nella (7) e nella (5) otteniamo infine anche i parametri della retta ai minimi quadrati:

$$\hat{\alpha}_2 = \frac{codev(X_1, X_2)}{dev(X_2)} = \frac{4888,202}{46810,806} = 0,104$$

e

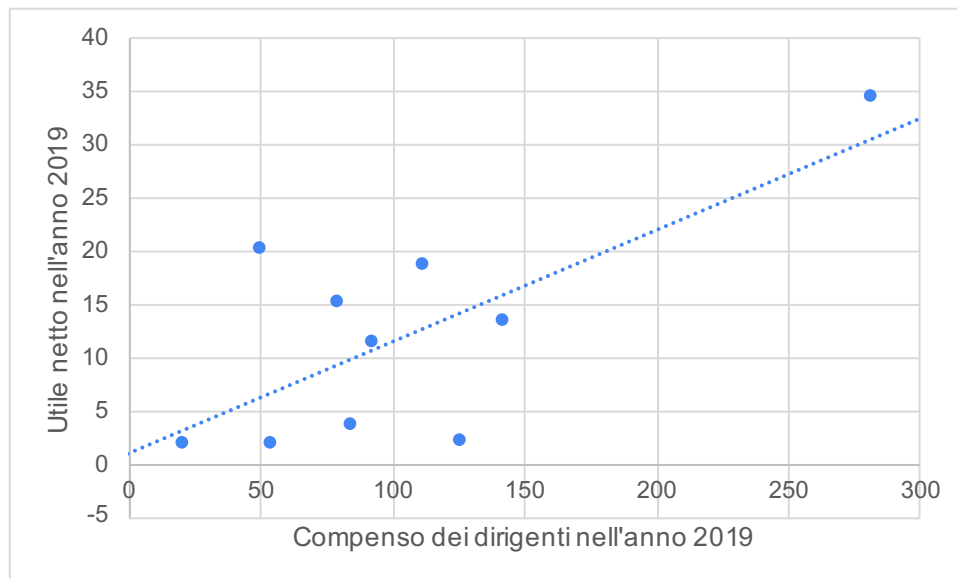
$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2 = 12,0566 - 0,104 \times 104,977 = 1,139.$$

L'equazione della retta ai minimi quadrati è dunque data da

$$\hat{x}_1 = 1,139 + 0,104 \times x_2.$$

Il grafico sottostante mostra la retta ai minimi quadrati sovrapposta alla nuvola di punti.

Figura 3.4:



Ora, per valutare la bontà d'adattamento di questa retta costruiremo innanzitutto il grafico dei residui. A tal fine costruiamo la tabella in Figura 3.3 che contiene

- nella colonna G i valori riprodotti

$$\hat{x}_{1i} = \hat{\alpha}_1 + \hat{\alpha}_2 x_{2i} = 1,139 + 0,104 \times x_{2i};$$

- nella colonna H i residui  $z_i = x_{1i} - \hat{x}_{1i}$ ;
- e nella colonna I i quadrati dei residui  $z_i^2$ .

Figura 3.5: Foglio "Retta ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni".

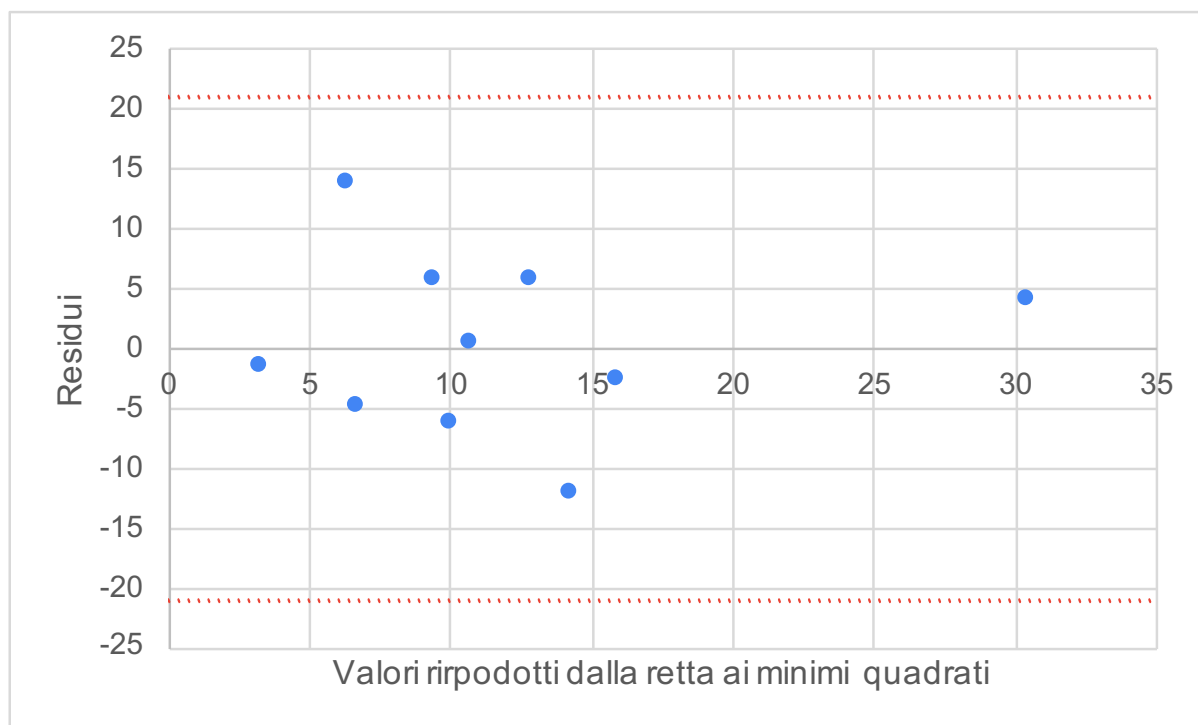
	D	E	F	G	H	I
22	<b>Valori riprodotti e residui</b>					
23		Variabile X1	Variabile X2	X1cap	Z	Z^2
24	GOOG	34,343	281,946	30,537	3,806	14,489
25	FB	18,485	112,844	12,878	5,607	31,437
26	DIS	11,054	92,795	10,785	0,269	0,073
27	VZ	19,913	51,095	6,430	13,483	181,791
28	CMCSA	13,057	142,706	15,996	-2,939	8,640
29	NFLX	1,867	126,379	14,291	-12,424	154,368
30	T	14,975	80,583	9,509	5,466	29,874
31	TMUS	3,468	85,494	10,022	-6,554	42,956
32	CHTR	1,668	21,511	3,341	-1,673	2,798
33	ATVI	1,736	54,417	6,777	-5,041	25,411
34						
35	<b>Somme</b>	<b>120,566</b>	<b>1049,77</b>	<b>120,566</b>	<b>0,000</b>	<b>491,838</b>

Poi riportiamo i punti  $(\hat{x}_{1i}, z_i)$  in un piano cartesiano e tracciamo due linee orizzontali in corrispondenza di  $\pm 3 \times A_2$  dove

$$A_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2} = \sqrt{\frac{491,838}{10}} = 7,013$$

(il valore di  $\sum_{i=1}^n z_i^2 = 491,838$  è il totale della colonna I). In questo modo otteniamo il grafico dei residui:

Figura 3.6:



Come si vede da questo grafico, la successione dei segni dei residui è casuale e tutti i residui rientrano nella fascia delimitata da  $\pm 3 \times A_2$ . Il grafico dei residui conferma quindi che l'andamento della retta ai minimi quadrati non sia troppo rigido (questo lo si poteva dedurre anche dal grafico in Figura 3.4) e non evidenzia la presenza di osservazioni anomale.

A questo punto procediamo anche al calcolo di un indice di adattamento relativo. Calcoleremo l'indice  $A_2^*$  definito nell'equazione (1). Notiamo che questo indice può essere espresso come

$$A_2^* = \sqrt{\frac{\sum_{i=1}^n z_i^2}{dev(X_1)}}$$

e quindi calcoliamo innanzitutto

$$\begin{aligned} dev(X_1) &= \sum_{i=1}^n x_{1i}^2 - n \times \bar{x}_1 \\ &= 2455,900 - 10 \times 12,0566^2 = 1002,284 \end{aligned}$$

(il valore di  $\sum_{i=1}^n x_{1i}^2 = 2455,900$  è il totale della colonna E nella tabella in Figura 3.3). Usando questo valore e la somma dei quadrati dei residui (riportata in fondo alla colonna I della tabella in Figura 3.5) otteniamo dunque

$$A_2^* = \sqrt{\frac{\sum_{i=1}^n z_i^2}{dev(X_1)}} = \sqrt{\frac{491,838}{1002,284}} = 0,701.$$

Questo valore ci dice che l'ordine di grandezza medio dei residui è pari al 70,1% dello scarto quadratico medio delle osservazioni dell'utile netto. Trattandosi di una percentuale abbastanza elevata dobbiamo dunque concludere che, in termini relativi, le osservazioni tendono a discostarsi abbastanza nettamente dalla retta ai minimi quadrati.

A questo punto possiamo procedere con l'interpretazione dell'andamento della retta ai minimi quadrati. Siccome l'andamento è determinato dal suo coefficiente angolare e dalla sua intercetta, basterà commentare i valori di questi due parametri. Il valore del coefficiente angolare  $\hat{\alpha}_2 = 0,104$  ci dice che se l'ammontare del compenso dei dirigenti (variabile  $X_2$ ) aumenta di un milione di USD, il valore dell'utile netto riprodotto dalla retta ai minimi quadrati aumenta di 0,104 miliardi di USD (ovvero di 104 milioni di USD). D'altra parte, il valore di  $\hat{\alpha}_1 = 1,139$  ci dice che l'utile

netto riprodotto dalla retta ai minimi quadrati sia pari a 1,139 miliardi di euro se il compenso dei dirigenti è nullo. Se i dati includessero delle società per le quali il compenso dei dirigenti (variabile  $X_2$ ) è prossimo a zero, questo valore potrebbe essere considerato come rappresentativo per l'utile netto di tali società.  $\triangle$

**Esempio 3.2** (Indice PMI e tasso di crescita del PIL). Per calcolare i parametri della retta ai minimi quadrati che descrive il tasso di crescita del PIL (variabile  $X_1$ ) in funzione del valore dell'indice PMI del primo mese di ciascun trimestre (variabile  $X_2$ ; vedi Figura 2.6) ci conviene innanzitutto calcolare le somme

$$\sum_{i=1}^{76} x_{1i} = 152,6, \quad \sum_{i=1}^{76} x_{2i} = 3989,5,$$

$$\sum_{i=1}^{76} x_{2i}^2 = 211439,93, \quad \sum_{i=1}^{76} x_{1i} \times x_{2i} = 8507,4.$$

Queste somme possono essere comodamente calcolate usando il [foglio di calcolo](#) in Figura 2.2 (basta procedere come nell'Esempio 3.1). A questo punto possiamo calcolare anche

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} = \frac{152,6}{76} = 2,01$$

$$\bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2i} = \frac{3989,5}{76} = 52,49$$

$$\begin{aligned} dev(X_2) &= \sum_{i=1}^n x_{2i}^2 - n \times \bar{x}_2^2 \\ &= 211439,93 - 76 \times 52,49^2 = 2044,7224 \end{aligned}$$



$$\begin{aligned} \text{codev}(X_1, X_2) &= \sum_{i=1}^n x_{1i} \times x_{2i} - n \times \bar{x}_1 \times \bar{x}_2 \\ &= 8507,4 - 76 \times 2,01 \times 52,49 = 489,0276 \end{aligned}$$

e sostituendo questi risultati nelle formule (7) e (5) otteniamo i valori dei parametri:

$$\hat{\alpha}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_2)} = \frac{489,0276}{2044,7224} = 0,239$$

e

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2 = 2,01 - 0,239 \times 52,49 = -10,535.$$

Come anticipato nella Sezione 2.3, l'equazione della retta ai minimi quadrati che descrive il tasso di crescita del PIL in funzione dell'indice PMI è dunque data da

$$\hat{x}_1 = -10,535 + 0,239x_2.$$

Il grafico della retta ai minimi quadrati sovrapposto alla nuvola di punti lo abbiamo già visto nella Figura 2.6. Il grafico dei residui lo abbiamo invece visto nella Figura 2.10. Come abbiamo già osservato nella Sezione 2.3, la successione dei segni dei residui è casuale e quindi possiamo concludere che l'andamento della retta ai minimi quadrati non sia troppo rigido (questo lo si poteva anche dedurre dal grafico in Figura 2.6). Per quanto concerne la verifica sull'esistenza di valori anomali notiamo che uno dei residui si trova leggermente al di sotto della retta orizzontale che interseca l'asse delle ordinate in corrispondenza di  $-3 \times A_2$ . Questo residuo è quindi sospetto. Ma visto che è molto prossimo alla fascia dei

residui non sospetti, non procederemo all'eliminazione della corrispondente coppia di osservazioni  $(x_{2i}, x_{1i})$ . Infine, per completare la fase di valutazione della bontà d'adattamento calcoliamo anche l'indice di adattamento relativo  $A_2^*$  definito nell'equazione (1). Usando il [foglio di calcolo](#) possiamo verificare che (per calcolare i residui bisogna calcolare prima i valori riprodotti  $\hat{x}_{1i}$ ; più avanti vedremo una formula che nel caso di rette ai minimi quadrati ci permette di calcolare la somma dei quadrati dei residui senza calcolare i valori riprodotti)

$$\sum_{i=1}^n x_{1i}^2 = 691,54 \quad \text{e che} \quad \sum_{i=1}^n z_i^2 = 262,853.$$

$$\Rightarrow \text{dev}(X_1) = \sum_{i=1}^n x_{1i}^2 - n \times \bar{x}_1^2$$

$$= 691,54 - 76 \times 2,01^2 = 384,492.$$

Dalla formula (1) otteniamo quindi

$$A_2^* = \sqrt{\frac{262,853}{384,492}} = 0,827.$$

Il valore di  $A_2^*$  ci dice che l'ordine di grandezza medio dei residui è pari all'82,7% dello scarto quadratico medio delle osservazioni del tasso di crescita del PIL. Questa percentuale è abbastanza elevata e ci segnala che, in termini relativi, la dispersione attorno alla funzione interpolante è molto elevata.

A questo punto possiamo procedere con l'interpretazione dell'andamento della retta ai minimi quadrati, ovvero con l'interpretazione dei suoi parametri. Il valore di  $\hat{\alpha}_2 = 0,239$  ci dice il valore

del tasso di crescita del PIL riprodotto dalla retta ai minimi quadrati cresce di 0,2463 punti percentuali se il valore dell'indice PMI aumenta di un punto. D'altra parte,  $\hat{\alpha}_1 = -10,535$  è il valore del tasso di crescita del PIL riprodotto dalla retta ai minimi quadrati quando il valore dell'indice PMI è nullo.  $\triangle$

### 3.3 Proprietà dei residui della retta ai minimi quadrati

Come accennato in precedenza, la bontà d'adattamento di una funzione interpolante è strettamente legata ai residui di interpolazione. I residui di rette ai minimi quadrati godono di alcune proprietà importanti. Per dimostrarle conviene tornare al sistema di equazioni nella (3) a partire dal quale abbiamo ricavato le formule per calcolare i parametri di una retta ai minimi quadrati. Come abbiamo visto, i parametri  $\hat{\alpha}_1$  e  $\hat{\alpha}_2$  di una retta ai minimi quadrati sono soluzioni del seguente sistema di equazioni:

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i}) \times x_{2i} = 0. \end{cases}$$

Siccome

$$\hat{x}_{1i} = \hat{\alpha}_1 + \hat{\alpha}_2 x_{2i}$$

è la formula per calcolare i valori riprodotti dalla retta ai minimi quadrati, il precedente sistema di equazioni può essere riscritto come

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{x}_{1i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{x}_{1i}) \times x_{2i} = 0. \end{cases} \quad (8)$$

Quest'ultimo sistema viene chiamato **sistema normale**. A partire da esso si ricavano le **tre proprietà dei residui**

$$z_i = x_{1i} - \hat{x}_{1i}.$$

Per esprimere le proprietà dei residui in forma compatta conviene introdurre due variabili "artificiali"  $\hat{X}_1$  e  $Z$  che rappresentano, rispettivamente, i valore riprodotti  $\hat{x}_{1i}$  e i residui  $z_i$ .

**Prima proprietà dei residui:** La somma dei residui di una retta ai minimi quadrati è sempre nulla:

$$\sum_{i=1}^n z_i = 0$$

**Dimostrazione.** La dimostrazione segue immediatamente dalla prima equazione nella (8). □

**Corollari della prima proprietà dei residui:**

- a) La media aritmetica dei residui è nulla:  $\bar{z} = 0$
- b) La media aritmetica dei valori riprodotti  $\hat{x}_{1i}$  coincide con la media aritmetica delle osservazioni della variabile dipendente:  $M_1(\hat{X}_1) = \bar{x}_1$ .

**Dimostrazione.** Per dimostrare il corollario a) è sufficiente osservare che la media aritmetica dei residui è il rapporto tra la loro somma (che è nulla) e il numero  $n$  di residui.

Per dimostrare il corollario b) osserviamo innanzitutto che

$$Z = X_1 - \hat{X}_1 \quad \Rightarrow \quad \hat{X}_1 = X_1 - Z.$$

Ne consegue che (vedi Sezione 3.1)

$$M_1(\hat{X}_1) = M_1(X_1 - Z) = \bar{x}_1 - \bar{z} = \bar{x}_1 - 0 = \bar{x}_1.$$

□

**Seconda proprietà dei residui:**  $codev(Z, X_2) = 0$ .

**Dimostrazione.** Usando la formula indiretta per il calcolo della codevarianza (si tenga presente che per la prima proprietà dei residui si ha  $\bar{z} = 0$ ) vediamo che

$$\begin{aligned} codev(Z, X_2) &= \sum_{i=1}^n z_i \times x_{2i} - n \times \bar{z} \times \bar{x}_2 \\ &= \sum_{i=1}^n (x_{1i} - \hat{x}_{1i})x_{2i} - 0 = 0 \end{aligned}$$

L'ultima uguaglianza segue dalla seconda equazione del sistema normale nella (8). □

**Terza proprietà dei residui:**  $codev(Z, \hat{X}_1) = 0$ .

**Dimostrazione.** Tenendo presente che  $\hat{X}_1 = \hat{\alpha}_1 + \hat{\alpha}_2 \times X_2$  e usando la proprietà della codevarianza che riguarda le trasformazioni lineari affini (vedi Sezione 3.1) otteniamo

$$codev(Z, \hat{X}_1) = codev(Z, \hat{\alpha}_1 + \hat{\alpha}_2 \times X_2) = \hat{\alpha}_2 \times codev(Z, X_2) = 0.$$

L'ultima uguaglianza segue dalla seconda proprietà dei residui che abbiamo appena dimostrato. □

**Esempio 3.3** (Verifica numerica delle proprietà dei residui della retta ai minimi quadrati). Per verificare numericamente che i residui di una retta ai minimi quadrati devono soddisfare le tre proprietà che abbiamo visto in questa sezione, faremo riferimento alla retta ai minimi quadrati che abbiamo ricavato nell'Esempio 3.1. Per effettuare la verifica integriamo la tabella in Figura 3.5 aggiungendo

- la colonna J con i prodotti  $z_i \times x_{2i}$
- e la colonna K con i prodotti  $z_i \times \hat{x}_{1i}$ .

La tabella in Figura 3.7 mostra le due nuove colonne.

Figura 3.7: Foglio "Retta ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni"

	D	E	F	G	H	I	J	K
22	<b>Valori riprodotti e residui</b>							
23		Variabile X1	Variabile X2	X1cap	Z	Z^2	Z*X2	Z*X1cap
24	GOOG	34,343	281,946	30,537	3,806	14,489	1073,226	116,237
25	FB	18,485	112,844	12,878	5,607	31,437	632,704	72,206
26	DIS	11,054	92,795	10,785	0,269	0,073	25,008	2,906
27	VZ	19,913	51,095	6,430	13,483	181,791	688,914	86,696
28	CMCSA	13,057	142,706	15,996	-2,939	8,640	-419,475	-47,020
29	NFLX	1,867	126,379	14,291	-12,424	154,368	-1570,195	-177,565
30	T	14,975	80,583	9,509	5,466	29,874	440,445	51,975
31	TMUS	3,468	85,494	10,022	-6,554	42,956	-560,336	-65,686
32	CHTR	1,668	21,511	3,341	-1,673	2,798	-35,982	-5,588
33	ATVI	1,736	54,417	6,777	-5,041	25,411	-274,310	-34,162
34								
35	<b>Somme</b>	<b>120,566</b>	<b>1049,77</b>	<b>120,566</b>	<b>0,000</b>	<b>491,838</b>	<b>0,000</b>	<b>0,000</b>

Questa tabella mostra che i valori delle tre somme

$$\sum_{i=1}^n z_i, \quad \sum_{i=1}^n z_i \times x_{2i}, \quad \sum_{i=1}^n z_i \times \hat{x}_{1i}$$

che si trovano in fondo alle colonne H, J e K sono tutti nulli esattamente come previsto dalle tre proprietà che i residui di tutte le rette ai minimi quadrati devono soddisfare.  $\triangle$

**Esercizio 3.1.** Si costruisca una tabella simile a quella in Figura 3.7 per verificare che anche i residui della retta ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione del primo indice PMI di ciascuno trimestre (vedi Esempio 3.2) soddisfano le tre proprietà che i residui di tutte le rette ai minimi quadrati devono soddisfare (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale").  $\triangle$

### 3.4 Scomposizione della devianza

Usando le tre proprietà dei residui delle rette ai minimi quadrati si dimostra facilmente la cosiddetta **scomposizione della devianza**, seconda la quale

$$\begin{aligned} \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 &= \sum_{i=1}^n (\hat{x}_{1i} - \bar{x}_1)^2 + \sum_{i=1}^n (x_{1i} - \hat{x}_{1i})^2 \\ &= \sum_{i=1}^n (\hat{x}_{1i} - \bar{x}_1)^2 + \sum_{i=1}^n z_i^2. \end{aligned}$$

Tenendo presente che  $M_1(\hat{X}_1) = \bar{x}_1$  e che  $\bar{z} = 0$  (prima proprietà dei residui), si vede che la scomposizione della devianza può essere espressa come

**Scomposizione della devianza:**

$$dev(X_1) = dev(\hat{X}_1) + dev(Z).$$

In parole: la **devianza totale**  $dev(X_1)$  è uguale alla somma tra la **devianza spiegata**  $dev(\hat{X}_1)$  e la **devianza residua**  $dev(Z)$ .

Dividendo per  $n$  si ottiene l'analogo risultato i termini di varianze:

$$var(X_1) = var(\hat{X}_1) + var(Z).$$

**Dimostrazione** (scomposizione della devianza). Siccome

$$Z = X_1 - \hat{X}_1 \quad \Rightarrow \quad X_1 = \hat{X}_1 + Z,$$

possiamo concludere che (vedi Sezione 3.1)

$$dev(X_1) = dev(\hat{X}_1) + dev(Z) + 2 \times codev(\hat{X}_1, Z).$$

Tenendo presente che  $codev(\hat{X}_1, Z) = 0$  (terza proprietà dei residui), si vede dunque che

$$dev(X_1) = dev(\hat{X}_1) + dev(Z)$$

come volevasi dimostrare. □

**Esempio 3.4** (Verifica numerica della scomposizione della devianza). Per effettuare una verifica numerica della scomposizione della devianza, considereremo ancora la retta ai minimi quadrati che abbiamo ricavato nell'Esempio 3.1. Per eseguire la verifica aggiungiamo alla tabella in Figura 3.7 anche la colonna L con i quadrati dei valori riprodotti dalla retta ai minimi quadrati:



Figura 3.8: Foglio "Retta ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni"

	D	E	F	G	H	I	J	K	L
22	<b>Valori riprodotti e residui</b>								
23		Variabile X1	Variabile X2	X1cap	Z	Z^2	Z*X2	Z*X1cap	X1cap^2
24	GOOG	34,343	281,946	30,537	3,806	14,489	1073,226	116,237	932,478
25	FB	18,485	112,844	12,878	5,607	31,437	632,704	72,206	165,846
26	DIS	11,054	92,795	10,785	0,269	0,073	25,008	2,906	116,305
27	VZ	19,913	51,095	6,430	13,483	181,791	688,914	86,696	41,345
28	CMCSA	13,057	142,706	15,996	-2,939	8,640	-419,475	-47,020	255,886
29	NFLX	1,867	126,379	14,291	-12,424	154,368	-1570,195	-177,565	204,247
30	T	14,975	80,583	9,509	5,466	29,874	440,445	51,975	90,426
31	TMUS	3,468	85,494	10,022	-6,554	42,956	-560,336	-65,686	100,442
32	CHTR	1,668	21,511	3,341	-1,673	2,798	-35,982	-5,588	11,160
33	ATVI	1,736	54,417	6,777	-5,041	25,411	-274,310	-34,162	45,926
34									
35	<b>Somme</b>	<b>120,566</b>	<b>1049,77</b>	<b>120,566</b>	<b>0,000</b>	<b>491,838</b>	<b>0,000</b>	<b>0,000</b>	<b>1964,062</b>

Usando il totale della colonna L otteniamo

$$\begin{aligned}
 dev(\hat{X}_1) &= \sum_{i=1}^n \hat{x}_{1i}^2 - n \times [M_1(\hat{X}_1)]^2 \\
 &= \sum_{i=1}^n \hat{x}_{1i}^2 - n \times \bar{x}_1^2 \\
 &= 1964,062 - 10 \times 12,0566^2 = 510,446.
 \end{aligned}$$

e ricordando i valori di

$$dev(X_1) = 1002,284$$

e

$$dev(Z) = \sum_{i=1}^n z_i^2 = 491,838$$

che abbiamo ricavato nel precedente Esempio 3.1, vediamo che

$$dev(\hat{X}_1) + dev(Z) = 510,446 + 491,838 = 1002,284 = dev(X_1)$$

esattamente come previsto dalla scomposizione della devianza.  $\triangle$

**Esercizio 3.2.** Si verifichi che la somma tra la devianza spiegata e la devianza residua della retta ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dell'indice PMI (vedi Esempio 3.2) coincide con la devianza totale (i dati sono contenuti nella cartella di lavoro "[Indice PMI e tasso di crescita del PIL reale](#)").  $\triangle$

### 3.5 Indice di determinazione

Consideriamo ancora la scomposizione della devianza

$$dev(X_1) = dev(\hat{X}_1) + dev(Z).$$

Notiamo che la devianza totale  $dev(X_1)$  al primo membro della scomposizione non dipende in alcun modo da come è posizionata la retta ai minimi quadrati, mentre i valori della devianza spiegata  $dev(\hat{X}_1)$  e della devianza residua  $dev(Z)$  variano a seconda della bontà d'adattamento della retta ai minimi quadrati. Infatti:

- Se i residui

$$z_i = x_{1i} - \hat{x}_{1i}$$

sono prossimi a zero, allora la devianza residua  $dev(Z)$  sarà piccola e (in virtù della scomposizione della devianza) la devianza spiegata  $dev(\hat{X}_1)$  dovrà quindi essere prossima alla devianza totale  $dev(X_1)$ . Nel caso estremo in cui tutti i residui  $z_i$  fossero nulli, avremmo

$$dev(Z) = 0 \quad \text{e} \quad dev(X_1) = dev(\hat{X}_1).$$

Ovviamente, questo caso si verifica se e solo se tutti i punti  $(x_{2i}, x_{1i})$  giacciono sulla retta ai minimi quadrati.

- Se d'altra parte i valori riprodotti  $\hat{x}_{1i}$  sono molto lontani dai valori osservati  $x_{1i}$ , allora i residui  $z_i$  saranno lontani da zero e quindi osserveremo una devianza residua  $dev(Z)$  elevata. Tuttavia, siccome  $dev(\hat{X}_1) \geq 0$ , la devianza residua  $dev(Z)$  non potrà mai superare la devianza totale  $dev(X_1)$  (altrimenti sarebbe violata l'uguaglianza prevista dalla scomposizione della devianza). Nel "peggiore" dei casi si potrà avere

$$dev(Z) = dev(X_1)$$

e in questo caso la devianza spiegata  $dev(\hat{X}_1)$  dovrà essere nulla. Ma questo accade se e solo se i valori interpolati  $\hat{x}_{1i}$  sono tutti uguali tra di loro e quindi uguali a  $\bar{x}_1$  (si ricordi che per il secondo corollario della prima proprietà dei residui la media dei valori riprodotti deve coincidere con la media delle osservazioni della variabile dipendente), ovvero se la retta ai minimi quadrati è orizzontale. In questo caso si dice che "la retta ai minimi quadrati non riesce a spiegare la variabilità della variabile dipendente  $X_1$ ".

Queste considerazioni sulla scomposizione della devianza suggeriscono che la bontà d'adattamento di una retta ai minimi quadrati potrebbe essere misurata attraverso l'indice

$$I^2 = \frac{dev(\hat{X}_1)}{dev(X_1)} = 1 - \frac{dev(Z)}{dev(X_1)}$$

che è noto come **indice di determinazione**. Infatti, dalla scomposizione della devianza discende che

$$0 \leq I^2 \leq 1$$

e le precedenti considerazioni sulla scomposizione della devianza implicano che

- 1)  $I^2 = 1$  se e solo se  $dev(Z) = 0$ , ovvero se e solo se tutti i residui sono nulli e la retta ai minimi quadrati passa per tutti i punti osservati.
- 2)  $I^2 = 0$  se e solo se  $dev(Z)$  raggiunge il suo massimo teorico, ovvero  $dev(X_1)$ , e  $dev(\hat{X}_1) = 0$ . In questo caso la retta ai minimi quadrati è orizzontale e quindi non riesce a descrivere la variabilità delle osservazioni  $X_1$ .
- 3)  $0 < I^2 < 1$  nei casi intermedi. In questo caso la bontà d'adattamento della retta ai minimi quadrati sarà tanto migliore quanto più elevato è il valore di  $I^2$ .

L'indice di determinazione è definito come

$$I^2 = \frac{\text{dev}(\hat{X}_1)}{\text{dev}(X_1)} = 1 - \frac{\text{dev}(Z)}{\text{dev}(X_1)}. \quad (9)$$

e fornisce una misura per la bontà d'adattamento di una retta ai minimi quadrati:

- $I^2 = 1$  se e solo se la retta ai minimi quadrati passa per tutti i punti osservati.
- $I^2 = 0$  se e solo se la retta ai minimi quadrati è orizzontale e non riesce a descrivere la variabilità delle osservazioni di  $X_1$ .
- $0 < I^2 < 1$  nei casi intermedi. L'indice di determinazione è tanto più prossimo a 1 quanto più la retta ai minimi quadrati passa vicino ai punti osservati.

**Osservazione 3.1.** Siccome il rapporto  $\text{dev}(Z)/\text{dev}(X_1)$  che compare nella seconda formula definitoria (9) è il quadrato dell'indice  $A_2^*$  definito nella (1), possiamo scrivere  $I^2 = 1 - [A_2^*]^2 \triangle$

**Esempio 3.5.** Usando i valori della devianza totale e della devianza spiegata (vedi Esempio 3.4) si vede che l'indice di determinazione della retta ai minimi quadrati che descrive gli utili in funzione dei compensi degli amministratori (vedi Esempio 3.1) è dato da

$$I^2 = \frac{\text{dev}(\hat{X}_1)}{\text{dev}(X_1)} = \frac{510,446}{1002,2084} = 0,509.$$

Per commentare il valore dell'indice di determinazione possiamo dire che la retta ai minimi quadrati spiega il 50,9% della variabilità degli utili.  $\triangle$

**Esercizio 3.3.** Si calcoli l'indice di determinazione della retta ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dell'indice PMI (vedi Esempio 3.2). I dati sono contenuti nella cartella di lavoro "[Indice PMI e tasso di crescita del PIL reale](#)".  $\triangle$

### 3.6 Formula indiretta per la devianza spiegata

Il calcolo della devianza spiegata attraverso la sua formula definitoria

$$dev(\hat{X}_1) = \sum_{i=1}^n (\hat{x}_{1i} - \bar{x}_1)^2$$

è molto oneroso perché richiede di calcolare tutti i valori riprodotti  $\hat{x}_{1i}$ . Per evitare il calcolo dei valori riprodotti si può ricorrere alla seguente formula indiretta:

**Formula indiretta per la devianza spiegata:**

$$dev(\hat{X}_1) = \hat{\alpha}_2 \times codev(X_1, X_2). \quad (10)$$

**Dimostrazione** (della formula indiretta per la devianza spiegata). Tenendo presente che

$$Z = X_1 - \hat{X}_1 \quad \Rightarrow \quad \hat{X}_1 = X_1 - Z$$

e usando le proprietà della codevianza (vedi Sezione 3.1) otteniamo

$$\begin{aligned} dev(\hat{X}_1) &= codev(\hat{X}_1, \hat{X}_1) \\ &= codev(\hat{X}_1, X_1 - Z) \\ &= codev(\hat{X}_1, X_1) - codev(\hat{X}_1, Z). \end{aligned}$$

Siccome  $codev(\hat{X}_1, Z) = 0$  (terza proprietà dei residui), possiamo dunque concludere che

$$dev(\hat{X}_1) = codev(\hat{X}_1, X_1). \quad (11)$$

Per ottenere la formula indiretta nella (10) ricordiamo ora che

$$\hat{X}_1 = \hat{\alpha}_1 + \hat{\alpha}_2 \times X_2$$

e che quindi (vedi le proprietà della codevianza in Sezione 3.1)

$$\begin{aligned} codev(\hat{X}_1, X_1) &= codev(\hat{\alpha}_1 + \hat{\alpha}_2 \times X_2, X_1) \\ &= codev(X_1, \hat{\alpha}_1 + \hat{\alpha}_2 \times X_2) \\ &= \hat{\alpha}_2 \times codev(X_1, X_2). \end{aligned}$$

Combinando questo risultato con l'uguaglianza nella (11) otteniamo infine la formula indiretta (10) che volevamo dimostrare.  $\square$

**Osservazione 3.2.** Si noti che nel corso della precedente dimostrazione abbiamo anche dimostrato il risultato intermedio

$$dev(\hat{X}_1) = codev(\hat{X}_1, X_1). \quad (12)$$

$\triangle$

**Esempio 3.6** (Verifica numerica della formula indiretta per il calcolo della devianza spiegata). Per effettuare una verifica numerica

della formula indiretta per il calcolo della devianza spiegata consideriamo ancora la retta ai minimi quadrati che descrive l'utile netto in funzione del compenso dei dirigenti (i dati sono contenuti nella cartella di lavoro "Settore GICS delle comunicazioni"). Nell'Esempio 3.1 abbiamo ricavato il valore della codevarianza

$$\text{codev}(X_1, X_2) = 4888,202$$

e quello del coefficiente angolare

$$\hat{\alpha}_2 = 0,104.$$

Applicando la formula indiretta per il calcolo della devianza spiegata (formula (10)) otteniamo dunque

$$\text{dev}(\hat{X}) = \hat{\alpha}_2 \times \text{codev}(X_1, X_2) = 0,104 \times 4888,202 = 508,373.$$

Questo valore differisce dal valore della devianza spiegata trovato nell'Esempio 3.4 solo a causa degli arrotondamenti.  $\triangle$

**Esercizio 3.4.** Si calcoli la devianza spiegata della retta ai minimi quadrati che descrive il tasso di crescita del PIL reale (variabile  $X_1$ ) in funzione del primo indice PMI del trimestre (variabile  $X_2$ ) utilizzando la formula (10). Si ricordi che nell'Esempio ?? abbiamo ottenuto

$$\text{codev}(X_1, X_2) = 489,0276 \quad \text{e} \quad \hat{\alpha}_2 = 0,239.$$

$\triangle$



### 3.7 Coefficiente di correlazione lineare

Sostituendo la formula indiretta per il calcolo della devianza spiegata nella definizione dell'indice di determinazione otteniamo

$$I^2 = \frac{\text{dev}(\hat{X}_1)}{\text{dev}(X_1)} = \frac{\hat{\alpha}_2 \times \text{codev}(X_1, X_2)}{\text{dev}(X_1)}$$

e ricordando che

$$\hat{\alpha}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_2)}$$

vediamo che

$$I^2 = \frac{[\text{codev}(X_1, X_2)]^2}{\text{dev}(X_1) \times \text{dev}(X_2)}.$$

Questa formula mostra che il valore dell'indice di determinazione rimane uguale se si scambiano i ruoli della variabile dipendente e di quella indipendente.

L'indice di determinazione può essere calcolato come

$$I^2 = \frac{[\text{codev}(X_1, X_2)]^2}{\text{dev}(X_1) \times \text{dev}(X_2)} \quad (13)$$

e rimane dunque invariato se si scambiano i ruoli della variabile dipendente e di quella indipendente.

Ricordando che

$$0 \leq I^2 = [[ \text{formula (13)} ]] = \frac{[\text{codev}(X_1, X_2)]^2}{\text{dev}(X_1) \times \text{dev}(X_2)} \leq 1$$

e che  $I^2 = 1$  se e solo se la retta ai minimi quadrati passa per tutti i punti  $(x_{2i}, x_{1i})$  (ovvero se e solo se i punti  $(x_{2i}, x_{1i})$  sono allineati lungo una retta), vediamo ora che la disuguaglianza di

Cauchy-Schwarz vale anche nel caso in cui le due devianze  $dev(X_1)$  e  $dev(X_2)$  sono entrambe positive:<sup>4</sup>

**Disuguaglianza di Cauchy-Schwarz:**

$$[codev(X_1, X_2)]^2 \leq dev(X_1) \times dev(X_2)$$

con uguaglianza se e solo se tutti i punti  $(x_{2i}, x_{1i})$  sono allineati lungo una retta. In questo caso si dice che le osservazioni di  $X_1$  e  $X_2$  sono "collineari".

La formula (13) per calcolare l'indice di determinazione suggerisce la definizione di un nuovo indice che descrive contemporaneamente sia il verso che l'intensità della relazione lineare tra le osservazioni di due variabili  $X_1$  e  $X_2$ . L'indice in questione è noto come **coefficiente di correlazione lineare** ed è definito come

$$r(X_1, X_2) = \frac{codev(X_1, X_2)}{\sqrt{dev(X_1) \times dev(X_2)}} = \frac{cov(X_1, X_2)}{\sqrt{var(X_1) \times var(X_2)}}.$$

---

<sup>4</sup>Si ricordi che nella Sezione 3.1 abbiamo già dimostrato la disuguaglianza di Cauchy-Schwarz per il caso in cui almeno una delle due devianze è nulla (vedi la dimostrazione della proprietà d) della codevianza)

Il **coefficiente di correlazione lineare** è definito come

$$\begin{aligned} r(X_1, X_2) &= \frac{\text{codev}(X_1, X_2)}{\sqrt{\text{dev}(X_1) \times \text{dev}(X_2)}} \\ &= \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1) \times \text{var}(X_2)}} \end{aligned} \quad (14)$$

e il suo quadrato coincide con il valore dell'indice di determinazione delle due rette ai minimi quadrati che si possono ottenere con le osservazioni delle variabili  $X_1$  e  $X_2$ .

Per l'interpretazione di un coefficiente di correlazione lineare osserviamo quanto segue:

- Il segno di  $r(X_1, X_2)$  indica il verso della relazione lineare tra le osservazioni di  $X_1$  e  $X_2$  in quanto coincide con il segno di  $\text{codev}(X_1, X_2)$  che a sua volta è uguale al segno del coefficiente angolare

$$\hat{\alpha}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_2)}$$

della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  e anche al segno del coefficiente angolare

$$\hat{\beta}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_1)}$$

della retta ai minimi quadrati che descrive  $X_2$  in funzione di  $X_1$ .

- Il valore assoluto  $|r(X_1, X_2)|$  è una misura per l'intensità della relazione lineare tra  $X_1$  e  $X_2$  in quanto  $|r(X_1, X_2)|^2 = I^2$ .

Se il valore di  $|r(X_1, X_2)|$  è prossimo a 1, allora dovrà essere prossimo a 1 anche il valore dell'indice di determinazione  $I^2$  e i punti  $(x_{2i}, x_{1i})$  saranno quindi molto prossimi alle due rette ai minimi quadrati che si possono ottenere con le coppie di osservazioni delle variabili  $X_1$  e  $X_2$ . D'altra parte, se il valore di  $|r(X_1, X_2)|$  è prossimo a 0, allora dovrà essere prossimo a zero anche il valore di  $I^2$  e questo significa che i punti  $(x_{2i}, x_{1i})$  sono molto dispersi attorno alle due rette ai minimi quadrati (nel senso che le devianze residue delle due rette ai minimi quadrati raggiungono i rispettivi valori massimi teorici, ovvero i valori delle rispettive devianze totali).

**Interpretazione del coefficiente di correlazione lineare:** Il valore di  $r(X_1, X_2)$  è sempre compreso tra  $-1$  e  $+1$  e indica sia il *verso* (attraverso il segno) che l'*intensità* (attraverso il modulo) della relazione lineare tra le osservazioni delle variabili  $X_1$  e  $X_2$ .

Dalle precedenti osservazioni possiamo dedurre alcune proprietà fondamentali del coefficiente di correlazione lineare:

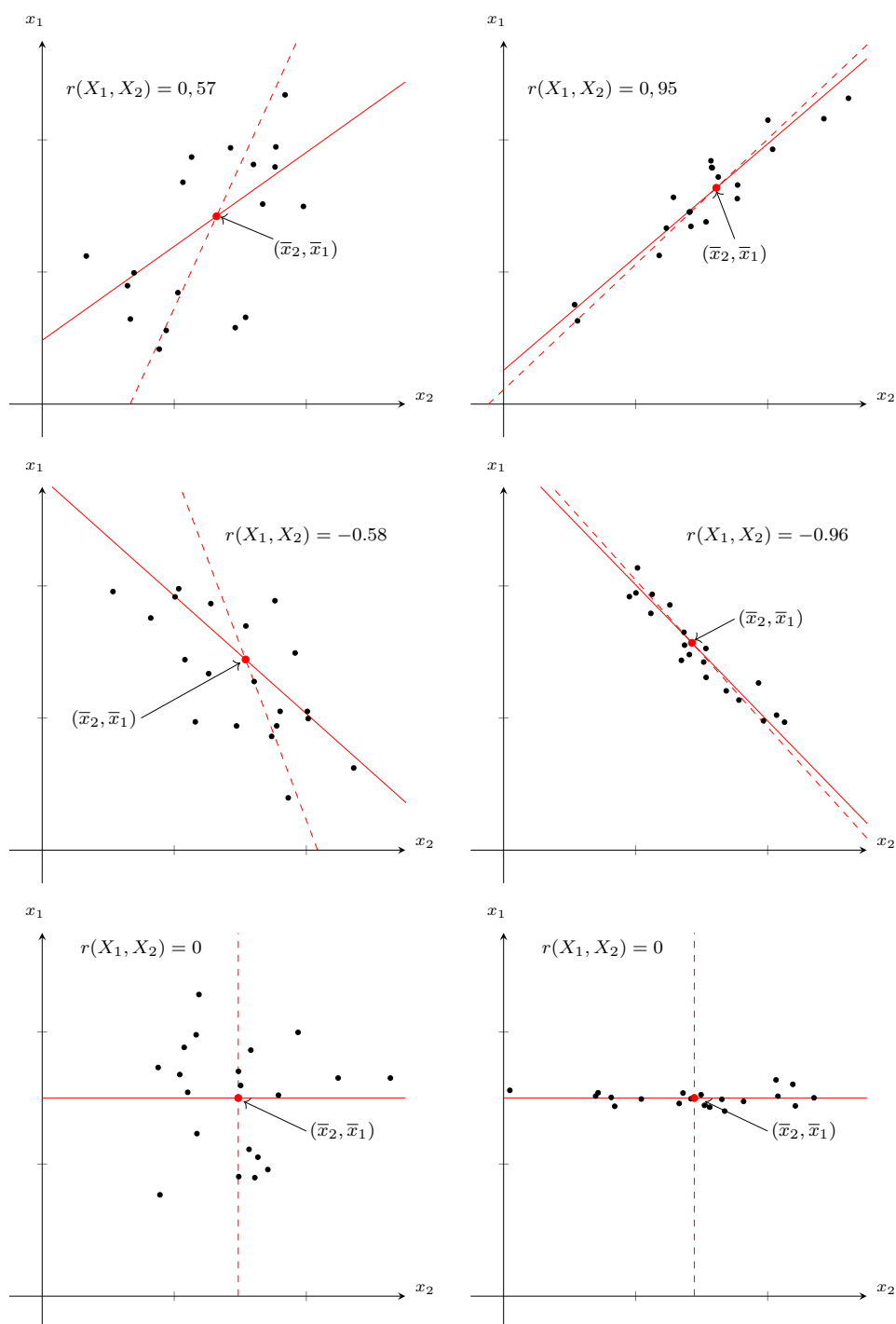
- $r(X_1, X_2) = 1$  se e solo se tutti i punti  $(x_{2i}, x_{1i})$  sono allineati lungo una retta crescente;
- $0 < r(X_1, X_2) < 1$  se e solo se le due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  sono entrambe crescenti e i punti  $(x_{2i}, x_{1i})$  sono dispersi attorno alle due rette ai minimi quadrati; più il valore di  $r(X_1, X_2)$  è prossimo a 1, più le

due rette ai minimi quadrati saranno prossime l'una all'altra e più i punti  $(x_{2i}, x_{1i})$  saranno prossimi ad entrambe le rette ai minimi quadrati;

- $r(X_1, X_2) = -1$  se e solo se tutti i punti  $(x_{2i}, x_{1i})$  sono allineati lungo una retta decrescente;
- $-1 < r(X_1, X_2) < 0$  se e solo se le due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  sono entrambe decrescenti e i punti  $(x_{2i}, x_{1i})$  sono dispersi attorno alle due rette ai minimi quadrati; più il valore di  $r(X_1, X_2)$  è prossimo a  $-1$ , più le due rette ai minimi quadrati saranno prossime l'una all'altra e più i punti  $(x_{2i}, x_{1i})$  saranno prossimi ad entrambe le rette ai minimi quadrati;
- $r(X_1, X_2) = 0$  se e solo se le due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  sono perpendicolari e i punti  $(x_{2i}, x_{1i})$  sono dispersi attorno alle due rette.

I grafici in Figura 3.9 illustrano alcuni casi dove il valore di  $r(X_1, X_2)$  è strettamente compreso tra  $-1$  e  $1$ .

Figura 3.9:



La formula defnitoria (14) per calcolare il coefficiente di correlazione lineare  $r(X_1, X_2)$  suggerisce che anche il suo valore numerico sia strettamente legato ai valori numerici dei coefficienti angolari

delle due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  (ricordiamo che in precedenza abbiamo già osservato che il segno di  $r(X_1, X_2)$  coincide sempre con il segno di  $codev(X_1, X_2)$  che a sua volta coincide con il segno dei coefficienti angolari delle due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$ ). Infatti,

$$\begin{aligned}\hat{\alpha}_2 &= \frac{codev(X_1, X_2)}{dev(X_2)} = \frac{cov(X_1, X_2)}{var(X_2)} \\ &= \frac{\sqrt{var(X_1)}}{\sqrt{var(X_2)}} \times \frac{cov(X_1, X_2)}{\sqrt{var(X_2) \times var(X_1)}} = \frac{\sigma_1}{\sigma_2} \times r(X_1, X_2)\end{aligned}$$

( $\sigma_1 = \sqrt{var(X_1)}$  e  $\sigma_2 = \sqrt{var(X_2)}$  indicano gli scarti quadratici medi delle osservazioni delle variabili  $X_1$  e  $X_2$ , rispettivamente) e allo stesso modo si ottiene anche

$$\hat{\beta}_2 = \frac{codev(X_1, X_2)}{dev(X_1)} = \frac{\sigma_2}{\sigma_1} \times r(X_1, X_2).$$

Ne consegue che  $segno(r(X_1, X_2)) = segno(\hat{\alpha}_2) = segno(\hat{\beta}_2)$  (questo già lo sapevamo) e che

$$r(X_1, X_2) = segno(\hat{\alpha}_2) \times \sqrt{\hat{\alpha}_2 \times \hat{\beta}_2} = segno(\hat{\beta}_2) \times \sqrt{\hat{\alpha}_2 \times \hat{\beta}_2}.$$

Queste osservazioni suggeriscono una definizione alternativa del coefficiente di correlazione lineare:

I coefficienti angolari delle due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  possono essere calcolati come

$$\hat{\alpha}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_2)} = \frac{\sigma_1}{\sigma_2} \times r(X_1, X_2) \quad (15)$$

e

$$\hat{\beta}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_1)} = \frac{\sigma_2}{\sigma_1} \times r(X_1, X_2),$$

rispettivamente, e il coefficiente di correlazione lineare  $r(X_1, X_2)$  può essere calcolato (e anche *definito*) come media geometrica dei coefficienti angolari  $\hat{\alpha}_2$  e  $\hat{\beta}_2$ .

Analizziamo ora come varierebbe il valore del coefficiente di correlazione lineare se al posto delle variabili  $X_1$  e  $X_2$  considerassimo delle loro trasformazioni lineari affini. Definiamo quindi  $Y_1 = a_1 + b_1X_1$  e  $Y_2 = a_2 + b_2X_2$  con  $b_1$  e  $b_2$  due costanti reali diverse da zero (una o entrambe le costanti reali  $a_1$  e  $a_2$  potrebbero invece essere anche nulle). In questo caso avremmo

$$\begin{aligned} \text{codev}(Y_1, Y_2) &= \text{codev}(a_1 + b_1X_1, a_2 + b_2X_2) \\ &= b_1b_2 \times \text{codev}(X_1, X_2), \end{aligned}$$

$$\text{dev}(Y_1) = \text{dev}(a_1 + b_1X_1) = b_1^2 \times \text{dev}(X_1)$$

e (in modo analogo)  $\text{dev}(Y_2) = b_2^2 \times \text{dev}(X_2)$ . Il coefficiente di correlazione lineare riferito alle variabili  $Y_1$  e  $Y_2$  sarebbe quindi



dato da

$$\begin{aligned}
 r(Y_1, Y_2) &= \frac{\text{codev}(Y_1, Y_2)}{\sqrt{\text{dev}(Y_1) \times \text{dev}(Y_2)}} \\
 &= \frac{b_1 b_2 \times \text{codev}(X_1, X_2)}{\sqrt{b_1^2 \times \text{dev}(X_1) \times b_2^2 \times \text{dev}(X_2)}} \\
 &= \frac{b_1 b_2}{|b_1 b_2|} \times \frac{\text{codev}(X_1, X_2)}{\sqrt{\text{dev}(X_1) \times \text{dev}(X_2)}} \\
 &= \text{segno}(b_1 b_2) \times r(X_1, X_2).
 \end{aligned}$$

Questo risultato ci permette di dedurre un'altra proprietà fondamentale del coefficiente di correlazione lineare:

**Coefficiente di correlazione lineare di trasformazioni lineari affini:** Se  $Y_1 = a_1 + b_1 X_1$  e  $Y_2 = a_2 + b_2 X_2$  sono due trasformazioni lineari affini con  $b_1, b_2 \neq 0$ , allora

$$r(Y_1, Y_2) = \text{segno}(b_1 b_2) \times r(X_1, X_2).$$

Ne consegue che:

- $r(Y_1, Y_2) = r(X_1, X_2)$  se  $b_1$  e  $b_2$  sono dello stesso segno;
- $r(Y_1, Y_2) = -r(X_1, X_2)$  se  $b_1$  e  $b_2$  sono di segno opposto;
- $r^2(X_1, X_2) = r^2(Y_1, Y_2)$ .

Dall'ultima proprietà legata alle trasformazioni lineari affini deduciamo che il valore dell'indice di determinazione delle due rette ai minimi quadrati che interpolano i punti  $(x_{2i}, x_{1i})$  non cambia se ad una o entrambe le variabili viene applicata una trasformazione lineare affine strettamente crescente o decrescente.

L'indice di determinazione è invariante rispetto a trasformazioni lineari affini strettamente crescenti o decrescenti.

**Esempio 3.7.** Usando i valori di  $codev(X_1, X_2)$ ,  $dev(X_1)$  e  $dev(X_2)$  calcolati nell'Esempio 3.1 si vede che il coefficiente di correlazione lineare tra gli utili e i compensi è dato da

$$\begin{aligned} r(X_1, X_2) &= \frac{codev(X_1, X_2)}{\sqrt{dev(X_1) \times dev(X_2)}} \\ &= \frac{4888,202}{\sqrt{1002,284 \times 46810,806}} = 0,714. \end{aligned}$$

Il valore di  $r(X_1, X_2)$  ci dice che il verso della relazione lineare tra gli utili e i compensi è positivo e che l'intensità della relazione lineare è pari all'71,4% del massimo teorico.  $\triangle$

**Esercizio 3.5.** Si calcoli il coefficiente di correlazione lineare tra le osservazioni del tasso di crescita del PIL e quelle dell'indice PMI (si veda l'Esempio 3.2).  $\triangle$

## 3.8 Applicazione in ambito finanziario

### 3.8.1 L'alpha di Jensen

In questa sezione deriveremo un noto indice per valutare la *performance* di un fondo di investimento (o più in generale una qualunque attività finanziaria) tenendo conto del suo livello di rischiosità. La costruzione di questo indice è basata sulla teoria del *Capital Asset Pricing Model* (CAPM), un modello che descrive la relazione tra il rendimento e il rischio di attività finanziarie. Sotto le

ipotesi di tale modello le uniche strategie di investimento che un investitore razionale dovrebbe seguire sono quelle che prevedono di investire una certa quota  $q \geq 0$  del proprio capitale complessivo nel cosiddetto "portafoglio di mercato" e la parte restante in un "titolo privo di rischio".

- Il "portafoglio di mercato" è un ipotetico portafoglio che è composto da tutte le attività finanziarie rischiose presenti sul mercato, ovvero da tutte le attività i cui rendimenti sono incerti nel momento in cui vengono acquistate.
- Un titolo privo di rischio è un titolo il cui rendimento è noto con certezza nel momento in cui lo si acquista.

Il valore di  $q$  scelto dal singolo investitore dipende dalle sue preferenze personali: investitori più inclini al rischio sceglieranno di investire una quota  $q$  maggiore nel portafoglio di mercato, mentre quelli più avversi al rischio preferiranno mantenere il valore di  $q$  prossimo a zero. Sotto le ipotesi del modello CAPM un investitore razionale potrebbe scegliere anche un valore di  $q$  maggiore di 1 (l'investitore prende a prestito al tasso di rendimento del titolo privo di rischio e investe il capitale preso in prestito e tutto il capitale di cui già disponeva nel portafoglio di mercato).

Per mettere in atto una strategia razionale CAPM con  $q > 0$  un investitore deve quindi conoscere la composizione del portafoglio di mercato. Ammesso che la composizione di tale portafoglio sia nota a qualcuno, gli investitori che operano nel mondo reale di

solito non hanno la possibilità di investire in qualunque attività finanziaria e per questo motivo sostituiscono il portafoglio di mercato con un indice di mercato che rappresenta il valore complessivo di tutti i titoli finanziari nei quali hanno effettivamente la possibilità di investire. Per esempio, i fondi di investimento che dichiarano di investire in società quotate nei mercati nordamericani di solito considerano come portafoglio di mercato uno degli indici S&P500, NASDAQ 100 oppure gli indici Russell 1000/2000/3000. Siccome la composizione di un indice di mercato deve essere pubblica, tutti gli investitori possono, almeno in teoria, implementare una delle corrispondenti strategie razionali del CAPM senza incorrere in costi.<sup>5</sup> Per valutare la *performance* di un fondo d'investimento non si dovrebbe quindi guardare solo ai rendimenti (giornalieri, settimanali, mensili o annuali) del fondo, ma bisognerebbe anche confrontarli con quelli di una opportuna strategia razionale CAPM. Di solito il portafoglio di mercato della strategia razionale CAPM di riferimento coincide con il cosiddetto *benchmark* del fondo, ovvero con l'indice di mercato che il fondo dichiara di voler replicare o addirittura battere.

Indichiamo dunque con

$$r_1^{(p)}, r_2^{(p)}, \dots, r_n^{(p)}$$

---

<sup>5</sup>Di fatto, implementare una strategia razionale CAPM con  $q > 0$  non è per nulla semplice. Infatti, per replicare la composizione di un indice di mercato bisogna ribilanciare continuamente i pesi di tutti i titoli contenuti nel portafoglio per assicurare che coincidano con i corrispondenti pesi nell'indice. Per questo motivo anche i fondi di investimento passivi, ovvero quelli che mirano a replicare i rendimenti di un indice di mercato, applicano delle commissioni.

i rendimenti generati da un fondo d'investimento durante un periodo storico composto da  $n$  *intervalli di rendimento*. Di solito, come intervalli di rendimento si considerano i giorni, le settimane, i mesi oppure gli anni che ricadono nel periodo storico sotto esame. Ricordiamo che i singoli rendimenti sono definiti come

$$r_i^{(p)} = \frac{\left( \begin{array}{c} \text{Valore del fondo} \\ \text{al termine dell}'i\text{-esimo} \\ \text{intervallo di rendimento} \end{array} \right) - \left( \begin{array}{c} \text{Valore del fondo} \\ \text{al termine} \\ \text{dell'intervallo di rendimento } i-1 \end{array} \right)}{\left( \begin{array}{c} \text{Valore del fondo} \\ \text{al termine} \\ \text{dell'intervallo di rendimento } i-1 \end{array} \right)}.$$

Come abbiamo visto, per valutare la performance di un fondo bisogna confrontare i rendimenti  $r_i^{(p)}$  del fondo con quelli di una strategia razionale CAPM. Vediamo dunque come si calcolano i rendimenti di una strategia razionale CAPM. Investendo all'inizio dell' $i$ -esimo intervallo di rendimento una data quota  $q$  di un dato capitale iniziale  $C_0$  nel portafoglio di mercato e investendo la parte restante nel titolo privo di rischio, un investitore otterrebbe un capitale finale di ammontare pari a

$$C_1 = C_0 + q \times C_0 \times r_i^{(m)} + (1 - q) \times C_0 \times r_i^{(f)},$$

dove  $r_i^{(m)}$  indica il rendimento del portafoglio di mercato e dove  $r_i^{(f)}$  indica il rendimento del titolo privo di rischio. Il rendimento che si ottiene con una strategia razionale CAPM è dunque dato da

$$\begin{aligned} r_i^{(CAPMq)} &= \frac{C_1 - C_0}{C_0} \\ &= q \times r_i^{(m)} + (1 - q) \times r_i^{(f)} \\ &= r_i^{(f)} + q \times (r_i^{(m)} - r_i^{(f)}) \\ \Rightarrow r_i^{(CAPMq)} &= r_i^{(f)} + q \times (r_i^{(m)} - r_i^{(f)}). \end{aligned} \quad (16)$$

Come ci aspettavamo, il rendimento di una strategia razionale CAPM dipende dal valore della quota  $q$  investita nel portafoglio di mercato e quindi si pone il problema di ***come scegliere il valore di  $q$  per calcolare i rendimenti della strategia razionale CAPM che dovrebbero essere confrontati con i rendimenti del fondo.*** Chiaramente il valore di  $q$  potrebbe essere scelto in tanti modi diversi, ma visto che il valore di  $q$  rappresenta il livello di esposizione al portafoglio di mercato, per fare un confronto equo bisognerà scegliere il valore di  $q$  in modo tale che la strategia ottimale CAPM presenti un livello di esposizione "equivalente" a quello del fondo. Per determinare il valore di  $q$  bisogna quindi definire una misura per il livello di esposizione tenuto dal fondo. Per pervenire ad una definizione ragionevole conviene approfondire l'effetto di  $q$  sui rendimenti della strategia razionale CAPM. Notiamo innanzitutto che la (16) può essere riscritta come

$$r_i^{(CAPMq)} - r_i^{(f)} = q \times (r_i^{(m)} - r_i^{(f)})$$

e che la relazione tra gli **eccessi di rendimento** di una strategia razionale CAPM (ovvero  $r_i^{(CAPMq)} - r_i^{(f)}$ ) e gli eccessi di rendimento del portafoglio di mercato (ovvero  $r_i^{(m)} - r_i^{(f)}$ ) è perfettamente lineare con coefficiente angolare  $q$ . Questo fatto suggerisce di definire il livello di esposizione tenuto dal fondo attraverso il coefficiente angolare della retta ai minimi quadrati che descrive l'eccesso di rendimento del fondo in funzione dell'eccesso di rendimento del portafoglio di mercato, ovvero il coefficiente angolare della retta ai

minimi quadrati che interpola i punti

$$(x_{2i}, x_{1i}) = \left( r_i^{(m)} - r_i^{(f)}, r_i^{(p)} - r_i^{(f)} \right), \quad i = 1, 2, \dots, n.$$

Nel modo della finanza il coefficiente angolare della retta ai minimi quadrati in questione viene chiamato "beta del fondo". Per questo motivo d'ora in poi lo indicheremo con  $\hat{\beta}$ . Se  $dev(X_2) > 0$  (come sempre accade nelle applicazioni), il *beta* del fondo sarà dato da (vedi la (7))

$$\hat{\beta} = \frac{codev(X_1, X_2)}{dev(X_2)}.$$

In ambito finanziario il *beta* non viene calcolato solo per fondi d'investimento, ma per qualsiasi attività finanziaria con rendimenti incerti. Siccome il *beta* di un'attività finanziaria descrive la sensibilità dei suoi eccessi di rendimento a variazioni degli eccessi di rendimento di un portafoglio di mercato di riferimento, spesso si dice che il "beta" misura l'esposizione al "rischio di mercato" oppure al "rischio sistemico".

Il **beta** di un'attività finanziaria è il coefficiente angolare della retta ai minimi quadrati che interpola i punti con ascisse date dagli eccessi di rendimento del portafoglio di mercato e ordinate date dagli eccessi di rendimento dell'attività finanziaria.

Il *beta* è una misura per l'esposizione al "rischio di mercato" che spesso viene anche chiamato "rischio sistemico".

A questo punto, avendo individuato un valore di  $q$  che identifica una strategia razionale CAPM di riferimento, possiamo calcolare i

rendimenti di tale strategia: ponendo  $q = \hat{\beta}$  nella (16) otteniamo

$$r_i^{(CAPM\hat{\beta})} = r_i^{(f)} + \hat{\beta} \times (r_i^{(m)} - r_i^{(f)}), \quad i = 1, 2, \dots, n.$$

Gli **extra rendimenti** del fondo rispetto ai rendimenti della corrispondente strategia razionale del CAPM sono quindi dati da

$$r_i^{(p)} - r_i^{(CAPM\hat{\beta})} = r_i^{(p)} - \left[ r_i^{(f)} + \hat{\beta} \times (r_i^{(m)} - r_i^{(f)}) \right] = x_{1i} - \hat{\beta} x_{2i}$$

e la media di questi extra rendimenti fornisce una misura per la *performance* del fondo:

$$\begin{aligned} \hat{\alpha} &= \frac{1}{n} \sum_{i=1}^n \left[ r_i^{(p)} - r_i^{(CAPM\hat{\beta})} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[ x_{1i} - \hat{\beta} x_{2i} \right] \\ &= \frac{1}{n} \sum_{i=1}^n x_{1i} - \hat{\beta} \times \frac{1}{n} \sum_{i=1}^n x_{2i} \\ &= \bar{x}_1 - \hat{\beta} \times \bar{x}_2. \end{aligned}$$

Notiamo che  $\hat{\alpha}$  è l'intercetta della retta ai minimi quadrati che si ottiene interpolando gli eccessi di rendimento  $(x_{2i}, x_{1i})$ . *In ambito finanziario questa media è nota come **alpha di Jensen** e viene usata per valutare la performance di qualsiasi attività finanziaria (non solo di fondi).*



L'**alpha di Jensen** è la media degli extra rendimenti di un'attività finanziaria rispetto ai rendimenti della strategia razionale CAPM con lo stesso *beta*. L'alpha di Jensen coincide con l'intercetta della retta ai minimi quadrati che interpola i punti le cui ascisse sono gli eccessi di rendimento del portafoglio di mercato e le cui ordinate sono i corrispondenti eccessi di rendimento dell'attività finanziaria.

### 3.8.2 Confronto tra fondi d'investimento

Per illustrare un'applicazione dell'alpha di Jensen in questa sezione confronteremo le performance di alcuni fondi comuni d'investimento con riferimento al periodo temporale compreso tra il 31 ottobre del 2017 e il 31 ottobre del 2020. I dati che verranno usati in questa sezione provengono dalla piattaforma FactSet. **Ovviamente lo scopo di questa esposizione è puramente didattico e non quello di promuovere o bocciare alternative di investimento. Tantomeno il contenuto di questa sezione deve essere interpretato come una sollecitazione all'investimento.**

Per illustrare le procedure di calcolo considereremo il fondo comune d'investimento AAA, un fondo che dichiara di investire soprattutto in azioni del settore della tecnologia elettronica con potenziale di crescita a lungo termine. Come portafoglio di mercato considereremo l'indice S&P500 che rappresenta il valore complessivo delle 500 aziende statunitensi con maggiore capitalizzazione

di mercato. Il grafico sottostante mostra l'andamento storico dei prezzi del fondo AAA e delle quotazioni dell'indice S&P500 alla fine di ciascun mese a partire da novembre 2017 fino a ottobre 2020 (i dati numerici sono contenuti in questo [foglio di calcolo](#)):

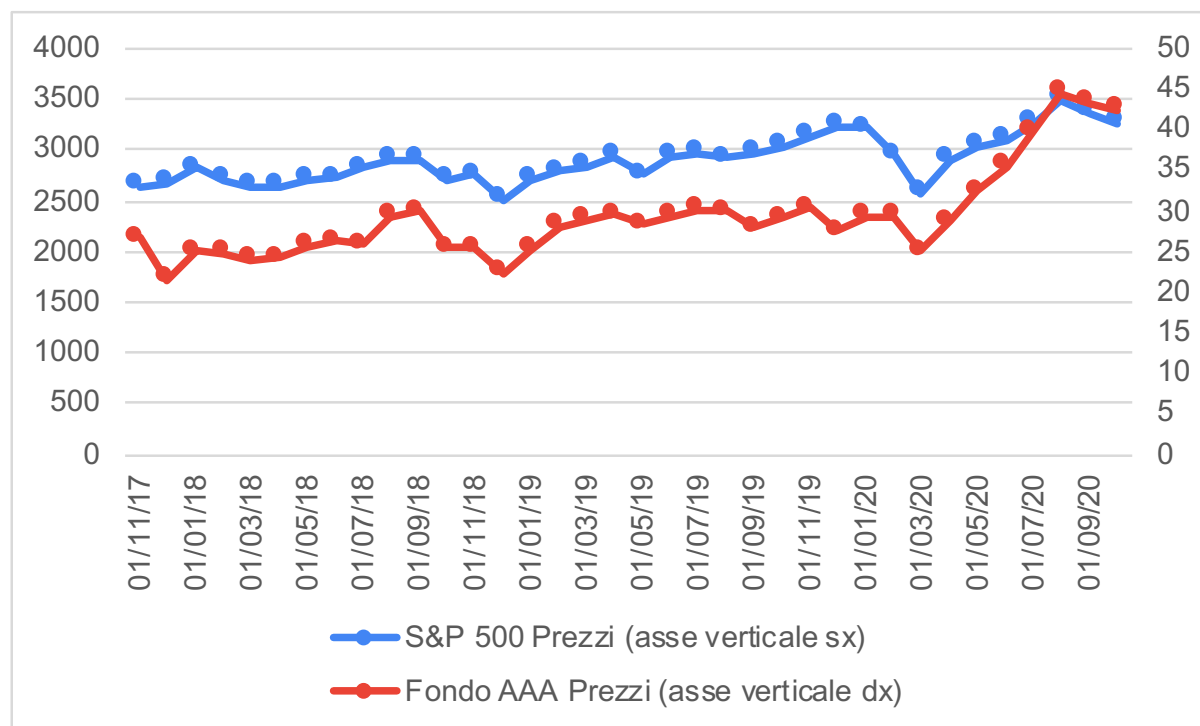


Figura 3.10:

Per procedere al calcolo del *beta* e dell'*alpha* di Jensen del fondo AAA (ovvero dei parametri della retta ai minimi quadrati che descrive gli eccessi di rendimento del fondo AAA in funzione degli eccessi di rendimento dell'indice S&P500) dobbiamo innanzitutto ricavare i rendimenti mensili del fondo AAA e quelli dell'indice S&P500. Durante il periodo storico di riferimento sono stati registrati 36 *prezzi di chiusura mensili* del fondo AAA e dell'indice S&P500 e a partire da tali prezzi si ottengono  $n = 36 - 1 = 35$  rendimenti mensili per ciascuna delle due serie storiche di prezzi.

Il grafico sottostante mostra le serie storiche dei rendimenti mensili (i valori numerici sono contenuti nelle colonne E e F del [foglio di calcolo](#)):

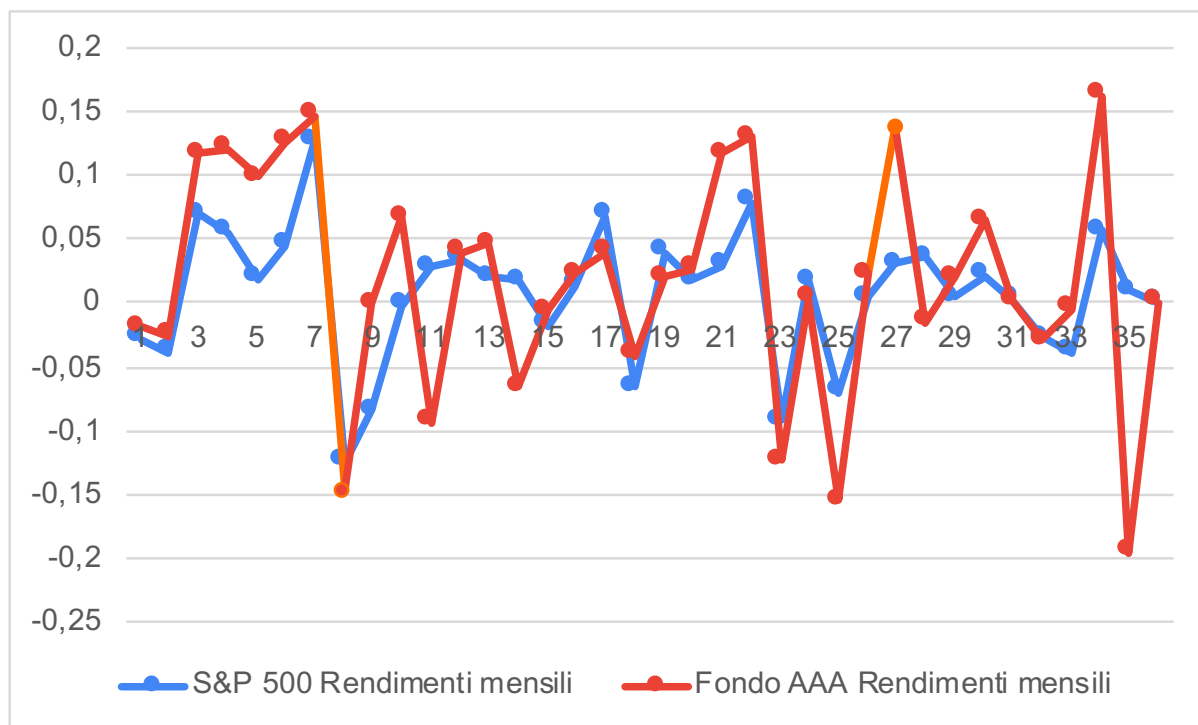


Figura 3.11:

Per calcolare i rendimenti dell' $i$ -esimo mese sono state usate le formule

$$r_i^{(p)} = \frac{\left( \text{Prezzo del fondo AAA} \right)_{\text{al termine del mese } i} - \left( \text{Prezzo del fondo AAA} \right)_{\text{al termine del mese } i-1}}{\left( \text{Prezzo del fondo AAA} \right)_{\text{al termine del mese } i-1}}$$

$$r_i^{(m)} = \frac{\left( \text{Prezzo dell'indice S\&P500} \right)_{\text{al termine del mese } i} - \left( \text{Prezzo dell'indice S\&P500} \right)_{\text{al termine del mese } i-1}}{\left( \text{Prezzo dell'indice S\&P500} \right)_{\text{al termine del mese } i-1}}$$

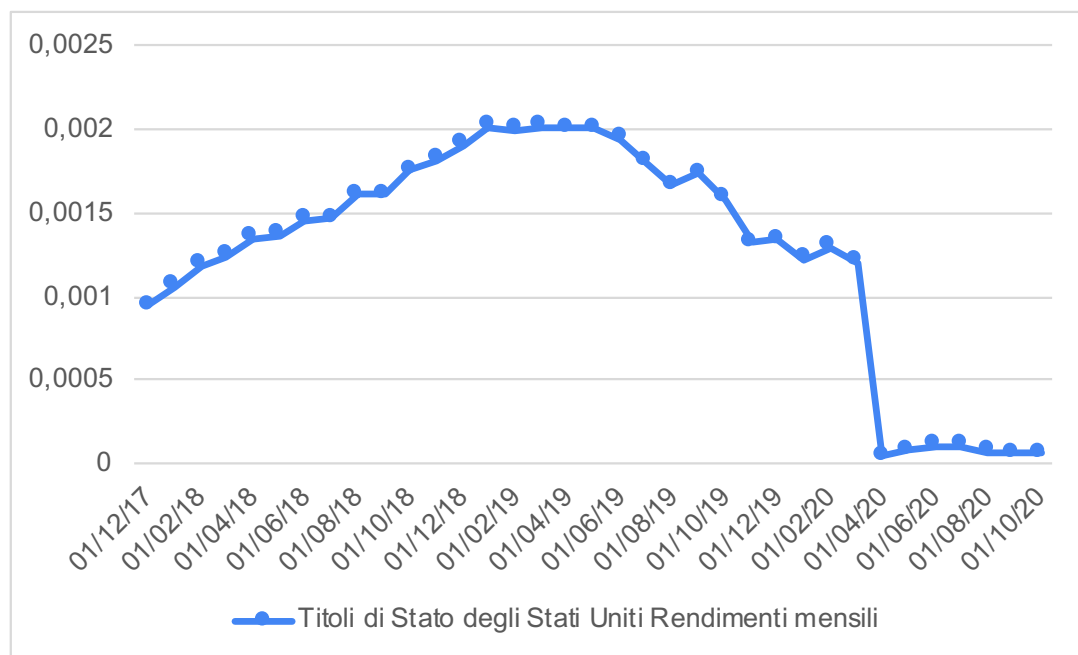
A questo punto, per ottenere gli eccessi di rendimento ci servono i rendimenti del titolo privo di rischio. In questa analisi considereremo come titoli privi di rischio i titoli di Stato degli Stati Uniti e al

posto dei rendimenti  $r_i^{(f)}$  useremo dunque i rendimenti mensili che si ricavano dalle curve di rendimento dei titoli di stato statunitensi. Siccome la curva dei rendimenti che si riferisce ad un dato istante temporale  $t$  restituisce i rendimenti annualizzati dei titoli di stato che scadranno negli istanti temporali  $t + h$  al variare di  $h > 0$ , per ricavare il rendimento dei titoli di stato statunitensi riferito al mese  $i$  bisogna prendere il valore del rendimento annualizzato che si trova sulla curva dei rendimenti che si riferisce all'ultimo giorno del mese precedente, che indicheremo con  $\rho_{i-1}^{(f)}$ , e applicargli la seguente trasformazione

$$r_i^{(f)} = \left(1 + \rho_{i-1}^{(f)}\right)^{1/12} - 1.$$

La colonna B del [foglio di calcolo](#) riporta i valori di  $\rho_i^{(f)}$  forniti da FactSet e la colonna G riporta i corrispondenti valori di  $r_i^{(f)}$ . La serie storica dei rendimenti  $r_i^{(f)}$  così ottenuti è riportata nel grafico sottostante.

Figura 3.12:



Confrontando gli assi verticali nelle Figure 3.11 e 3.12 notiamo immediatamente la spiccata differenza tra la variabilità dei rendimenti "rischiosi" del fondo AAA e dell'indice S&P500 e la variabilità dei rendimenti "non rischiosi" dei titoli di stato degli Stati Uniti. Sottraendo i rendimenti  $r_i^{(f)}$  dai rendimenti del fondo AAA e da quelli dell'indice S&P500 si ottengono finalmente le serie storiche degli eccessi di rendimento (i valori numerici sono contenuti nelle colonne H e I del [foglio di calcolo](#)):

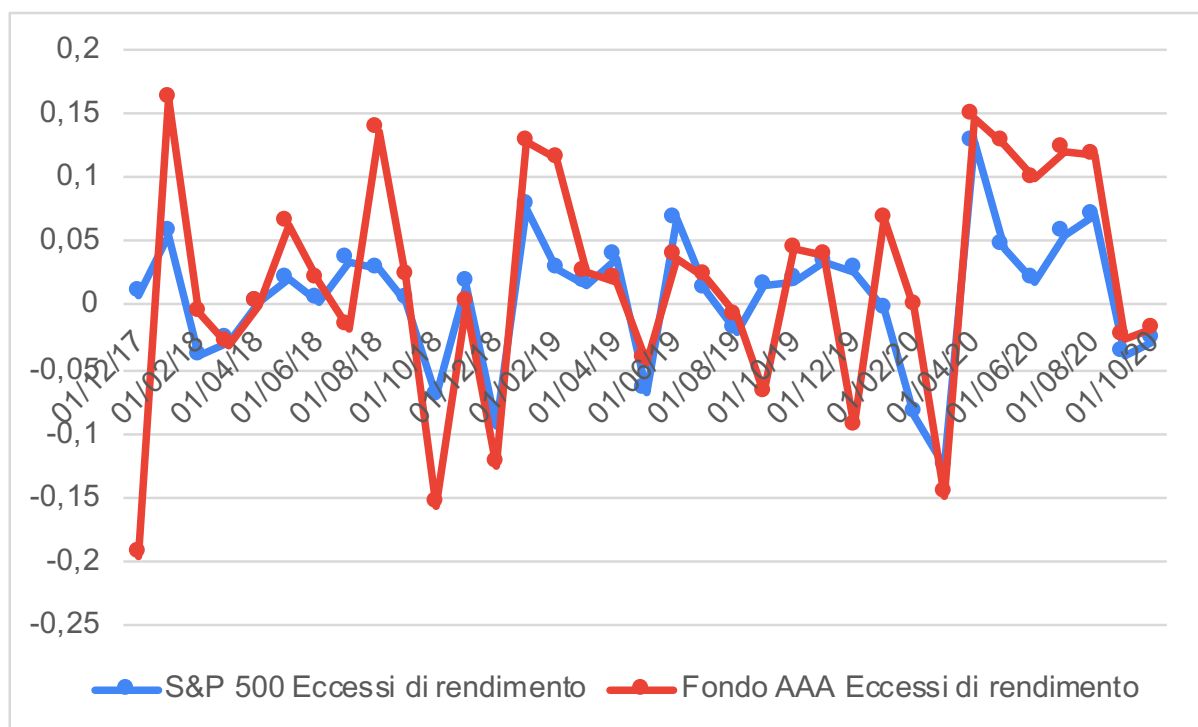


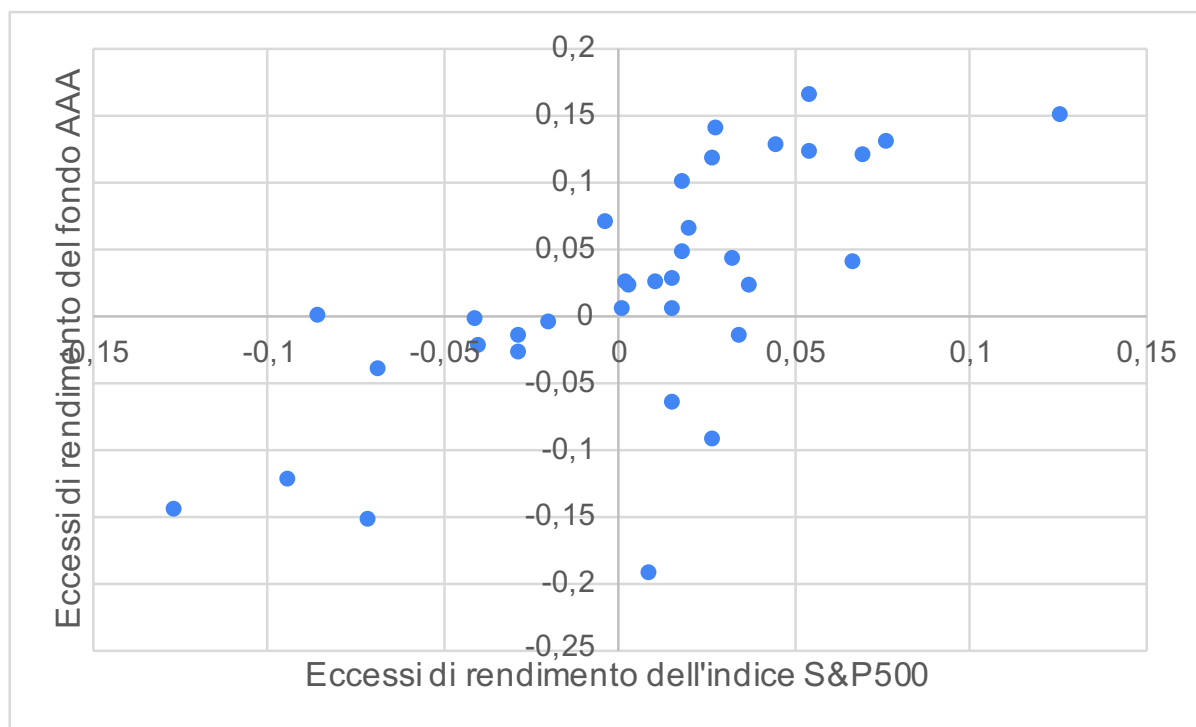
Figura 3.13:

A questo punto ci conviene riportare i punti

$$(x_{2i}, x_{1i}) = (r_i^{(m)} - r_i^{(f)}, r_i^{(p)} - r_i^{(f)}), \quad i = 1, 2, \dots, 35,$$

in un sistema di assi cartesiani per farci una prima impressione sulla relazione di fondo tra gli eccessi di rendimento dell'indice S&P500 e quelli del fondo AAA:

Figura 3.14:



Notiamo subito che il grafico evidenzia una correlazione positiva tra gli eccessi di rendimento dell'indice S&P500 e quelli del fondo AAA. A questo punto possiamo procedere con il calcolo del *beta* e dell'*alpha* di Jensen del fondo AAA. In primo luogo calcoliamo (si vedano i totali delle colonne H - K del [foglio di calcolo](#))

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} = \frac{0,5639}{35} = 0,0161 \quad [[ \text{colonna I} ]],$$

$$\bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2i} = \frac{0,2142}{35} = 0,0061 \quad [[ \text{colonna H} ]],$$

$$\sum_{i=1}^n x_{2i}^2 = 0,094313 \quad [[ \text{colonna J} ]]$$

$$\begin{aligned}\Rightarrow \text{dev}(X_2) &= \sum_{i=1}^n x_{2i}^2 - n \times \bar{x}_2^2 \\ &= 0,0943 - 35 \times 0,0061^2 = 0,0930\end{aligned}$$

e

$$\begin{aligned}\sum_{i=1}^n x_{1i} \times x_{2i} &= 0,1155 \quad [[ \text{colonna K} ]] \\ \Rightarrow \text{codev}(X_1, X_2) &= \sum_{i=1}^n x_{1i} \times x_{2i} - n \times \bar{x}_1 \times \bar{x}_2 \\ &= 0,1155 - 35 \times 0,0161 \times 0,0061 \\ &= 0,1121.\end{aligned}$$

Poi sostituiamo questi valori nelle formule per calcolare i parametri della retta ai minimi quadrati onde ottenere il valore del *beta*

$$\hat{\beta} = \hat{\alpha}_2 = \frac{\text{codev}(X_1, X_2)}{\text{dev}(X_2)} = \frac{0,1121}{0,0930} = 1,2043$$

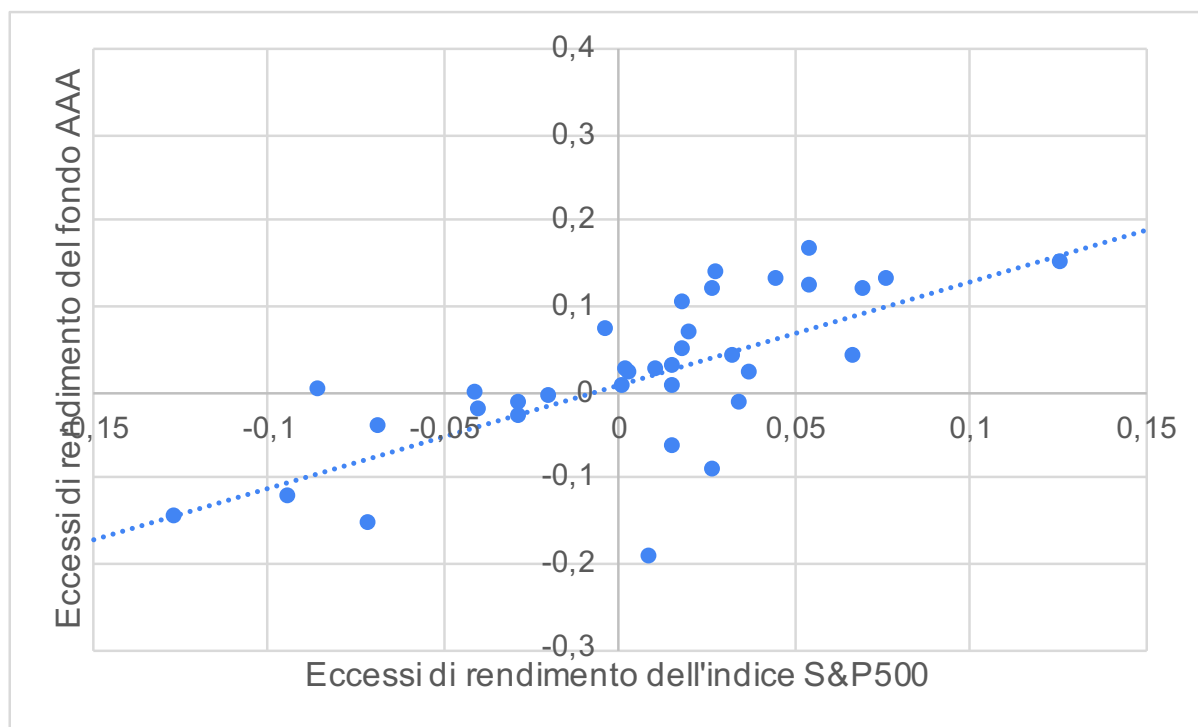
e quello dell'*alpha* di Jensen

$$\begin{aligned}\hat{\alpha} = \hat{\alpha}_1 &= \bar{x}_1 - \hat{\alpha}_2 \times \bar{x}_2 \\ &= 0,0161 - 1,2043 \times 0,0061 = 0,0088.\end{aligned}$$

Il grafico sottostante mostra la retta ai minimi quadrati sovrapposta alla nuvola di punti generata degli eccessi di rendimento:



Figura 3.15:



Ricordiamo che il valore dell'*alpha* di Jensen ci dice che durante il periodo sottoposto a valutazione il fondo AAA ha generato mediamente un extrarendimento mensile dello 0,88% rispetto alla strategia razionale CAPM con lo stesso *beta*.

Per agevolare il confronto tra le performance di più fondi spesso questi ultimi vengono identificati con dei punti all'interno di un piano cartesiano le cui ascisse sono date dai rispettivi *beta* e le cui ordinate sono date dalle medie dei rispettivi rendimenti mensili (oppure giornalieri, settimanali, annuali, ecc.). A questo diagramma viene poi aggiunta la cosiddetta *Security Market Line* (SML), ovvero la retta che in corrispondenza di ogni valore di  $\beta$  sull'asse delle ascisse restituisce la media dei rendimenti della strategia razionale CAPM con livello di esposizione al rischio di mercato

$q = \beta$ , ovvero il valore di

$$\begin{aligned}\bar{r}^{(CAPM\beta)} &= \frac{1}{n} \sum_{i=1}^n r_i^{(CAPM\beta)} \\ \text{[[ equ. (16) ]]} &= \frac{1}{n} \sum_{i=1}^n \left[ r_i^{(f)} + \beta \times (r_i^{(m)} - r_i^{(f)}) \right] \\ &= \frac{1}{n} \sum_{i=1}^n r_i^{(f)} + \beta \times \left( \frac{1}{n} \sum_{i=1}^n r_i^{(m)} - \frac{1}{n} \sum_{i=1}^n r_i^{(f)} \right) \\ &= \bar{r}^{(f)} + \beta \times \left( \bar{r}^{(m)} - \bar{r}^{(f)} \right).\end{aligned}$$

La *Security Market Line* (SML) è la retta che ad ogni valore di  $\beta$  sull'asse delle ascisse associa la media dei rendimenti della strategia razionale CAPM con livello di esposizione al rischio di mercato  $q = \beta$ . L'intercetta della SML è uguale a  $\bar{r}^{(f)}$  e il suo coefficiente angolare è dato dalla differenza  $\bar{r}^{(m)} - \bar{r}^{(f)}$ :

$$\bar{r}^{(CAPM\beta)} = \bar{r}^{(f)} + \beta \times \left( \bar{r}^{(m)} - \bar{r}^{(f)} \right).$$

In corrispondenza di  $\beta = 1$  la SML restituisce la media dei rendimenti del portafoglio di mercato.

Siccome l'*alpha* di Jensen di un fondo può essere espresso come

$$\begin{aligned}\hat{\alpha} &= \frac{1}{n} \sum_{i=1}^n \left[ r_i^{(p)} - r_i^{(CAPM\hat{\beta})} \right] \\ &= \frac{1}{n} \sum_{i=1}^n r_i^{(p)} - \frac{1}{n} \sum_{i=1}^n r_i^{(CAPM\hat{\beta})} \\ &= \bar{r}^{(p)} - \bar{r}^{(CAPM\hat{\beta})},\end{aligned}$$

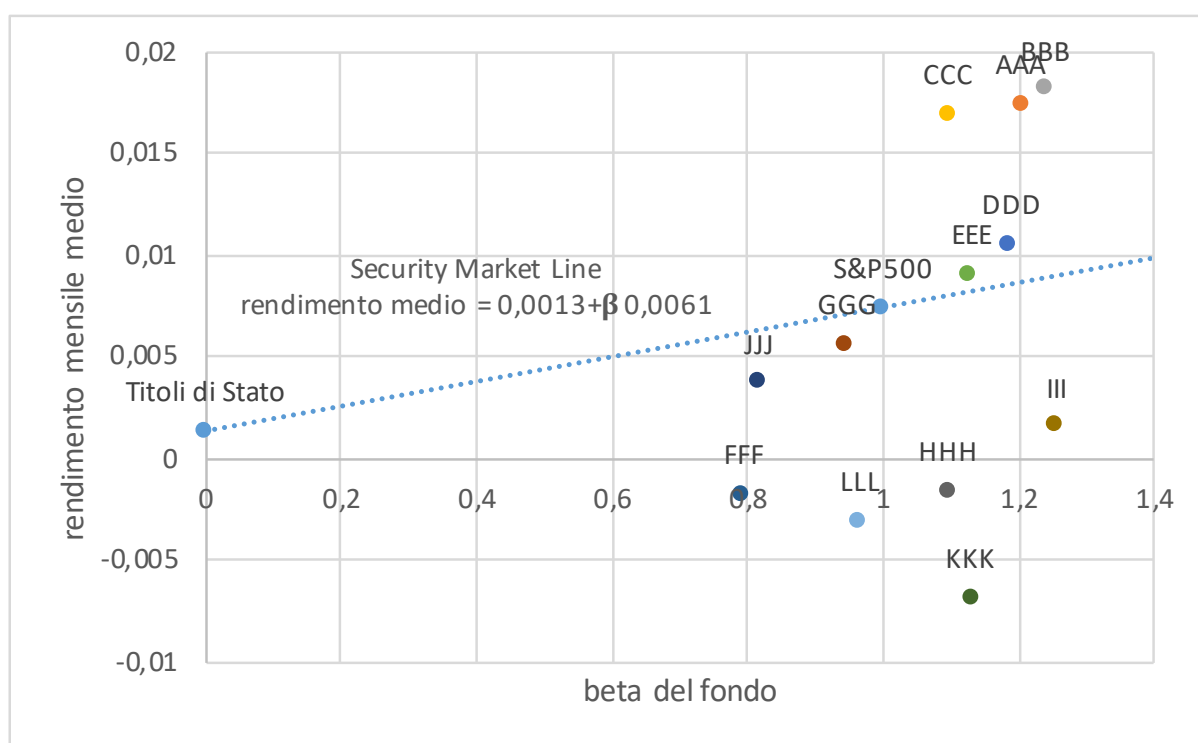
i fondi con un *alpha* di Jensen positivo sono quelli identificati da punti  $(\hat{\beta}, \bar{r}^{(p)})$  che si trovano al di sopra della SML, mentre quelli

con un *alpha* di Jensen negativo si trovano al di sotto della SML. Inoltre, le distanze verticali tra i punti e la SML coincidono con i valori assoluti degli *alpha* di Jensen dei fondi. Il grafico con la SML permette quindi di confrontare contemporaneamente

- i livelli di esposizione al rischio di mercato dei fondi (confrontando i valori dei loro *beta* sull'asse delle ascisse),
- le medie dei loro rendimenti (sull'asse delle ordinate)
- e i loro *alpha* di Jensen (le distanze verticali tra i punti e la SML).

Il grafico sottostante mostra la posizione del fondo AAA e di altri 11 fondi rispetto alla SML:

Figura 3.16:



E' importante tenere presente che mentre il *beta* di un fondo (o di una qualsiasi attività finanziaria) è un numero puro, ovvero un numero privo di unità di misura, l'unità di misura dell'*alpha* di Jensen è la stessa degli eccessi di rendimento che sono stati usati per calcolarlo. Per esempio, in questa sezione abbiamo considerato intervalli di rendimento di *durata mensile* e pertanto gli eccessi di rendimento del fondo AAA e dell'indice S&P500 erano anch'essi *mensili* e quindi abbiamo ottenuto un *alpha* di Jensen che rappresenta un extrarendimento *mensile*. Il suo valore può essere *annualizzato* attraverso la formula

$$\hat{\alpha}_{\text{annualizzato}} = (1 + \hat{\alpha}_{\text{mensile}})^{12} - 1.$$

Ovviamente, anche il valore dell'*alpha* di Jensen *annualizzato* (così come quello del *beta*) dipende ancora dalla scelta della *durata* degli intervalli di rendimento. Con intervalli di rendimento *giornalieri* (oppure settimanali o annuali) avremmo ottenuto un *alpha* di Jensen *annualizzato* (e anche un *beta*) diverso. In ogni caso, anche se al variare della durata degli intervalli di rendimento si ottengono valori del *beta* e dell'*alpha* di Jensen *annualizzato* che possono essere anche molto diversi,<sup>6</sup> di solito queste differenze non stravolgono i risultati di confronti tra attività finanziarie.

---

<sup>6</sup>Salvo in casi particolari che non si verificano mai nelle applicazioni, i valori del *beta* e dell'*alpha* di Jensen variano al variare della scelta della durata degli intervalli di rendimento.

## 4 Interpolazione con la famiglia delle funzioni potenza

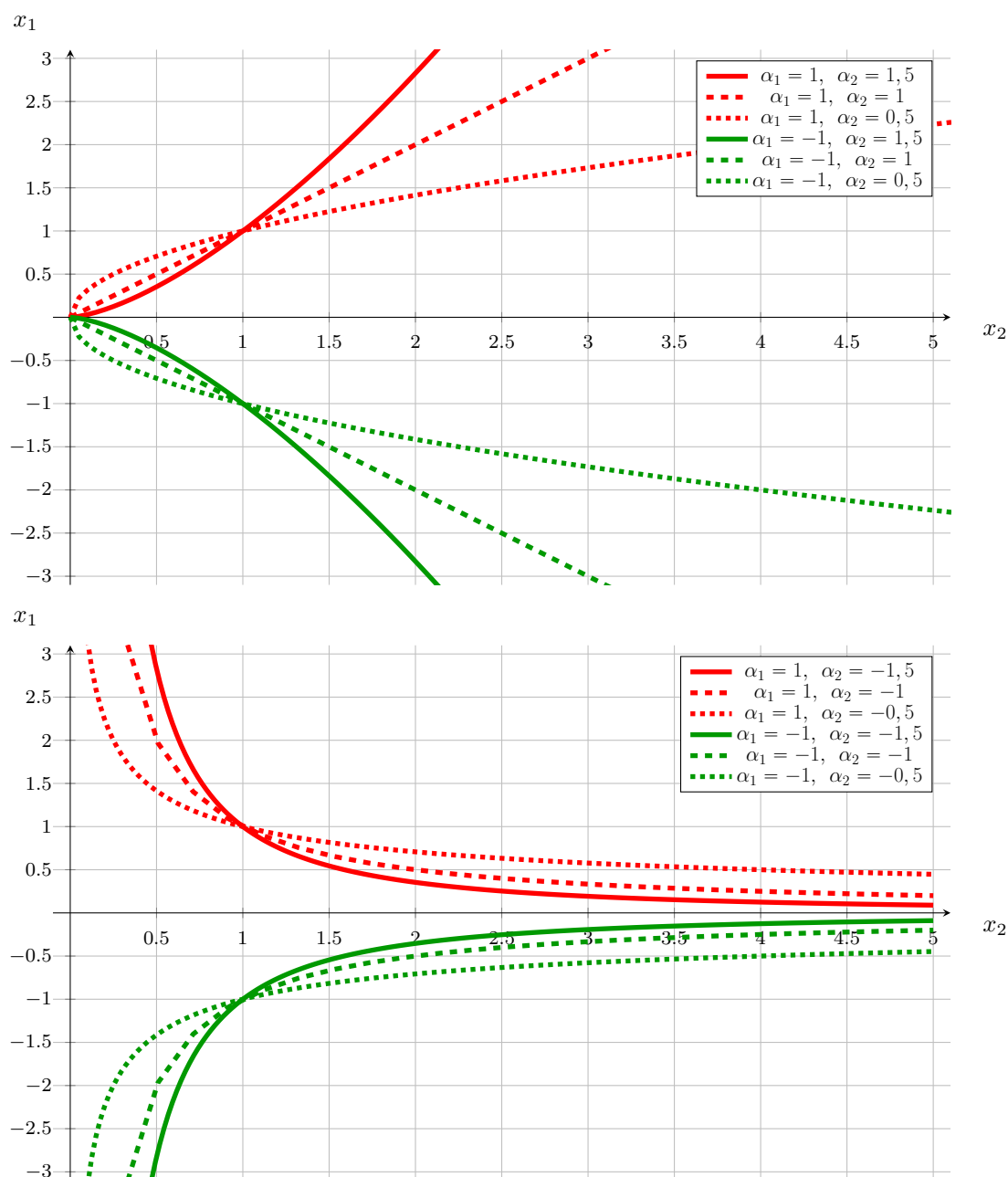
### 4.1 Proprietà delle funzioni potenza

Come già visto nella Sezione [2.2](#), le funzioni potenza sono definite come

$$x_1 = f(x_2) = \alpha_1 \times x_2^{\alpha_2} \quad \text{per } x_2 > 0, \quad (17)$$

dove  $\alpha_1$  e  $\alpha_2$  sono due parametri (numeri reali). I grafici sottostanti mostrano l'andamento di alcune funzioni potenza.

Figura 4.1:



Notiamo immediatamente che le funzioni potenza con parametro  $\alpha_1$  negativo possono assumere solo valori negativi e che il loro andamento è speculare rispetto a quelle che hanno parametro  $\alpha_1$  positivo. Le proprietà delle funzioni potenza con parametro  $\alpha_1$  negativo possono quindi essere facilmente dedotte da quelle con

parametro  $\alpha_1$  positivo e visto che di solito trattiamo variabili dipendenti che possono assumere solo valori positivi, *d'ora in poi assumeremo sempre che il parametro  $\alpha_1$  di una funzione potenza sia positivo*. Le principali proprietà delle funzioni potenza con parametro  $\alpha_1$  positivo sono qui di seguito elencate:

- Il dominio delle funzioni potenza è ristretto all'insieme dei numeri reali positivi e il loro codominio è dato dall'insieme dei numeri reali positivi.
- Il parametro  $\alpha_1$  (positivo) è un parametro di scala: per ottenere il grafico di una funzione potenza con un qualunque valore (positivo) del parametro  $\alpha_1$  basta disegnare il grafico della funzione potenza

$$x_1 = f(x_2) = 1 \times x_2^{\alpha_1} \quad \text{per } x_2 > 0$$

e poi modificare i valori riportati sull'asse delle ordinate moltiplicandoli per il valore di  $\alpha_1$ .

- Il parametro  $\alpha_2$  è un parametro di forma: modificando il valore del parametro  $\alpha_2$  si possono ottenere funzioni potenza
  - concave (se  $\alpha_2 \in (0, 1)$ ),
  - lineari (se  $\alpha_2 = 0$  oppure  $\alpha_2 = 1$ )
  - oppure convesse (se  $\alpha_2 \notin [0, 1]$ ).

**Dimostrazione.** Per dimostrare queste affermazioni riguardo il parametro  $\alpha_2$  calcoliamo la derivata seconda di una funzione potenza:

$$x_1 = f(x_2) = \alpha_1 \times x_2^{\alpha_2} \quad \text{per } x_2 > 0$$

$$\begin{aligned} &\Rightarrow \frac{dx_1}{dx_2} = f'(x_2) = \alpha_1 \times \alpha_2 \times x_2^{\alpha_2-1} \quad \text{per } x_2 > 0 \\ &\Rightarrow \frac{d^2x_1}{dx_2^2} = f''(x_2) = \alpha_1 \times \alpha_2 \times (\alpha_2 - 1) \times x_2^{\alpha_2-2} \quad \text{per } x_2 > 0. \end{aligned}$$

Siccome

$$x_2 > 0 \quad \Rightarrow \quad x_2^{\alpha_2-2} > 0,$$

e siccome per ipotesi  $\alpha_1 > 0$ , possiamo concludere che

- $f''(x_2) < 0$  (la funzione potenza è concava) se e solo se  $\alpha_2 \in (0, 1)$ ;
- $f''(x_2) = 0$  (la funzione potenza è lineare) se e solo se  $\alpha_2 = 0$  oppure  $\alpha_2 = 1$ ;
- $f''(x_2) > 0$  (la funzione potenza è convessa) se e solo se  $\alpha_2 \notin [0, 1]$ .

□

Prendendo il logaritmo (naturale o in qualunque altra base) ad ambo i membri dell'equazione che definisce una funzione potenza vediamo che

$$x_1 = \alpha_1 \times x_2^{\alpha_2} \quad \Leftrightarrow \quad \ln x_1 = \ln \alpha_1 + \alpha_2 \times \ln x_2 \quad (18)$$

(si ricordi che consideriamo solo funzioni potenza con  $\alpha_1 > 0$ ). Questo vuol dire che:

La relazione tra due variabili positive è descritta da una funzione potenza se e solo se la relazione tra i loro logaritmi (in qualunque base) è una relazione lineare affine.

Come si vede dalla (18), il parametro  $\alpha_2$  di una funzione potenza ci dice di quanto aumenta di logaritmo della variabile dipendente  $x_1$  se il logaritmo della variabile indipendente  $x_2$  aumenta di una unità. Volendo essere più precisi (vedi Figura 4.2) possiamo affermare che  $\alpha_2$  è il valore per il quale bisogna moltiplicare

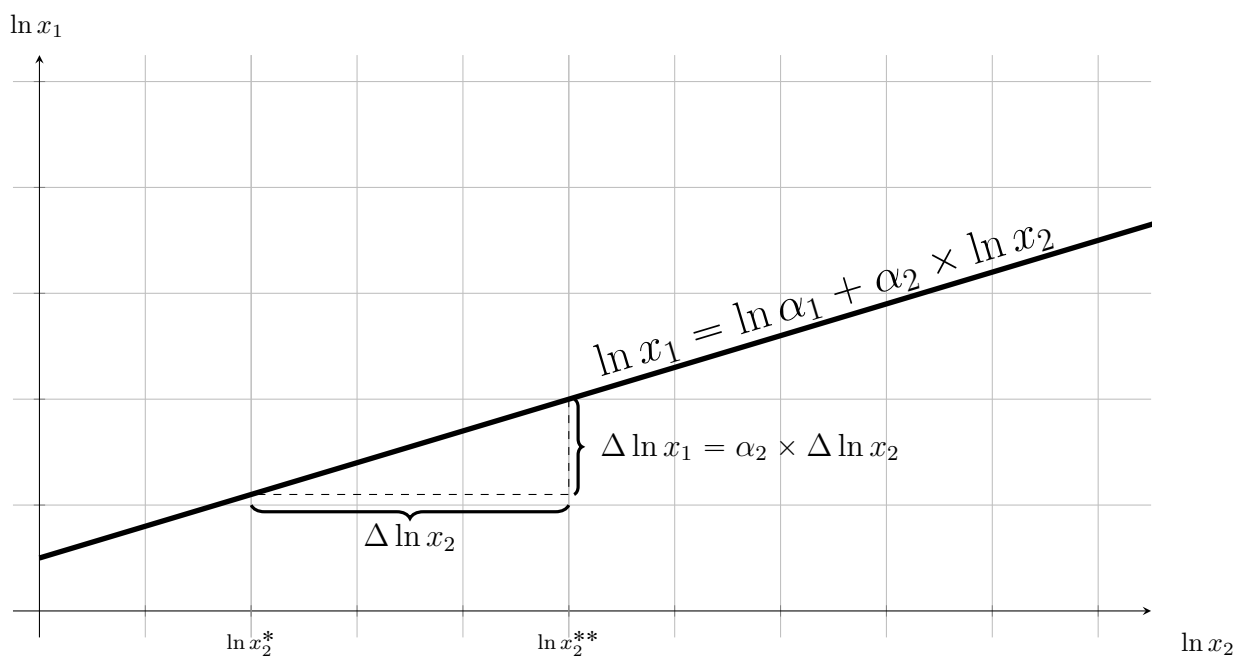


una variazione del logaritmo della variabile indipendente  $x_2$  per ottenere la corrispondente variazione del logaritmo della variabile dipendente  $x_1$ :

$$\Delta \ln x_1 = \alpha_2 \times \Delta \ln x_2. \quad (19)$$

Per comprendere il ruolo del parametro  $\alpha_2$  conviene dunque chiedersi che cosa succede con le variabili  $x_1$  e  $x_2$  al variare dei loro logaritmi. La risposta a questa domanda la vedremo nella prossima sezione dove vedremo che l'entità di una **variazione logaritmica** è strettamente legata a quella della corrispondente **variazione relativa**.

Figura 4.2:



## 4.2 Variazioni logaritmiche e variazioni relative

Sia  $x$  una variabile che può assumere solo valori positivi e analizziamo che cosa accade se il valore di  $x$  subisce una variazione passando da  $x^*$  a  $x^{**}$ . In questo caso la **variazione assoluta** di  $x$  sarà data da  $\Delta x = x^{**} - x^*$  e la corrispondente **variazione logaritmica** sarà invece data da

$$\Delta \ln x = \ln x^{**} - \ln x^* = \ln \frac{x^{**}}{x^*} = \ln \left( 1 + \frac{x^{**} - x^*}{x^*} \right). \quad (20)$$

Indicando con

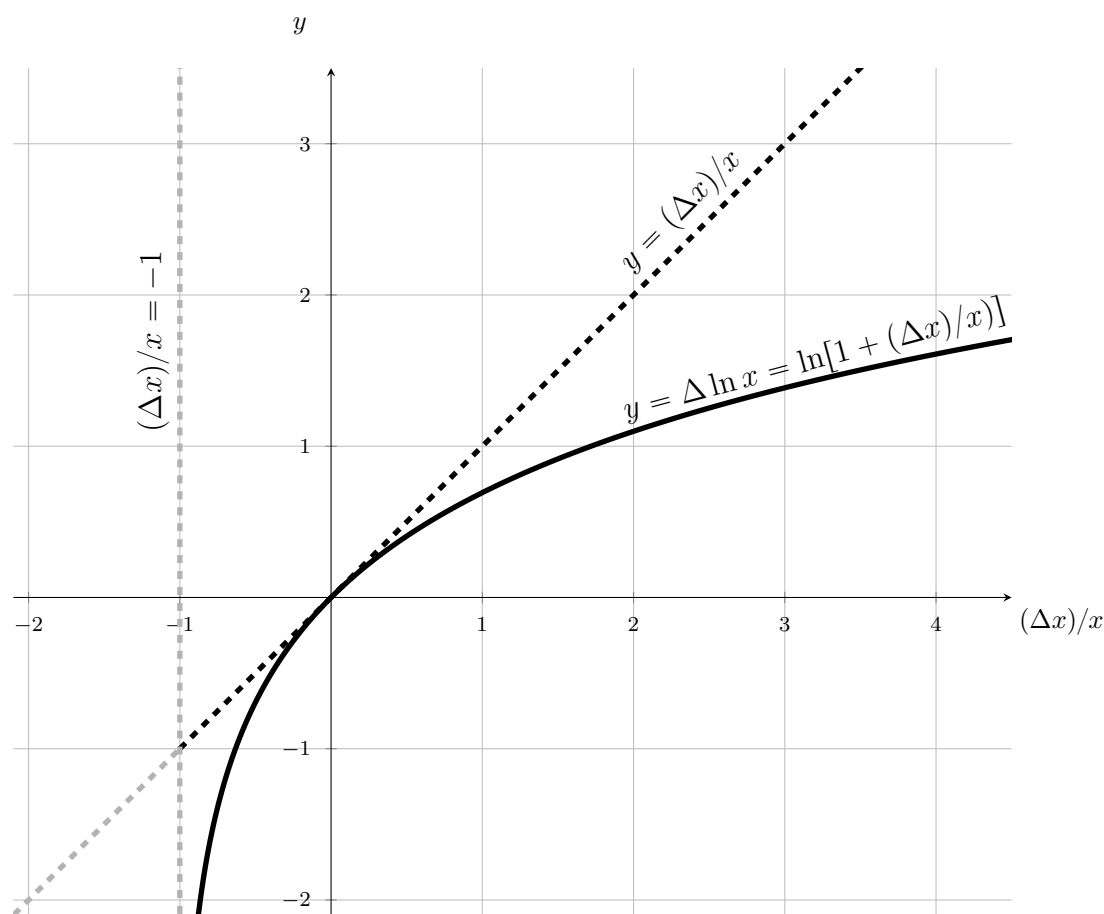
$$(\Delta x)/x = \frac{x^{**} - x^*}{x^*}$$

la **variazione relativa** di  $x$ , vediamo che la (20) può essere espressa come

$$\Delta \ln x = \ln [1 + (\Delta x)/x]. \quad (21)$$

Quest'ultima equazione ci dice che la variazione logaritmica  $\Delta \ln x$  è una funzione monotona crescente della corrispondente variazione relativa  $(\Delta x)/x$  (si ricordi che abbiamo ipotizzato che  $x$  possa assumere solo valori positivi e che la variazione relativa  $(\Delta x)/x$  deve quindi essere sempre maggiore di  $-1$ ). Il grafico sottostante mostra l'andamento della variazione logaritmica al variare della corrispondente variazione relativa:

Figura 4.3:



Dal grafico si desume che

$$\Delta \ln x \simeq (\Delta x)/x \quad \text{se e solo se} \quad (\Delta x)/x \simeq 0,$$

ovvero che la variazione logaritmica e la corrispondente variazione relativa sono molto prossime l'una all'altra se una di esse è prossima a zero (e in tal caso dovrà dunque essere prossima a zero anche l'altra). Il Teorema 4.1 formalizza questa affermazione:

**Teorema 4.1** (Variazioni logaritmiche e variazioni relative).

Siano  $x^*$  e  $x^{**}$  due numeri reali positivi, sia

$$(\Delta x)/x = \frac{x^{**} - x^*}{x^*}$$

la **variazione relativa** che subisce una variabile  $x$  nel passaggio da  $x^*$  a  $x^{**}$  e sia  $\Delta \ln x = \ln x^{**} - \ln x^*$  la corrispondente **variazione logaritmica**. Allora si ha:

- a)  $\Delta \ln x = \ln [1 + (\Delta x)/x]$  e  $(\Delta x)/x = e^{\Delta \ln x} - 1$ ;
- b)  $\Delta \ln x = (\Delta x)/x + o((\Delta x)/x)$ , dove  $o((\Delta x)/x)$  è trascurabile rispetto alla variazione relativa  $(\Delta x)/x$  se quest'ultima è prossima a zero;
- c)  $(\Delta x)/x = \Delta \ln x + o(\Delta \ln x)$ , dove  $o(\Delta \ln x)$  è trascurabile rispetto alla variazione logaritmica  $\Delta \ln x$  se quest'ultima è prossima a zero.

**Dimostrazione** (del Teorema 4.1). a) La prima equazione nella conclusione al punto a) è l'equazione (21) che è già stata dimostrata prima dell'enunciato del teorema. Risolvendo la prima equazione rispetto alla variazione relativa  $(\Delta x)/x$  si ottiene la seconda equazione.

b) La conclusione al punto b) è solo un modo diverso per affermare che

$$\lim_{(\Delta x)/x \rightarrow 0} \frac{\Delta \ln x}{(\Delta x)/x} = 1.$$

Per calcolare il limite basta applicare la regola di de l'Hopital (teniamo presente che  $\ln 1 = 0$ ). Per convenienza di notazione poniamo  $q = \Delta \ln x$  e  $r = (\Delta x)/x$ . In questo modo la prima equazione nella (21) può essere riscritta come  $q = \ln(1 + r)$  e quindi si ottiene

$$\lim_{(\Delta x)/x \rightarrow 0} \frac{\Delta \ln x}{(\Delta x)/x} = \lim_{r \rightarrow 0} \frac{q}{r} = \lim_{r \rightarrow 0} \frac{\ln(1 + r)}{r} = \text{[[de l'Hopital]]} = \lim_{r \rightarrow 0} \frac{\frac{1}{1+r}}{1} = 1$$

come volevasi dimostrare.

- c) La dimostrazione della conclusione al punto c) è analoga a quella della conclusione al punto b). Questa volta bisogna dimostrare che

$$\lim_{\Delta \ln x \rightarrow 0} \frac{(\Delta x)/x}{\Delta \ln x} = 1.$$

Usando la notazione appena introdotta vediamo che questo limite può essere riscritto come

$$\lim_{q \rightarrow 0} \frac{r}{q} = 1.$$

Esprimendo ora la seconda equazione della conclusione al punto a) come  $r = e^q - 1$  e applicando la regola di de l'Hopital otteniamo (teniamo presente che  $e^0 = 1$ )

$$\lim_{\Delta \ln x \rightarrow 0} \frac{(\Delta x)/x}{\Delta \ln x} = \lim_{q \rightarrow 0} \frac{r}{q} = \lim_{q \rightarrow 0} \frac{e^q - 1}{q} = [[\text{de l'Hopital}]] = \lim_{q \rightarrow 0} \frac{e^q}{1} = 1$$

come volevasi dimostrare. □

Così come variazioni relative vengono spesso espresse in termini percentuali, lo stesso accade anche per le variazioni logaritmiche. Per dire che una determinata variabile abbia subito una variazione *relativa* di ammontare pari a  $+0,01$ , si dice quindi spesso che la variabile abbia subito una variazione relativa positiva dell'1%. Siccome  $+0,01$  è un numero molto prossimo a 0, questo significa che l'ammontare della corrispondente variazione *logaritmica* sia anch'essa all'incirca pari a  $+0,01$ , oppure a  $+1\%$ . Tuttavia, a volte si sente anche dire che una variabile abbia subito una variazione "percentuale" positiva dell'1%. In questo caso, se non diversamente specificato, di solito si intende che la variabile abbia subito una variazione *relativa* di  $+1\%$ , anche se, in virtù del Teorema 4.1, la situazione non cambierebbe di molto se la variazione venisse interpretata come logaritmica. Tuttavia, se al posto di una variazione "percentuale" così modesta si avesse una variazione "percentuale" (per esempio) del 50%, allora la differenza tra i due tipi di

variazione sarebbe rilevante. Infatti,

$$(\Delta x)/x = +0,50 = +50\% \Rightarrow$$

$$\Rightarrow \Delta \ln x = \ln[1 + 0,50] = +0,4055 = +40,55\%,$$

mentre

$$\Delta \ln x = +0,50 = +50\% \Rightarrow$$

$$\Rightarrow (\Delta x)/x = e^{0,50} - 1 = +0,6487 = +64,87\%.$$

Questo esempio ci fa capire che nel caso di variazioni "percentuali" di entità rilevante è importante verificare con attenzione se si tratta di variazioni relative o logaritmiche. La tabella sottostante ci aiuta a comprendere l'entità dell'errore in cui si potrebbe incorrere confondendo una variazione logaritmica con una variazione relativa o viceversa:

$(\Delta x)/x$	$\Delta \ln x$	$[(\Delta x)/x] - \Delta \ln x$	$(\Delta x)/x$	$\Delta \ln x$	$[(\Delta x)/x] - \Delta \ln x$
-99%	-460,5170%	361,5170%	1%	0,9950%	0,0050%
-50%	-69,3147%	19,3147%	2%	1,9803%	0,0197%
-20%	-22,3144%	2,3144%	3%	2,9559%	0,0441%
-10%	-10,5361%	0,5361%	4%	3,9221%	0,0779%
-5%	-5,1293%	0,1293%	5%	4,8790%	0,1210%
-4%	-4,0822%	0,0822%	10%	9,5310%	0,4690%
-3%	-3,0459%	0,0459%	20%	18,2322%	1,7678%
-2%	-2,0203%	0,0203%	50%	40,5465%	9,4535%
-1%	-1,0050%	0,0050%	99%	68,8135%	30,1865%

### 4.3 Il parametro di elasticità

Torniamo ora al parametro  $\alpha_2$  di una funzione potenza e supponiamo che la variabile indipendente  $x_2$  subisca una variazione

*logaritmica* di ammontare pari a

$$\Delta \ln x_2 = 0,01 = 1\%.$$

Trattandosi di una variazione logaritmica prossima a zero, questo significa che la variabile  $x_2$  sta subendo una variazione *relativa* dell'1% circa. E l'equazione (19) ci dice che la variabile dipendente  $x_1$  subir  dunque una variazione *logaritmica* di ammontare pari a

$$\Delta \ln x_1 = \alpha_2 \times \Delta \ln x_2 = \alpha_2 \times 0,01 = \alpha_2 \times 1\% = \alpha_2\%.$$

Se il valore di  $\alpha_2 \times 0,01 = \alpha_2\%$    anch'esso prossimo a zero, questo significa che la variabile dipendente sta subendo una variazione *relativa* di circa  $\alpha_2$  punti percentuali.

Il parametro  $\alpha_2$  di una funzione potenza  $x_1 = \alpha_1 \times x_2^{\alpha_2}$    il valore per il quale si deve moltiplicare una piccola *variazione relativa* della variabile indipendente per ottenere (con buona approssimazione) la corrispondente *variazione relativa* della variabile dipendente. Per questo motivo si dice che  $\alpha_2$  misura l'**elasticit ** della variabile  $x_1$  rispetto alla variabile  $x_2$ .

#### 4.4 Determinazione dei parametri di una funzione interpolante potenza

Tranne in casi banali, i parametri di una funzione potenza "ai minimi quadrati" non possono essere espressi in termini di funzioni elementari e devono essere calcolati tramite metodi numerici iterativi, ovvero tramite metodi che partono da una coppia di valori

iniziali  $(\tilde{\alpha}_1, \tilde{\alpha}_2)$  e che ad ogni iterazione calcolano dei nuovi valori per i parametri che dovrebbero avvicinarsi sempre di più ad una coppia di valori  $(\hat{\alpha}_1, \hat{\alpha}_2)$  che minimizza la distanza quadratica

$$D(\alpha_1, \alpha_2) = \sum_{i=1}^n (x_{1i} - \alpha_1 x_{2i}^{\alpha_2})^2.$$

Per evitare di ricorrere a metodi numerici iterativi (oppure per altri motivi sui quali sorvoleremo), nelle applicazioni spesso si determinano i parametri di una funzione interpolante potenza secondo un metodo alternativo che comunque è strettamente legato a quello dei minimi quadrati. Il metodo in questione parte dal presupposto che la relazione tra due variabili  $x_1$  e  $x_2$  è descritta da una funzione potenza se e solo se la relazione tra i loro logaritmi (in qualunque base) è descritta da una funzione lineare affine (ovvero da una retta):

$$x_1 = \alpha_1 x_2^{\alpha_2} \quad \Leftrightarrow \quad \ln x_1 = \ln \alpha_1 + \alpha_2 \ln x_2. \quad (22)$$

Questo fatto ci suggerisce di determinare i parametri di una funzione interpolante potenza in modo tale che  $\ln \alpha_1$  (si noti il logaritmo!!!) e  $\alpha_2$  siano, rispettivamente, l'intercetta e il coefficiente angolare della retta ai minimi quadrati che considera  $Y_1 = \ln X_1$  come variabile dipendente e  $Y_2 = \ln X_2$  come variabile esplicativa. In quanto segue indicheremo questi parametri con  $\tilde{\alpha}_1$  e  $\tilde{\alpha}_2$ .



Se  $dev(Y_2) > 0$  (come sempre accade nelle applicazioni), allora si possono calcolare i parametri  $\alpha_1 = \tilde{\alpha}_1$  e  $\alpha_2 = \tilde{\alpha}_2$  di una funzione interpolante potenza attraverso le formule

$$\tilde{\alpha}_2 = \frac{codev(Y_1, Y_2)}{dev(Y_2)} = \frac{codev(\ln X_1, \ln X_2)}{dev(\ln X_2)} \quad (23)$$

e

$$\begin{aligned} \ln \tilde{\alpha}_1 &= \bar{y}_1 - \tilde{\alpha}_2 \bar{y}_2 = M_1(\ln X_1) - \tilde{\alpha}_2 M_1(\ln X_2) \\ \Rightarrow \tilde{\alpha}_1 &= e^{\bar{y}_1 - \tilde{\alpha}_2 \bar{y}_2}. \end{aligned} \quad (24)$$

E' importante tenere presente che, tranne in casi banali che non sono d'interesse nelle applicazioni, i parametri  $\tilde{\alpha}_2$  e  $\tilde{\alpha}_1$  determinati secondo le formule (23) e (24) non minimizzano la distanza quadratica

$$D(\alpha_1, \alpha_2) = \sum_{i=1}^n (x_{1i} - \alpha_1 x_{2i}^{\alpha_2})^2$$

bensì una funzione di distanza diversa che è definita come

$$\tilde{D}(\alpha_1, \alpha_2) = \sum_{i=1}^n (\ln x_{1i} - \ln \alpha_1 + \alpha_2 \ln x_{2i})^2.$$

Per questo motivo, la funzioni interpolante potenza

$$\hat{x}_1 = f(x_2) = \tilde{\alpha}_1 x_2^{\tilde{\alpha}_2}, \quad x_2 > 0,$$

non può essere chiamata funzione potenza "ai minimi quadrati".

Per descrivere un frequente malinteso legato alla funzione interpolante potenza determinata secondo il metodo qui descritto, indicheremo ora con  $\hat{Y}_1$  la variabile che rappresenta i valori

riprodotti

$$\hat{y}_{1i} = \ln \tilde{\alpha}_1 + \tilde{\alpha}_2 y_{2i}, \quad i = 1, 2, \dots, n,$$

(si noti il logaritmo!!!) e con  $Z_Y$  la variabile che rappresenta i corrispondenti residui

$$z_{Yi} = y_{1i} - \hat{y}_{1i}, \quad i = 1, 2, \dots, n.$$

Siccome  $\ln \tilde{\alpha}_1$  e  $\tilde{\alpha}_2$  sono i parametri della retta ai minimi quadrati che considera la variabile  $Y_1 = \ln X_1$  come variabile dipendente, e la variabile  $Y_2 = \ln X_2$  come variabile esplicativa, i residui  $z_{Yi}$  devono soddisfare le tre proprietà che i residui di qualunque retta ai minimi quadrati devono soddisfare (si veda la Sezione 3.3), ovvero

$$1) \quad \sum_{i=1}^n z_{Yi} = 0 \quad \Rightarrow \quad M_1(Z_Y) = 0,$$

$$2) \quad \text{codev}(Z_Y, Y_2) = 0,$$

$$3) \quad \text{codev}(Z_Y, \hat{Y}_1) = 0.$$

Come abbiamo visto nelle Sezioni 3.4, 3.5 e 3.6, queste proprietà implicano

a) la scomposizione della devianza

$$\text{dev}(Y_1) = \text{dev}(\hat{Y}_1) + \text{dev}(Z_Y) \quad (25)$$

che a sua volta implica che l'indice di determinazione

$$I^2 = \frac{\text{dev}(\hat{Y}_1)}{\text{dev}(Y_1)}$$

sia compreso nell'intervallo  $[0, 1]$ ;

b) e la formula indiretta per il calcolo della devianza spiegata secondo la quale

$$dev(\hat{Y}_1) = \tilde{\alpha}_2 \times codev(Y_1, Y_2). \quad (26)$$

D'altra parte, se consideriamo i valori riprodotti dalla funzione interpolante potenza, ovvero i valori

$$\hat{x}_{1i} = e^{\hat{y}_{1i}} = e^{\ln \tilde{\alpha}_1 + \tilde{\alpha}_2 \times y_{2i}} = \tilde{\alpha}_1 \times x_{2i}^{\tilde{\alpha}_2}, \quad i = 1, 2, \dots, n,$$

e i corrispondenti residui

$$z_i = x_{1i} - \hat{x}_{1i}, \quad i = 1, 2, \dots, n,$$

allora, salvo in casi particolari che non sono di interesse nelle applicazioni, si avrà

$$1) \sum_{i=1}^n z_i \neq 0 \Rightarrow M_1(Z) \neq 0,$$

$$2) codev(Z, X_2) \neq 0,$$

$$3) codev(Z, \hat{X}_1) \neq 0$$

e di conseguenza si avrà anche

$$dev(X_1) \neq dev(\hat{X}_1) + dev(Z)$$

e

$$dev(\hat{X}_1) \neq \tilde{\alpha}_2 \times codev(X_1, X_2).$$

Siccome si possono costruire degli esempi dove  $dev(\hat{X}_1) > dev(X_1)$ , non si può escludere che il rapporto tra la devianza dei valori riprodotti  $dev(\hat{X}_1)$  e la devianza della variabile dipendente  $dev(X_1)$

(che nel caso di una retta ai minimi quadrati viene chiamato indice di determinazione) assuma valori maggiori di 1.

I valori riprodotti dalla funzione interpolante

$$\hat{x}_1 = f(x_2) = \tilde{\alpha}_1 x_2^{\tilde{\alpha}_2}, \quad x_2 > 0,$$

*non* soddisfano la scomposizione della devianza, e il rapporto tra la devianza dei valori riprodotti  $dev(\hat{X}_1)$  e la devianza della variabile dipendente  $dev(X_1)$  può assumere valori maggiori di 1.

## 4.5 Applicazione: la regressione di Cobb e Douglas

Nel 1928 Cobb e Douglas pubblicarono i risultati di uno studio sull'economia degli Stati Uniti. In questo studio analizzarono la relazione di fondo tra le quantità di lavoro e capitale impegnate nella produzione e il volume della produzione durante il periodo storico compreso tra l'anno 1899 e l'anno 1922. Per quantificare il lavoro, capitale e il volume della produzione usarono degli indici annuali che Douglas aveva costruito a partire da varie fonti. Prendendo poi spunto da un teorema dell'economista britannico Wicksteed, ipotizzarono che la relazione di fondo tra il volume della produzione  $P$  e le quantità di lavoro  $L$  e capitale  $C$  impegnate nella produzione fosse descritta da una funzione del tipo

$$P = \alpha_1 \times L^{\alpha_2} \times C^{1-\alpha_2} \quad \text{con } \alpha_1 > 0 \text{ e con } 0 < \alpha_2 < 1. \quad (27)$$

Si osservi che questa ipotesi implica **rendimenti di scala costanti**: raddoppiando entrambi i fattori produttivi  $L$  e  $C$ , si raddoppia anche il volume della produzione  $P$  e, più in generale, moltiplicando entrambi i fattori produttivi  $L$  e  $C$  per una qualsiasi costante positiva  $h$ , risulterà moltiplicato per  $h$  anche il volume della produzione  $P$ .

**Dimostrazione.** Sostituendo  $L$  con  $h \times L$  e  $C$  con  $h \times C$ , il secondo membro della (27) diventa

$$\begin{aligned} \alpha_1 \times (h \times L)^{\alpha_2} \times (h \times C)^{1-\alpha_2} &= \alpha_1 \times h^{\alpha_2} \times L^{\alpha_2} \times h^{1-\alpha_2} \times C^{1-\alpha_2} = \\ &= h^{\alpha_2} \times h^{1-\alpha_2} \times \alpha_1 \times L^{\alpha_2} \times C^{1-\alpha_2} = h^{\alpha_2+1-\alpha_2} \times \alpha_1 \times L^{\alpha_2} \times C^{1-\alpha_2} = \\ &= h \times \alpha_1 \times L^{\alpha_2} \times C^{1-\alpha_2} = h \times P \end{aligned}$$

□

Dividendo ambo i membri della (27) per  $C$  vediamo che

$$P = \alpha_1 \times L^{\alpha_2} \times C^{1-\alpha_2} \quad \Leftrightarrow \quad (P/C) = \alpha_1 \times (L/C)^{\alpha_2}$$

e questa equazione mostra che secondo l'ipotesi fatta da Cobb e Douglas la relazione di fondo tra le variabili  $X_1 = P/C$  e  $X_2 = L/C$  dovrebbe essere descritta da una funzione potenza. In virtù di questo fatto Cobb e Douglas usarono le formule (23) e (24) per calcolare i valori dei parametri  $\alpha_1$  e  $\alpha_2$  di una funzione interpolante per i dati costruiti da Douglas. Nel resto di questa sezione replicheremo la regressione di Cobb e Douglas con dati trimestrali estratti dal sito della *Federal Reserve Bank of St. Louis* (vedi Figura 4.4). Come variabile che rappresenta il volume della produzione  $P$  considereremo il PIL reale degli Stati Uniti,<sup>7</sup> e come

<sup>7</sup>fonte: U.S. Bureau of Economic Analysis, Real Gross Domestic Product [GDPC1], serie storica scaricata dal sito della FRED, Federal Reserve Bank of St. Louis in data 24 febbraio 2021 (<https://fred.stlouisfed.org/series/GDPC1>).

variabili che rappresentano la quantità di lavoro  $L$  e la quantità di capitale  $C$  considereremo il numero di occupati<sup>8</sup> e gli investimenti fissi lordi,<sup>9</sup> rispettivamente.<sup>10</sup> Figura 4.4 mostra la parte iniziale e la parte finale del foglio di lavoro che contiene i dati che andremo ad analizzare.

Figura 4.4: Foglio "Dati rielaborati" della cartella di lavoro "Regressione di Cobb e Douglas".

	A	B	C	D	E	F
1	Dati rielaborati (cambio dell'unità di misura degli investimenti fissi lordi da USD a miliardi di USD)					
2	P = Prodotto interno lordo espresso in miliardi di USD concatenati con anno di riferimento 2012					
3	L = numero di occupati (in migliaia di unità)					
4	C = Investimenti fissi lordi espressi in miliardi di USD concatenati con anno di riferimento 2012					
5	Data	P	L	C	X1=P/C	X2=L/C
6	2010-01-01	15415,145	138590	2757,249	5,591	50,264
7	2010-04-01	15557,277	139226	2845,626	5,467	48,926
8	2010-07-01	15671,967	139338	2856,404	5,487	48,781
9	2010-10-01	15750,625	139155	2892,01	5,446	48,117
	A	B	C	D	E	F
47	2020-04-01	17302,511	137565,000	3795,484	4,559	36,244
48	2020-07-01	18596,521	146199,000	4016,493	4,630	36,400
49	2020-10-01	18780,325	149769	4163,881	4,510	35,969
50						
51	<b>Somme</b>	<b>762210,461</b>	<b>6505605</b>	<b>154909,647</b>	<b>217,770</b>	<b>1864,899</b>

I dati coprono il periodo storico che va dal primo trimestre del 2010 fino all'ultimo trimestre del 2020 per un totale di

<sup>8</sup>fonte: U.S. Bureau of Labor Statistics, Employment Level [CE16OV], serie storica scaricata dal sito della FRED, Federal Reserve Bank of St. Louis in data 24 febbraio 2021 (<https://fred.stlouisfed.org/series/CE16OV>).

<sup>9</sup>fonte: Organization for Economic Co-operation and Development, Gross Fixed Capital Formation in United States [USAGFCFQDSNAQ], serie storica scaricata dal sito della FRED, Federal Reserve Bank of St. Louis in data 24 febbraio 2021 (<https://fred.stlouisfed.org/series/USAGFCFQDSNAQ>).

<sup>10</sup>Sia il PIL reale che gli investimenti fissi lordi sono espressi in miliardi di USD concatenati con l'anno 2012 come anno di riferimento.

$n = 44$  trimestri. Per replicare la regressione di Cobb e Douglas costruiamo un nuovo foglio di calcolo che riporta i valori delle variabili derivate  $Y_1 = \ln X_1 = \ln(P/C)$  e  $Y_2 = \ln X_2 = \ln(L/C)$  nonché quelli di alcune quantità di sintesi che ci serviranno per calcolare le medie  $\bar{y}_1$  e  $\bar{y}_2$ , le devianze  $dev(Y_1)$  e  $dev(Y_2)$ , e la codevianza  $codev(Y_1, Y_2)$ . Figura 4.5 mostra le parti rilevanti di questo foglio di calcolo:

Figura 4.5: Foglio "Regressione CD" della cartella di lavoro "Regressione di Cobb e Douglas".

	E	F	G	H	I
1	Y1=ln(X1)=ln(P/C)	Y2=ln(X2)=ln(L/C)	Y1^2	Y2^2	Y1*Y2
2	1,7211	3,9173	2,9622	15,3451	6,7421
3	1,6987	3,8903	2,8857	15,1346	6,6087
4	1,7023	3,8873	2,8979	15,1114	6,6175
5	1,6949	3,8736	2,8728	15,0051	6,5655
6	1,6984	3,8815	2,8845	15,0658	6,5922
	E	F	G	H	I
43	1,5170	3,5903	2,3014	12,8901	5,4466
44	1,5326	3,5946	2,3488	12,9209	5,5089
45	1,5064	3,5826	2,2691	12,8354	5,3968
46					
47	<b>70,3054</b>	<b>164,7093</b>	<b>112,4591</b>	<b>616,8675</b>	<b>263,3676</b>

Usando i totali di colonna riportati nella riga 47 otteniamo

$$\bar{y}_1 = \frac{1}{n} \ln y_{1i} = \frac{70,3054}{44} = 1,5979,$$

$$\bar{y}_2 = \frac{1}{n} \ln y_{2i} = \frac{164,7093}{44} = 3,7434,$$

$$dev(Y_1) = \sum_{i=1}^n y_{1i}^2 - n \times \bar{y}_1^2$$

$$= 112,4591 - 44 \times 1,5979^2 = 0,1146,$$

$$\begin{aligned} dev(Y_2) &= \sum_{i=1}^n y_{2i}^2 - n \times \bar{y}_2^2 \\ &= 616,8675 - 44 \times 3,7434^2 = 0,2936, \end{aligned}$$

e

$$\begin{aligned} codev(Y_1, Y_2) &= \sum_{i=1}^n y_{1i} \times y_{2i} - n \times \bar{y}_1 \times \bar{y}_2 \\ &= 263,3676 - 44 \times 1,5979 \times 3,7434 = 0,1781. \end{aligned}$$

Sostituendo questi valori nelle formule (23) e (24) otteniamo

$$\tilde{\alpha}_2 = \frac{codev(Y_1, Y_2)}{dev(Y_2)} = \frac{0,1781}{0,2936} = 0,6066$$

e

$$\begin{aligned} \ln \tilde{\alpha}_1 &= \bar{y}_1 - \tilde{\alpha}_2 \times \bar{y}_2 \\ &= 1,5979 - 0,6066 \times 3,7434 = -0,6728 \\ \Rightarrow \tilde{\alpha}_1 &= e^{\bar{y}_1 - \tilde{\alpha}_2 \times \bar{y}_2} = e^{-0,6728} = 0,5103. \end{aligned}$$

L'equazione della retta ai minimi quadrati che interpola i logaritmi delle osservazioni delle variabili  $X_1 = P/C$  e  $X_2 = L/C$  è quindi data da

$$\hat{y}_1 = -0,6728 + 0,6066 \times y_2 \quad (28)$$

Sostituendo in questa equazione  $\hat{y}_1 = \ln \hat{x}_1$  e  $y_2 = \ln x_2$  (ricordiamo che  $Y_1$  e  $Y_2$  rappresentano i logaritmi delle variabili  $X_1$  e  $X_2$ ) e applicando la funzione esponenziale  $g(x) = e^x$  ad ambo i membri dell'equazione così ottenuta, otteniamo anche l'equazione della corrispondente funzione interpolante potenza:

$$\hat{x}_1 = 0,5103 \times x_2^{0,6066}, \quad x_2 > 0. \quad (29)$$



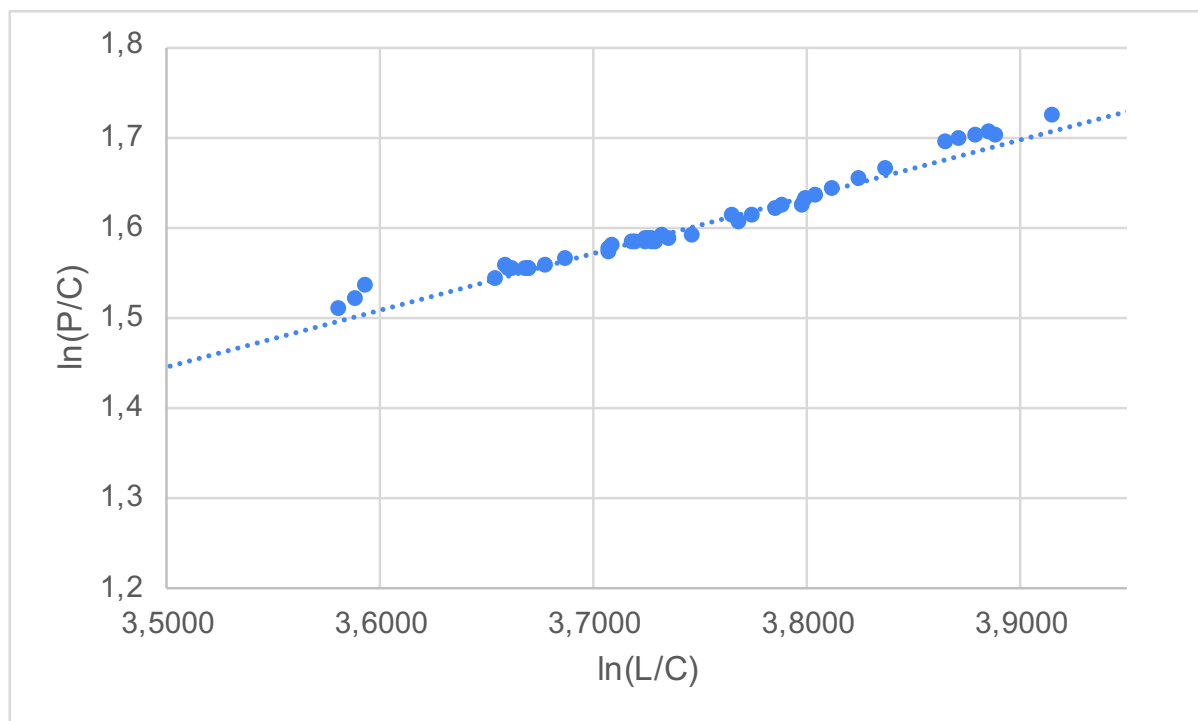
Ponendo poi  $\hat{x}_1 = \hat{p}/c$  e  $x_2 = l/c$  (ricordiamo che  $X_1 = P/C$  e che  $X_2 = L/C$ ) e risolvendo per  $\hat{p}$ , otteniamo infine anche l'equazione di una funzione interpolante che descrive la relazione di fondo tra le osservazioni del volume della produzione  $P$  e le osservazioni delle quantità di lavoro  $L$  e capitale  $C$  impegnate nella produzione:

$$\hat{p} = 0,5103 \times l^{0,6066} \times c^{1-0,6066}, \quad l, c > 0. \quad (30)$$

Ovviamente la bontà d'adattamento di ciascuna delle tre funzioni interpolanti definite dalle equazioni (28), (29) e (30) è strettamente legata alla bontà d'adattamento delle altre due. Per semplicità in quanto segue analizzeremo in modo dettagliato solo la bontà d'adattamento della funzione interpolante lineare nella (28), mentre per le altre due funzioni interpolanti ci limiteremo al calcolo degli indici di adattamento  $A_2$  e  $A_2^*$ .

Figura 4.6 mostra l'andamento della funzione interpolante (28) sovrapposto alla nuvola di punti generata dalle osservazioni delle variabili  $Y_1 = \ln X_1 = \ln(P/C)$  e  $Y_2 = \ln X_2 = \ln(L/C)$ :

Figura 4.6: Foglio "Regressione CD" della cartella di lavoro "Regressione di Cobb e Douglas".



Questo grafico suggerisce un ottimo adattamento ma per avere un quadro più chiaro procederemo comunque alla costruzione del grafico dei residui. A tal fine aggiungiamo nel foglio di calcolo in Figura 4.4

- la colonna J con i valori riprodotti

$$\hat{y}_{1i} = \ln \tilde{\alpha}_1 + \tilde{\alpha}_2 \times y_{2i} = -0,6728 + 0,6066 \times y_{2i}$$

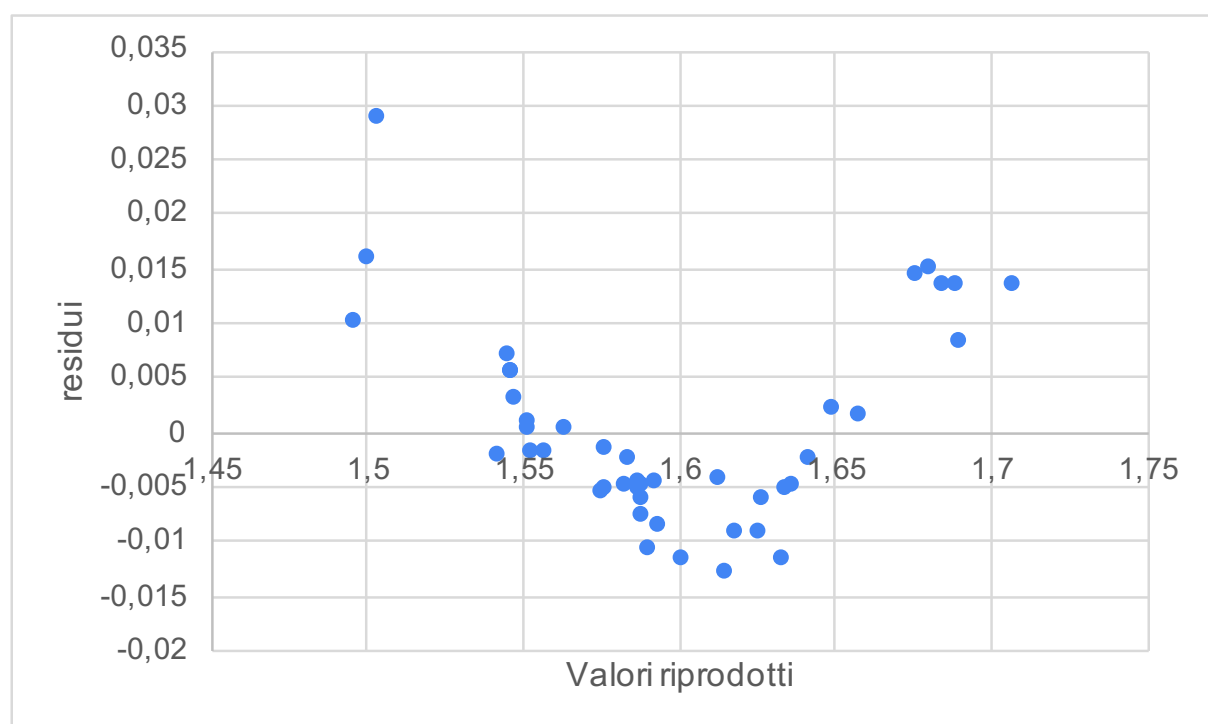
- e la colonna K con i residui  $z_{Yi} = y_{1i} - \hat{y}_{1i}$ .

Figura 4.7:

	E	F	G	H	I	J	K
1	$Y1=\ln(X1)=\ln(P/C)$	$Y2=\ln(X2)=\ln(L/C)$	$Y1^2$	$Y2^2$	$Y1*Y2$	$Y1_{cap}$	$Zy = Y1-Y1_{cap}$
2	1,7211	3,9173	2,9622	15,3451	6,7421	1,7077	0,0134
3	1,6987	3,8903	2,8857	15,1346	6,6087	1,6906	0,0081
4	1,7023	3,8873	2,8979	15,1114	6,6175	1,6888	0,0135
5	1,6949	3,8736	2,8728	15,0051	6,5655	1,6801	0,0148
	E	F	G	H	I	J	K
43	1,5170	3,5903	2,3014	12,8901	5,4466	1,5011	0,0159
44	1,5326	3,5946	2,3488	12,9209	5,5089	1,5038	0,0287
45	1,5064	3,5826	2,2691	12,8354	5,3968	1,4963	0,0100
46							
47	<b>70,3054</b>	<b>164,7093</b>	<b>112,4591</b>	<b>616,8675</b>	<b>263,3676</b>	<b>70,3054</b>	<b>0,0000</b>

Riportando i valori riprodotti  $\hat{y}_{1i}$  in ascissa e i residui  $z_{Yi}$  in ordinata otteniamo il grafico dei residui:

Figura 4.8:



Guardando il grafico dei residui notiamo che la successione dei segni dei residui non è casuale e questo fatto suggerisce che la retta

ai minimi quadrati abbia un andamento troppo rigido. Questo problema potrebbe essere affrontato mediante il metodo descritto nella Sezione 2.4, ma per non appesantire troppo la trattazione sorvoleremo su questo problema.

Passiamo dunque all'identificazione di eventuali residui "sospetti". Per calcolare l'indice  $A_2$  associato alla retta ai minimi quadrati nella (28) dobbiamo innanzitutto calcolare la devianza residua  $dev(Z_Y)$ . Dalla scomposizione della devianza nella (25) vediamo che

$$dev(Z_Y) = dev(Y_1) - dev(\hat{Y}_1)$$

e usando la formula (26) otteniamo

$$\begin{aligned} dev(\hat{Y}_1) &= \tilde{\alpha}_2 \times codev(Y_1, Y_2) \\ &= 0,6066 \times 0,1781 = 0,1080 \end{aligned}$$

Ricordando che  $dev(Y_1) = 0,1146$  otteniamo dunque

$$\begin{aligned} dev(Z_Y) &= dev(Y_1) - dev(\hat{Y}_1) \\ &= 0,1146 - 0,1080 = 0,0066 \end{aligned}$$

$$\Rightarrow A_{Y_2} = \sqrt{\frac{dev(Z_Y)}{n}} = \sqrt{\frac{0,0066}{44}} = 0,0122$$

Moltiplicando per 3 il valore dell'indice  $A_{Y_2}$  vediamo ora che le due linee orizzontali che delimitano l'area dei residui non "sospetti" si trovano in corrispondenza delle ordinate  $\pm 3 \times A_{Y_2} = \pm 0,0366$ , ovvero al di fuori dell'area visibile nel grafico in Figura 4.7. Quindi non ci sono residui che ricadono nell'area "sospetta" e pertanto non c'è motivo di ritenere che tra i dati analizzati ci siano delle osservazioni anomale. Per concludere la fase di valu-

tazione della bontà d'adattamento della retta ai minimi quadrati definita dall'equazione (28), calcoliamo infine anche l'indice di determinazione:

$$I^2 = \frac{dev(\hat{Y}_1)}{dev(Y_1)} = \frac{0,1080}{0,1146} = 0,9424.$$

Questo valore è molto elevato e ci dice che la retta ai minimi quadrati spiega il 94,24% della variabilità dei valori di  $Y_1 = \ln X_1 = \ln(P/C)$ .

Come promesso, procederemo ora anche al calcolo degli indici  $A_2$  e  $A_2^*$  per le funzioni interpolanti nella (29) e nella (30). A tal fine dobbiamo calcolare

- i valori riprodotti (vedi la colonna L della tabella in Figura 4.9)

$$\hat{x}_{1i} = e^{\hat{y}_{1i}} = e^{\hat{\beta}_1} \times x_{2i}^{\hat{\beta}_2} = 0,5103 \times x_{2i}^{0,6066};$$

- i residui  $z_i = x_{1i} - \hat{x}_{1i}$  (colonna M) e i loro quadrati  $z_i^2$  (colonna N);
- i quadrati dei valori  $x_{1i}$  (colonna O);
- i valori riprodotti (colonna P)

$$\hat{p}_i = \hat{x}_{1i} \times c_i = e^{\hat{\beta}_1} \times l_i^{\hat{\beta}_2} \times c_i^{1-\hat{\beta}_2} = 0,5103 \times l_i^{0,6066} \times c_i^{1-0,6066},$$

dove  $l_i$  e  $c_i$  sono le osservazioni delle quantità di lavoro e capitale  $L$  e  $C$ , rispettivamente;

- i residui  $z_{P_i} = p_i - \hat{p}_i$  (colonna Q;  $p_i$  indica le osservazioni per il volume della produzione  $P$ ) e i loro quadrati  $z_{P_i}^2$  (colonna R);
- e infine i quadrati delle osservazioni  $p_i$  (colonna S).

Figura 4.9:

	L	M	N	O	P	Q	R	S
1	X1cap=exp(Y1cap)	Z=X1-X1cap	Z^2	X1^2	Pcap	Zp=P-Pcap	Zp^2	P^2
2	5,516	0,075	0,006	31,257	15209	206	42344	237626695
3	5,423	0,044	0,002	29,889	15432	126	15758	242028868
4	5,413	0,074	0,005	30,103	15461	211	44463	245610550
5	5,266	0,090	0,008	29,667	15510	222	52676	248927188
	L	M	N	O	P	Q	R	S
43	4,487	0,072	0,005	20,782	17030	273	74402	299376887
44	4,499	0,131	0,017	21,437	18070	526	277120	345830593
45	4,465	0,045	0,002	20,343	18593	188	35176	352700607
46								
47	<b>217,758</b>	<b>0,012</b>	<b>0,088</b>	<b>1080,905</b>	<b>762242</b>	<b>-31</b>	<b>1035621</b>	<b>13263458983</b>

Usando la somma dei quadrati dei residui in fondo alla colonna N vediamo che il valore dell'indice  $A_2$  associato alla funzione interpolante potenza nella (29) è dato da

$$A_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2} = \sqrt{\frac{0,088}{44}} = 0,0447$$

Questo valore ci dice che l'ordine di grandezza medio dei residui della funzione interpolante (29) è pari a 0,0447 unità (in questo caso l'indice  $A_2$  è un numero privo di unità di misura perché le osservazioni della variabile dipendente  $X_1 = P/C$  non hanno un'unità di misura). Per calcolare il valore dell'indice  $A_2^*$  associato alla funzione interpolante (29), dobbiamo innanzitutto calcolare

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i} = \frac{217,770}{44} = 4,949$$

(il valore di  $\sum_{i=1}^n x_{1i} = 217,770$  si trova in fondo alla colonna E nella tabella in Figura 4.4)

$$\begin{aligned} \Rightarrow \text{dev}(X_1) &= \sum_{i=1}^n x_{1i}^2 - n \times \bar{x}_1^2 \\ &= 1080,905 - 44 \times 4,949^2 = 3,231. \end{aligned}$$

(il valore di  $\sum_{i=1}^n x_{1i}^2 = 1080,905$  si trova in fondo alla colonna O). Usando i valori di  $\sum_{i=1}^n z_i^2 = 0,088$  (in fondo alla colonna N) e di  $\text{dev}(X_1) = 3,231$  otteniamo

$$A_2^* = \sqrt{\frac{\sum_{i=1}^n z_i^2}{\text{dev}(X_1)}} = \sqrt{\frac{0,088}{3,231}} = 0,165.$$

Questo valore ci dice che l'ordine di grandezza medio dei residui della funzione interpolante (29) è pari al 16,5% del valore dello scarto quadratico medio delle osservazioni della variabile  $X_1 = P/C$ .

Procedendo in modo analogo otteniamo anche i valori degli indici  $A_2$  e  $A_2^*$  associati alla funzione interpolante nella (30):

$$A_{P_2} = \sqrt{\frac{1}{n} \sum_{i=1}^n z_{P_i}^2} = \sqrt{\frac{1035621}{44}} = 153,417$$

(la somma  $\sum_{i=1}^n z_{P_i}^2 = 1035621$  si trova in fondo alla colonna R); e

$$\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i = \frac{762210,461}{44} = 17322,965 \text{ miliardi di USD}$$

(il valore di  $\sum_{i=1}^n p_i = 762210,461$  si trova in fondo alla colonna B nella tabella in Figura 4.4)

$$\begin{aligned} \Rightarrow \text{dev}(P) &= \sum_{i=1}^n p_i^2 - n \times \bar{p}^2 \\ &= 13263458983 - 44 \times 17322,965^2 = 597138612 \end{aligned}$$

(il valore di  $\sum_{i=1}^n p_i^2 = 13263458983$  si trova in fondo alla colonna S)

$$\Rightarrow A_{P_2}^* = \sqrt{\frac{\sum_{i=1}^n z_{P_i}^2}{dev(P)}} = \sqrt{\frac{1035621}{597138612}} = 0,042.$$

Il valore di  $A_{P_2} = 153,417$  ci dice che i valori riprodotti dalla funzione interpolante (30) si discostano mediamente di 153,417 miliardi di USD dai valori osservati per il volume della produzione  $P$ , mentre il valore di  $A_{P_2}^* = 0,042$  ci dice che lo scarto quadratico medio dei residui della funzione interpolante (30) è pari al 4,2% del valore dello scarto quadratico medio delle osservazioni del volume della produzione  $P$ .

Passiamo ora alla fase interpretativa. Siccome siamo interessati solo alle variabili  $P$ ,  $L$ ,  $C$ ,  $X_1 = P/C$  e  $X_2 = L/C$  commenteremo solo l'andamento delle funzioni interpolanti (29) e (30). Ricordiamo che le equazioni di queste due funzioni interpolanti sono rispettivamente date da

$$\hat{x}_1 = 0,5103 \times x_2^{0,6066}, \quad x_2 > 0.$$

e da

$$\hat{p} = 0,5103 \times l^{0,6066} \times c^{1-0,6066}, \quad l, c > 0.$$

In entrambe queste funzioni interpolanti il parametro  $\tilde{\alpha}_1 = 0,5103$  è un parametro di scala che ci dice a quanto ammontano i loro valori riprodotti quando le rispettive variabili indipendenti assumono il valore 1. Il parametro  $\tilde{\alpha}_2 = 0,6066$  descrive invece la sensibilità dei valori riprodotti a variazioni della o delle variabili indipendenti. Nella prima funzione interpolante  $\tilde{\alpha}_2 = 0,6066$  rappresenta l'elasticità di  $\hat{x}_1$  rispetto a  $x_2$ . Quindi possiamo dire che secondo la funzione interpolante potenza, a fronte di un incremento dell'1% del rapporto tra la quantità di lavoro e la quantità capitale (questo rapporto è rappresentato dalla variabile  $x_2 = l/c$ ), il valore del rapporto tra il volume della produzione e la quantità di capitale

(questo rapporto è rappresentato dalla variabile  $\hat{x}_1$ ) aumenta dello 0,6066%.

Per comprendere il ruolo di  $\tilde{\alpha}_2 = 0,6066$  nella seconda funzione interpolante conviene prendere il logaritmo ad ambo i membri della sua equazione. In questo modo si ottiene

$$\ln \hat{p} = \ln 0,5103 + 0,6066 \times \ln l + (1 - 0,6066) \times \ln c$$

e da questa equazione evince che

$$\Delta \ln \hat{p} = 0,6066 \times \Delta \ln l + (1 - 0,6066) \times \Delta \ln c.$$

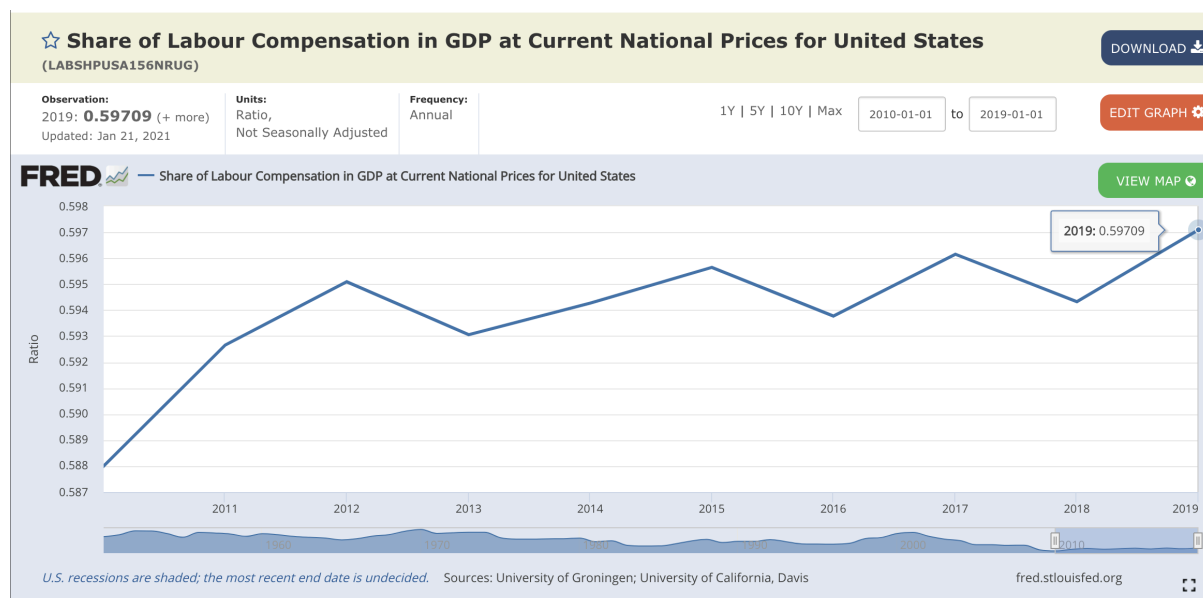
Ricordando la relazione tra variazioni logaritmiche e variazioni relative (vedi Sezione 4.2), possiamo quindi interpretare il valore di  $\tilde{\alpha}_2 = 0,6066$  come l'ammontare (espresso in punti percentuali) della variazione relativa subita da  $\hat{p}$  a fronte di un aumento della variabile  $l$  dell'1%, fermo restando il valore della variabile  $c$  (se la variabile  $c$  rimane costante, si ha  $\Delta \ln c = 0$ ).

Sotto le ipotesi della teoria economica neoclassica o marginalista, il valore del parametro  $\alpha_2$  nella funzione di produzione (27) dovrebbe rappresentare la quota della produzione  $P$  che viene usata per la remunerazione del lavoro. In questa dispensa omettiamo la giustificazione di questa interpretazione. Tuttavia, cogliamo l'occasione per osservare che negli anni 2010 - 2020 la quota dei salari sul PIL statunitense oscillava tra il 58,8% e il 59,709% (vedi Figura 4.10), percentuali che sono molto prossime al valore  $\tilde{\alpha}_2 = 0,6066$  che abbiamo ottenuto nella nostra regressione.

Anche il valore di  $\tilde{\alpha}_2$  che trovarono Cobb e Douglas è molto prossimo ad una stima "diretta" della quota dei salari sul valore della produzione. Infatti, Cobb e Douglas ottennero  $\tilde{\alpha}_2 = 0,75$ , un valore che è praticamente identico alla stima del *National Bureau of Economic Research* per la quota dei salari sul valore aggiunto del settore manifatturiero durante il periodo 1909-1918.



Figura 4.10: Quota dei salari sul PIL statunitense. Fonte: University of Groningen and University of California, Davis; Share of Labour Compensation in GDP at Current National Prices for United States [LABSHPUSA156NRUG]; serie storica scaricata dal sito della FRED, Federal Reserve Bank of St. Louis in data 22 febbraio 2021 (<https://fred.stlouisfed.org/series/LABSHPUSA156NRUG>).



**Esercizio 4.1.** Si replichi l'analisi di regressione presentata in questa sezione usando i dati contenuti nel foglio "Italia" della cartella di lavoro "Regressione di Cobb e Douglas".  $\triangle$

## 5 Il piano ai minimi quadrati

Supponiamo ora di disporre di  $n$  terne di osservazioni di *tre* variabili  $X_1$ ,  $X_2$  e  $X_3$  e di essere interessati ad una funzione interpolante

$$\hat{x}_1 = f(x_2, x_3; \hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_g)$$

che descriva l'andamento della relazione di fondo tra le osservazioni delle tre variabili. Come osservavamo nella Sezione 2, per

inferire l'andamento della relazione di fondo si potrebbe tentare di rappresentare le  $n$  terne di osservazioni

$$(x_{11}, x_{21}, x_{31}), \quad \dots, \quad (x_{1i}, x_{2i}, x_{3i}), \quad \dots, \quad (x_{1n}, x_{2n}, x_{3n})$$

come dei punti in uno spazio tridimensionale, ma a partire da una tale rappresentazione grafica è spesso molto difficile trarre indicazioni utili. Per questo motivo, nelle analisi di regressione che coinvolgono più di due variabili solitamente si parte dalla famiglia dei piani

$$x_1 = f(x_2, x_3; \alpha_1, \alpha_2, \alpha_3) = \alpha_1 + \alpha_2 \times x_2 + \alpha_3 \times x_3,$$

$$x_1, x_2 \in \mathbb{R}, \quad \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R},$$

rimandando possibili miglioramenti ad una fase successiva (quando si potranno sfruttare le indicazioni fornite dal grafico dei residui). Visto il ruolo centrale dei piani, in questa sezione studieremo dunque le proprietà dei cosiddetti *piani ai minimi quadrati*, ovvero le proprietà dei piani

$$\hat{x}_1 = \hat{\alpha}_1 + \hat{\alpha}_2 \times x_2 + \hat{\alpha}_3 \times x_3$$

che si ottengono applicando il metodo di accostamento dei minimi quadrati.

## 5.1 Determinazione dei parametri del piano ai minimi quadrati

Volendo applicare il metodo dei minimi quadrati per determinare i parametri  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$  e  $\hat{\alpha}_3$  di un piano ai minimi quadrati ci si trova

di fronte al problema di minimizzare la distanza quadratica

$$D(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^N (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \alpha_3 x_{3i})^2. \quad (31)$$

rispetto ai parametri  $\alpha_1$ ,  $\alpha_2$  e  $\alpha_3$ .

Per determinare i parametri  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$  e  $\hat{\alpha}_3$  che identificano un piano ai minimi quadrati procederemo in modo analogo a come abbiamo fatto nel caso della retta ai minimi quadrati (vedi Sezione 3.2). Innanzitutto osserviamo che anche la funzione  $D(\alpha_1, \alpha_2, \alpha_3)$  è differenziabile e che i suoi punti di minimo devono quindi trovarsi in corrispondenza di una terna  $(\alpha_1, \alpha_2, \alpha_3)$  che annulla le tre derivate parziali

$$\begin{aligned} \frac{\partial D}{\partial \alpha_1} &= 2 \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \alpha_3 x_{3i}) \times (-1) \\ \frac{\partial D}{\partial \alpha_2} &= 2 \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \alpha_3 x_{3i}) \times (-x_{2i}) \\ \frac{\partial D}{\partial \alpha_3} &= 2 \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \alpha_3 x_{3i}) \times (-x_{3i}). \end{aligned} \quad (32)$$

Inoltre osserviamo che anche in questo caso la funzione  $D(\alpha_1, \alpha_2, \alpha_3)$  è convessa (omettiamo la dimostrazione) e che quindi qualunque terna  $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$  che annulla le tre derivate parziali deve essere un punto di minimo della funzione  $D(\alpha_1, \alpha_2, \alpha_3)$ . Risolvendo il sistema di equazioni che si ottiene annullando le tre derivate parziali, ovvero risolvendo il sistema

$$\begin{cases} 2 \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) \times (-1) = 0 \\ 2 \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) \times (-x_{2i}) = 0 \\ 2 \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) \times (-x_{3i}) = 0 \end{cases}$$

troviamo dunque tutte le terne  $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$  che identificano un piano ai minimi quadrati.

Come primo passo per trovare le soluzioni di questo sistema dividiamo tutte le sue equazioni per  $(-2)$ . In questo modo possiamo riscrivere il sistema nella forma

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) \times x_{2i} = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 x_{2i} - \hat{\alpha}_3 x_{3i}) \times x_{3i} = 0. \end{cases} \quad (33)$$

A questo punto spezziamo la sommatoria nella prima equazione onde ottenere

$$\begin{aligned} \sum_{i=1}^n x_{1i} - n \times \hat{\alpha}_1 - \sum_{i=1}^n \hat{\alpha}_2 x_{2i} - \sum_{i=1}^n \hat{\alpha}_3 x_{3i} &= 0 \quad \Rightarrow \\ \frac{1}{n} \sum_{i=1}^n x_{1i} - \hat{\alpha}_1 - \hat{\alpha}_2 \times \frac{1}{n} \sum_{i=1}^n x_{2i} - \hat{\alpha}_3 \times \frac{1}{n} \sum_{i=1}^n x_{3i} &= 0 \quad \Rightarrow \\ \bar{x}_1 &= \hat{\alpha}_1 + \hat{\alpha}_2 \bar{x}_2 + \hat{\alpha}_3 \bar{x}_3. \end{aligned} \quad (34)$$

L'equazione (34) ci dice che in corrispondenza di  $x_2 = \bar{x}_2$  e  $x_3 = \bar{x}_3$  ogni piano ai minimi quadrati deve assumere il valore  $\bar{x}_1$ .<sup>11</sup>

Ogni piano ai minimi quadrati passa per il punto  $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ .

Risolvendo l'equazione (34) per  $\hat{\alpha}_1$  vediamo che l'intercetta di un piano ai minimi quadrati è data da

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \bar{x}_2 - \hat{\alpha}_3 \bar{x}_3. \quad (35)$$

<sup>11</sup>Si ricordi che per la retta ai minimi quadrati vale un risultato analogo: ogni retta ai minimi quadrati deve necessariamente passare per il punto determinato dalle medie delle due variabili.

Sostituendo ora questo risultato nelle ultime due equazioni del sistema (33) otteniamo il sistema

$$\begin{cases} \sum_{i=1}^n [x_{1i} - (\bar{x}_1 - \hat{\alpha}_2\bar{x}_2 - \hat{\alpha}_3\bar{x}_3) - \hat{\alpha}_2x_{2i} - \hat{\alpha}_3x_{3i}] \times x_{2i} = 0 \\ \sum_{i=1}^n [x_{1i} - (\bar{x}_1 - \hat{\alpha}_2\bar{x}_2 - \hat{\alpha}_3\bar{x}_3) - \hat{\alpha}_2x_{2i} - \hat{\alpha}_3x_{3i}] \times x_{3i} = 0 \end{cases} \quad (36)$$

che non contiene più l'incognita  $\hat{\alpha}_1$ . Sviluppando i prodotti e spezzando le sommatorie possiamo riscrivere questo sistema nella forma

$$\begin{cases} \hat{\alpha}_2 \times \sum_{i=1}^n (x_{2i} - \bar{x}_2) \times x_{2i} + \hat{\alpha}_3 \times \sum_{i=1}^n (x_{3i} - \bar{x}_3) \times x_{2i} = \\ = \sum_{i=1}^n (x_{1i} - \bar{x}_1) \times x_{2i} \\ \hat{\alpha}_2 \times \sum_{i=1}^n (x_{2i} - \bar{x}_2) \times x_{3i} + \hat{\alpha}_3 \times \sum_{i=1}^n (x_{3i} - \bar{x}_3) \times x_{3i} = \\ = \sum_{i=1}^n (x_{1i} - \bar{x}_1) \times x_{3i}. \end{cases}$$

A questo punto notiamo che le sommatorie in quest'ultimo sistema sono delle formule alternative per calcolare le devianze e codevianze delle osservazioni delle variabili coinvolte (vedi Sezione 3.1). Quindi riscriviamo il sistema nella forma

$$\begin{cases} \hat{\alpha}_2 \times dev(X_2) + \hat{\alpha}_3 \times codev(X_2, X_3) = codev(X_1, X_2) \\ \hat{\alpha}_2 \times codev(X_2, X_3) + \hat{\alpha}_3 \times dev(X_3) = codev(X_1, X_3) \end{cases}$$

e dividendo per  $n$  vediamo che il sistema può essere riscritto anche in termini di varianze e covarianze:

$$\begin{cases} \hat{\alpha}_2 \times \sigma_{22} + \hat{\alpha}_3 \times \sigma_{23} = \sigma_{12} \\ \hat{\alpha}_2 \times \sigma_{23} + \hat{\alpha}_3 \times \sigma_{33} = \sigma_{13}. \end{cases} \quad (37)$$

Trattandosi di un sistema lineare, possiamo esprimerlo in forma matriciale scrivendo

$$\begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{bmatrix} \times \begin{bmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} \sigma_{12} \\ \sigma_{13} \end{bmatrix}, \quad (38)$$

Secondo il teorema di Rouché-Capelli questo sistema ha un' *unica soluzione* per il vettore che contiene le due incognite  $\hat{\alpha}_2$  e  $\hat{\alpha}_3$  se e solo se il determinante della matrice dei termini noti è diverso da zero:<sup>12</sup>

$$\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix} = \sigma_{22}\sigma_{33} - \sigma_{23}^2 \neq 0$$

In questo caso, secondo la regola di Cramer, la soluzione sarà data dal vettore con componenti date da

$$\hat{\alpha}_2 = \frac{\begin{vmatrix} \sigma_{12} & \sigma_{23} \\ \sigma_{13} & \sigma_{33} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix}} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \quad (39)$$

$$\hat{\alpha}_3 = \frac{\begin{vmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{23} & \sigma_{13} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix}} = \frac{\sigma_{22}\sigma_{13} - \sigma_{23}\sigma_{12}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}. \quad (40)$$

---

<sup>12</sup>Si noti che in questo caso il determinante deve necessariamente essere positivo. Infatti, secondo la disuguaglianza di Cauchy-Schwartz si ha sempre  $\sigma_{22}\sigma_{33} \geq \sigma_{23}^2$  ovvero  $\sigma_{22}\sigma_{33} - \sigma_{23}^2 \geq 0$ .

Se il determinante della matrice di varianza e covarianza delle variabili esplicative non è nullo, ovvero se

$$\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix} = \sigma_{22}\sigma_{33} - \sigma_{23}^2 \neq 0, \quad (41)$$

allora esiste un unico piano ai minimi quadrati i cui parametri sono dati da

$$\hat{\alpha}_2 = \frac{\begin{vmatrix} \sigma_{12} & \sigma_{23} \\ \sigma_{13} & \sigma_{33} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix}} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \quad (42)$$

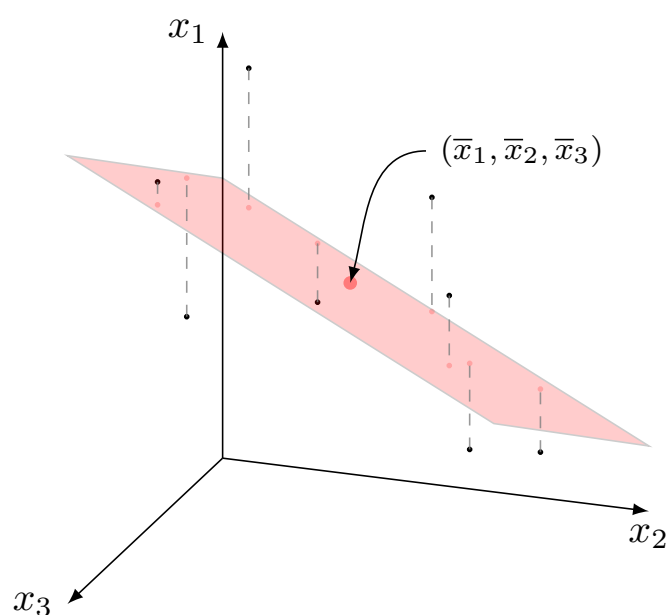
$$\hat{\alpha}_3 = \frac{\begin{vmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{23} & \sigma_{13} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix}} = \frac{\sigma_{22}\sigma_{13} - \sigma_{23}\sigma_{12}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}. \quad (43)$$

e

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2\bar{x}_2 - \hat{\alpha}_3\bar{x}_3. \quad (44)$$

Il grafico in Figura 5.1 illustra una situazione dove il determinante della matrice di varianza e covarianza non è nullo (e quindi positivo).

Figura 5.1:



Si noti che i punti sono dispersi nelle tre dimensioni dello spazio e che il piano ai minimi quadrati è sospeso in una sorta di posizione di equilibrio in mezzo ai punti. Come previsto dall'equazione (34), il piano ai minimi quadrati passa per il punto determinato dalle medie delle tre variabili.

Consideriamo ora invece il caso in cui il determinante nella (41) è nullo, ovvero il caso in cui

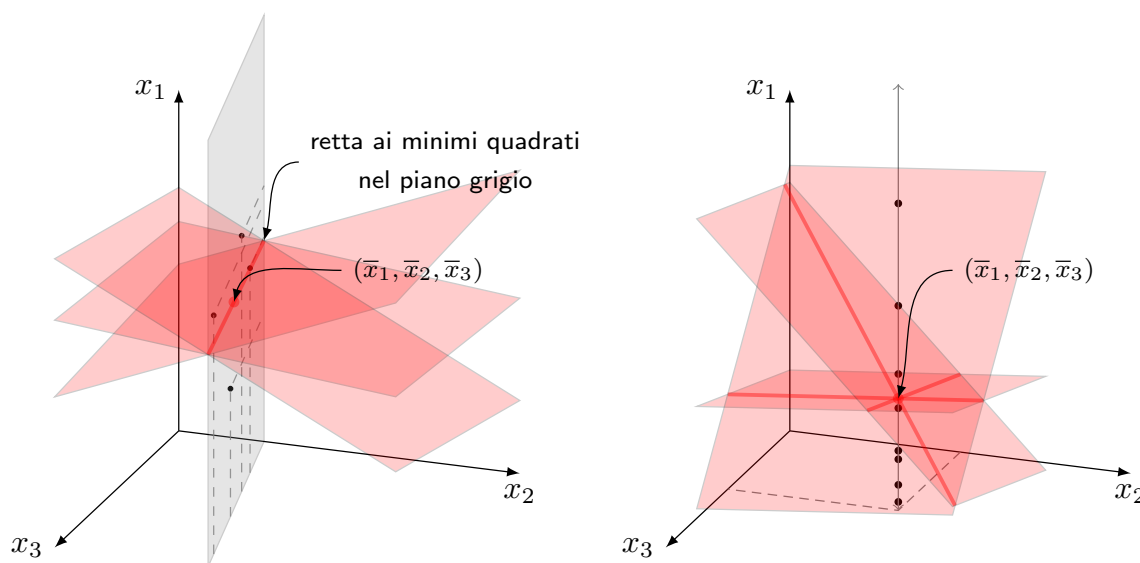
$$\sigma_{22} \times \sigma_{33} - \sigma_{23}^2 = 0 \quad \Leftrightarrow \quad \sigma_{23}^2 = \sigma_{22} \times \sigma_{33}. \quad (45)$$

Come vedremo tra breve, questo caso è analogo al caso in cui la varianza della variabile esplicativa di una retta ai minimi quadrati è nulla. Per non appesantire troppo la trattazione, analizzeremo questo caso soltanto dal punto di vista geometrico.



Dalla [disuguaglianza di Cauchy-Schwarz](#) (espressa di termini di varianze e covarianze) discende immediatamente che la condizione (45) è soddisfatta se e solo se le osservazioni delle due variabili esplicative  $X_2$  e  $X_3$  sono *collineari*, ovvero se e solo le osservazioni delle due variabili esplicative sono allineate lungo una retta nel piano  $(x_2, x_3)$ . Non è difficile rendersi conto che in questo caso tutti i punti  $(x_{1i}, x_{2i}, x_{3i})$  devono trovarsi all'interno di un piano che è verticale rispetto al piano  $(x_2, x_3)$  o addirittura allineati lungo una retta verticale (nel caso particolare in cui si avesse  $dev(X_2) = dev(X_3) = 0$ ). I due grafici in [Figura 5.2](#) illustrano questi due casi:

Figura 5.2:



Come si vede in [Figura 5.2](#), in entrambi i casi esistono infiniti piani ai minimi quadrati. Nel primo caso sono dei piani ai minimi quadrati tutti i piani che intersecano il piano verticale che contiene la nuvola di punti in corrispondenza della retta ai minimi quadrati

che interpola la nuvola di punti all'interno del piano verticale che la contiene (ovvero in corrispondenza della retta rossa nel piano grigio), mentre nel secondo caso è un piano ai minimi quadrati qualunque piano che passa per il punto determinato dalle medie delle tre variabili. Si osservi che in entrambi i casi ogni piano ai minimi quadrati passa per il punto determinato dalle medie delle tre variabili così come previsto dall'equazione (34).

Ovviamente, casi di perfetta collinearità tra le osservazioni di due variabili non si verificano praticamente mai nelle applicazioni<sup>13</sup> e qualora si verificassero non avrebbe comunque senso procedere all'analisi di un piano ai minimi quadrati perché quest'ultimo non sarebbe univocamente determinato. Per questo motivo, onde evitare inutili complicazioni, (se non diversamente indicato) d'ora in poi assumeremo sempre che le osservazioni di qualunque coppia di variabili non siano collineari. Non è difficile verificare che questa ipotesi escluda la possibilità che la devianza di una variabile possa essere nulla e che escluda anche la possibilità che un coefficiente di correlazione lineare possa assumere i valori  $\pm 1$ .

Tuttavia, non di rado nelle applicazioni si verificano casi di *quasi-collinearità*, ovvero nuvole di punti che sono assai prossime ad un piano verticale. Come nel caso dove la devianza della variabile esplicativa di una retta ai minimi quadrati è molto piccola, anche in caso di *quasi-collinearità* gli errori di rilevazione (oppure gli arrotondamenti) possono avere un'influenza notevole sui parametri di un piano ai minimi quadrati. Per assicurarsi che gli errori di rilevazione o gli arrotondamenti non abbiano causato delle variazioni drastiche nei valori dei parametri bisogna, per ciascuna variabile esplicativa, verificare che l'entità di eventuali errori di rilevazione sia trascurabile rispetto allo scarto quadratico medio dei residui della retta ai minimi quadrati che considera la variabile esplicativa in questione come variabile dipendente, e l'altra variabile esplicativa come

---

<sup>13</sup>I motivi per i quali si potrebbero verificare casi di perfetta collinearità sono analoghi a quelli per i quali la devianza di una variabile potrebbe essere nulla (vedi Sezione 3.2)

variabile indipendente. Quindi, l'entità di un eventuale errore di rilevazione nelle osservazioni della variabile  $X_2$  deve essere confrontata con  $\sqrt{\text{var}(Z_{2.3})}$ , dove  $Z_{2.3}$  è la variabile che rappresenta i residui della retta ai minimi quadrati che descrive  $X_2$  in funzione di  $X_3$ . Allo stesso modo, l'entità di un eventuale errore di rilevazione nelle osservazioni della variabile  $X_3$  deve essere confrontata con  $\sqrt{\text{var}(Z_{3.2})}$ , dove  $Z_{3.2}$  è la variabile che rappresenta i residui della retta ai minimi quadrati che descrive  $X_3$  in funzione di  $X_2$ . Siccome  $\sqrt{\text{var}(Z_{2.3})} \leq \sqrt{\text{var}(X_2)}$ , potrebbe succedere che un errore di rilevazione in una osservazione della variabile  $X_2$  abbia un impatto trascurabile sui valori dei parametri della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$ , ma che abbia un impatto rilevante sui parametri del piano ai minimi quadrati che oltre a  $X_2$  considera anche  $X_3$  come variabile esplicativa. Allo stesso modo, siccome  $\sqrt{\text{var}(Z_{3.2})} \leq \sqrt{\text{var}(X_3)}$ , potrebbe succedere che un errore di rilevazione in una osservazione della variabile  $X_3$  abbia un impatto trascurabile sui valori dei parametri della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_3$ , ma che abbia un impatto rilevante sui parametri del piano ai minimi quadrati che oltre a  $X_3$  considera anche  $X_2$  come variabile esplicativa.

Prima di concludere questa sezione conviene aggiungere un'ultima osservazione di carattere tecnico. L'osservazione riguarda il ragionamento che abbiamo seguito per risolvere il sistema di equazioni che si ottiene annullando le derivate parziali della funzione  $D(\alpha_1, \alpha_2, \alpha_3)$ , ovvero il sistema (33). Ovviamente, anche questo sistema è lineare e quindi avremmo potuto applicare il teorema di Rouché-Capelli e la regola di Cramer direttamente a questo sistema piuttosto che procedere prima all'eliminazione dell'incognita  $\alpha_1$ . Non abbiamo seguito questa strada perché poi sarebbero stati necessari comunque molti passaggi algebrici per esprimere la condizione di esistenza e unicità della soluzione solo in termini di varianze e covarianze (si veda la (41)) e per ottenere le formule (35), (39) e (40) che ci permettono di calcolare i parametri di un piano ai minimi quadrati usando solo le medie, le varianze e le covarianze.

**Esempio 5.1** (Capitalizzazione, Compensi e Utili). Nell'Esempio 3.1 abbiamo analizzato i dati sulle  $n = 10$  più grandi società del settore delle comunicazioni che sono incluse nell'indice di borsa S&P500 e abbiamo ricavato i parametri della retta ai minimi quadrati che descrive gli utili societari (variabile  $X_1$ ) in funzione dei compensi dei loro dirigenti (variabile  $X_2$ ). In questo esempio ricaveremo invece il piano ai minimi quadrati che tiene anche conto della capitalizzazione di mercato (variabile  $X_3$ ). Figura 5.3 mostra il foglio di calcolo con i dati delle tre variabili:

Figura 5.3: Foglio "Piano ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni"

	A	B	C	D	E
1			UTILE NETTO DELL'ANNO 2019 (in miliardi di USD)	COMPENSO DIRIGENTI ANNO 2019 (in milioni di USD)	CAPITALIZZAZIONE DI MERCATO (in miliardi di USD)
2	TICKER	NOME SOCIETÀ			
3			Variabile X1	Variabile X2	Variabile X3
4	GOOG	Alphabet Inc. Class C	34,343	281,946	1096,723
5	FB	Facebook, Inc. Class A	18,485	112,844	648,350
6	DIS	Walt Disney Company	11,054	92,795	251,246
7	VZ	Verizon Communications Inc.	19,913	51,095	248,451
8	CMCSA	Comcast Corporation Class A	13,057	142,706	224,504
9	NFLX	Netflix, Inc.	1,867	126,379	215,702
10	T	AT&T Inc.	14,975	80,583	201,808
11	TMUS	T-Mobile US, Inc.	3,468	85,494	160,957
12	CHTR	Charter Communications, Inc. Class A	1,668	21,511	131,000
13	ATVI	Activision Blizzard, Inc.	1,736	54,417	59,363
14					
15		<b>SOMME</b>	<b>120,566</b>	<b>1049,77</b>	<b>3238,106</b>
16		<b>MEDIE</b>	<b>12,0566</b>	<b>104,977</b>	<b>323,811</b>

Usando le consuete formule otteniamo la matrice di varianza e covarianza in Figura 5.4:

Figura 5.4:

<b>Matrice di varianza e covarianza</b>			
	<b>X1</b>	<b>X2</b>	<b>X3</b>
<b>X1</b>	100,228		
<b>X2</b>	488,818	4681,065	
<b>X3</b>	2556,957	17383,135	88372,821

Applicando le formule (42) e (43) vediamo ora che

$$\begin{aligned}\hat{\alpha}_2 &= \frac{\begin{vmatrix} 488,818 & 17383,135 \\ 2556,957 & 88372,821 \end{vmatrix}}{\begin{vmatrix} 4681,065 & 17383,135 \\ 17383,135 & 88372,821 \end{vmatrix}} \\ &= \frac{488,818 \times 88372,821 - 2556,957 \times 17383,135}{4681,065 \times 88372,821 - 17383,135^2} \\ &= -0,011\end{aligned}$$

$$\begin{aligned}\hat{\alpha}_3 &= \frac{\begin{vmatrix} 4681,065 & 488,818 \\ 17383,135 & 2556,957 \end{vmatrix}}{\begin{vmatrix} 4681,065 & 17383,135 \\ 17383,135 & 88372,821 \end{vmatrix}} \\ &= \frac{4681,065 \times 2556,957 - 17383,135 \times 488,818}{4681,065 \times 88372,821 - 17383,135^2} \\ &= 0,031\end{aligned}$$

e che

$$\hat{\alpha}_1 = 12,057 - (-0,011) \times 104,977 - 0,031 \times 323,811 = 3,150.$$

L'equazione del piano ai minimi quadrati è dunque data da

$$\hat{x}_1 = 3,150 + (-0,011) \times x_2 + 0,031 \times x_3.$$

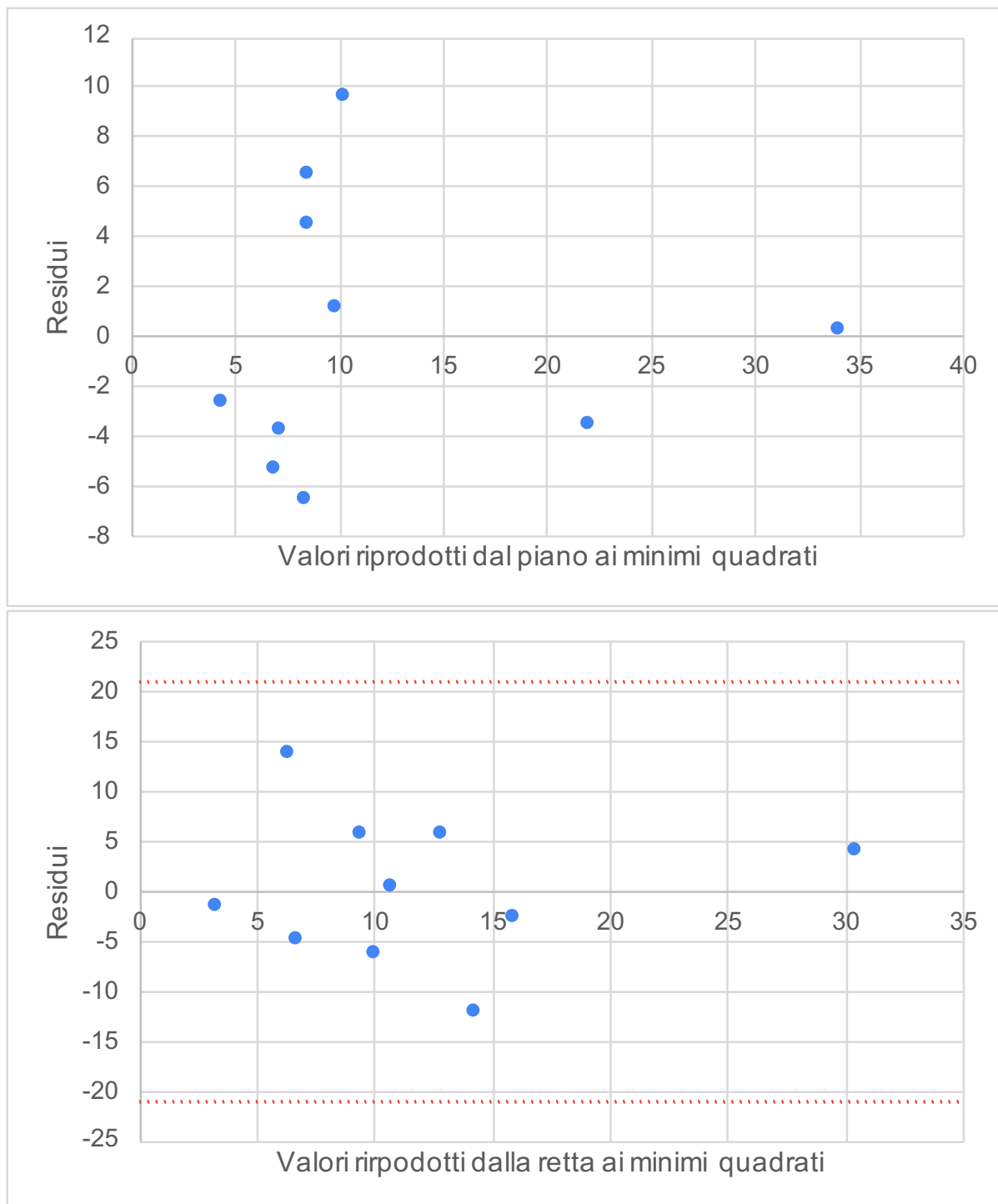
Per valutare la bontà d'adattamento di questo piano costruiamo innanzitutto il grafico dei residui. A tal fine calcoliamo i valori riprodotti  $\hat{x}_i$  e i residui  $z_i$  (vedi le colonne L e M del foglio di calcolo in Figura 5.5) ...

Figura 5.5: Foglio "Piano ai minimi quadrati" della cartella di lavoro "Settore GICS delle comunicazioni"

L	M	N
X1_cap	residui	residui^2
34,140	0,203	0,041
22,074	-3,589	12,881
9,934	1,120	1,255
10,314	9,599	92,142
8,542	4,515	20,390
8,450	-6,583	43,341
8,531	6,444	41,524
7,204	-3,736	13,958
6,988	-5,320	28,305
4,389	-2,653	7,037
<b>120,566</b>	<b>0,000</b>	<b>260,875</b>
<b>12,057</b>	<b>0,000</b>	<b>26,088</b>

... e riportiamo i punti  $(\hat{x}_{1i}, z_i)$  all'interno di un piano cartesiano. Il primo dei due grafici in Figura 5.6 è il grafico dei residui del piano ai minimi quadrati. Per confronto, riportiamo sotto anche il grafico dei residui della retta ai minimi quadrati che tiene conto solo dei compensi dei dirigenti (variabile  $X_2$ ), ovvero della retta ai minimi quadrati che avevamo ricavato nell'Esempio 3.1.

Figura 5.6: Grafico dei residui del piano ai minimi quadrati (in alto) e della retta ai minimi quadrati (in basso)



Confrontando il grafico dei residui del piano ai minimi quadrati

con quello della retta ai minimi quadrati, notiamo che, rispetto al secondo, il primo grafico dei residui evidenzia una certa sistematicità nella successione dei segni dei residui. Infatti, i primi quattro punti sul lato sinistro del primo grafico si trovano tutti al di sotto dell'asse delle ascisse e poco a destra seguono quattro punti al di sopra dell'asse delle ascisse, mentre la successione dei segni dei residui della retta ai minimi quadrati sembra perfettamente casuale. Questo fatto segnala un problema di rigidità del piano ai minimi quadrati che potrebbe essere risolto mediante l'introduzione di una variabile *dummy* ma per non appesantire la trattazione sorvoliamo su questa questione (del resto questo esempio ha solo scopo didattico; infatti, di solito le analisi di regressione vengono eseguite su *dataset* molto più numerosi).

Passiamo ora al problema dei valori anomali. Confrontando le scale sugli assi delle ordinate dei due grafici dei residui in Figura 5.6, notiamo subito che l'ordine di grandezza medio dei residui del piano ai minimi quadrati è nettamente più piccolo (come vedremo più avanti, questa non è una sorpresa!!!). Infatti, mentre l'indice  $A_2$  della retta ai minimi quadrati è dato da

$$A_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n z_{retta,i}^2} = 7,013$$

(vedi Esempio 3.1), quello del piano ai minimi quadrati risulta pari a (la media dei quadrati dei residui è riportata in fondo alla colonna N in Figura 5.5)

$$A_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n z_{piano,i}^2} = \sqrt{26,088} = 5,108$$



ed è quindi nettamente più piccolo. Moltiplicando per  $\pm 3$  il valore dell'indice  $A_2$  otteniamo le ordinate delle due linee rette orizzontali che separano i residui "sospetti" da quelli "non sospetti". Per il piano ai minimi quadrati otteniamo

$$\pm 3 \times A_2 = \pm 3 \times 5,108 = 15,324$$

e le due linee rette orizzontali si trovano quindi al di fuori dell'area rappresentata nel primo grafico in Figura 5.6. Anche rispetto al piano ai minimi quadrati non ci sono quindi delle osservazioni anomale.

Dividendo il valore dell'indice  $A_2$  per lo scarto quadratico medio degli utili otteniamo anche il valore dell'indice  $A_2^*$  del piano:

$$A_2^* = \sqrt{\frac{\sum_{i=1}^n z_{piano,i}^2}{dev(X_1)}} = \frac{A_2}{\sigma_1} = \frac{5,108}{\sqrt{100,228}} = 0,510.$$

Questo valore ci dice che l'ordine di grandezza medio dei residui del piano ai minimi quadrati ammonta al 51% del valore dello scarto quadratico medio degli utili. Ricordiamo che l'indice  $A_2^*$  delle rette ai minimi quadrati è pari a

$$A_2^* = \sqrt{\frac{\sum_{i=1}^n z_{retta,i}^2}{dev(X_1)}} = \frac{A_2}{\sigma_1} = \frac{7,013}{\sqrt{100,228}} = 0,701.$$

A questo punto possiamo procedere all'interpretazione dei parametri del piano ai minimi quadrati. Il coefficiente angolare  $\hat{\alpha}_2 = -0,011$  ci dice che, a parità di capitalizzazione di mercato (variabile  $X_3$ ), il valore che il piano ai minimi quadrati riproduce per l'utile netto diminuisce di 0,011 miliardi di USD (ovvero

di 11 milioni di USD) se il compenso per i dirigenti (variabile  $X_2$ ) aumenta di 1 milione di USD. L'interpretazione del coefficiente angolare  $\hat{\alpha}_3 = 0,031$  è analoga. Il valore dell'intercetta  $\hat{\alpha}_1 = 3,150$  ci dice che se il compenso dei dirigenti e la capitalizzazione di mercato sono entrambi nulli, il piano ai minimi quadrati riproduce un utile netto di ammontare pari a 3,150 miliardi di USD.

Vale la pena osservare che il valore  $\hat{\alpha}_2 = -0,011$  del coefficiente angolare del piano ai minimi quadrati è molto diverso dal valore del coefficiente angolare della retta ai minimi quadrati che ricordiamo essere pari a  $\hat{\alpha}_2 = 0,104$  (vedi Esempio 3.1). Volendo inferire qualcosa sul processo che ha generato i dati,<sup>14</sup> che cosa possiamo dire sulla relazione tra i compensi degli amministratori e gli utili societari???

- Il coefficiente angolare  $\hat{\alpha}_2 = 0,104$  della retta ai minimi quadrati (che è di segno positivo) suggerisce che gli utili societari aumentino all'aumentare dei compensi. Quindi si potrebbe pensare che ...
  - ...all'aumentare dei loro compensi gli amministratori siano più motivati nel loro lavoro e che per questo motivo gli utili societari aumentino, ...
  - ...oppure che per accaparrarsi gli amministratori più bravi le società debbano pagare di più ...

---

<sup>14</sup>Per poter inferire qualcosa sul processo che genera determinati dati di solito serve un numero di osservazioni molto più elevato di quello che stiamo usando in questo esempio, ovvero di  $n = 10$ . Tuttavia, il problema di fondo che viene affrontato nel resto di questo esempio rimane lo stesso indipendentemente dal numero di osservazioni.

– ... oppure che ...

- D'altra parte, il coefficiente angolare  $\hat{\alpha}_2 = -0,011$  del piano ai minimi quadrati (che è di segno negativo) suggerisce che, a parità di capitalizzazione di mercato, gli utili societari diminuiscano all'aumentare dei compensi.

Dal confronto tra il valore del coefficiente angolare  $\hat{\alpha}_2 = 0,104$  che si riferisce alla retta ai minimi quadrati, e il coefficiente angolare  $\hat{\alpha}_2 = -0,011$  che si riferisce al piano ai minimi quadrati, emerge dunque che ...

*... la correlazione positiva tra i compensi dei dirigenti e gli utili societari potrebbe essere "spuria", ovvero potrebbe essere solo una conseguenza del fatto che abbiamo analizzato una popolazione di società con capitalizzazioni di mercato molto diverse!!!*

Più avanti aggiungeremo qualche approfondimento sul fenomeno della cosiddetta "correlazione spuria" (vedi Esempio 5.4).

△

**Esercizio 5.1** (Indice PMI e tasso di crescita del PIL). Si calcolino i parametri del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (vedi Esempio 3.2; i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")

△

## 5.2 Formule alternative per calcolare i coefficienti angolari del piano ai minimi quadrati

In questa sezione ricaveremo alcune formule alternative per calcolare i coefficienti angolari di un piano ai minimi quadrati. Prima di procedere conviene introdurre una notazione generale che ci permetterà di indicare in modo univoco parametri, valori riprodotti e residui associati a rette e piani ai minimi quadrati. D'ora in poi la retta ai minimi quadrati che considera  $X_1$  come variabile dipendente e  $X_2$  come variabile esplicativa verrà dunque indicata con

$$\hat{x}_{1.2} = \hat{c}_{1.2} + \hat{\alpha}_{12}x_2$$

Come abbiamo già dimostrato, se  $\sigma_{22} > 0$ , i parametri di questa retta saranno dati da

$$\hat{\alpha}_{12} = \frac{\sigma_{12}}{\sigma_{22}} \quad \text{e} \quad \hat{c}_{1.2} = \bar{x}_1 - \hat{\alpha}_{12}\bar{x}_2. \quad (46)$$

Con riferimento alla suddetta retta ai minimi quadrati, indicheremo con  $\hat{X}_{1.2}$  la variabile "artificiale" che rappresenta i valori riprodotti

$$\hat{x}_{1.2i} = \hat{c}_{1.2} + \hat{\alpha}_{12}x_{2i}, \quad i = 1, 2, \dots, n,$$

e indicheremo con  $Z_{1.2}$  la variabile "artificiale" che rappresenta i residui

$$z_{1.2i} = x_{1i} - \hat{x}_{1.2i}, \quad i = 1, 2, \dots, n.$$

Consideriamo ora invece il piano ai minimi quadrati che considera  $X_1$  come variabile dipendente e  $X_2$  e  $X_3$  come variabili

esplicative. D'ora in poi indicheremo questo piano con

$$\hat{x}_{1.23} = \hat{c}_{1.23} + \hat{\alpha}_{12.3} \times x_2 + \hat{\alpha}_{13.2} \times x_3.$$

Come abbiamo visto nella sezione precedente, l'intercetta di questo piano è data da

$$\hat{c}_{1.23} = \bar{x}_1 - \hat{\alpha}_{12.3} \times \bar{x}_2 - \hat{\alpha}_{13.2} \times \bar{x}_3$$

e (in assenza di collinearità) gli altri due parametri possono essere calcolati attraverso le formule

$$\hat{\alpha}_{12.3} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}, \quad (47)$$

e

$$\hat{\alpha}_{13.2} = \frac{\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}.$$

Nel caso del piano ai minimi quadrati in questione scriveremo  $\hat{X}_{1.23}$  per indicare la variabile "artificiale" che rappresenta gli  $n$  valori riprodotti

$$\hat{x}_{1.23i} = \hat{c}_{1.23} + \hat{\alpha}_{12.3}x_{2i} + \hat{\alpha}_{13.2}x_{3i}, \quad i = 1, 2, \dots, n,$$

e scriveremo  $Z_{1.23}$  per indicare la variabile "artificiale" che rappresenta i residui

$$z_{1.23i} = x_{1i} - \hat{x}_{1.23i}, \quad i = 1, 2, \dots, n.$$

Usando questa nuova notazione saremo ora in grado di esprimere molti concetti in modo più agevole. Cominceremo ricavando tre formule alternative per calcolare i parametri  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{13.2}$  di un

piano ai minimi quadrati. Per comodità scriveremo solo le formule per calcolare  $\hat{\alpha}_{12.3}$ . Scambiando opportunamente i pedici "1", "2" e "3" si ottengono le analoghe formule per calcolare  $\hat{\alpha}_{ij.k}$  per qualsiasi permutazione dei pedici.

**Prima formula alternativa:**

$$\hat{\alpha}_{12.3} = \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \quad (48)$$

dove  $\sigma_1$  e  $\sigma_2$  (con un solo pedice) indicano gli scarti quadratici medi delle osservazioni delle variabili  $X_1$  e  $X_2$ , rispettivamente.

**Esempio 5.2** (Verifica numerica della prima formula alternativa). Per effettuare una verifica numerica della prima formula alternativa faremo riferimento ai dati contenuti nella cartella di lavoro "[Settore GICS delle comunicazioni](#)". Dalla matrice di varianza e covarianza in Figura 5.4 ricaviamo facilmente i tre coefficienti di correlazione lineare

$$r_{12} = \frac{488,818}{\sqrt{100,228 \times 4681,065}} = 0,714$$

$$r_{13} = \frac{2556,957}{\sqrt{100,228 \times 88372,821}} = 0,859$$

$$r_{23} = \frac{17383,135}{\sqrt{4681,065 \times 88372,821}} = 0,855$$

e gli scarti quadratici medi

$$\sigma_1 = \sqrt{100,228} = 10,011, \quad \sigma_2 = \sqrt{4681,065} = 68,418$$

$$\sigma_3 = \sqrt{88372,821} = 297,276.$$

Usando la formula alternativa (48) otteniamo dunque

$$\begin{aligned}\hat{\alpha}_{12.3} &= \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \\ &= \frac{10,011}{68,418} \times \frac{0,714 - 0,859 \times 0,855}{1 - 0,855^2} = -0,011\end{aligned}$$

$$\begin{aligned}\hat{\alpha}_{13.2} &= \frac{\sigma_1}{\sigma_3} \times \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \\ &= \frac{10,011}{297,276} \times \frac{0,859 - 0,714 \times 0,855}{1 - 0,855^2} = 0,031.\end{aligned}$$

Questi valori coincidono con i valori dei coefficienti di regressione parziale che avevamo già calcolato nell'Esempio 5.1.  $\triangle$

**Esercizio 5.2.** Si usi la formula alternativa (48) per calcolare i coefficienti angolari del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (vedi Esempio 3.2; i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

**Dimostrazione** (della formula alternativa (48)). Nella Sezione 5.1 abbiamo dimostrato che

$$\hat{\alpha}_{12.3} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2}.$$

Dividendo il numeratore ed il denominatore per il prodotto  $\sigma_{22} \times \sigma_{33}$  otteniamo

$$\begin{aligned}\hat{\alpha}_{12.3} &= \frac{(\sigma_{12}\sigma_{33} - \sigma_{13} \times \sigma_{23})/\sigma_{22}\sigma_{33}}{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)/\sigma_{22}\sigma_{33}} \\ &= \frac{\frac{\sigma_{12}}{\sigma_{22}} - \frac{\sigma_{13} \times \sigma_{23}}{\sigma_{22}\sigma_{33}}}{1 - \frac{\sigma_{23}^2}{\sigma_{22}\sigma_{33}}} \\ &= \frac{\frac{\sigma_{12}}{\sigma_2 \times \sigma_2} - \frac{\sigma_{13} \times \sigma_{23}}{\sigma_2 \times \sigma_2 \times \sigma_3 \times \sigma_3}}{1 - \frac{\sigma_{23}^2}{\sigma_{22}\sigma_{33}}} \quad (49) \\ &= \frac{\frac{\sigma_1}{\sigma_2} \left( \frac{\sigma_{12}}{\sigma_1 \times \sigma_2} - \frac{\sigma_{13}}{\sigma_1 \times \sigma_3} \times \frac{\sigma_{23}}{\sigma_2 \times \sigma_3} \right)}{1 - \frac{\sigma_{23}^2}{\sigma_{22}\sigma_{33}}} \\ &= \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{13} \times r_{23}}{1 - r_{23}^2}\end{aligned}$$

e l'espressione nell'ultima riga coincide con la formula che volevamo dimostrare.  $\square$

### Seconda formula alternativa:

$$\hat{\alpha}_{12.3} = \frac{\hat{\alpha}_{12} - \frac{\sigma_1}{\sigma_2} \times r_{13} \times r_{23}}{1 - r_{23}^2} \quad (50)$$

Ne consegue che  $\hat{\alpha}_{12.3} = \hat{\alpha}_{12}$  se  $r_{23} = 0$ .

**Esempio 5.3** (Verifica numerica della seconda formula alternativa). Per effettuare una verifica numerica della seconda formula alternativa faremo ancora riferimento ai dati della cartella di lavoro "Settore GICS delle comunicazioni". Nell'Esempio 5.2 abbiamo già ricavato

- i valori dei coefficienti di correlazione lineare, che ricordiamo essere dati da

$$r_{12} = 0,714, \quad r_{13} = 0,859, \quad r_{23} = 0,855;$$

- e anche i valori degli scarti quadratici medi, che ricordiamo essere dati da

$$\sigma_1 = 10,011, \quad \sigma_2 = 68,418, \quad \sigma_3 = 297,276.$$

Sostituendo questi dati nella formula (15) otteniamo i valori dei coefficienti di regressione grezzi

$$\hat{\alpha}_{12} = \frac{\sigma_1}{\sigma_2} \times r_{12} = \frac{10,011}{68,418} \times 0,714 = 0,104$$

e

$$\hat{\alpha}_{13} = \frac{\sigma_1}{\sigma_3} \times r_{13} = \frac{10,011}{297,276} \times 0,859 = 0,029$$



(il valore di  $\hat{\alpha}_{12} = 0,104$  lo avevamo già calcolato come rapporto tra  $codev(X_1, X_2)$  e  $dev(X_2)$  nell'Esempio 3.1).

Per effettuare una verifica numerica della formula alternativa (50) osserviamo ora che i valori di

$$\begin{aligned}\hat{\alpha}_{12.3} &= \frac{\hat{\alpha}_{12} - \frac{\sigma_1}{\sigma_2} \times r_{13} \times r_{23}}{1 - r_{23}^2} \\ &= \frac{0,104 - \frac{10,011}{68,418} \times 0,859 \times 0,855}{1 - 0,855^2} = -0,011\end{aligned}$$

e

$$\begin{aligned}\hat{\alpha}_{13.2} &= \frac{\hat{\alpha}_{13} - \frac{\sigma_1}{\sigma_3} \times r_{12} \times r_{23}}{1 - r_{23}^2} \\ &= \frac{0,029 - \frac{10,011}{297,276} \times 0,714 \times 0,855}{1 - 0,855^2} = 0,031\end{aligned}$$

coincidono con quelli che avevamo già trovato nell'Esempio 5.1  $\triangle$

**Esercizio 5.3.** Si usi la formula alternativa (50) per calcolare i coefficienti angolari del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (vedi Esempio 3.2; i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

Le formula alternativa (50) evidenzia la relazione tra  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{12}$ . Entrambi questi parametri sono dei coefficienti angolari. Infatti,  $\hat{\alpha}_{12}$  descrive l'inclinazione della retta ai minimi quadrati

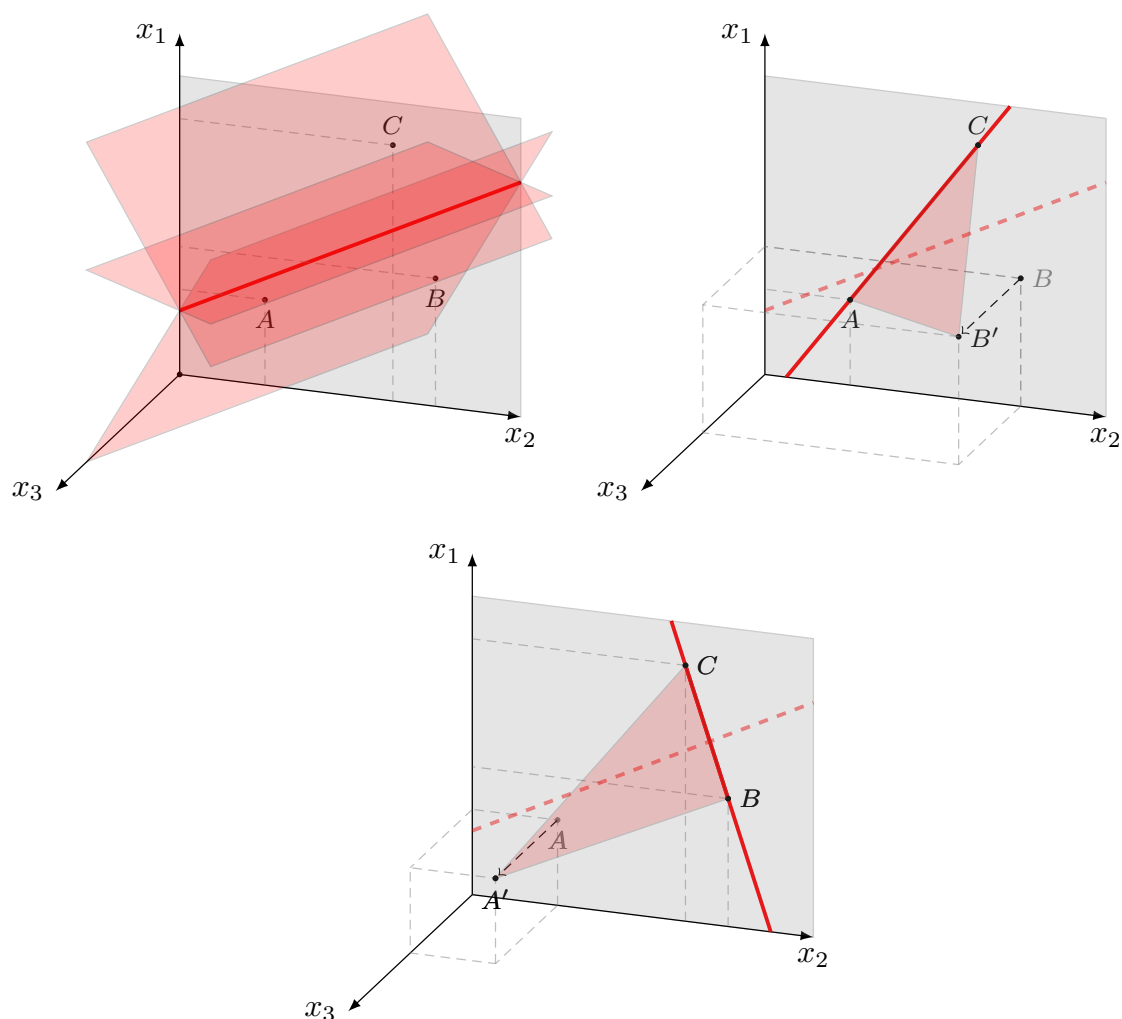
$$\hat{x}_{1.2} = \hat{c}_{1.2} + \hat{\alpha}_{12} \times x_2$$

rispetto all'asse delle ascisse (ovvero l'asse assegnato alla variabile  $X_2$ ), mentre  $\hat{\alpha}_{12.3}$  descrive l'inclinazione del piano ai minimi quadrati rispetto all'asse assegnato alla variabile  $X_2$ . Siccome i due coefficienti angolari di un piano ai minimi quadrati sono delle derivate parziali, spesso vengono chiamati "coefficienti di regressione parziale". D'altro canto, il coefficiente angolare di una retta ai minimi quadrati viene chiamato "coefficiente di regressione grezzo". In questo caso l'aggettivo "grezzo" è legato al fatto che il coefficiente angolare di una retta ai minimi quadrati descrive solo la relazione tra due variabili e che ignora completamente le osservazioni di qualsiasi altra variabile.

Dalla formula (50) si evince immediatamente che  $\hat{\alpha}_{12.3} = \hat{\alpha}_{12}$  se  $r_{23} = 0$ . Scambiando i pedici "2" e "3" si vede che  $r_{23} = 0$  implica anche  $\hat{\alpha}_{13.2} = \hat{\alpha}_{13}$ . Quindi possiamo concludere i due coefficienti angolari del piano ai minimi quadrati coincidono con i coefficienti angolari delle corrispondenti rette ai minimi quadrati se le osservazioni delle due variabili esplicative hanno correlazione nulla.

Dalla formula (50) si evince inoltre che i coefficienti angolari  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{12}$  potrebbero essere di segno opposto se  $r_{23} \neq 0$ . Questo fatto non deve sorprendere. Infatti, per calcolare il valore del coefficiente angolare  $\hat{\alpha}_{12}$  si usano solo le osservazioni delle variabili  $X_1$  e  $X_2$ , mentre per calcolare  $\hat{\alpha}_{12.3}$  servono anche le osservazioni della variabile  $X_3$ . Per illustrare come i valori delle osservazioni di  $X_3$  possono incidere sul valore del coefficiente di regressione parziale  $\hat{\alpha}_{12.3}$  conviene fare riferimento ai tre grafici in Figura 5.7:

Figura 5.7:



Nel primo grafico in alto a sinistra si vede una nuvola di punti composta dai tre punti  $A$ ,  $B$  e  $C$  che si trovano tutti e tre all'interno del piano verticale dove  $x_3 = 0$  (il piano di colore grigio). Come abbiamo visto nella sezione precedente, in questo caso esistono infiniti piani ai minimi quadrati e ogni piano ai minimi quadrati deve intersecare il piano verticale grigio nella retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  (ovvero la retta rossa). I valori dei coefficienti angolari  $\hat{\alpha}_{12.3}$  di tutti i piani ai minimi quadrati de-

sono quindi coincidere con il valore di  $\hat{\alpha}_{12}$ , ovvero con il valore del coefficiente angolare della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$ . Come si vede nel grafico questa retta è crescente e quindi si ha  $\hat{\alpha}_{12.3} = \hat{\alpha}_{12} > 0$ . Analizziamo ora che cosa accade se il punto  $B$  si spostasse nel punto  $B'$ , ovvero se la coordinata  $x_3$  del punto  $B$  passasse da 0 ad un valore positivo (vedi il grafico in alto a destra; la seguente analisi rimarrebbe immutata anche nel caso in cui il valore della coordinata  $x_3$  diventasse negativo). In questo caso la nuvola di punti composta dai punti  $A$ ,  $B'$  e  $C$  non sarebbe più concentrata all'interno di un piano verticale e quindi esisterebbe un unico piano ai minimi quadrati. Siccome per tre punti passa sempre un unico piano, il piano ai minimi quadrati passerebbe per tutti e tre i punti  $A$ ,  $B'$  e  $C$  e coinciderebbe quindi con il prolungamento dell'area del triangolo  $AB'C$ . Si osservi che la retta di intersezione di questo piano ai minimi quadrati con il piano verticale di colore grigio è più ripida della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  (la linea rossa tratteggiata) e quindi possiamo concludere che  $\hat{\alpha}_{12.3} > \hat{\alpha}_{12} > 0$ . Analizziamo infine che cosa accadrebbe se al posto del punto  $B$  spostassimo il punto  $A$  facendo passare la corrispondente coordinata  $x_3$  da 0 ad un valore positivo (la seguente analisi rimarrebbe immutata anche al caso in cui il valore della coordinata  $x_3$  diventasse negativo). Supponiamo quindi che il punto  $A$  si sposti nel punto  $A'$  così come mostrato nel grafico in basso. Anche in questo si otterrebbe una nuvola di punti che non è più concentrata all'interno di un piano verticale e a questa nuvola di punti corrisponderebbe quindi un unico piano ai

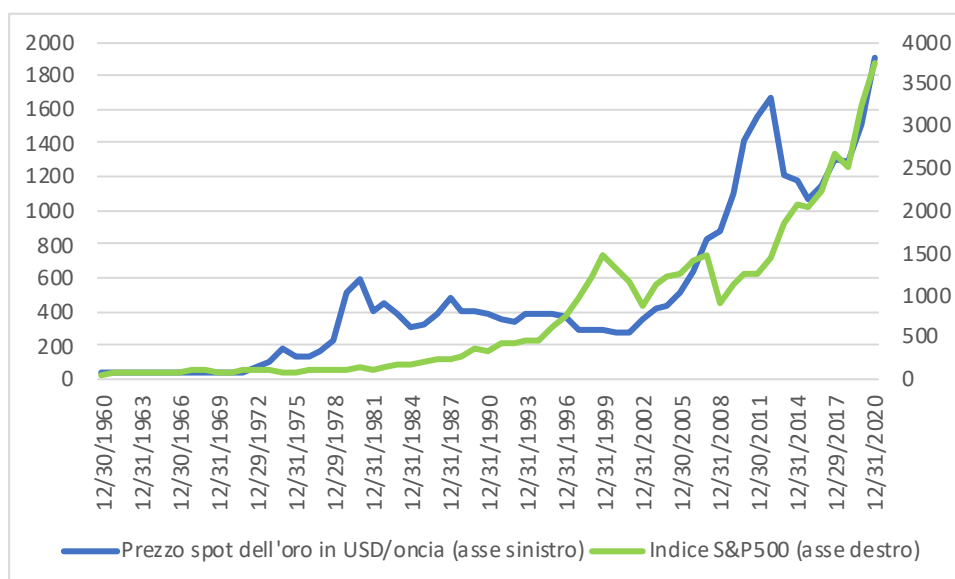
minimi quadrati. Il piano ai minimi quadrati in questione sarebbe dato dall'unico piano che passa per i punti  $A'$ ,  $B$  e  $C$ . Nel grafico in basso si vede che questo piano interseca il piano verticale grigio in una retta decrescente e il coefficiente angolare  $\hat{\alpha}_{12.3}$  di questo piano è quindi negativo. Questo esempio, purché poco realistico, è comunque molto istruttivo perché ci mostra come l'aggiunta di una variabile esplicativa potrebbe modificare la nostra opinione sulla relazione di fondo tra due variabili  $X_1$  e  $X_2$ . Infatti, i tre grafici in Figura 5.7 mostrano come a diversi valori della variabile  $X_3$  potrebbe corrispondere un ampio insieme di valori diversi del coefficiente di regressione parziale  $\hat{\alpha}_{12.3}$ . Per questo motivo, *quando si cerca di inferire qualcosa sul processo che potrebbe aver generato i dati, bisogna essere molto prudenti nell'interpretazione di coefficienti di regressione grezzi e bisogna sempre chiedersi se un'apparente relazione crescente o decrescente (o anche un'apparente assenza di relazione) tra le osservazioni di due variabili non sia in realtà dovuta alle variazioni di una terza variabile che è stata esclusa dall'analisi (per esempio perché non disponiamo delle sue osservazioni)*. Se ciò fosse il caso, la correlazione tra le osservazioni delle due variabili incluse nell'analisi verrebbe chiamata "*correlazione spuria*" e la terza variabile esclusa dall'analisi verrebbe chiamata "*variabile confondente*".

La correlazione (positiva, negativa o anche nulla) tra le osservazioni di due variabili  $X_1$  e  $X_2$  viene chiamata "**spuria**" se *si ritiene* che essa sia indotta dalla variabilità dei valori (valori che potrebbero anche essere ignoti) di una terza variabile  $X_3$  che viene chiamata "**variabile confondente**".

Ricordiamo che nell'Esempio 5.1 abbiamo già incontrato un caso che potrebbe essere interpretato come caso di correlazione spuria. Il prossimo esempio suggerisce che casi di correlazione spuria siano abbastanza frequenti.

**Esempio 5.4** (Correlazione spuria). Il grafico in Figura 5.8 mostra l'andamento delle quotazioni dell'oro (variabile  $X_1$ ) e l'andamento dell'indice di borsa S&P500 (variabile  $X_2$ ) che riflette il valore complessivo delle 500 società statunitensi a maggiore capitalizzazione. Il grafico si riferisce al periodo storico compreso tra l'anno 1960 e l'anno 2020. I dati sottostanti sono contenuti nella cartella di lavoro "[Esempio di correlazione spuria](#)" e provengono dalla piattaforma *Bloomberg*.

Figura 5.8:



Durante il periodo storico considerato entrambe le variabili hanno avuto un andamento tendenzialmente crescente e per questo motivo non c'è da stupirsi che la correlazione tra le loro osservazioni sia molto elevata. Infatti, usando i dati contenuti nel foglio di calcolo possiamo facilmente verificare che

$$r_{12} = r(X_1, X_2) = 0,831,$$

$$\sigma_1 = \sqrt{\text{var}(X_1)} = 477,88, \quad \sigma_2 = \sqrt{\text{var}(X_2)} = 859,27$$

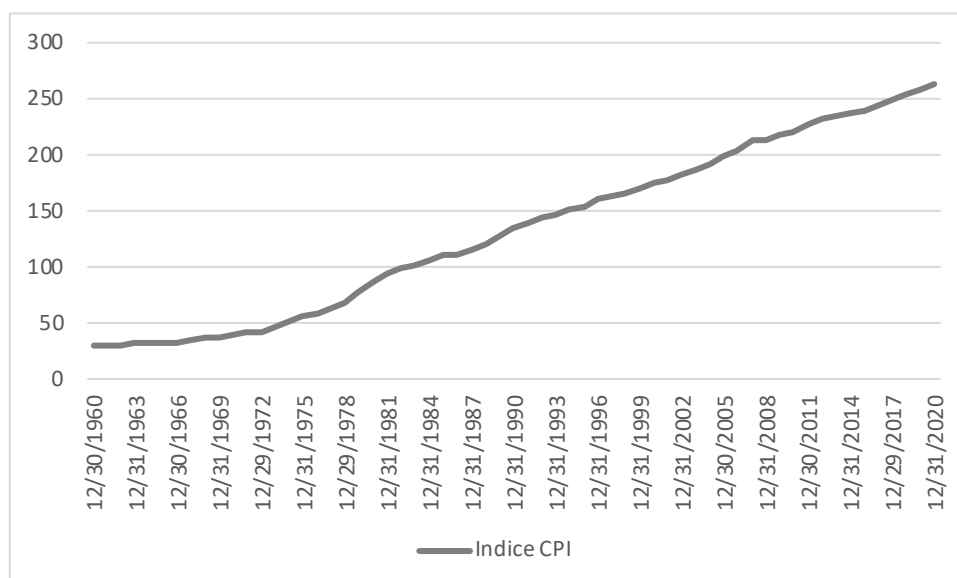
e che il coefficiente angolare della retta ai minimi quadrati che descrive  $X_1$  (il prezzo dell'oro) in funzione di  $X_2$  (indice S&P500) è quindi dato da

$$\hat{\alpha}_{12} = \frac{\sigma_1}{\sigma_2} \times r_{12} = \frac{477,88}{859,27} \times 0,831 = 0,462$$

(qui abbiamo utilizzato la formula (15)). Il valore positivo di  $\hat{\alpha}_{12}$  conferma che durante il periodo preso in considerazione la relazione

di fondo tra le quotazioni dell'oro e il valore dell'indice S&P500 fosse crescente. Tuttavia, come si vede in Figura 5.9, durante il periodo in questione è cresciuto anche l'indice CPI (variabile  $X_3$ ) che misura il livello generale dei prezzi dei beni di largo consumo e questa crescita potrebbe spiegare l'elevata correlazione positiva tra le osservazioni delle variabili  $X_1$  e  $X_2$ . Il sospetto è quindi che l'elevata correlazione positiva tra le quotazioni dell'oro e il valore dell'indice S&P500 sia spuria e che la variabile  $X_3$  abbia agito come variabile confondente.

Figura 5.9:



Per verificare se i dati supportano questa ipotesi possiamo calcolare il valore del coefficiente di regressione parziale  $\hat{\alpha}_{12.3}$  e confrontarlo con il valore di  $\hat{\alpha}_{12}$ . Per calcolare il valore di  $\hat{\alpha}_{12.3}$  calcoliamo innanzitutto i valori dei coefficienti di correlazione

$$r_{13} = 0,853 \quad \text{e} \quad r_{23} = 0,877$$



(la cartella di lavoro "Esempio di correlazione spuria" contiene anche i dati sull'indice CPI), onde poi applicare la formula alternativa (50):

$$\begin{aligned}\hat{\alpha}_{12.3} &= \frac{\hat{\alpha}_{12} - \frac{\sigma_1}{\sigma_2} \times r_{13} \times r_{23}}{1 - r_{23}^2} \\ &= \frac{0,462 - \frac{477,88}{859,27} \times 0,853 \times 0,877}{1 - 0,877^2} = 0,199.\end{aligned}$$

Notiamo immediatamente il valore di  $\hat{\alpha}_{12.3}$  è molto diverso da quello di  $\hat{\alpha}_{12}$ . Questo significa che l'andamento crescente dell'indice CPI potrebbe effettivamente aver inciso sulla relazione tra le quotazioni dell'oro e i valori dell'indice S&P500. Tuttavia, questo confronto *non dimostra* che l'andamento crescente dell'indice CPI abbia *causato* l'elevata correlazione tra le quotazioni dell'oro e i valori dell'indice S&P500. Infatti, se al posto dell'indice CPI avessimo considerato una qualsiasi altra variabile che durante il periodo considerato abbia avuto un andamento crescente simile a quello dell'indice CPI, avremmo ottenuto un valore del coefficiente di regressione parziale simile a quello di  $\hat{\alpha}_{12.3}$ . Per esempio, se al posto dell'indice CPI (variabile  $X_3$ ) come possibile variabile confondente avessimo considerato semplicemente il numero di anni trascorsi a partire dall'anno 1960 (variabile  $X_4$ ), allora avremmo ottenuto (i calcoli li lasciamo per esercizio)

$$\hat{\alpha}_{12.4} = 0,223,$$

un valore molto simile a quello di  $\hat{\alpha}_{12.3}$ .

Siccome nel presente contesto è abbastanza ovvio che la quotazione dell'oro (variabile  $X_1$ ) e il valore dell'indice S&P500 (variabile  $X_2$ ) siano entrambi fortemente legati al

livello generale dei prezzi misurato dall'indice CPI, sembra ragionevole attribuire l'elevato grado di correlazione tra le quotazioni dell'oro e i valori dell'indice S&P500 all'andamento crescente dell'indice CPI. Tuttavia, se le variabili  $X_1$  e  $X_2$  fossero di natura completamente diversa, come per esempio nel caso in cui  $X_1$  rappresentasse ancora la quotazione dell'oro ma  $X_2$  rappresentasse la numerosità della popolazione mondiale, allora potrebbe non esistere una candidata ovvia per il ruolo di variabile confondente se non una qualche variabile che misura il tempo che è trascorso dall'inizio delle serie storiche analizzate come per esempio la variabile  $X_4$  di prima. Se, come potrebbe sembrare ovvio, l'andamento crescente comune alle variabili  $X_1$  e  $X_2$  fosse dovuto solo ad una coincidenza storica, allora dovremmo aspettarci che il valore di  $\hat{\alpha}_{12,4}$  sia molto più prossimo a zero del valore di  $\hat{\alpha}_{12}$  e confrontando i valori di questi due coefficienti di regressione potremmo quindi verificare se i dati supportano l'ipotesi della coincidenza storica.

△

**Dimostrazione** (della formula alternativa (50)). Per ottenere la formula (50) osserviamo innanzitutto che nella nuova notazione la formula (15) diventa

$$\hat{\alpha}_{12} = \frac{\sigma_{12}}{\sigma_{22}} = \frac{\sigma_1}{\sigma_2} \times \frac{\sigma_{12}}{\sigma_1 \times \sigma_2} = \frac{\sigma_1}{\sigma_2} \times r_{12}.$$

Da queste uguaglianze deduciamo che

$$r_{12} = \hat{\alpha}_{12} \times \frac{\sigma_2}{\sigma_1}$$

e sostituendo questo risultato nella formula (48) otteniamo la formula alternativa (50):

$$\hat{\alpha}_{12,3} = \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{13} \times r_{23}}{1 - r_{23}^2} = \frac{\frac{\sigma_1}{\sigma_2} \times r_{12} - \frac{\sigma_1}{\sigma_2} \times r_{13} \times r_{23}}{1 - r_{23}^2} = \frac{\hat{\alpha}_{12} - \frac{\sigma_1}{\sigma_2} \times r_{13} \times r_{23}}{1 - r_{23}^2}.$$

□

### Terza formula alternativa:

$$\hat{\alpha}_{12,3} = \frac{\text{codev}(Z_{1,3}, Z_{2,3})}{\text{dev}(Z_{2,3})} \quad (51)$$

e il valore di  $\hat{\alpha}_{12,3}$  coincide quindi con il coefficiente angolare della retta ai minimi quadrati che descrive  $Z_{1,3}$  in funzione di  $Z_{2,3}$ .

**Esempio 5.5** (Verifica numerica della terza formula alternativa). Per effettuare una verifica numerica della terza formula alternativa faremo ancora riferimento ai dati contenuti nella cartella di lavoro "Settore GICS delle comunicazioni". Nel foglio "Verifica della terza formula alternativa" (vedi Figura 5.10) abbiamo ricavato

- i residui della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_3$  (ovvero i residui  $z_{13i} = x_{1i} - \hat{x}_{1.3i}$ )
- e i residui della retta ai minimi quadrati che descrive  $X_2$  in funzione di  $X_3$  (ovvero i residui  $z_{23i} = x_{2i} - \hat{x}_{2.3i}$ ).

Figura 5.10: Foglio "Verifica della terza formula alternativa" della cartella di lavoro "Settore GICS delle comunicazioni"

	A	B	C	D	E	F	G	H
1	<b>Retta ai minimi quadrati che descrive X1 in funzione di X3</b>							
2	TICKER	X1	X3	X1^2	X3^2	X1*X3	Xcap1.3	Z1.3
3	GOOG	34,343	1096,723	1179,442	1202801,953	37664,768	34,420	-0,077
4	FB	18,485	648,350	341,695	420358,328	11984,758	21,447	-2,962
5	DIS	11,054	251,246	122,191	63124,393	2777,270	9,957	1,097
6	VZ	19,913	248,451	396,528	61727,990	4947,408	9,876	10,037
7	CMCSA	13,057	224,504	170,485	50402,161	2931,352	9,183	3,874
8	NFLX	1,867	215,702	3,486	46527,350	402,716	8,929	-7,062
9	T	14,975	201,808	224,251	40726,597	3022,080	8,527	6,448
10	TMUS	3,468	160,957	12,027	25907,185	558,199	7,345	-3,877
11	CHTR	1,668	131,000	2,782	17161,039	218,508	6,478	-4,810
12	ATVI	1,736	59,363	3,014	3523,985	103,054	4,405	-2,669
13								
14	Somme	120,566	3238,106	2455,900	1932260,981	64610,113	120,566	0,000
15	Medie	12,0566	323,811	245,590	193226,098	6461,011	12,057	0,000
16								
17	Parametri:	alpha13	0,029					
18		intercetta	2,688					
19								
20	<b>Retta ai minimi quadrati che descrive X2 in funzione di X3</b>							
21	TICKER	X2	X3	X1^2	X3^2	X1*X3	Xcap2.3	Z2.3
22	GOOG	281,946	1096,723	79493,547	1202801,953	309216,742	257,011	24,935
23	FB	112,844	648,350	12733,768	420358,328	73162,460	168,815	-55,971
24	DIS	92,795	251,246	8610,912	63124,393	23314,343	90,703	2,092
25	VZ	51,095	248,451	2610,699	61727,990	12694,613	90,154	-39,059
26	CMCSA	142,706	224,504	20365,002	50402,161	32038,104	85,443	57,263
27	NFLX	126,379	215,702	15971,652	46527,350	27260,202	83,712	42,667
28	T	80,583	201,808	6493,620	40726,597	16262,320	80,979	-0,396
29	TMUS	85,494	160,957	7309,224	25907,185	13760,866	72,943	12,551
30	CHTR	21,511	131,000	462,723	17161,039	2817,944	67,051	-45,540
31	ATVI	54,417	59,363	2961,210	3523,985	3230,365	52,960	1,457
32								
33	Somme	1049,77	3238,106	157012,357	1932260,981	513757,959	1049,770	0,000
34	Medie	104,977	323,811	15701,236	193226,098	51375,796	104,977	0,000
35								
36	Parametri:	alpha23	0,197					
37		intercetta	41,283					

Le due serie di residui sono riportate nella colonna H. Per calcolare i valori di  $codev(Z_{1.3}, Z_{2.3})$  e di  $dev(Z_{2.3})$  ci conviene costruire una nuova tabella che riporta le due serie di residui fianco a fianco:

Figura 5.11: Foglio "Verifica della terza formula alternativa" della cartella di lavoro "Settore GICS delle comunicazioni"

	A	B	C	D	E	F
39	<b>Calcolo di codev(Z1.3, Z2.3) e di dev(Z2.3)</b>					
40	TICKER	Z1.3	Z2.3	Z1.3^2	Z2.3^2	Z1.3*Z2.3
41	GOOG	-0,077	24,935	0,006	621,770	-1,917
42	FB	-2,962	-55,971	8,772	3132,724	165,772
43	DIS	1,097	2,092	1,203	4,375	2,295
44	VZ	10,037	-39,059	100,738	1525,577	-392,025
45	CMCSA	3,874	57,263	15,006	3279,025	221,819
46	NFLX	-7,062	42,667	49,866	1820,490	-301,299
47	T	6,448	-0,396	41,582	0,157	-2,553
48	TMUS	-3,877	12,551	15,028	157,519	-48,654
49	CHTR	-4,810	-45,540	23,135	2073,869	219,040
50	ATVI	-2,669	1,457	7,124	2,124	-3,890
51						
52	<b>Somme</b>	<b>0,000</b>	<b>0,000</b>	<b>262,460</b>	<b>12617,630</b>	<b>-141,412</b>
53	<b>Medie</b>	<b>0,000</b>	<b>0,000</b>	<b>26,246</b>	<b>1261,763</b>	<b>-14,141</b>

Siccome la media dei residui di una retta ai minimi quadrati è sempre nulla (si ricordi la prima proprietà dei residui della retta ai minimi quadrati), le somme in fondo alle colonne E e F in Figura 5.11 forniscono direttamente i valori di

$$dev(Z_{2.3}) = \sum_{i=1}^n z_{2.3i}^2 = 12617,630$$

e di

$$codev(Z_{1.3}, Z_{2.3}) = \sum_{i=1}^n z_{1.3i} \times z_{2.3i} = -141,412.$$

Applicando la formula alternativa (51) otteniamo dunque

$$\hat{\alpha}_{12.3} = \frac{codev(Z_{1.3}, Z_{2.3})}{dev(Z_{2.3})} = \frac{-141,412}{12617,630} = -0,011$$

e questo valore coincide esattamente con il valore di  $\hat{\alpha}_{12.3}$  che avevamo già trovato nell'Esempio 5.1.

Procedendo in modo analogo potremmo anche verificare che la formula

$$\hat{\alpha}_{13.2} = \frac{\text{codev}(Z_{1.2}, Z_{3.2})}{\text{dev}(Z_{3.2})}$$

fornisce il valore  $\hat{\alpha}_{13.2} = 0,031$  che avevamo già trovato nell'Esempio 5.1. I dettagli di questa verifica li lasciamo per esercizio (attenzione: questa volta bisogna calcolare i residui della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  e quelli della retta ai minimi quadrati che descrive  $X_3$  in funzione di  $X_2$ ).  $\triangle$

**Esercizio 5.4.** Si usi la formula alternativa (51) per calcolare i coefficienti angolari del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (vedi Esempio 3.2; i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

**Dimostrazione** (della formula alternativa (51)). In primo luogo osserviamo che

$$\begin{aligned} \hat{c}_{1.3} &= \bar{x}_1 - \hat{\alpha}_{13}\bar{x}_3 \quad \Rightarrow \\ \Rightarrow \hat{X}_{1.3} &= \hat{c}_{1.3} + \hat{\alpha}_{13}X_3 = (\bar{x}_1 - \hat{\alpha}_{13}\bar{x}_3) + \hat{\alpha}_{13}X_3 = \bar{x}_1 + \hat{\alpha}_{13}(X_3 - \bar{x}_3) \quad \Rightarrow \\ &\Rightarrow Z_{1.3} = X_1 - \hat{X}_{1.3} = X_1 - \bar{x}_1 + \hat{\alpha}_{13}(X_3 - \bar{x}_3) \end{aligned}$$

e che (basta sostituire il pedice "2" al posto del pedice "1")

$$Z_{2.3} = X_2 - \hat{X}_{2.3} = X_2 - \bar{x}_2 + \hat{\alpha}_{23}(X_3 - \bar{x}_3).$$

Usando le proprietà della covarianza vediamo quindi che

$$\begin{aligned} \text{codev}(Z_{1.3}, Z_{2.3}) &= \text{codev}(X_1 - \bar{x}_1 + \hat{\alpha}_{13}(X_3 - \bar{x}_3), X_2 - \bar{x}_2 + \hat{\alpha}_{23}(X_3 - \bar{x}_3)) \\ &= \dots = \text{codev}(X_1, X_2) + \hat{\alpha}_{13} \times \text{codev}(X_2, X_3) + \\ &\quad + \hat{\alpha}_{23} \times \text{codev}(X_1, X_3) + \hat{\alpha}_{13} \times \hat{\alpha}_{23} \times \text{dev}(X_3) \end{aligned}$$

A questo punto sostituiamo  $\hat{\alpha}_{13} = \text{codev}(X_1, X_3)/\text{dev}(X_3)$  e  $\hat{\alpha}_{23} = \text{codev}(X_2, X_3)/\text{dev}(X_3)$  onde ottenere

$$\text{codev}(Z_{1.3}, Z_{2.3}) = \text{codev}(X_1, X_2) - \frac{\text{codev}(X_1, X_3) \times \text{codev}(X_2, X_3)}{\text{dev}(X_3)}. \quad (52)$$

Siccome questa equazione rimane valida anche se sostituiamo il pedice "1" con il pedice "2", possiamo concludere che

$$\text{dev}(Z_{2.3}) = \text{dev}(X_2) - \frac{[\text{codev}(X_2, X_3)]^2}{\text{dev}(X_3)}. \quad (53)$$

Usando la (52) e la (53) vediamo dunque che

$$\begin{aligned} \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\text{dev}(Z_{2.3})} &= \frac{\text{codev}(X_1, X_2) - \frac{\text{codev}(X_1, X_3) \times \text{codev}(X_2, X_3)}{\text{dev}(X_3)}}{\text{dev}(X_2) - \frac{\text{codev}(X_2, X_3)^2}{\text{dev}(X_3)}} \\ &= \frac{\text{codev}(X_1, X_2) \times \text{dev}(X_3) - \text{codev}(X_1, X_3) \times \text{codev}(X_2, X_3)}{\text{dev}(X_2) \times \text{dev}(X_3) - \text{codev}(X_2, X_3)^2} \end{aligned}$$

e dividendo per  $n^2$  il numeratore ed il denominatore dell'ultima espressione otteniamo l'uguaglianza che volevamo dimostrare:

$$\frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\text{dev}(Z_{2.3})} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = \hat{\alpha}_{12.3}.$$

□

### 5.3 Le proprietà dei residui del piano ai minimi quadrati

In questa sezione vedremo che i residui  $z_{1.23i} = x_{1i} - \hat{x}_{1.23i}$  di un piano ai minimi quadrati soddisfano tre proprietà che sono perfettamente analoghe alle tre proprietà dei residui di una retta ai minimi quadrati (vedi Sezione 3.3). Per dimostrare queste proprietà procederemo come nel caso della retta ai minimi quadrati, ovvero partendo dall'osservazione che i parametri di un piano ai minimi quadrati sono delle soluzioni del sistema di equazioni che si ottiene annullando le derivate parziali della distanza quadratica. Come abbiamo visto nella Sezione 5.1, nel caso dei piani ai minimi quadrati si ottiene il sistema (33), che nella nuova notazione

diventa

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{c}_{1.23} - \hat{\alpha}_{12.3}x_{2i} - \hat{\alpha}_{13.2}x_{3i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{c}_{1.23} - \hat{\alpha}_{12.3}x_{2i} - \hat{\alpha}_{13.2}x_{3i}) \times x_{2i} = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{c}_{1.23} - \hat{\alpha}_{12.3}x_{2i} - \hat{\alpha}_{13.2}x_{3i}) \times x_{3i} = 0. \end{cases}$$

Ricordando che i valori riprodotti dal piano sono dati da

$$\hat{x}_{1.23i} = \hat{c}_{1.23} + \hat{\alpha}_{12.3}x_{2i} + \hat{\alpha}_{13.2}x_{3i}$$

vediamo che le tre equazioni possono essere espresse in modo più conciso come

$$\begin{cases} \sum_{i=1}^n (x_{1i} - \hat{x}_{1.23i}) = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{x}_{1.23i}) \times x_{2i} = 0 \\ \sum_{i=1}^n (x_{1i} - \hat{x}_{1.23i}) \times x_{3i} = 0. \end{cases}$$

Quest'ultimo sistema di equazioni è analogo al sistema (8) e viene anch'esso chiamato **sistema normale**. Partendo da questo sistema si deducono le tre proprietà che i residui di qualunque piano ai minimi quadrati devono soddisfare. Le dimostrazioni di queste tre proprietà sono perfettamente analoghe a quelle già viste in Sezione 3.3 con riferimento a rette ai minimi quadrati.



### Proprietà dei residui di piani ai minimi quadrati:

1) La somma dei residui è nulla:  $\sum_{i=1}^n z_{1.23i} = 0$ .

Corollari:

a)  $M_1(Z_{1.23}) = \bar{z}_{1.23} = 0$ .

b)  $M_1(\hat{X}_{1.23}) = \bar{x}_1$ .

2) La covarianza tra i residui e le osservazioni di ciascuna delle due variabili esplicative è nulla:  
 $codev(Z_{1.23}, X_2) = codev(Z_{1.23}, X_3) = 0$ .

3) La covarianza tra i residui e i valori riprodotti è nulla:  
 $codev(Z_{1.23}, \hat{X}_{1.23}) = 0$ .

**Dimostrazione.** Le dimostrazioni delle prime due proprietà sono identiche (parola per parola) alle dimostrazioni delle corrispondenti proprietà dei residui di rette ai minimi quadrati (vedi Sezione 3.3; per ottenere la dimostrazione di  $codev(Z_{1.23}, X_3) = 0$  basta sostituire il pedice 3 con il pedice 2).

La dimostrazione della terza proprietà si ottiene applicando le proprietà della covarianza (vedi Sezione 3.1) e la seconda proprietà dei residui di piani ai minimi quadrati:

$$\begin{aligned} codev(Z_{1.23}, \hat{X}_{1.23}) &= codev(Z_{1.23}, \hat{c}_{1.23} + \hat{\alpha}_{12.3}\hat{X}_2 + \hat{\alpha}_{13.2}\hat{X}_3) \\ &= \hat{\alpha}_{12.3} \times codev(Z_{1.23}, \hat{X}_2) + \hat{\alpha}_{13.2} \times codev(Z_{1.23}, \hat{X}_3) \\ &= \hat{\alpha}_{12.3} \times 0 + \hat{\alpha}_{13.2} \times 0 = 0. \end{aligned}$$

□

**Esempio 5.6** (Verifica numerica delle proprietà dei residui). Consideriamo ancora il piano ai minimi quadrati che avevamo ricavato nell'Esempio 5.1. Per verificare che i suoi residui  $z_{1.23i}$  soddisfino le tre proprietà che i residui di tutti i piani ai minimi quadrati devono soddisfare, ci conviene costruire una tabella come quella in Figura 5.12:

Figura 5.12: Foglio "Proprietà dei residui del piano MQ" contenuto nel foglio di calcolo "Settore GICS delle comunicazioni"

	A	B	C	D	E	F	G	H	I
1	<b>Verifica delle proprietà dei residui del piano ai minimi quadrati:</b>								
2	TICKER	Variabile X1	Variabile X2	Variabile X3	X1cap1.23	Z1.23	Z1.23*X2	Z1.23*X3	Z1.23*X1cap1.23
3	GOOG	34,343	281,946	1096,723	34,140	0,203	57,122	222,197	6,917
4	FB	18,485	112,844	648,350	22,074	-3,589	-405,003	-2326,962	-79,225
5	DIS	11,054	92,795	251,246	9,934	1,120	103,969	281,500	11,130
6	VZ	19,913	51,095	248,451	10,314	9,599	490,465	2384,902	99,004
7	CMCSA	13,057	142,706	224,504	8,542	4,515	644,386	1013,744	38,569
8	NFLX	1,867	126,379	215,702	8,450	-6,583	-832,006	-1420,057	-55,633
9	T	14,975	80,583	201,808	8,531	6,444	519,272	1300,442	54,974
10	TMUS	3,468	85,494	160,957	7,204	-3,736	-319,404	-601,332	-26,914
11	CHTR	1,668	21,511	131,000	6,988	-5,320	-114,444	-696,955	-37,179
12	ATVI	1,736	54,417	59,363	4,389	-2,653	-144,358	-157,479	-11,643
13									
14	<b>Somma</b>	<b>120,566</b>	<b>1049,770</b>	<b>3238,106</b>	<b>120,566</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
15	<b>Media</b>	<b>12,057</b>	<b>104,977</b>	<b>323,811</b>	<b>12,057</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>

Le somme in fondo alle colonne F, G, H e I forniscono i valori di  $\sum_{i=1}^n z_{1.23i}$ ,  $codev(Z_{1.23}, X_2)$ ,  $codev(Z_{1.23}, X_3)$  e di  $codev(Z_{1.23}, \hat{X}_{1.23})$ , rispettivamente. Come previsto dalle tre proprietà dei residui, queste somme sono tutte nulle.  $\triangle$

**Esercizio 5.5.** Si verifichi che anche i residui del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento soddisfano le tre proprietà che i residui di tutti i piani ai minimi quadrati devono soddisfare (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

## 5.4 La scomposizione della devianza

Nella Sezione 3.4 abbiamo dimostrato che nel caso di una retta ai minimi quadrati si ha

$$dev(X_1) = dev(\hat{X}_{1.2}) + dev(Z_{1.2}),$$

dove  $dev(\hat{X}_{1.2})$  e  $dev(Z_{1.2})$  sono, rispettivamente, la devianza spiegata e la devianza residua della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$ . Siccome la stessa dimostrazione può essere applicata anche alle variabili  $\hat{X}_{1.23}$  e  $Z_{1.23}$ , possiamo concludere che la scomposizione della devianza rimanga valida anche nel caso di un piano ai minimi quadrati.

### Scomposizione della devianza:

$$dev(X_1) = dev(\hat{X}_{1.23}) + dev(Z_{1.23})$$

Dividendo per  $n$  si ottiene l'analoga scomposizione della varianza:

$$\sigma_{11} = var(X_1) = var(\hat{X}_{1.23}) + var(Z_{1.23}).$$

**Osservazione 5.1.** Siccome  $M_1(\hat{X}_{1.23}) = \bar{x}_1$  e  $M_1(Z_{1.23}) = 0$  (prima proprietà dei residui), la devianza spiegata e la devianza residua possono essere calcolate secondo le stesse formule che usavamo nel caso di una retta ai minimi quadrati, ovvero come

$$dev(\hat{X}_{1.23}) = \sum_{i=1}^n (\hat{x}_{1.23i} - \bar{x}_1)^2$$

e

$$dev(Z_{1.23}) = \sum_{i=1}^n z_{1.23i}^2.$$

△

**Esempio 5.7** (Verifica numerica della scomposizione della devianza). Consideriamo ancora il piano ai minimi quadrati che avevamo ricavato nell'Esempio 5.1. Per verificare la scomposizione

della devianza ci conviene costruire una tabella come quella in Figura 5.13.

Figura 5.13: Foglio "Scomposizione della devianza" contenuto nel foglio di calcolo "Settore GICS delle comunicazioni"

	A	B	C	D	E	F	G	H
1	<b>Verifica della scomposizione della devianza (piano ai minimi quadrati)</b>							
2	TICKER	Variabile X1	Variabile X2	Variabile X3	X1cap1.23	Z1.23	X1cap1.23^2	Z1.23^2
3	GOOG	34,343	281,946	1096,723	34,140	0,203	1165,567	0,041
4	FB	18,485	112,844	648,350	22,074	-3,589	487,264	12,881
5	DIS	11,054	92,795	251,246	9,934	1,120	98,676	1,255
6	VZ	19,913	51,095	248,451	10,314	9,599	106,377	92,142
7	CMCSA	13,057	142,706	224,504	8,542	4,515	72,958	20,390
8	NFLX	1,867	126,379	215,702	8,450	-6,583	71,410	43,341
9	T	14,975	80,583	201,808	8,531	6,444	72,779	41,524
10	TMUS	3,468	85,494	160,957	7,204	-3,736	51,897	13,958
11	CHTR	1,668	21,511	131,000	6,988	-5,320	48,836	28,305
12	ATVI	1,736	54,417	59,363	4,389	-2,653	19,262	7,037
13								
14	<b>Somma</b>	<b>120,566</b>	<b>1049,770</b>	<b>3238,106</b>	<b>120,566</b>	<b>0,000</b>	<b>2195,024</b>	<b>260,875</b>
15	<b>Media</b>	<b>12,057</b>	<b>104,977</b>	<b>323,811</b>	<b>12,057</b>	<b>0,000</b>	<b>219,502</b>	<b>26,088</b>

La colonna E contiene i valori riprodotti  $\hat{x}_{1.23i}$ , la colonna F contiene i residui  $z_{1.23i}$ , la colonna G contiene i quadrati dei valori riprodotti e la colonna H contiene i quadrati dei residui. Usando i dati della tabella otteniamo

$$\begin{aligned} dev(\hat{X}_{1.23}) &= \sum_{i=1}^n \hat{x}_{1.23i}^2 - n \times \bar{x}_1^2 \\ &= 2195,024 - 10 \times 12,057^2 = 741,312, \end{aligned}$$

e sommando il valore di  $dev(\hat{X}_{1.23}) = 741,312$  al valore di

$$dev(Z_{1.23}) = \sum_{i=1}^n z_{1.23i}^2 = 260,875$$

( $dev(Z_{1.23})$  è la somma della colonna H) otteniamo

$$dev(\hat{X}_{1.23}) + dev(Z_{1.23}) = 741,312 + 260,875 = 1002,187. \quad (54)$$

Secondo la formula della scomposizione della devianza questa somma dovrebbe coincidere con il valore di  $dev(X_1)$ . Per ottenere il valore di  $dev(X_1)$  basta moltiplicare per  $n = 10$  il valore di  $\sigma_{11} = 100,228$  contenuto nella matrice di varianza covarianza in Figura 5.4. La piccola differenza tra il valore di

$$dev(X_1) = n \times \sigma_{11} = 10 \times 100,228 = 1002,28$$

e la somma nella (54) è dovuta solo agli arrotondamenti.  $\triangle$

**Esercizio 5.6.** Si verifichi che anche il piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento soddisfa la scomposizione della devianza (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

## 5.5 Indice di determinazione

Nella Sezione 3.5 abbiamo definito l'indice di determinazione di una retta ai minimi quadrati come rapporto tra la devianza spiegata e la devianza totale della variabile dipendente. D'ora in poi, per evitare ambiguità, indicheremo l'indice di determinazione della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  con  $I_{1.2}^2$ . Nella nuova notazione (ovvero nella notazione che abbiamo introdotto all'inizio della Sezione 5.2) la definizione di questo indice di determinazione è data da

$$I_{1.2}^2 = \frac{dev(\hat{X}_{1.2})}{dev(X_1)}.$$

Ovviamente, il rapporto tra la devianza spiegata e la devianza totale è ben definito anche nel caso di un piano ai minimi quadrati (a patto che la devianza totale sia positiva) e quindi possiamo definire l'indice di determinazione del piano ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  e  $X_3$  come

$$I_{1.23}^2 = \frac{dev(\hat{X}_{1.23})}{dev(X_1)}.$$

Siccome tutte le proprietà degli indici di determinazione di rette ai minimi quadrati sono semplici conseguenze della scomposizione della devianza (vedi Sezione 3.5), e siccome la scomposizione della devianza rimane valida anche per i piani ai minimi quadrati, possiamo concludere che le proprietà degli indici di determinazione di rette ai minimi quadrati (vedi Sezione 3.5) rimangono valide anche per gli indici di determinazione di piani ai minimi quadrati. Le proprietà in questione sono riportate nel riquadro sottostante:

L'indice di determinazione del piano ai minimo quadrati che descrive  $X_1$  in funzione di  $X_2$  e  $X_3$  è definito come

$$I_{1.23}^2 = \frac{\text{dev}(\hat{X}_{1.23})}{\text{dev}(X_1)} = 1 - \frac{\text{dev}(Z_{1.23})}{\text{dev}(X_1)}$$

e misura la bontà d'adattamento del piano ai minimi quadrati:

- $I_{1.23}^2 = 1$  se e solo se il piano ai minimi quadrati passa per tutti i punti osservati.
- $I_{1.23}^2 = 0$  se e solo se il piano ai minimi quadrati è orizzontale (ovvero parallelo al piano  $x_1 = 0$ ) e quindi non riesce a descrivere la variabilità di  $X_1$ .
- $0 < I_{1.23}^2 < 1$  nei casi intermedi. L'indice di determinazione è tanto più prossimo a 1 quanto più il piano ai minimi quadrati passa vicino ai punti osservati.

**Osservazione 5.2.** Anche per l'indice di determinazione di un piano ai minimi quadrati vale la relazione  $I_{1.23}^2 = 1 - [A_2^*]^2$ , dove  $A_2^*$  è l'indice di adattamento definito nella (1) (si confronti con l'Osservazione 3.1).  $\triangle$

**Esempio 5.8** (Capitalizzazione, Compensi e Utili). Nell'Esempio 5.1 abbiamo analizzato i dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione di mercato (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS delle comunicazioni e abbiamo ricavato i parametri del piano ai minimi quadrati che descrive  $X_1$  in funzione delle altre due variabili.

Per calcolare l'indice di determinazione di tale piano, ricordiamo innanzitutto che

$$\text{dev}(\widehat{X}_{1.23}) = 741,312 \quad \text{e che} \quad \text{dev}(X_1) = 1002,28$$

(vedi Esempio 5.7). L'indice di determinazione è quindi dato da

$$I_{1.23}^2 = \frac{\text{dev}(\widehat{X}_{1.23})}{\text{dev}(X_1)} = \frac{741,312}{1002,28} = 0,740.$$

Questo valore ci dice che il piano ai minimi quadrati spiega il 74% della variabilità degli utili (variabile  $X_1$ ). Ricordiamo che l'indice di determinazione della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  è solo pari a (vedi Esempio 3.5)

$$I_{1.2}^2 = \frac{\text{dev}(\widehat{X}_{1.2})}{\text{dev}(X_1)} = \frac{510,446}{1002,2084} = 0,509.$$

△

**Esercizio 5.7.** Si calcoli l'indice di determinazione del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (i dati sono contenuti nella cartella di lavoro ["Indice PMI e tasso di crescita del PIL reale"](#))

△

## 5.6 Miglioramento della bontà d'adattamento nel passaggio da una retta ai minimi quadrati ad un piano ai minimi quadrati

Consideriamo ora la retta ai minimi che descrive  $X_1$  in funzione di  $X_2$ , e il piano ai minimi quadrati che si ottiene aggiungendo la



variabile esplicativa  $X_3$ . Quale tra queste due funzioni interpolanti si adatta meglio ai dati? La risposta a questa domanda dipende dai dati oppure si può rispondere anche senza conoscere i dati?

Per rispondere a queste domande conviene ricordare che, *per definizione*, la coppia  $(\hat{c}_{1.2}, \hat{\alpha}_{12})$  che identifica una retta ai minimi quadrati deve essere una coppia  $(\alpha_1, \alpha_2)$  che minimizza la distanza quadratica

$$D_r(\alpha_1, \alpha_2) = \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i})^2,$$

e che la terna  $(\hat{c}_{1.23}, \hat{\alpha}_{12.3}, \hat{\alpha}_{13.2})$  che identifica un piano ai minimi quadrati deve essere una terna  $(\alpha_1, \alpha_2, \alpha_3)$  che minimizza la distanza quadratica

$$D_p(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^n (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \alpha_3 x_{3i})^2.$$

Siccome in corrispondenza di  $\alpha_1 = \hat{c}_{1.2}$  e  $\alpha_2 = \hat{\alpha}_{12}$  la funzione di  $D_r(\alpha_1, \alpha_2)$  raggiunge il suo punto di minimo, e siccome

$$D_p(\hat{c}_{1.2}, \hat{\alpha}_{12}, \alpha_3 = 0) = D_r(\hat{c}_{1.2}, \hat{\alpha}_{12}),$$

possiamo concludere che il valore minimo della funzione  $D_p(\alpha_1, \alpha_2, \alpha_3)$  non può mai essere maggiore di quello della funzione  $D_r(\alpha_1, \alpha_2)$ , ovvero che

$$D_p(\hat{c}_{1.23}, \hat{\alpha}_{12.3}, \hat{\alpha}_{13.2}) \leq D_r(\hat{c}_{1.2}, \hat{\alpha}_{12}). \quad (55)$$

Ma

$$\begin{aligned}
 D_p(\hat{c}_{1.23}, \hat{\alpha}_{12.3}, \hat{\alpha}_{13.2}) &= \sum_{i=1}^n (x_{1i} - \hat{c}_{1.23} - \hat{\alpha}_{12.3}x_{2i} - \hat{\alpha}_{13.2}x_{3i})^2 \\
 &= \sum_{i=1}^n (x_{1i} - \hat{x}_{1.23i})^2 \\
 &= \sum_{i=1}^n z_{1.23i}^2 = dev(Z_{1.23})
 \end{aligned}$$

e

$$\begin{aligned}
 D_r(\hat{c}_{1.2}, \hat{\alpha}_{12}) &= \sum_{i=1}^n (x_{1i} - \hat{c}_{1.2} - \hat{\alpha}_{12}x_{2i})^2 \\
 &= \sum_{i=1}^n (x_{1i} - \hat{x}_{1.2i})^2 \\
 &= \sum_{i=1}^n z_{1.2i}^2 = dev(Z_{1.2}).
 \end{aligned}$$

La disuguaglianza (55) può quindi essere espressa come

$$dev(Z_{1.23}) \leq dev(Z_{1.2}). \quad (56)$$

Questo ragionamento dimostra quindi che la devianza residua del piano ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  e  $X_3$  non può mai essere maggiore della devianza residua della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$ .

Dalla (56) discende immediatamente che (scomposizione della devianza)

$$dev(\hat{X}_{1.23}) \geq dev(\hat{X}_{1.2}) \quad \Rightarrow \quad I_{1.23}^2 \geq I_{1.2}^2$$

e scambiando i ruoli delle variabili esplicative  $X_2$  e  $X_3$  otteniamo anche

$$dev(\hat{X}_{1.23}) \geq dev(\hat{X}_{1.3}) \quad \Rightarrow \quad I_{1.23}^2 \geq I_{1.3}^2$$

Passando da una retta ai minimi quadrati ad un piano ai minimi quadrati mediante l'aggiunta di una variabile esplicativa il valore dell'indice di determinazione non può diminuire:

$$I_{1.23}^2 \geq I_{1.2}^2 \quad \text{e} \quad I_{1.23}^2 \geq I_{1.3}^2. \quad (57)$$

Come abbiamo appena dimostrato, nel passaggio da una retta ai minimi quadrati ad un piano ai minimi quadrati mediante l'aggiunta di una variabile esplicativa l'indice di determinazione può solo aumentare o al più rimanere costante. Per misurare l'entità del miglioramento della bontà d'adattamento di solito si calcolano uno o entrambi i seguenti due indici:

$$MVS_{1.2 \rightarrow 1.23} = \frac{\text{dev}(\hat{X}_{1.23}) - \text{dev}(\hat{X}_{1.2})}{\text{dev}(X_1)} = I_{1.23}^2 - I_{1.2}^2,$$

$$MVR_{1.2 \rightarrow 1.23} = \frac{\text{dev}(Z_{1.2}) - \text{dev}(Z_{1.23})}{\text{dev}(Z_{1.2})} = \frac{I_{1.23}^2 - I_{1.2}^2}{1 - I_{1.2}^2}. \quad (58)$$

L'indice MVS (acronimo di "Miglioramento della Varianza Spiegata") ci dice di quanto aumenta la devianza spiegata in rapporto alla devianza totale, mentre l'indice MVR (acronimo di "Miglioramento della Varianza Residua") ci informa di quanto si riduce la devianza residua in rapporto alla devianza residua iniziale, ovvero in rapporto alla devianza residua della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$ .

Usando la disuguaglianza (57) non è difficile dimostrare che gli indici MVS e MVR possono assumere solo valori appartenenti all'intervallo  $[0, 1]$ .

**Esempio 5.9** (Verifica numerica del miglioramento della bontà d'adattamento). Consideriamo ancora i dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione di mercato (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS. Nell'Esempio 3.5 abbiamo visto che l'indice di determinazione della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  è dato da

$$I_{1.2}^2 = \frac{\text{dev}(\hat{X}_{1.2})}{\text{dev}(X_1)} = \frac{510,446}{1002,2084} = 0,509,$$

mentre nell'Esempio 5.8 abbiamo visto che l'indice di determinazione del piano ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  e  $X_3$  è dato da

$$I_{1.23}^2 = \frac{\text{dev}(\hat{X}_{1.23})}{\text{dev}(X_1)} = \frac{741,312}{1002,28} = 0,740.$$

Come previsto dalla disuguaglianza (57) l'indice di determinazione del piano ai minimi quadrati è maggiore. Per valutare l'entità del miglioramento della bontà d'adattamento che si ottiene passando dalla retta al piano, calcoliamo

$$\begin{aligned} MVS_{1.2 \rightarrow 1.23} &= \frac{\text{dev}(\hat{X}_{1.23}) - \text{dev}(\hat{X}_{1.2})}{\text{dev}(X_1)} \\ &= I_{1.23}^2 - I_{1.2}^2 \\ &= 0,740 - 0,509 = 0,231 \end{aligned}$$

e

$$\begin{aligned}
 MVR_{1.2 \rightarrow 1.23} &= \frac{\text{dev}(Z_{1.2}) - \text{dev}(Z_{1.23})}{\text{dev}(Z_{1.2})} \\
 &= \frac{I_{1.23}^2 - I_{1.2}^2}{1 - I_{1.2}^2} \\
 &= \frac{0,740 - 0,509}{1 - 0,509} = 0,470.
 \end{aligned}$$

Il valore di  $MVS_{1.2 \rightarrow 1.23} = 0,231$  ci dice che in rapporto alla devianza totale la devianza spiegata aumenta di 23,1 punti percentuali e il valore di  $MVR_{1.2 \rightarrow 1.23} = 0,470$  ci dice che la devianza residua si riduce del 47%.

Ovviamente, il valore di  $I_{1.23}^2$  non può nemmeno essere più piccolo del valore di  $I_{1.3}^2$ . Per verificare numericamente anche questa disuguaglianza ricordiamo che nell'Esempio 5.2 abbiamo ricavato il valore del coefficiente di correlazione grezzo  $r_{13} = 0,859$  e che il suo quadrato è uguale all'indice di determinazione  $I_{1.3}^2$  (vedi Sezione 3.7):

$$I_{1.3}^2 = r_{13}^2 = 0,859^2 = 0,738.$$

Come previsto dalla disuguaglianza (57) (con il pedice "3" al posto del pedice "2"), il valore di  $I_{1.23}^2 = 0,740$  è anche maggiore di quello di  $I_{1.3}^2 = 0,738$ . Volendo calcolare anche gli indici  $MVS_{1.3 \rightarrow 1.23}$  e  $MVR_{1.3 \rightarrow 1.23}$  che misurano l'entità del miglioramento della bontà d'adattamento nel passaggio dalla retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_3$ , al piano ai minimi quadrati che

tiene conto anche della variabile esplicativa  $X_2$ , otteniamo

$$\begin{aligned} MVS_{1.3 \rightarrow 1.23} &= \frac{\text{dev}(\hat{X}_{1.23}) - \text{dev}(\hat{X}_{1.3})}{\text{dev}(X_1)} \\ &= I_{1.23}^2 - I_{1.3}^2 \\ &= 0,740 - 0,738 = 0,002 \end{aligned}$$

e

$$\begin{aligned} MVR_{1.3 \rightarrow 1.23} &= \frac{\text{dev}(Z_{1.3}) - \text{dev}(Z_{1.23})}{\text{dev}(Z_{1.3})} \\ &= \frac{I_{1.23}^2 - I_{1.2}^3}{1 - I_{1.3}^2} \\ &= \frac{0,740 - 0,738}{1 - 0,738} = 0,008. \end{aligned}$$

L'interpretazione di questi due indici è analoga a quella degli indici  $MVS_{1.2 \rightarrow 1.23}$  e  $MVR_{1.2 \rightarrow 1.23}$ .  $\triangle$

**Esercizio 5.8.** Si calcolino gli indici  $MVS_{1.2 \rightarrow 1.23}$  e  $MVR_{1.2 \rightarrow 1.23}$  per valutare il miglioramento della bontà d'adattamento nel passaggio dalla retta ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione del primo indice PMI del trimestre di riferimento, al piano ai minimi quadrati che include anche il secondo indice PMI del trimestre come variabile esplicativa (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

## 5.7 Formula indiretta per il calcolo della devianza spiegata

Nella Sezione 3.6 abbiamo visto che la devianza spiegata della retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_2$  può essere

calcolata attraverso la formula indiretta

$$dev(\hat{X}_{1.2}) = \hat{\alpha}_{12} \times codev(X_1, X_2)$$

che ci risparmia il calcolo dei valori riprodotti  $\hat{x}_{1.2i}$ . In questa sezione dimostreremo che la devianza spiegata di un piano ai minimi quadrati può essere calcolata mediante una formula analoga:

**Formula indiretta per la devianza spiegata:**

$$dev(\hat{X}_{1.23}) = \hat{\alpha}_{12.3} \times codev(X_1, X_2) + \hat{\alpha}_{13.2} \times codev(X_1, X_3) \quad (59)$$

Dividendo per  $n$  si ottiene l'analoga formula in termini di varianze e covarianze:

$$var(\hat{X}_{1.23}) = \hat{\alpha}_{12.3}\sigma_{12} + \hat{\alpha}_{13.2}\sigma_{13}. \quad (60)$$

La dimostrazione di questa formula indiretta è quasi identica a quella della formula indiretta che si riferisce alle rette ai minimi quadrati:

**Dimostrazione** (della formula indiretta (59)). Tenendo presente che

$$Z_{1.23} = X_1 - \hat{X}_{1.23} \quad \Rightarrow \quad \hat{X}_{1.23} = X_1 - Z_{1.23}$$

e usando le proprietà della codevianza (vedi Sezione 3.1) otteniamo

$$\begin{aligned} dev(\hat{X}_{1.23}) &= codev(\hat{X}_{1.23}, \hat{X}_{1.23}) \\ &= codev(\hat{X}_{1.23}, X_1 - Z_{1.23}) \\ &= codev(\hat{X}_{1.23}, X_1) - codev(\hat{X}_{1.23}, Z_{1.23}). \end{aligned}$$

Siccome  $codev(\hat{X}_{1.23}, Z_{1.23}) = 0$  (terza proprietà dei residui), possiamo dunque concludere che

$$dev(\hat{X}_{1.23}) = codev(\hat{X}_{1.23}, X_1). \quad (61)$$

Per ottenere la formula indiretta nella (59) ricordiamo ora che

$$\hat{X}_{1.23} = \hat{c}_{1.23} + \hat{\alpha}_{12.3} \times X_2 + \hat{\alpha}_{13.2} \times X_3$$

e che quindi (vedi le proprietà della covarianza in Sezione 3.1)

$$\begin{aligned} \text{codev}(\hat{X}_{1.23}, X_1) &= \text{codev}(\hat{c}_{1.23} + \hat{a}_{12.3} \times X_2 + \hat{a}_{13.2} \times X_3, X_1) \\ &= \hat{a}_{12.3} \times \text{codev}(X_1, X_2) + \hat{a}_{13.2} \times \text{codev}(X_1, X_3) \end{aligned}$$

Combinando questo risultato con l'uguaglianza (61) otteniamo la formula indiretta (59) che volevamo dimostrare.  $\square$

**Osservazione 5.3.** Nel corso della precedente dimostrazione abbiamo anche dimostrato che

$$\text{dev}(\hat{X}_{1.23}) = \text{codev}(\hat{X}_{1.23}, X_1) \quad (62)$$

(vedi l'equazione (61)). Ricordiamo che lo stesso risultato vale anche per i valori riprodotti da una retta ai minimi quadrati (vedi Osservazione 3.2).  $\triangle$

**Osservazione 5.4.** Non è difficile verificare che la varianza spiegata può essere anche espressa attraverso l'uso di determinanti come

$$\text{var}(\hat{X}_{1.23}) = \sigma_{11} - \frac{\begin{vmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix}} \quad (63)$$

$\triangle$

**Esempio 5.10** (Verifica numerica della formula indiretta per il calcolo della varianza spiegata). Per verificare numericamente la formula indiretta per il calcolo della devianza spiegata consideriamo ancora il piano ai minimi quadrati che abbiamo ricavato



nell'Esempio 5.1. Ricordiamo che i coefficienti angolari di questo piano sono dati da

$$\hat{\alpha}_{12.3} = -0,011 \quad \text{e} \quad \hat{\alpha}_{13.2} = 0,031$$

e che le covarianze  $\sigma_{12}$  e  $\sigma_{13}$  sono date da

$$\sigma_{12} = 488,818, \quad \text{e} \quad \sigma_{13} = 2556,957.$$

Applicando la formula indiretta per il calcolo della varianza spiegata (formula 60) otteniamo dunque

$$\begin{aligned} \text{var}(\hat{X}_{1.23}) &= \hat{\alpha}_{12.3}\sigma_{12} + \hat{\alpha}_{13.2}\sigma_{13} \\ &= -0,011 \times 488,818 + 0,031 \times 2556,957 = 73,8887. \end{aligned}$$

Moltiplicando per  $n = 10$  otteniamo

$$\text{dev}(\hat{X}_{1.23}) = n \times \text{var}(\hat{X}_{1.23}) = 10 \times 73,8887 = 738,887.$$

Questo valore è praticamente identico al valore  $\text{dev}(\hat{X}_{1.23}) = 741,312$  che avevamo calcolato nell'Esempio 5.7. La differenza è dovuta solo agli arrotondamenti.  $\triangle$

**Esercizio 5.9.** Si calcoli la varianza spiegata del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento usando la formula indiretta nella (60) (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale").  $\triangle$

## 5.8 Formula per calcolare l'indice di determinazione a partire dai coefficienti di correlazione lineare

Partendo dalla (63) si può dimostrare che l'indice di determinazione di un piano ai minimi quadrati può essere calcolato attraverso la formula

$$I_{1,23}^2 = 1 - \frac{\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \quad (64)$$

che coinvolge solo i coefficienti di correlazione lineare. Da questa formula si evince

- a) che anche l'indice di determinazione di un piano ai minimi quadrati non cambia se al posto delle variabili  $X_1$ ,  $X_2$ ,  $X_3$  consideriamo delle loro trasformazioni lineari affini  $Y_1 = a_1 + b_1X_1$ ,  $Y_2 = a_2 + b_2X_2$  e  $Y_3 = a_3 + b_3X_3$  strettamente crescenti o strettamente decrescenti (se tutte e tre le trasformazioni lineari affini sono strettamente crescenti, questa affermazione è un'immediata conseguenza del fatto che tutti e tre i coefficienti di correlazione lineare coinvolti rimangono invariati; un semplice ragionamento dimostra che il valore di  $I_{1,23}^2$  rimane invariato anche se una o più di una delle tre trasformazioni lineari affini sono strettamente decrescenti);<sup>15</sup>

<sup>15</sup>Usando direttamente la (63) si può dimostrare una proprietà di invarianza più forte.

b) che  $I_{1.23}^2 = r_{12}^2 + r_{13}^2 = I_{1.2}^2 + I_{1.3}^2$  se  $r_{23} = 0$ .

$$I_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}. \quad (65)$$

Ne consegue che

- anche l'indice di determinazione di un piano ai minimi quadrati è invariante rispetto a trasformazioni lineari affini;
- $I_{1.23}^2 = I_{1.2}^2 + I_{1.3}^2$  se  $r_{23} = 0$ .

**Esempio 5.11** (Verifica numerica della formula per calcolare il coefficiente di determinazione a partire dai coefficienti di correlazione lineare). Per verificare numericamente la formula (65) consideriamo ancora il piano ai minimi quadrati che abbiamo ricavato nell'Esempio 5.1. Nell'Esempio 5.2 abbiamo visto che i coefficienti di correlazione lineare coinvolti sono dati da

$$r_{12} = 0,714, \quad r_{13} = 0,859, \quad r_{23} = 0,855.$$

Applicando la formula (65) otteniamo dunque

$$\begin{aligned} I_{1.23}^2 &= \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \\ &= \frac{0,714^2 + 0,859^2 - 2 \times 0,714 \times 0,859 \times 0,855}{1 - 0,855^2} \\ &= 0,739. \end{aligned}$$

Questo valore è molto prossimo al valore  $I_{1.23}^2 = 0,740$  che avevamo già calcolato nell'Esempio 5.8. La differenza è dovuta solo agli arrotondamenti.  $\triangle$

**Esercizio 5.10.** Si calcoli l'indice di determinazione del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento usando la formula (65) (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")  $\triangle$

## 5.9 Il coefficiente di correlazione multiplo

Consideriamo ora il coefficiente di correlazione lineare

$$r(\hat{X}_{1.23}, X_1) = \frac{\text{codev}(\hat{X}_{1.23}, X_1)}{\sqrt{\text{dev}(\hat{X}_{1.23}) \times \text{dev}(X_1)}}.$$

Ricordando che

$$\text{codev}(\hat{X}_{1.23}, X_1) = \text{dev}(\hat{X}_{1.23})$$

(vedi l'Osservazione 5.3), vediamo che questo coefficiente di correlazione lineare può essere calcolato come

$$\begin{aligned} r(\hat{X}_{1.23}, X_1) &= \frac{\text{codev}(\hat{X}_{1.23}, X_1)}{\sqrt{\text{dev}(\hat{X}_{1.23}) \times \text{dev}(X_1)}} \\ &= \frac{\text{dev}(\hat{X}_{1.23})}{\sqrt{\text{dev}(\hat{X}_{1.23}) \times \text{dev}(X_1)}} \quad (66) \\ &= \sqrt{\frac{\text{dev}(\hat{X}_{1.23})}{\text{dev}(X_1)}} \\ &= \sqrt{I_{1.23}^2}. \end{aligned}$$

Il coefficiente di correlazione lineare  $r(\hat{X}_{1.23}, X_1)$  è noto come **coefficiente di correlazione multiplo** perché la variabile  $\hat{X}_{1.23}$  è una

combinazione lineare di *due* variabili: di  $X_2$  e  $X_3$ . Il coefficiente di correlazione multiplo  $r(X_1, \hat{X}_{1.23})$  viene solitamente indicato con  $R_{1.23}$ .

**Coefficiente di correlazione multiplo:** Il coefficiente di correlazione multiplo è definito come

$$R_{1.23} = r(\hat{X}_{1.23}, X_1) = \frac{\text{codev}(\hat{X}_{1.23}, X_1)}{\sqrt{\text{dev}(\hat{X}_{1.23}) \times \text{dev}(X_1)}}$$

e coincide con la radice quadrata principale (ovvero quella non negativa) dell'indice di determinazione  $I_{1.23}^2$ .

**Esempio 5.12** (Capitalizzazione, Compensi e Utili). Per calcolare il coefficiente di correlazione multiplo del piano ai minimi quadrati che abbiamo ricavato nell'Esempio 5.1 ricordiamo innanzitutto che il suo indice di determinazione è dato da  $I_{1.23}^2 = 0,740$  (vedi Esempio 5.8). Il valore del coefficiente di correlazione multiplo è quindi dato da

$$R_{1.23} = \sqrt{I_{1.23}^2} = \sqrt{0,740} = 0,860.$$

△

**Esercizio 5.11.** Si calcoli il coefficiente di correlazione multiplo del piano ai minimi quadrati che descrive il tasso di crescita del PIL reale in funzione dei primi due indici PMI del trimestre di riferimento (i dati sono contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale")

△

## 5.10 Coefficienti di correlazione parziale

Come abbiamo visto nella Sezione 3.7, il coefficiente di correlazione lineare  $r_{12}$  fornisce una misura per il *verso* e l'*intensità* della relazione lineare tra le osservazioni delle due variabili  $X_1$  e  $X_2$ . Come il coefficiente di regressione grezzo  $\hat{\alpha}_{12}$ , al quale è strettamente legato, il coefficiente di correlazione lineare  $r_{12}$  tiene conto solo delle osservazioni delle variabili  $X_1$  e  $X_2$  e ignora completamente i valori di tutte le altre variabili. Per questo motivo anche il coefficiente di correlazione lineare  $r_{12}$  viene chiamato "grezzo".

In questa sezione introdurremo la definizione di un indice statistico che fornisce una valutazione della correlazione lineare tra le osservazioni delle variabili  $X_1$  e  $X_2$  tenendo anche conto delle variazioni di una terza variabile  $X_3$ . L'indice in questione è noto come **coefficiente di correlazione "parziale"** ed è definito come

$$\begin{aligned} r_{12.3} &= r(Z_{1.3}, Z_{2.3}) \\ &= \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\sqrt{\text{dev}(Z_{1.3}) \times \text{dev}(Z_{2.3})}}. \end{aligned}$$

Per *motivare* la definizione del coefficiente di correlazione parziale  $r_{12.3}$  conviene ricordare che

$$x_{1i} = \hat{x}_{1.3i} + z_{1.3i}, \quad i = 1, 2, \dots, n,$$

e che la variabilità delle osservazioni  $x_{1i}$  può essere scomposta come (scomposizione della devianza)

$$\text{dev}(X_1) = \text{dev}(\hat{X}_{1.3}) + \text{dev}(Z_{1.3}).$$

Questa scomposizione ci dice che la variabilità di  $X_1$  può essere divisa in due parti:

- $dev(\widehat{X}_{1.3})$ , che rappresenta la parte della variabilità delle osservazioni  $x_{1i}$  che può essere spiegata attraverso una trasformazione lineare affine delle osservazioni  $x_{3i}$  (infatti  $\widehat{x}_{1.3i} = \widehat{c}_{1.3} + \widehat{a}_{13}x_{3i}$  è una trasformazione lineare di  $x_{3i}$  e nessun'altra trasformazione lineare affine delle osservazioni  $x_{3i}$  riesce a spiegare una quota maggiore della variabilità delle osservazioni  $x_{1i}$ );
- e  $dev(Z_{1.3})$ , che è la parte di variabilità residua delle osservazioni  $x_{1i}$  che non può essere spiegata attraverso trasformazioni lineari affini delle osservazioni  $x_{3i}$ .

Quindi possiamo pensare ai residui  $z_{1.3i} = x_{1i} - \widehat{x}_{1.3i}$  come la parte "residua" delle osservazioni  $x_{1i}$  dopo che da quest'ultima sia stata sottratta la parte di variabilità che può essere spiegata attraverso una trasformazione lineare affine delle osservazioni  $x_{3i}$ , e i residui  $z_{2.3i} = x_{2i} - \widehat{x}_{2.3i}$  possono essere interpretati in modo analogo. Interpretando i residui  $z_{1.3i}$  e  $z_{2.3i}$  in questo modo, possiamo interpretare il coefficiente di correlazione parziale  $r_{12.3}$  come un indice che misura la correlazione lineare tra le osservazioni delle variabili  $X_1$  e  $X_2$  *dopo che da queste ultime sia stata sottratta tutta la variabilità che può essere spiegata attraverso trasformazioni lineari affini delle osservazioni della variabile  $X_3$ .*

Il coefficiente di correlazione parziale  $r_{12.3}$  è *definito* come

$$\begin{aligned} r_{12.3} &= r(Z_{1.3}, Z_{2.3}) \\ &= \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\sqrt{\text{dev}(Z_{1.3}) \times \text{dev}(Z_{2.3})}}. \end{aligned} \quad (67)$$

Questo indice statistico fornisce una misura per la correlazione lineare tra le osservazioni di due variabili  $X_1$  e  $X_2$  che tiene anche in considerazione la variabilità delle osservazioni di una terza variabile  $X_3$ .

Come si vede nel prossimo esempio, coefficienti di correlazione parziale sono lo strumento "naturale" per indagare su casi di correlazione spuria.

**Esempio 5.13.** Per illustrare l'impiego dei coefficienti di correlazione parziale consideriamo ancora i dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS delle comunicazioni. Nella tabella in Figura 5.11 troviamo i valori di

$$\text{dev}(Z_{1.3}) = \sum_{i=1}^n z_{1.3i}^2 = 262,460,$$

$$\text{dev}(Z_{2.3}) = \sum_{i=1}^n z_{2.3i}^2 = 12617,630$$

e di

$$\text{codev}(Z_{1.3}, Z_{2.3}) = \sum_{i=1}^n z_{1.3i} \times z_{2.3i} = -141,412.$$



Il valore del coefficiente di correlazione parziale  $r_{12.3}$  è quindi dato da

$$\begin{aligned} r_{12.3} = r(Z_{1.3}, Z_{2.3}) &= \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\sqrt{\text{dev}(Z_{1.3}) \times \text{dev}(Z_{2.3})}} \\ &= \frac{-141,412}{\sqrt{262,460 \times 12617,630}} = -0,077. \end{aligned}$$

Questo valore ci dice che *al netto della variabilità che può essere spiegata attraverso trasformazioni lineari affini dei valori della capitalizzazione di mercato (variabile  $X_3$ )*, il verso della correlazione tra gli utili (variabile  $X_1$ ) e i compensi dei dirigenti (variabile  $X_2$ ) è negativo e l'*intensità* della correlazione è pari al 7% del massimo teorico. Notiamo immediatamente il valore di  $r_{12.3}$  è molto diverso da quello di  $r_{12} = 0,714$  che avevamo calcolato nell'Esempio 5.2. La differenza è dovuta al fatto che il coefficiente di correlazione grezzo  $r_{12} = r(X_1, X_2)$  non tiene conto della variabilità dei valori della capitalizzazione (variabile  $X_3$ ). Come abbiamo visto nell'Esempio 5.4 dove abbiamo illustrato il fenomeno della correlazione spuria, le variazioni nei valori di una terza variabile (in questo caso  $X_3$ ) potrebbero sempre essere interpretate come la "*causa*" dell'elevata o bassa correlazione tra le osservazioni di due variabili (in questo caso  $X_1$  e  $X_2$ ). Per verificare se i dati supportano questa ipotesi, si può quindi ricorrere ad un confronto tra il valore di un opportuno coefficiente di correlazione parziale (in questo caso  $r_{12.3}$ ) e il corrispondente coefficiente di correlazione grezzo (in questo caso  $r_{12}$ ), così come nell'Esempio 5.4 abbiamo confrontato il coefficiente di regressione parziale  $\hat{\alpha}_{12.3}$  con il corrispondente coefficiente di regressione grezzo  $\hat{\alpha}_{12}$ . In ogni

caso, bisogna sempre tenere presente che anche se il coefficiente di correlazione (o regressione) grezzo e quello parziale sono molto diversi, questo *non dimostra* che l'elevata o bassa correlazione sia stata necessariamente "*causata*" delle variazioni nei valori della terza variabile. Infatti, potrebbe sempre esistere un'altra variabile  $X_4$  con la quale otteniamo un valore di  $r_{12.4}$  molto simile a quello di  $r_{12.3}$ .  $\triangle$

**Esercizio 5.12.** Si considerino ancora i dati sulle quotazioni dell'oro (variabile  $X_1$ ), dell'indice S&P500 (variabile  $X_2$ ) e dell'indice CPI (variabile  $X_3$ ) contenuti nella cartella di lavoro "[Esempio di correlazione spuria](#)". Si calcoli il valore di  $r_{12.3}$  e lo si confronti con il valore di  $r_{12} = 0,831$  che avevamo già calcolato nell'[Esempio 5.4](#). Che cosa emerge dal confronto?  $\triangle$

**Esercizio 5.13.** Si considerino ancora di dati sul tasso di crescita del PIL reale e sugli indici PMI contenuti nella cartella di lavoro "[Indice PMI e tasso di crescita del PIL reale](#)". Sia  $X_1$  la variabile che rappresenta i tassi di crescita del PIL reale e siano  $X_2$  e  $X_3$  le variabili che rappresentano i primi due indici PMI di ciascun trimestre. Si calcoli il valore di  $r_{12.3}$  e lo si confronti con quello di  $r_{12}$ .  $\triangle$

In alternativa (o in aggiunta), la definizione del coefficiente di correlazione parziale  $r_{12.3}$  può essere motivata attraverso l'osservazione che  $r_{12.3}$  è la media geometrica dei due coefficienti di regressione parziale  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$ , così come  $r_{12}$  è la media geometrica dei due coefficienti di regressione grezzi  $\hat{\alpha}_{12}$  e  $\hat{\alpha}_{21}$  (vedi Sezione

3.7). Per dimostrare questa affermazione ricordiamo innanzitutto che secondo la formula alternativa (51) si ha

$$\hat{\alpha}_{12.3} = \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\text{dev}(Z_{2.3})} \quad \text{e} \quad \hat{\alpha}_{21.3} = \frac{\text{codev}(Z_{1.3}, Z_{2.3})}{\text{dev}(Z_{1.3})}$$

e notiamo che i due coefficienti di regressione parziale  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$  hanno sempre lo stesso segno che coincide con il segno della covarianza  $\text{codev}(Z_{1.3}, Z_{2.3})$ . La media geometrica dei due coefficienti di regressione parziali  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$  è quindi ben definita e, come è facile verificare, coincide con il valore del coefficiente di correlazione parziale  $r_{12.3}$ .

Il coefficiente di correlazione parziale  $r_{12.3}$  coincide con la media geometrica dei due coefficienti di regressione parziale  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$ :

$$\begin{aligned} r_{12.3} &= \text{segno}(\hat{\alpha}_{12.3}) \times \sqrt{\hat{\alpha}_{12.3} \times \hat{\alpha}_{21.3}} \\ &= \text{segno}(\hat{\alpha}_{21.3}) \times \sqrt{\hat{\alpha}_{12.3} \times \hat{\alpha}_{21.3}}. \end{aligned} \quad (68)$$

**Esempio 5.14** (Verifica numerica della formula (68)). Per effettuare una verifica numerica della formula (68), faremo ancora riferimento ai dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS delle comunicazioni. Ricordiamo che nell'Esempio 5.1 abbiamo già calcolato il valore di

$$\hat{\alpha}_{12.3} = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = -0,011.$$

Scambiando i ruoli delle variabili  $X_1$  e  $X_2$  (ovvero scambiando i pedici "1" e "2") otteniamo anche

$$\begin{aligned}\hat{\alpha}_{21.3} &= \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{11}\sigma_{33} - \sigma_{13}^2} \\ &= \frac{488,818 \times 88372,821 - 2556,957 \times 17383,135}{-2556,957^2} \\ &= -0,539\end{aligned}$$

(le varianze e covarianze sono contenute nella matrice di varianza e covarianza in Figura 5.4). La media geometrica di  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$  è quindi data da

$$-\sqrt{\hat{\alpha}_{12.3} \times \hat{\alpha}_{21.3}} = -\sqrt{(-0,011) \times (-0,539)} = -0,077.$$

Come previsto dalla formula (68), questo valore coincide esattamente con il valore di  $r_{12.3}$  che avevamo calcolato nell'Esempio 5.13.  $\triangle$

**Esercizio 5.14.** Si considerino ancora di dati sul tasso di crescita del PIL reale e sugli indici PMI contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale"). Sia  $X_1$  la variabile che rappresenta i tassi di crescita del PIL reale e siano  $X_2$  e  $X_3$  le variabili che rappresentano i primi due indici PMI di ciascun trimestre. Si calcoli il valore di  $r_{12.3}$  come media geometrica di due opportuni coefficienti di regressione parziale.  $\triangle$

Usando la formula (68) possiamo anche ricavare una formula che ci permette di calcolare  $r_{12.3}$  a partire dai coefficienti di correlazione grezzi. Per ottenere questa formula ricordiamo innanzitutto

che secondo la formula alternativa (48) i coefficienti di regressione parziale  $\hat{\alpha}_{12.3}$  e  $\hat{\alpha}_{21.3}$  possono essere espressi come

$$\hat{\alpha}_{12.3} = \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{23}r_{13}}{1 - r_{23}^2} \quad \text{e} \quad \hat{\alpha}_{21.3} = \frac{\sigma_2}{\sigma_1} \times \frac{r_{12} - r_{13}r_{23}}{1 - r_{13}^2}.$$

Da queste formule deduciamo che

$$\text{segno}(\hat{\alpha}_{12.3}) = \text{segno}(\hat{\alpha}_{21.3}) = \text{segno}(r_{12} - r_{13}r_{23}),$$

e sostituendo nella formula (68) otteniamo la formula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}.$$

Il coefficiente di correlazione parziale  $r_{12.3}$  può essere calcolato a partire dai coefficienti di correlazione grezzi attraverso la formula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}. \quad (69)$$

**Esempio 5.15** (Verifica numerica della formula (69)). Per effettuare una verifica numerica della formula (69), faremo ancora riferimento ai dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS delle comunicazioni. Ricordiamo che nell'Esempio 5.2 abbiamo già ricavato i valori dei coefficienti di correlazione grezzi

$$r_{12} = 0,714, \quad r_{13} = 0,859, \quad r_{23} = 0,855.$$

Sostituendoli nella formula (69) otteniamo

$$\begin{aligned} r_{12.3} &= \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \\ &= \frac{0,714 - 0,859 \times 0,855}{\sqrt{(1 - 0,859^2)(1 - 0,855^2)}} = -0,077 \end{aligned}$$

e questo valore coincide esattamente con il valore di  $r_{12.3}$  che avevamo già calcolato nell'Esempio 5.13.  $\triangle$

**Esercizio 5.15.** Si considerino ancora di dati sul tasso di crescita del PIL reale e sugli indici PMI contenuti nella cartella di lavoro "Indice PMI e tasso di crescita del PIL reale". Sia  $X_1$  la variabile che rappresenta i tassi di crescita del PIL reale e siano  $X_2$  e  $X_3$  le variabili che rappresentano i primi due indici PMI di ciascun trimestre. Si calcolino i valori di  $r_{12.3}$  e di  $r_{13.2}$  a partire dai coefficienti di correlazione grezzi.  $\triangle$

Un'ultima importante proprietà del coefficiente di correlazione parziale  $r_{12.3}$  è quella che lo lega all'indice  $MVR_{1.3 \rightarrow 1.23}$  che misura il miglioramento della varianza residua nel passaggio dalla retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_3$ , al piano ai minimi quadrati che si ottiene aggiungendo la variabile esplicativa  $X_2$ . Infatti, si può dimostrare che

$$r_{12.3}^2 = \frac{I_{1.23}^2 - I_{1.3}^2}{1 - I_{1.3}^2} = MVR_{1.3 \rightarrow 1.23}$$

Il quadrato del coefficiente di correlazione parziale  $r_{12.3}^2$  coincide con l'indice  $MVR_{1.3 \rightarrow 1.23}$  che misura il miglioramento della varianza residua nel passaggio dalla retta ai minimi quadrati che descrive  $X_1$  in funzione di  $X_3$ , al piano ai minimi quadrati che si ottiene aggiungendo la variabile esplicativa  $X_2$ :

$$r_{12.3}^2 = \frac{I_{1.23}^2 - I_{1.3}^2}{1 - I_{1.3}^2} = MVR_{1.3 \rightarrow 1.23} \quad (70)$$

**Esempio 5.16** (Verifica numerica della formula (70)). Per effettuare una verifica numerica della formula (70), faremo ancora riferimento ai dati sugli utili (variabile  $X_1$ ), i compensi dei dirigenti (variabile  $X_2$ ) e la capitalizzazione (variabile  $X_3$ ) delle  $n = 10$  più grandi società del settore GICS delle comunicazioni. Ricordiamo che nell'Esempio 5.13 abbiamo calcolato il valore di  $r_{12.3} = -0,077$ . Elevandolo al quadrato otteniamo

$$r_{12.3}^2 = (-0,077)^2 = 0,006.$$

Secondo la formula (70) questo valore dovrebbe coincidere con quello di  $MVR_{1.3 \rightarrow 1.23}$ , che nell'Esempio 5.9 abbiamo visto essere dato da

$$MVR_{1.3 \rightarrow 1.23} = 0,008.$$

La piccola differenza tra il valore di  $r_{12.3}^2$  e quello di  $MVR_{1.3 \rightarrow 1.23}$  è dovuta solo agli arrotondamenti.  $\triangle$

**Esercizio 5.16.** Si considerino ancora di dati sul tasso di crescita del PIL reale e sugli indici PMI contenuti nella cartella di lavoro

”Indice PMI e tasso di crescita del PIL reale”). Sia  $X_1$  la variabile che rappresenta i tassi di crescita del PIL reale e siano  $X_2$  e  $X_3$  le variabili che rappresentano i primi due indici PMI di ciascun trimestre. Si calcolino i valori di  $MVR_{1.2 \rightarrow 1.23}$  e di  $MVR_{1.3 \rightarrow 1.23}$  come quadrati di due opportuni coefficienti di correlazione parziale.

△

**Dimostrazione** (della formula (70)). Usando la formula  $r_{12.3}$  usando la formula (69) otteniamo

$$\begin{aligned}
 r_{12.3}^2 &= \frac{(r_{12} - r_{13}r_{23})^2}{(1 - r_{13}^2) \times (1 - r_{23}^2)} \\
 &= \frac{r_{12}^2 + r_{13}^2 r_{23}^2 - 2r_{12}r_{13}r_{23}}{(1 - r_{13}^2) \times (1 - r_{23}^2)} \\
 &= \frac{(r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}) - r_{13}^2 + r_{13}^2 r_{23}^2}{(1 - r_{13}^2) \times (1 - r_{23}^2)} \\
 &= \frac{(r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}) - r_{13}^2(1 - r_{23}^2)}{(1 - r_{13}^2) \times (1 - r_{23}^2)} \\
 &= \frac{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} - r_{13}^2}{(1 - r_{13}^2)}.
 \end{aligned}$$

Siccome il rapporto al numeratore nell’ultima riga (quello evidenziato in verde) coincide con il secondo membro della formula (65) per calcolare  $I_{1.23}^2$  a partire dai coefficienti di correlazione lineare grezzi, e siccome  $r_{13}^2 = I_{1.3}^2$ , possiamo concludere che

$$r_{12.3}^2 = \frac{I_{1.23}^2 - I_{1.3}^2}{1 - I_{1.3}^2}$$

A questo punto non è difficile verificare che il secondo membro di questa formula coincide con la formula definitoria dell’indice  $MVR_{1.3 \rightarrow 1.23}$  (basta scambiare i pedici ”2” e ”3” nella formula (58)). □



## 6 Interpolazione con la famiglia delle funzioni di produzione di Cobb-Douglas generalizzata

### 6.1 La famiglia delle funzioni di produzione di Cobb-Douglas generalizzata: definizione e proprietà

Come abbiamo visto nella Sezione 4.5, secondo uno studio effettuato da Cobb e Douglas, durante il periodo storico compreso tra l'anno 1899 e l'anno 1922 la relazione di fondo tra il volume della produzione  $P$  e le quantità di lavoro  $L$  e di capitale  $C$  impegnate nella produzione poteva essere descritta attraverso una funzione del tipo

$$P = \alpha_1 L^{\alpha_2} C^{1-\alpha_2}$$

con parametro  $\alpha_2$  compreso tra 0 e 1. Come abbiamo visto nella Sezione 4.5, questa forma funzionale implica *rendimenti di scala costanti*: moltiplicando  $L$  e  $C$  per una costante  $h > 0$ , il volume della produzione  $P$  risulterà anch'esso moltiplicato per la stessa costante  $h$ .

Ammettendo anche la possibilità che l'esponente  $1 - \alpha_2$  di  $C$  sia slegato dall'esponente  $\alpha_2$  di  $L$ , possiamo ottenere anche rendimenti di scala crescenti e decrescenti. Infatti, se

$$P = \alpha_1 L^{\alpha_2} C^{\alpha_3},$$

allora non è difficile verificare che moltiplicando  $L$  e  $C$  per una stessa quantità positiva  $h$ , il volume della produzione  $P$  risulterà moltiplicato per  $h^{\alpha_2 + \alpha_3}$ . Questo significa che un aumento di  $L$  e  $C$  in proporzioni identiche comporti

- un aumento in proporzione maggiore del volume della produzione  $P$  nel caso in cui  $\alpha_2 + \alpha_3 > 1$  (rendimenti di scala crescenti),
- un aumento nella stessa proporzione del volume della produzione  $P$  nel caso in cui  $\alpha_2 + \alpha_3 = 1$  (rendimenti di scala costanti),
- un aumento in proporzione minore del volume della produzione  $P$  nel caso in cui  $\alpha_2 + \alpha_3 < 1$  (rendimenti di scala decrescenti).

Queste considerazioni ci suggeriscono di considerare anche la famiglia delle funzioni di produzione di Cobb-Douglas *generalizzata*, ovvero la famiglia che contiene tutte le funzioni del tipo

$$x_1 = f(x_2, x_3; \alpha_1, \alpha_2, \alpha_3) = \alpha_1 x_2^{\alpha_2} x_3^{\alpha_3}, \quad x_2, x_3 > 0$$

dove  $\alpha_1$  un parametro reale positivo e dove  $\alpha_2$  e  $\alpha_3$  sono due parametri reali liberi di variare in modo indipendente l'uno dall'altro. In quanto segue, per semplicità, useremo il termine "funzione di Cobb-Douglas" per indicare una qualunque funzione di questo tipo, anche se a volte questo termine viene riservato solo al caso in cui  $\alpha_3 = 1 - \alpha_2$ .

La **famiglia delle funzioni di produzione di Cobb-Douglas generalizzata** è l'insieme di tutte le funzioni che possono essere definite come

$$x_1 = f(x_2, x_3; \alpha_1, \alpha_2, \alpha_3) = \alpha_1 x_2^{\alpha_2} x_3^{\alpha_3}, \quad x_2, x_3 > 0 \quad (71)$$

con  $\alpha_1$  un parametro reale positivo e con  $\alpha_2$  e  $\alpha_3$  due parametri reali qualunque.

A seconda del valore di  $\alpha_2 + \alpha_3$ , la corrispondente funzione di Cobb-Douglas può presentare

- rendimenti di scala crescenti (se  $\alpha_2 + \alpha_3 > 1$ ),
- rendimenti di scala costanti (se  $\alpha_2 + \alpha_3 = 1$ ),
- oppure rendimenti di scala decrescenti (se  $\alpha_2 + \alpha_3 < 1$ ).

Come si vede dall'equazione (71) che definisce una funzione di Cobb-Douglas, la famiglia di queste funzioni può essere considerata come controparte della famiglia delle funzioni potenza per il caso in cui si avessero *due* variabili esplicative. Notiamo immediatamente che anche nella famiglia delle funzioni di Cobb-Douglas generalizzata il parametro  $\alpha_1$  è un *parametro di scala* che restituisce il valore assunto da una funzione di Cobb-Douglas quando entrambe le variabili indipendenti assumono il valore 1. Per comprendere il ruolo dei parametri  $\alpha_2$  e  $\alpha_3$  conviene prendere il logaritmo naturale ad ambo i membri dell'equazione che definisce una funzione di Cobb-Douglas. Così facendo si ottiene

$$\ln x_1 = \alpha_1 + \alpha_2 \ln x_2 + \alpha_3 \ln x_3,$$

e ricordando la relazione tra variazioni logaritmiche a variazioni relative, possiamo dunque concludere che

- il parametro  $\alpha_2$  è il valore per il quale bisogna moltiplicare l'ammontare di una piccola variazione relativa della variabile  $x_2$  per ottenere l'ammontare della corrispondente variazione relativa di  $x_1$  . . .
- . . . e che  $\alpha_3$  è il valore per il quale bisogna moltiplicare l'ammontare di una piccola variazione relativa della variabile  $x_3$  per ottenere l'ammontare della corrispondente variazione relativa di  $x_1$ .

L'interpretazione dei parametri  $\alpha_2$  e  $\alpha_3$  di una funzione di Cobb-Douglas è quindi del tutto analoga all'interpretazione del parametro  $\alpha_2$  di una funzione potenza. Tuttavia, siccome i parametri  $\alpha_2$  e  $\alpha_3$  di una funzione di Cobb-Douglas sono le derivate *parziali* di una funzione di due variabili (infatti,  $\alpha_2 = \partial \ln x_1 / \partial \ln x_2$  e  $\alpha_3 = \partial \ln x_1 / \partial \ln x_3$ ), mentre il parametro  $\alpha_2$  di una funzione potenza è la derivata ordinaria di una funzione di un'unica variabile (infatti, se  $x_1 = \alpha_1 x_2^{\alpha_2}$ , allora  $\alpha_2 = d \ln x_1 / d \ln x_2$ ), i parametri  $\alpha_2$  e  $\alpha_3$  di una funzione di Cobb-Douglas vengono chiamati "*elasticità parziali*".

- Il parametro  $\alpha_1$  di una funzione di Cobb-Douglas è un *parametro di scala* che restituisce il valore assunto dalla funzione di Cobb-Douglas quando entrambe le variabili indipendenti assumono il valore 1.
- Il parametro  $\alpha_2$  di una funzione di Cobb-Douglas è il valore per il quale bisogna moltiplicare l'ammontare di una piccola variazione relativa della variabile indipendente  $x_2$  per ottenere l'ammontare della corrispondente variazione relativa della variabile dipendente  $x_1$ .

Il parametro  $\alpha_2$  descrive la relazione tra variazioni relative delle variabili  $x_1$  e  $x_2$  *sotto l'ipotesi che l'altra variabile indipendente  $x_3$  rimanga costante*.

- L'interpretazione del parametro  $\alpha_3$  è analoga a quella di  $\alpha_2$ .

## 6.2 Determinazione dei parametri di una funzione interpolante di Cobb-Douglas.

Come abbiamo già osservato nella sezione precedente, la famiglia delle funzioni di Cobb-Douglas generalizzata può essere considerata la controparte della famiglia delle funzioni potenza per il caso in cui si avessero *due* variabili esplicative.

Per determinare i parametri di una funzione interpolante di Cobb-Douglas possiamo ricorrere ad un metodo analogo a quello che abbiamo già visto nella Sezione 4.4 per determinare i parametri

di una funzione interpolante potenza. In questo caso il metodo parte dal presupposto che

$$x_1 = \alpha_1 x_2^{\alpha_2} x_3^{\alpha_3} \quad \Leftrightarrow \quad \ln x_1 = \ln \alpha_1 + \alpha_2 \ln x_2 + \alpha_3 \ln x_3$$

e prevede di determinare i parametri  $\alpha_1 = \tilde{\alpha}_1$ ,  $\alpha_2 = \tilde{\alpha}_2$  e  $\alpha_3 = \tilde{\alpha}_3$  di una funzione interpolante di Cobb-Douglas ponendo

- $\ln \tilde{\alpha}_1$  (attenzione al logaritmo!!!) uguale all'intercetta del piano ai minimi quadrati che considera  $Y_1 = \ln X_1$  come variabile dipendente e  $Y_2 = \ln X_2$  e  $Y_3 = \ln X_3$  come variabili esplicative,
- e ponendo  $\tilde{\alpha}_2$  e  $\tilde{\alpha}_3$  uguale ai coefficienti angolari dello stesso piano.

Si noti che questa funzione interpolante è univocamente determinata se e solo se i punti  $(\ln x_{2i}, \ln x_{3i})$  non sono collineari.

Per ottenere i parametri  $\alpha_1 = \tilde{\alpha}_1$ ,  $\alpha_2 = \tilde{\alpha}_2$  e  $\alpha_3 = \tilde{\alpha}_3$  di una funzione interpolante di Cobb-Douglas, basta porre

- $\ln \tilde{\alpha}_1$  (attenzione al logaritmo) uguale all'intercetta del piano ai minimi quadrati che considera  $Y_1 = \ln X_1$  come variabile dipendente e  $Y_2 = \ln X_2$  e  $Y_3 = \ln X_3$  come variabili esplicative,
- e porre  $\tilde{\alpha}_2$  e  $\tilde{\alpha}_3$  uguale ai coefficienti angolari dello stesso piano.

I valori di questi parametri sono univocamente determinati se e solo se i punti  $(\ln x_{2i}, \ln x_{3i})$  non sono collineari (come sempre accade nelle applicazioni).

E' importante tenere presente che, tranne in casi banali che non sono d'interesse nelle applicazioni, i parametri  $\tilde{\alpha}_1$ ,  $\tilde{\alpha}_2$  e  $\tilde{\alpha}_3$  determinati secondo il metodo appena descritto *non* minimizzano la distanza quadratica

$$D(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^n (x_{1i} - \alpha_1 x_{2i}^{\alpha_2} x_{3i}^{\alpha_3})^2$$

bensì una funzione di distanza diversa che è definita come

$$\tilde{D}(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^n (\ln x_{1i} - \ln \alpha_1 + \alpha_2 \ln x_{2i} + \alpha_3 \ln x_{3i})^2.$$

Per questo motivo, la funzioni interpolante

$$\hat{x}_1 = f(x_2, x_3) = \tilde{\alpha}_1 x_2^{\tilde{\alpha}_2} x_3^{\tilde{\alpha}_3}, \quad x_2, x_3 > 0,$$

non può essere chiamata funzione di Cobb-Douglas "ai minimi quadrati".

Anche per le funzioni interpolanti di Cobb-Douglas determinate secondo il metodo qui descritto valgono le stesse considerazioni che abbiamo già visto per le funzioni interpolanti potenza determinate secondo il metodo descritto nella Sezione 4.4. In particolare, i valori riprodotti

$$\hat{x}_{1i} = \tilde{\alpha}_1 x_{2i}^{\tilde{\alpha}_2} x_{3i}^{\tilde{\alpha}_3}$$

e i corrispondenti residui  $z_i = x_{1i} - \hat{x}_{1i}$  non soddisfano la scomposizione della devianza e possono verificarsi situazioni dove il rapporto tra la devianza spiegata  $dev(\hat{X}_1)$  e la devianza totale  $dev(X_1)$  è maggiore di 1.

### 6.3 Applicazione

In questa sezione illustreremo il procedimento per determinare una funzione interpolante di Cobb-Douglas secondo il metodo che abbiamo descritto nella sezione precedente. Considereremo ancora i dati sul PIL, sul numero di occupati e sugli investimenti fissi lordi che abbiamo già analizzato nella Sezione 4.5. A differenza di quanto abbiamo fatto nella 4.5, questa volta indicheremo con  $X_1$  la variabile che rappresenta il PIL, con  $X_2$  la variabile che rappresenta il numero di occupati e con  $X_3$  la variabile che rappresenta gli investimenti fissi lordi. I dati con le nuove definizioni delle variabili  $X_1$ ,  $X_2$  e  $X_3$  sono contenuti nel foglio "Regressione CD generalizzata" della cartella di lavoro "[Regressione di Cobb e Douglas](#)".



Per determinare la funzione interpolante di Cobb-Douglas, dobbiamo innanzitutto trasformare tutte le osservazioni nei loro logaritmi naturali così come nella tabella in Figura 6.1.

Figura 6.1: foglio "Regressione CD generalizzata" della cartella di lavoro "Regressione di Cobb e Douglas"

	A	B	C	D	E	F	G
1	Dati rielaborati (cambio dell'unità di misura degli investimenti fissi lordi da USD a miliardi di USD)						
2	X1 = Prodotto interno lordo espresso in miliardi di USD concatenati con anno di riferimento 2012						
3	X2 = numero di occupati (in migliaia di unità)						
4	X3 = Investimenti fissi lordi espressi in miliardi di USD concatenati con anno di riferimento 2012						
5	Data	X1	X2	X3	Y1=ln(X1)	Y2=ln(X2)	Y3=ln(X3)
6	2010-01-01	15415,145	138590	2757,249	9,643	11,839	7,922
7	2010-04-01	15557,277	139226	2845,626	9,652	11,844	7,954
8	2010-07-01	15671,967	139338	2856,404	9,660	11,845	7,957
	A	B	C	D	E	F	G
47	2020-04-01	17302,511	137565,000	3795,484	9,759	11,832	8,242
48	2020-07-01	18596,521	146199,000	4016,493	9,831	11,893	8,298
49	2020-10-01	18780,325	149769	4163,881	9,841	11,917	8,334
50							
51	<b>Somme</b>	<b>762210,461</b>	<b>6505605</b>	<b>154909,647</b>	<b>429,331</b>	<b>523,735</b>	<b>359,026</b>

Usando i totali delle colonne E, F e G otteniamo (ricordiamo che il numero di trimestri è  $n = 44$ )

$$\bar{y}_1 = \frac{1}{n} \sum_{i=1}^n y_{1i} = \frac{429,331}{44} = 9,758,$$

$$\bar{y}_2 = \frac{1}{n} \sum_{i=1}^n y_{2i} = \frac{523,735}{44} = 11,903,$$

$$\bar{y}_3 = \frac{1}{n} \sum_{i=1}^n y_{3i} = \frac{359,026}{44} = 8,160$$

e applicando le consuete formule otteniamo anche la matrice di varianza e covarianza in Figura 6.2.

Figura 6.2: Matrice di varianza e covarianza delle variabili  $Y_1 = \ln X_1$ ,  $Y_2 = \ln X_2$  e  $Y_3 = \ln X_3$ .

	<b>Y1</b>	<b>Y2</b>	<b>Y3</b>
<b>Y1</b>	0,00454		
<b>Y2</b>	0,00269	0,00184	
<b>Y3</b>	0,00773	0,00440	0,01369

A questo punto possiamo calcolare i due coefficienti angolari del piano ai minimi quadrati che considera  $Y_1$  come variabile dipendente e  $Y_2$  e  $Y_3$  come variabili esplicative:

$$\begin{aligned}\tilde{\alpha}_2 &= \frac{\sigma_{Y12} \times \sigma_{Y33} - \sigma_{Y13} \times \sigma_{Y23}}{\sigma_{Y22} \times \sigma_{Y33} - \sigma_{Y23}^2} \\ &= \frac{0,00269 \times 0,01369 - 0,00773 \times 0,00440}{0,00184 \times 0,01369 - 0,00440^2} = 0,483\end{aligned}$$

e

$$\begin{aligned}\tilde{\alpha}_3 &= \frac{\sigma_{Y13} \times \sigma_{Y22} - \sigma_{Y12} \times \sigma_{Y23}}{\sigma_{Y22} \times \sigma_{Y33} - \sigma_{Y23}^2} \\ &= \frac{0,00773 \times 0,00184 - 0,00269 \times 0,00440}{0,00184 \times 0,01369 - 0,00440^2} = 0,409.\end{aligned}$$

L'intercetta del piano ai minimi quadrati, che corrisponde al logaritmo di  $\tilde{\alpha}_1$ , è data da

$$\begin{aligned}\ln \tilde{\alpha}_1 &= \bar{y}_1 - \tilde{\alpha}_2 \times \bar{y}_2 - \tilde{\alpha}_3 \times \bar{y}_3 \\ &= 9,758 - 0,483 \times 11,903 - 0,409 \times 8,160 = 0,671 \\ &\Rightarrow \tilde{\alpha}_1 = e^{0,671} = 1,956.\end{aligned}$$

L'equazione del piano ai minimi quadrati che considera  $Y_1 = \ln X_1$  come variabile dipendente e  $Y_2 = \ln X_2$  e  $Y_3 = \ln X_3$  come variabili

esplicative è quindi data da

$$\hat{y}_1 = 0,671 + 0,483y_2 + 0,409y_3.$$

Applicando la funzione esponenziale ad ambo i membri di questa equazione (ricordiamo che  $\hat{y}_1 = \ln \hat{x}_1$ ,  $y_2 = \ln x_2$  e che  $y_3 = \ln x_3$ ) otteniamo la funzione interpolante di Cobb-Douglas di cui nella sezione precedente:

$$\hat{x}_1 = 1,956 \times x_2^{0,483} \times x_3^{0,409}.$$

Ovviamente, la bontà d'adattamento di questa funzione interpolante è strettamente legata a quella del piano ai minimi quadrati (e viceversa) e per comodità analizzeremo dunque solo la bontà d'adattamento del piano. Come al solito cominciamo costruendo il grafico dei residui. A tal fine dobbiamo calcolare i valori riprodotti

$$\hat{y}_{1i} = 0,671 + 0,483y_{2i} + 0,409y_{3i}$$

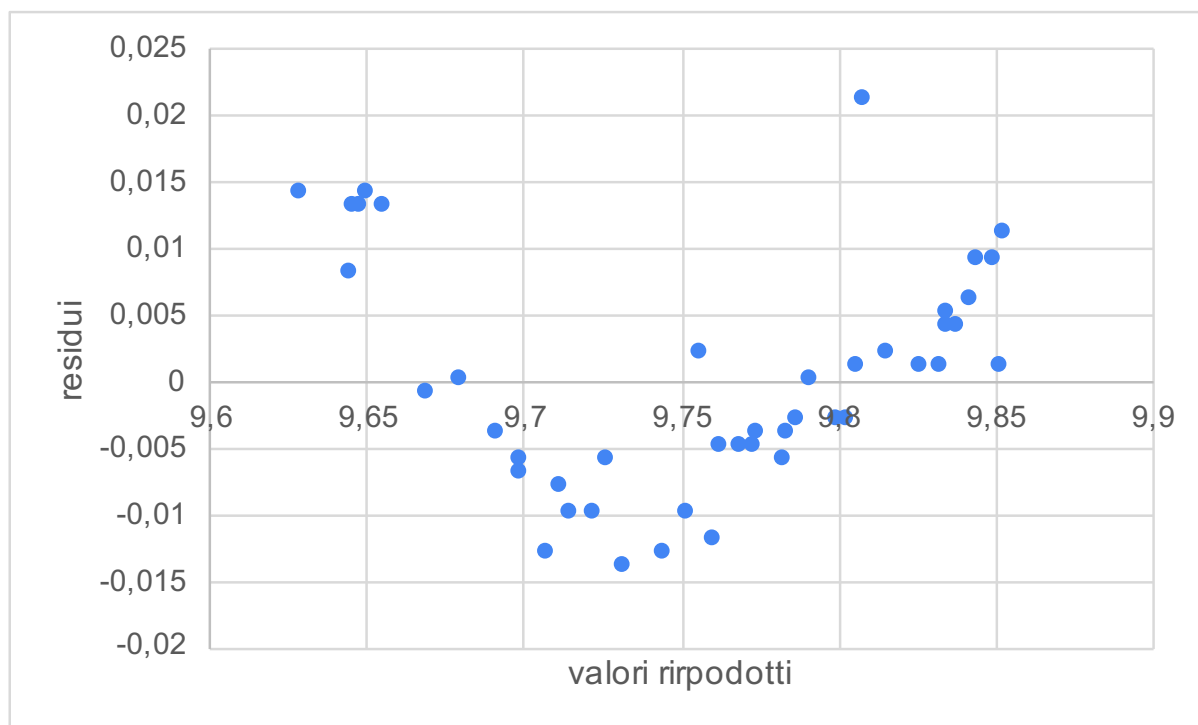
e i corrispondenti residui  $z_{Y_i} = y_{1i} - \hat{y}_{1i}$  (vedi Figura 6.3).

Figura 6.3: foglio "Regressione CD generalizzata" della cartella di lavoro "Regressione di Cobb e Douglas"

	E	F	G	H	I
5	Y1=ln(X1)	Y2=ln(X2)	Y3=ln(X3)	Y1cap	Zy
6	9,643	11,839	7,922	9,629	0,014
7	9,652	11,844	7,954	9,645	0,008
8	9,660	11,845	7,957	9,646	0,013
9	9,665	11,842	7,970	9,651	0,014
	E	F	G	H	I
48	9,831	11,893	8,298	9,809	0,021
49	9,841	11,917	8,334	9,836	0,005
50					
51	<b>429,331</b>	<b>523,735</b>	<b>359,026</b>	<b>429,331</b>	<b>0,000</b>

A questo punto possiamo riportare i punti  $(\hat{y}_{1i}, z_{Yi})$  all'interno di un piano cartesiano onde ottenere il grafico dei residui in Figura 6.4:

Figura 6.4:



Anche in questo grafico (si confronti con il grafico dei residui in Figura 4.8) notiamo la presenza di sistematicità nella successione dei segni dei residui: mentre i residui sui due bordi sono tutti positivi, quelli al centro del grafico sono tutti negativi. Anche in questo caso il problema potrebbe essere risolto usando il metodo descritto nella Sezione 2.4, ma per non appesantire troppo la trattazione sorvoleremo su questo problema.

Passiamo dunque all'identificazione di eventuali residui "sospetti". Per calcolare l'indice  $A_2$  associato al piano ai minimi quadrati dobbiamo innanzitutto calcolare la varianza residua  $var(Z_Y)$ . A tal fine sfruttiamo la scomposizione della varianza secondo la quale

$$var(Y_1) = var(\hat{Y}_1) + var(Z_Y) \quad \Rightarrow \quad var(Z_Y) = var(Y_1) - var(\hat{Y}_1).$$

Usando la formula indiretta (60) (ovvero la formula indiretta per il calcolo della varianza spiegata di un piano ai minimi quadrati), otteniamo

$$\begin{aligned} \text{var}(\hat{Y}_1) &= \tilde{\alpha}_2\sigma_{Y_{12}} + \tilde{\alpha}_3\sigma_{Y_{13}} \\ &= 0,483 \times 0,00269 + 0,409 \times 0,00773 = 0,00446 \end{aligned}$$

$$\Rightarrow \text{var}(Z_Y) = \text{var}(Y_1) - \text{var}(\hat{Y}_1) = 0,00454 - 0,00446 = 0,00008$$

(il valore di  $\text{var}(Y_1)$  è contenuto nella matrice di varianza e covarianza in Figura 6.2). Il valore dell'indice  $A_2$  è dunque dato da

$$A_2 = \sqrt{\text{var}(\hat{Y}_1)} = \sqrt{0,00008} = 0,00894.$$

Moltiplicando per  $\pm 3$  questo valore otteniamo le ordinate delle due linee orizzontali che delimitano l'area nella quale si trovano i residui "non sospetti". In questo caso otteniamo  $\pm 3 \times A_2 = \pm 3 \times 0,00894 = 0,02682$  e siccome queste ordinate si trovano al di fuori dell'area rappresentata in Figura 6.4, possiamo concludere che non ci sono residui sospetti.

Per concludere la fase di valutazione della bontà d'adattamento possiamo a questo punto calcolare anche l'indice di determinazione che è dato da

$$I^2 = \frac{\text{var}(\hat{Y}_1)}{\text{var}(Y_1)} = \frac{0,00446}{0,00454} = 0,982$$

e il cui valore conferma che il piano ai minimi quadrati riesce a cogliere una quota molto elevata della variabilità dei valori della variabile  $Y_1 = \ln X_1$ .

A questo punto possiamo passare alla fase interpretativa. Siccome siamo interessati solo alle variabili  $X_1$ ,  $X_2$  e  $X_3$  commenteremo solo l'andamento della funzione interpolante di Cobb-Douglas la cui equazione abbiamo visto essere data da

$$\hat{x}_1 = 1,956 \times x_2^{0,483} \times x_3^{0,409}.$$

Per interpretare l'andamento di questa funzione interpolante ci conviene tenere presente che la sua equazione è equivalente all'equazione

$$\ln \hat{x}_1 = \ln 1,956 + 0,483 \ln x_2 + 0,409 \times \ln x_3.$$

Ricordando la relazione tra variazioni logaritmiche e variazioni relative (vedi Sezione 4.2), possiamo affermare che  $\tilde{\alpha}_2 = 0,483\%$  sia la variazione relativa che subisce il valore  $\hat{x}_1$  se il valore di  $x_2$  aumenta di un punto percentuale e il valore di  $x_3$  rimane costante. Con riferimento al contenuto delle variabili questo significa che il valore riprodotto per il PIL aumenta del  $\tilde{\alpha}_2 = 0,483\%$  se il numero di occupati aumenta di un punto percentuale e gli investimenti fissi lordi rimangono invariati. L'interpretazione del parametro  $\tilde{\alpha}_3 = 0,409$  è analoga. Per ultimo osserviamo che

$$\tilde{\alpha}_2 + \tilde{\alpha}_3 = 0,483 + 0,409 = 0,892 < 1$$

e questo fatto suggerisce che durante il decennio 2010 - 2020 l'economia statunitense si trovasse in una fase di rendimenti di scala decrescenti.

**Esercizio 6.1.** Si replichi questa analisi di regressione con i dati contenuti nel foglio "Italia" della cartella di lavoro "[Regressione di Cobb e Douglas](#)". △

## 7 L'iperpiano ai minimi quadrati

### 7.1 Definizione di iperpiano

Come abbiamo visto, rette e piani sono delle funzioni che permettono di descrivere in modo semplice e intuitivo il comportamento della variabile dipendente a fronte di variazioni della o delle variabili indipendenti. Per questo motivo, nelle analisi di regressione che coinvolgono una o due variabili esplicative la funzione interpolante viene spesso cercata all'interno di una di queste due famiglie di funzioni.

La controparte di una retta o di un piano nel caso di *più di due variabili indipendenti* sono i cosiddetti *iperpiani*, ovvero le funzioni definite attraverso un'equazione del tipo

$$x_1 = \alpha_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots + \alpha_k x_k \quad (72)$$

dove  $\alpha_1, \alpha_2, \dots, \alpha_k$  sono  $k$  parametri reali.

Com'è facile verificare, nei casi  $k = 2$  e  $k = 3$  l'equazione che definisce un iperpiano si riduce all'equazione che definisce una retta o un piano, rispettivamente. Come vedremo tra breve, l'analogia tra le equazioni che definiscono rette, piani e iperpiani si estende anche a molte formule per calcolare indici statistici legati rette, piani e iperpiani *ai minimi quadrati*.



## 7.2 Determinazione dei parametri di un iperpiano ai minimi quadrati

Supponiamo ora di disporre di  $n$   $k$ -uple

$$(x_{1i}, x_{2i}, \dots, x_{ki}), \quad i = 1, 2, \dots, n,$$

di osservazioni di  $k$  variabili  $X_1, X_2, \dots, X_k$ . Volendo applicare il [metodo dei minimi quadrati](#) per trovare un iperpiano che descriva la relazione di fondo tra le osservazioni delle  $k$  variabili, ci troviamo di fronte al problema di minimizzare la distanza quadratica

$$D(\alpha_1, \dots, \alpha_k) = \sum_{i=1}^N (x_{1i} - \alpha_1 - \alpha_2 x_{2i} - \dots - \alpha_k x_{ki})^2.$$

Come nel caso della distanza quadratica che si riferisce a rette e piani,  $D(\alpha_1, \dots, \alpha_k)$  è una funzione differenziabile e convessa e pertanto l'insieme dei suoi punti di minimo è dato dall'insieme di tutte le  $k$ -uple  $(\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_k)$  in corrispondenza delle quali risultano annullate tutte le  $k$  derivate parziali  $\partial D / \partial \alpha_j$ ,  $j = 1, 2, \dots, k$ . Seguendo ragionamenti simili a quelli che abbiamo visto nelle Sezioni [3.2](#) e [5.1](#) dove abbiamo risolto gli analoghi sistemi riferiti alla retta e al piano ai minimi quadrati, si può dimostrare che ogni iperpiano ai minimi quadrati deve passare per il punto

$$(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$$

e che l'intercetta di un iperpiano ai minimi quadrati deve dunque essere data da

$$\hat{\alpha}_1 = \bar{x}_1 - \hat{\alpha}_2 \bar{x}_2 - \dots - \hat{\alpha}_k \bar{x}_k. \quad (73)$$

Non è difficile verificare che con  $k = 2$  e  $k = 3$  questa formula si riduca alla formula per calcolare l'intercetta di una retta ai minimi quadrati e alla formula per calcolare l'intercetta di un piano ai minimi quadrati, rispettivamente.

Ogni iperpiano ai minimi quadrati passa per il punto  $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ .

Dopo aver trovato la formula (73) per calcolare l'intercetta, possiamo portare avanti il ragionamento esposto nella Sezione 5.1 (oppure quello esposto nella Sezione 3.2) onde dimostrare che il sistema che annulla le  $k$  derivate parziali  $\partial D / \partial \alpha_j$  ha un'unica soluzione se e solo se il determinante della matrice di varianze e covarianze delle variabile esplicative non è nullo,<sup>16</sup> ovvero se e solo se

$$\begin{vmatrix} \sigma_{22} & \cdots & \cdots & \sigma_{2k} \\ \vdots & \ddots & \cdots & \vdots \\ \sigma_{k2} & \cdots & \cdots & \sigma_{kk} \end{vmatrix} \neq 0. \quad (74)$$

In questo caso i  $k - 1$  coefficienti angolari dell'unico iperpiano ai minimi quadrati (che vengono anch'essi chiamati coefficienti di

---

<sup>16</sup>Anche nel caso  $k > 3$  si può dimostrare che il determinante della matrice di varianze e covarianze delle variabile esplicative non può mai essere negativo.

regressione "parziale") possono essere calcolati secondo la formula

$$\hat{\alpha}_j = \frac{\begin{vmatrix} \sigma_{22} & \cdots & \sigma_{2(j-1)} & \sigma_{12} & \sigma_{2(j+1)} & \cdots & \sigma_{2k} \\ \vdots & \ddots & \sigma_{3(j-1)} & \sigma_{13} & \sigma_{3(j+1)} & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \sigma_{k2} & \cdots & \sigma_{k(j-1)} & \sigma_{1k} & \sigma_{k(j+1)} & \cdots & \sigma_{kk} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \cdots & \cdots & \sigma_{2k} \\ \vdots & \ddots & \cdots & \vdots \\ \sigma_{k2} & \cdots & \cdots & \sigma_{kk} \end{vmatrix}}$$

per  $j = 2, 3, \dots, k$ . Non è difficile verificare che con  $k = 2$  questa formula si riduca alla formula per calcolare l'unico coefficiente angolare di una retta ai minimi quadrati, e che con  $k = 3$  questa formula dia luogo alle formule (42) e (43) per calcolare i due coefficienti angolari di un piano ai minimi quadrati.

Se invece il determinante nella (74) è nullo, allora ci saranno infiniti iperpiani ai minimi quadrati così come accade nel caso di piani ai minimi quadrati quando il determinante della matrice di varianza e covarianza delle due variabili esplicative è nullo (e così come accade nel caso di rette ai minimi quadrati quando la varianza dell'unica variabile esplicativa è nulla). Tuttavia, il caso in cui il determinante nella (74) è esattamente nullo non si verifica praticamente mai nelle applicazioni. Infatti, si può dimostrare che il determinante nella (74) è nullo se e solo se esistono  $k$  costanti reali  $b_1, b_2, b_3, \dots, b_k$ , delle quali almeno una diversa da zero, tali che

$$b_1 + b_2 x_{2i} + b_3 x_{3i} + \cdots + b_k x_{ki} = 0 \quad \text{per ogni } i = 1, 2, \dots, n.$$

In questo caso si dice che le osservazioni delle  $k - 1$  variabili esplicative sono *multicollineari* (ricordiamo che, secondo la definizione che abbiamo dato in precedenza,  $k$  è il numero complessivo di variabili coinvolte in un iperpiano ai minimi quadrati e che questo numero include anche la variabile dipendente  $X_1$ ). Si osservi che con  $k = 3$  (ovvero nel caso di due variabili esplicative) la definizione di multicollinearità si riduce alla definizione di collinearità che abbiamo introdotto nella Sezione 5.1 con riferimento a piani ai minimi quadrati, e che con  $k = 2$  (ovvero nel caso di una singola variabile esplicative) la definizione di multicollinearità è soddisfatta se e solo se la devianza (ovvero la varianza) delle osservazioni della variabile  $X_2$  è nulla.

Siccome nelle applicazioni casi di perfetta multicollinearità non si verificano praticamente mai, d'ora in poi, se non diversamente specificato, assumeremo sempre che le osservazioni delle variabili  $X_1, X_2, \dots, X_k$  non siano multicollineari (da notare che abbiamo aggiunto anche la variabile dipendente  $X_1$ ). In altre parole, assumeremo sempre che il determinante di qualsiasi matrice di varianza e covarianza che coinvolge tutte o solo alcune delle variabili  $X_1, X_2, \dots, X_k$  sia strettamente positivo.

Tuttavia, non di rado nelle applicazioni si verificano casi di *quasi-multicollinearità* tra le variabili esplicative, ovvero casi dove il determinante nella (74) è molto prossimo a zero. Così come nei casi di *quasi-collinearità* gli errori di rilevazione (o le approssimazioni) possono causare variazioni drastiche nei valori di uno o più parametri di un piano ai minimi quadrati, lo stesso può anche accadere con i parametri di un iperpiano ai minimi quadrati in caso di *quasi-multicollinearità*. Per assicurarsi che gli errori di rilevazione non abbiano avuto un impatto rilevante sui valori dei parametri di un iperpiano ai minimi quadrati bisogna quindi verificare, per ciascuna variabile esplicative, che l'entità di un possibile errore di rilevazione sia trascurabile rispetto allo scarto quadratico medio

dei residui dell'iperpiano che si ottiene considerando la variabile esplicativa in questione come variabile dipendente, e considerando le altre variabili esplicative come variabili indipendenti. L'entità di un possibile errore di rilevazione nelle osservazioni della variabile esplicativa  $X_j$  deve quindi essere confrontata con  $\sqrt{\text{var}(Z_{j.23\dots(j-1)(j+1)\dots k})}$ , dove  $Z_{j.23\dots(j-1)(j+1)\dots k}$  è la variabile che rappresenta i residui dell'iperpiano che descrive la variabile esplicativa  $X_j$  in funzione di tutte le altre variabili esplicative. Come vedremo tra breve, il valore di  $\sqrt{\text{var}(Z_{j.23\dots(j-1)(j+1)\dots k})}$  non può aumentare all'aumentare del numero di variabili esplicative (che è uguale a  $k - 1$ ) e di solito diminuisce all'aumentare del numero di variabili esplicative. Un dato errore di rilevazione su una osservazione della variabile esplicativa  $X_j$  potrebbe quindi avere effetti trascurabili sui parametri di un dato iperpiano ai minimi quadrati con  $k - 1$  variabili esplicative, ma potrebbe avere effetti drastici sui valori dei parametri di un iperpiano ai minimi quadrati che oltre alle  $k - 1$  variabili esplicative dell'iperpiano precedente ne considera una aggiuntiva. In questo senso possiamo dunque dire che all'aumentare del numero di variabili esplicative il problema della *quasi-multicollinearità* diventa sempre più stringente.

D'ora in poi, onde avere una notazione univoca e coerente con quella introdotta nella Sezione 5.2, indicheremo l'intercetta di un iperpiano ai minimi quadrati con  $\hat{c}_{1.23\dots k}$ , e indicheremo il coefficiente angolare associato alla variabile esplicativa  $X_j$  con  $\hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k}$ . Con questa notazione l'equazione di un iperpiano ai minimi quadrati diventa

$$\begin{aligned} \hat{x}_{1.23\dots k} = & \hat{c}_{1.23\dots k} + \hat{\alpha}_{12.3\dots k}x_2 + \hat{\alpha}_{13.24\dots k}x_3 + \dots \\ & + \hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k}x_j + \dots + \hat{\alpha}_{1k.2\dots(k-1)}x_k. \end{aligned}$$

Siccome i coefficienti angolari  $\hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k}$  sono delle derivate *parziali*, anche essi vengono chiamati coefficienti di regressione "parziale".

Se il determinante della matrice di varianza e covarianza delle variabili esplicative è diverso da zero (e quindi positivo), allora esiste un unico iperpiano ai minimi quadrati i cui coefficienti angolari sono dati da

$$\hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k} = \frac{\begin{vmatrix} \sigma_{22} & \cdots & \sigma_{2(j-1)} & \sigma_{12} & \sigma_{2(j+1)} & \cdots & \sigma_{2k} \\ \vdots & \ddots & \sigma_{3(j-1)} & \sigma_{13} & \sigma_{3(j+1)} & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \sigma_{k2} & \cdots & \sigma_{k(j-1)} & \sigma_{1k} & \sigma_{k(j+1)} & \cdots & \sigma_{kk} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \cdots & \cdots & \sigma_{2k} \\ \vdots & \ddots & \cdots & \vdots \\ \sigma_{k2} & \cdots & \cdots & \sigma_{kk} \end{vmatrix}}, \quad (75)$$

$j = 2, 3, \dots, k$ , e la cui intercetta è data da

$$\hat{c}_{1.23\dots k} = \bar{x}_1 - \hat{\alpha}_{12.3\dots k} \bar{x}_2 - \cdots - \hat{\alpha}_{1k.23\dots(k-1)} \bar{x}_k. \quad (76)$$

Se invece il determinante della matrice di varianza e covarianza delle variabili esplicative è nullo, allora esistono infiniti iperpiani ai minimi quadrati. Questo caso si verifica se e solo se le osservazioni delle variabili esplicative sono multicollineari.

### 7.3 Formule alternative per calcolare i coefficienti angolari di un iperpiano ai minimi quadrati

In questa sezione presenteremo tre formule alternative per calcolare i coefficienti angolari di un iperpiano ai minimi quadrati che possono essere applicate quando quest'ultimo è unico. Le formule presentate in questa sezione sono le controparti delle formule (48), (50) e (51) che si riferiscono ai piani ai minimi quadrati.

#### Prima formula alternativa:

$$\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k} = \frac{\sigma_1}{\sigma_j} \times \frac{\begin{vmatrix} 1 & \cdots & r_{2(j-1)} & r_{12} & r_{2(j+1)} & \cdots & r_{2k} \\ \vdots & \ddots & r_{3(j-1)} & r_{13} & r_{3(j+1)} & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{k2} & \cdots & r_{k(j-1)} & r_{1k} & r_{k(j+1)} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & \cdots & r_{2k} \\ \vdots & \ddots & \cdots & \vdots \\ r_{k2} & \cdots & \cdots & 1 \end{vmatrix}}, \quad (77)$$

Questa formula ci permette di usare la matrice dei coefficienti di correlazione lineare (o "matrice di correlazione") al posto della matrice di varianza e covarianza (si ricordi che  $r_{jj} = 1$  per ogni  $j = 1, 2, \dots, k$ ). Tuttavia, così facendo bisogna moltiplicare il rapporto tra i determinanti per il rapporto tra lo scarto quadratico medio della variabile dipendente e lo scarto quadratico medio della variabile esplicativa associata al coefficiente angolare. Non è

difficile verificare che con  $j = 2$  e  $k = 3$  questa formula si riduce alla formula alternativa (48) che si riferisce al piano ai minimi quadrati.

La seconda formula alternativa si ottiene sostituendo il coefficiente di correlazione lineare  $r_{1j}$  che appare nella matrice al numeratore della formula (77) con  $\hat{\alpha}_{1j} \times \sigma_j / \sigma_1$ :

### Seconda formula alternativa:

$$\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k} =$$

$$= \frac{\sigma_1}{\sigma_j} \times \frac{\begin{vmatrix} 1 & \cdots & r_{2(j-1)} & r_{12} & r_{2(j+1)} & \cdots & r_{2k} \\ \vdots & \ddots & r_{3(j-1)} & r_{13} & r_{3(j+1)} & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{j2} & \cdots & \cdots & \hat{\alpha}_{1j} \frac{\sigma_j}{\sigma_1} & \cdots & \cdot & r_{jk} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{k2} & \cdots & r_{k(j-1)} & r_{1k} & r_{k(j+1)} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & \cdots & r_{2k} \\ \vdots & \ddots & \cdots & \vdots \\ r_{k2} & \cdots & \cdots & 1 \end{vmatrix}}. \quad (78)$$

Anche in questo caso non è difficile verificare che con  $j = 2$  e  $k = 3$  questa formula si riduce alla formula alternativa (50) che si riferisce al piano ai minimi quadrati.

Usando la formula alternativa (78) non è difficile dimostrare che

$$\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k} = \hat{\alpha}_{1j}$$

se la variabile esplicativa  $X_j$  è incorrelata con tutte le altre



variabili esplicative, ovvero se  $r_{ij} = r(X_i, X_j) = 0$  per ogni  $i = 2, 3, \dots, j-1, j+1, \dots, k$ .

Se la variabile esplicativa  $X_j$  è incorrelata con tutte le altre variabili esplicative (ovvero se  $r_{ij} = r(X_i, X_j) = 0$  per ogni  $i = 2, 3, \dots, j-1, j+1, \dots, k$ ), allora

$$\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k} = \hat{\alpha}_{1j}.$$

Per integrare l'osservazione nel precedente riquadro, aggiungiamo che così come il valore di  $\hat{\alpha}_{12}$  può essere molto diverso dal valore di  $\hat{\alpha}_{12.3}$  (si ricordi il fenomeno della correlazione spuria di cui abbiamo parlato nella Sezione 5.2), anche il valore del coefficiente angolare  $\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k}$  può essere molto diverso dal valore del coefficiente angolare  $\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k(k+1)}$  che si riferisce ad un iperpiano ai minimi quadrati che considera una ulteriore variabile esplicativa  $X_{k+1}$ . Questa considerazione suggerisce di usare sempre e comunque molta cautela quando a partire dal valore di un coefficiente di regressione parziale cerchiamo di inferire qualcosa sul processo che ha generato i dati: la variabilità nei valori di una ulteriore variabile  $X_{k+1}$  che non abbiamo incluso nell'analisi (per esempio perché non disponiamo delle sue osservazioni) potrebbe sempre essere la "causa" di valori sorprendenti dei coefficienti di regressione parziale  $\hat{\alpha}_{1j.23\dots(j-1)(j+1)\dots k}$ .

La terza formula alternativa per calcolare i coefficienti angolari di un iperpiano ai minimi quadrati è data da

**Terza formula alternativa:**

$$\begin{aligned} \hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k} &= \\ &= \frac{\text{codev}(Z_{1.2\dots(j-1)(j+1)\dots k}, Z_{j.2\dots(j-1)(j+1)\dots k})}{\text{dev}(Z_{j.2\dots(j-1)(j+1)\dots k})}, \end{aligned} \quad (79)$$

dove  $Z_{1.2\dots(j-1)(j+1)\dots k}$  è la variabile che rappresenta i residui dell'iperpiano ai minimi quadrati che descrive  $X_1$  in funzione di tutte le variabili esplicative tranne  $X_j$ , e dove  $Z_{j.2\dots(j-1)(j+1)\dots k}$  è la variabile che rappresenta i residui dell'iperpiano ai minimi quadrati che descrive la variabile esplicativa  $X_j$  in funzione di tutte le altre variabili esplicative.

Questa formula ci dice che il coefficiente angolare  $\hat{\alpha}_{1j.2\dots(j-1)(j+1)\dots k}$  coincide con il coefficiente angolare della retta ai minimi quadrati che descrive  $Z_{1.2\dots(j-1)(j+1)\dots k}$  in funzione di  $Z_{j.2\dots(j-1)(j+1)\dots k}$ . Anche in questo caso non è difficile verificare che con  $j = 2$  e  $k = 3$  la formula alternativa (79) si riduce alla formula (51).

## 7.4 Proprietà dei residui di iperpiani ai minimi quadrati

I residui di un iperpiano ai minimi quadrati godono di proprietà analoghe a quelle dei residui di un piano ai minimi quadrati:

**Proprietà dei residui di iperpiani ai minimi quadrati:**

- 1)  $M_1(Z_{1.23\dots k}) = 0 \Rightarrow M_1(\hat{X}_{1.23\dots k}) = 0$ , dove  $\hat{X}_{1.23\dots k}$  è la variabile che rappresenta i valori riprodotti dall'iperpiano;
- 2)  $codev(Z_{1.23\dots k}, X_j) = 0$  per  $j = 2, 3, \dots, k$ ;
- 3)  $codev(Z_{1.23\dots k}, \hat{X}_{1.23\dots k}) = 0$ .

Queste proprietà possono essere dimostrate seguendo la falsariga delle dimostrazioni delle proprietà dei residui di rette e piani ai minimi quadrati.

## 7.5 Scomposizione della devianza e indice di determinazione

Anche la scomposizione della devianza

$$Dev(X_1) = Dev(\hat{X}_{1.23\dots k}) + Dev(Z_{1.23\dots k}),$$

rimane valida per gli iperpiani ai minimi quadrati e l'indice di determinazione

$$I_{1.23\dots k}^2 = \frac{Dev(\hat{X}_{1.23\dots k})}{Dev(X_1)}$$

ha significato analogo a quello di  $I_{1.23}^2$ .

**Scomposizione della devianza:**

$$Dev(X_1) = Dev(\hat{X}_{1.23\dots k}) + Dev(Z_{1.23\dots k}),$$

$$\Rightarrow I_{1.23\dots k}^2 = \frac{Dev(\hat{X}_{1.23\dots k})}{Dev(X_1)} \in [0, 1].$$

L'interpretazione dell'indice di determinazione  $I_{1.23\dots k}^2$  è analoga a quella dell'indice di determinazione di una retta o di un piano ai minimi quadrati.

## 7.6 Miglioramento della bontà d'adattamento nel caso di aggiunta di variabili esplicative

Così come la devianza residua non può aumentare quando, mediante l'aggiunta di una variabile esplicativa, si passa da una retta ad un piano ai minimi quadrati (vedi Sezione 5.6), la devianza residua non può nemmeno aumentare passando da un iperpiano ai minimi quadrati ad un'altro aggiungendo una nuova variabile esplicativa:

$$I_{1.23\dots(k-1)}^2 \leq I_{1.23\dots k}^2.$$

Il miglioramento della bontà d'adattamento può essere misurato attraverso l'indice

$$\begin{aligned} MVS_{1.23\dots(k-1) \rightarrow 1.23\dots k} &= \frac{dev(\hat{X}_{1.23\dots k}) - dev(\hat{X}_{1.23\dots(k-1)})}{dev(X_1)} \\ &= I_{1.23\dots k}^2 - I_{1.23\dots(k-1)}^2, \end{aligned}$$

che ci dice di quanto aumenta la varianza spiegata in rapporto a quella totale, e/o attraverso l'indice

$$\begin{aligned} MVR_{1.23\dots(k-1)\rightarrow 1.23\dots k} &= \frac{dev(Z_{1.23\dots(k-1)}) - dev(Z_{1.23\dots k})}{dev(Z_{1.23\dots(k-1)})} \\ &= \frac{I_{1.23\dots k}^2 - I_{1.23\dots(k-1)}^2}{1 - I_{1.23\dots(k-1)}^2} \end{aligned}$$

che ci dice di quanto si riduce la devianza residua in rapporto alla devianza residua dell'iperpiano di partenza.

## 7.7 Formula indiretta per il calcolo della devianza spiegata

Anche nel caso dell'iperpiano ai minimi quadrati si può dimostrare che

$$codev(X_1, \hat{X}_{1.23\dots k}) = dev(\hat{X}_{1.23\dots k}) \quad (80)$$

e usando questa uguaglianza si ottiene la seguente formula indiretta per il calcolo della devianza spiegata (basta usare le proprietà della codevianza):

$$\begin{aligned} dev(\hat{X}_{1.23\dots k}) &= \hat{\alpha}_{12.3\dots k} codev(X_1, X_2) + \dots \\ &\quad \dots + \hat{\alpha}_{1k.2\dots(k-1)} codev(X_1, X_k). \end{aligned}$$

Con  $k = 2$  e  $k = 3$  questa formula si riduce alle formule indirette che abbiamo visto con riferimento alla retta e al piano ai minimi quadrati.

**Formula indiretta per il calcolo della devianza spiegata:**

$$\begin{aligned} dev(\hat{X}_{1.23\dots k}) = \hat{\alpha}_{12.3\dots k} codev(X_1, X_2) + \dots \\ \dots + \hat{\alpha}_{1k.2\dots(k-1)} codev(X_1, X_k). \end{aligned} \quad (81)$$

**Osservazione 7.1.** Anche nel caso di un iperpiano ai minimi quadrati la varianza spiegata può essere espressa attraverso l'uso di determinanti esattamente come nel caso di un piano ai minimi quadrati (vedi l'equazione (63)). Infatti, usando le proprietà dei determinanti non è difficile verificare (omettiamo i dettagli) che il secondo membro della (81) può essere espresso come

$$var(\hat{X}_{1.23\dots k}) = \sigma_{11} - \frac{\begin{vmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2k} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \cdots & \sigma_{3k} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \sigma_{3k} & \cdots & \sigma_{kk} \end{vmatrix}}{\begin{vmatrix} \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2k} \\ \sigma_{23} & \sigma_{33} & \cdots & \sigma_{3k} \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{2k} & \sigma_{3k} & \cdots & \sigma_{kk} \end{vmatrix}} \quad (82)$$

△

## 7.8 Formula per calcolare l'indice di determinazione a partire dai coefficienti di correlazione lineare

Partendo dalla formula (82) si può dimostrare che

$$I_{1.23\dots k}^2 = 1 - \frac{\begin{vmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{3k} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ r_{1k} & r_{2k} & \cdots & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} & \cdots & r_{2k} \\ r_{23} & 1 & \cdots & r_{3k} \\ \cdots & \cdots & \ddots & \vdots \\ r_{2k} & \cdots & \cdots & 1 \end{vmatrix}}.$$

Come è facile constatare, con  $k = 3$  questa formula si riduce alla formula (64) che si riferisce al piano ai minimi quadrati. Usando questa formula si può dimostrare che anche l'indice di determinazione di un iperpiano ai minimi quadrati è invariante a trasformazioni lineari affini delle variabili oggetto di analisi.<sup>17</sup> Si ricordi che lo stesso vale anche per gli indici di determinazione di rette e piani ai minimi quadrati.

---

<sup>17</sup>Usando direttamente la formula (82) si potrebbe dimostrare una proprietà di invarianza più forte.

L'indice di determinazione di un iperpiano ai minimi quadrati è il complemento a 1 del rapporto tra il determinante della matrice di correlazione di tutte le variabili coinvolte (anche la variabile dipendente) e il determinante della matrice di correlazione delle sole variabili esplicative.

L'indice di determinazione di un iperpiano ai minimi quadrati è invariante a trasformazioni lineari affini delle variabili oggetto di analisi.

## 7.9 Coefficiente di correlazione multiplo

Dalla (80) discende che il coefficiente di correlazione multiplo

$$R_{1.23\dots k} = \frac{\text{codev}(X_1, \hat{X}_{1.23\dots k})}{\sqrt{\text{dev}(X_1) \times \text{dev}(\hat{X}_{1.23\dots k})}}$$

non può mai essere negativo e che il suo quadrato è uguale all'indice di determinazione  $I_{1.23\dots k}^2$ :

$$R_{1.23\dots k}^2 = I_{1.23\dots k}^2.$$



**Coefficiente di correlazione multiplo:** Il coefficiente di correlazione multiplo di un iperpiano ai minimi quadrati è definito come

$$R_{1.23\dots k} = \frac{\text{codev}(X_1, \hat{X}_{1.23\dots k})}{\sqrt{\text{dev}(X_1) \times \text{dev}(\hat{X}_{1.23\dots k})}}$$

e non può mai essere negativo.

$$R_{1.23\dots k}^2 = I_{1.23\dots k}^2.$$

## 7.10 Coefficienti di correlazione parziale

Si possono definire anche dei **coefficienti di correlazione lineare parziale di ordine superiore al primo** che misurano la correlazione tra due variabili **tenendo conto delle variazioni di più di una variabile**. D'ora in poi indicheremo dunque con

$$r_{12.34\dots k}$$

il coefficiente di correlazione lineare parziale tra le variabili  $X_1$  e  $X_2$  che tiene conto delle variazioni delle variabili  $X_3, X_4, \dots, X_k$ .

Il coefficiente di correlazione parziale  $r_{12.34\dots k}$  è definito come coefficiente di correlazione lineare tra i residui dell'iperpiano ai minimi quadrati

$$\begin{aligned} \hat{x}_{1.34\dots k} &= \hat{e} + \hat{\alpha}_{13.45\dots k}x_3 + \dots \\ &\quad \dots + \hat{\alpha}_{14.35\dots k}x_4 + \dots \\ &\quad \dots + \hat{\alpha}_{1k.34\dots(k-1)}x_k \end{aligned}$$

che descrive  $X_1$  in funzione di  $X_3, X_4, \dots, X_k$  (si osservi che tra le variabili esplicative manca  $X_2$ ), ed i residui dell'iperpiano ai minimi quadrati

$$\begin{aligned}\hat{x}_{2.34\dots k} &= \hat{g} + \hat{\alpha}_{23.45\dots k}x_3 + \dots \\ &\quad \dots + \hat{\alpha}_{24.35\dots k}x_4 + \dots \\ &\quad \dots + \hat{\alpha}_{2k.34\dots(k-1)}x_k\end{aligned}$$

che invece descrive  $X_2$  in funzione di  $X_3, X_4, \dots, X_k$  (si osservi che tra le variabili esplicative manca  $X_1$ ).

Per calcolare dei coefficienti di correlazione parziale conviene a volte fare riferimento alla seguente formula ricorsiva:

Si può dimostrare che i coefficienti di correlazione parziale di ordine<sup>a</sup>  $k - 2$  possono essere ricavati da quelli di ordine  $k - 3$  attraverso la formula

$$r_{12.34\dots(k-1)k} = \frac{r_{12.34\dots(k-1)} - r_{1k.34\dots(k-1)}r_{2k.34\dots(k-1)}}{\sqrt{(1 - r_{1k.34\dots(k-1)}^2)(1 - r_{2k.34\dots(k-1)}^2)}}.$$

<sup>a</sup>Per ordine di un coefficiente di correlazione parziale intendiamo il numero di variabili dopo il punto che compare nel pedice.

Nel caso  $k = 3$  la precedente formula diventa

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}},$$

ovvero la formula (69) per calcolare i coefficienti di correlazione parziale del primo ordine partendo da quelli grezzi (ovvero di ordine  $k - 3 = 0$ ).

Ponendo invece  $k = 4$ , si ottiene

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1 - r_{14.3}^2)(1 - r_{24.3}^2)}}.$$

Tenendo presente la definizione di  $r_{12.34\dots k}$  come coefficiente di correlazione lineare grezzo tra due serie di residui, ci si rende conto che il valore di un coefficiente di correlazione parziale non cambia permutando l'ordine degli indici che identificano le variabili prima e/o dopo il punto (senza scambiare la posizione di un indice prima del punto con quella di un indice dopo il punto). Quindi si ha

$$r_{14.23} = r_{41.23} = r_{14.32} = r_{41.32},$$

ma in generale

$$r_{14.23} \neq r_{12.43}.$$

Così come

$$\text{segno}(\hat{\alpha}_{12}) = \text{segno}(\hat{\alpha}_{21})$$

e così come

$$\text{segno}(\hat{\alpha}_{12.3}) = \text{segno}(\hat{\alpha}_{21.3}),$$

si può dimostrare che in generale

$$\text{segno}(\hat{\alpha}_{12.34\dots k}) = \text{segno}(\hat{\alpha}_{21.34\dots k})$$

per ogni  $k = 3, 4, \dots$ ; e così come

$$r_{12} = \text{segno}(\hat{\alpha}_{12}) \times \sqrt{\hat{\alpha}_{12} \times \hat{\alpha}_{21}}$$

e

$$r_{12.3} = \text{segno}(\hat{\alpha}_{12.3}) \times \sqrt{\hat{\alpha}_{12.3} \times \hat{\alpha}_{21.3}},$$

si può dimostrare che in generale

$$r_{12.34\dots k} = \text{segno}(\hat{\alpha}_{12.34\dots k}) \times \sqrt{\hat{\alpha}_{12.34\dots k} \times \hat{\alpha}_{21.34\dots k}}.$$

In altre parole, tutti i coefficienti di correlazione (grezzi e parziali) possono essere espressi come media geometrica di opportuni coefficienti di regressione (grezzi o parziali).

*Definiamo* ora l'indice di determinazione di una funzione interpolante con zero variabili esplicative (ovvero di una funzione interpolante costante) ponendo  $I_{1.\bullet}^2 = 0$ . In questo modo possiamo scrivere

$$MVR_{1.\bullet \rightarrow 1.2} = \frac{I_{1.2}^2 - I_{1.\bullet}^2}{1 - I_{1.\bullet}^2} = I_{1.2}^2 = r_{12}^2.$$

Ricordiamo ora che secondo la formula (70)

$$MVR_{1.3 \rightarrow 1.23} = \frac{I_{1.23}^2 - I_{1.3}^2}{1 - I_{1.3}^2} = r_{12.3}^2.$$

Domanda: Possiamo estendere questa formula anche al caso di più di due variabili esplicative? Risposta: Sì!!! Infatti, si può dimostrare che

$$MVR_{1.3\dots k \rightarrow 1.234\dots k} = \frac{I_{1.234\dots k}^2 - I_{1.3\dots k}^2}{1 - I_{1.34\dots k}^2} = r_{12.34\dots k}^2.$$

## 7.11 Esempio

**Esempio 7.1.** La seguente tabella riporta alcuni dati sui mercati di sbocco di una multinazionale. In particolare, per ciascun mercato la tabella riporta il fatturato annuale  $X_1$  (in milioni di

euro), l'ammontare degli investimenti in attività promozionali  $X_2$  (in milioni di euro), il numero di concorrenti  $X_3$  e la popolazione di potenziali acquirenti  $X_4$  (in milioni di individui):

mercato	$x_{1i}$	$x_{2i}$	$x_{3i}$	$x_{4i}$
A	98	12	7	20
B	123	23	7	26
C	235	12	6	34
D	360	12	6	55
E	465	42	6	61
F	567	34	5	64
G	599	35	5	58
H	698	35	5	76
I	780	70	2	83
J	2043	68	1	126
TOT	5968	343	50	603

- Si ricavino i parametri dell'iperpiano ai minimi quadrati che descrive  $X_1$  in funzione delle altre tre variabili.
- Si calcoli l'indice di determinazione dell'iperpiano.
- Si valuti il miglioramento della bontà d'adattamento nel passaggio dal piano che spiega  $X_1$  in funzione di  $X_2$  e  $X_3$ , all'iperpiano ricavato al punto a).
- Si ricavi la matrice dei coefficienti di correlazione grezzi.
- Si ricavino le 4 matrici con i coefficienti di correlazione parziale di ordine 1.
- Si ricavino le 6 matrici con i coefficienti di correlazione parziale di ordine 2.

- g) Si calcoli il valore di  $r_{14.23}^2$ ?
- h) Si verifichi la formula che esprime  $r_{14.23}$  come media geometrica di due opportuni coefficienti di regressione parziale.
- i) Si calcoli  $r_{14.23}$  partendo dai residui di due opportuni piani ai minimi quadrati.

Soluzione:

- a) Per ricavare i parametri dell'iperpiano

$$\hat{x}_{1.234} = \hat{c}_{1.234} + \hat{\alpha}_{12.34}x_2 + \hat{\alpha}_{13.24}x_3 + \hat{\alpha}_{14.23}x_4,$$

occorre innanzitutto ricavare la matrice delle varianze e covarianze. Usando le consuete formule si ottiene:

$$\begin{bmatrix} 281382,36 \\ 8416,96 & 409,01 \\ -901,00 & -34,60 & 3,60 \\ 14767,86 & 502,91 & -51,10 & 861,81 \end{bmatrix}$$

I coefficienti di regressione parziale sono rapporti tra determinanti di opportune sottomatrici. Cominciamo da  $\hat{\alpha}_{12.23}$ :

$$\hat{\alpha}_{12.23} = \frac{\begin{vmatrix} 8416,96 & -34,60 & 502,91 \\ -901,00 & 3,60 & -51,10 \\ 14767,86 & -51,10 & 861,81 \end{vmatrix}}{\begin{vmatrix} 409,01 & -34,60 & 502,91 \\ -34,60 & 3,60 & -51,10 \\ 502,91 & -51,10 & 861,81 \end{vmatrix}}$$

Per il calcolo del determinante di una matrice  $3 \times 3$  si può ricorrere alla cosiddetta **regola di Sarrus**. Si consideri per esempio il determinante della matrice al denominatore: ripetendo alla destra della matrice le sue prime due colonne si ottiene una nuova matrice  $3 \times 5$  data da

$$\begin{bmatrix} 409,01 & -34,60 & 502,91 & 409,01 & -34,60 \\ -34,60 & 3,60 & -51,10 & -34,60 & 3,60 \\ 502,91 & -51,10 & 861,81 & 502,91 & -51,10 \end{bmatrix}$$

Questa matrice ha tre **diagonali discendenti** che partono in alto a sinistra e scendono in basso verso destra:

$$\begin{bmatrix} 409,01 & -34,60 & 502,91 & 409,01 & -34,60 \\ -34,60 & 3,60 & -51,10 & -34,60 & 3,60 \\ 502,91 & -51,10 & 861,81 & 502,91 & -51,10 \end{bmatrix}$$

Per ottenere il determinante della matrice  $3 \times 3$  di partenza bisogna innanzitutto calcolare la somma dei prodotti degli elementi sulle tre diagonali discendenti:

$$\begin{aligned} & 409,01 \times 3,60 \times 861,81 + \\ & + (-34,60) \times (-51,10) \times 502,91 + \\ & + 502,91 \times -34,60 \times -51,10 = \\ & = 3047310,178 \end{aligned}$$

A questo punto bisogna applicare la stessa operazione alle tre **diagonali ascendenti** della matrice  $3 \times 5$ :

$$\begin{bmatrix} 409,01 & -34,60 & 502,91 & 409,01 & -34,60 \\ -34,60 & 3,60 & -51,10 & -34,60 & 3,60 \\ 502,91 & -51,10 & 861,81 & 502,91 & -51,10 \end{bmatrix}$$

Si ottiene:

$$\begin{aligned} & 502,91 \times 3,60 \times 502,91 + \\ & + (-51,10) \times (-51,10) \times 409,01 + \\ & + 861,81 \times (-34,60) \times (-34,60) = \\ & = 3010241,947 \end{aligned}$$

Sottraendo la somma ricavata dalle diagonali ascendenti dalla somma ricavata dalle diagonali discendenti si ottiene finalmente il determinante della matrice  $3 \times 3$ :

$$\begin{aligned} & \begin{vmatrix} 409,01 & -34,60 & 502,91 \\ -34,60 & 3,60 & -51,10 \\ 502,91 & -51,10 & 861,81 \end{vmatrix} = \\ & = 3047310,178 - 3010241,947 = \\ & = 37068,231. \end{aligned}$$

Ripetendo questa procedura per calcolare anche il determinante al numeratore di  $\hat{\alpha}_{12.34}$  si ottiene finalmente

$$\hat{\alpha}_{12.34} = \frac{-203143,025}{37068,231} = -5,480.$$



Usando una procedura analoga a quella appena vista per calcolare  $\hat{\alpha}_{12.34}$ , si calcolano anche gli altri coefficienti di regressione parziale dell'iperpiano:

$$\hat{\alpha}_{13.24} = \frac{\begin{vmatrix} 409,01 & 8416,96 & 502,91 \\ -34,60 & -901,00 & -51,10 \\ 502,91 & 14767,86 & 861,81 \end{vmatrix}}{\begin{vmatrix} 409,01 & -34,60 & 502,91 \\ -34,60 & 3,60 & -51,10 \\ 502,91 & -51,10 & 861,81 \end{vmatrix}} = \frac{-3352273,674}{37068,231} = -90,435.$$

e

$$\hat{\alpha}_{14.23} = \frac{\begin{vmatrix} 409,01 & -34,60 & 8416,96 \\ -34,60 & 3,60 & -901,00 \\ 502,91 & -51,10 & 14767,86 \end{vmatrix}}{\begin{vmatrix} 409,01 & -34,60 & 502,91 \\ -34,60 & 3,60 & -51,10 \\ 502,91 & -51,10 & 861,81 \end{vmatrix}} = \frac{554971,4290}{37068,231} = 14,972.$$

Infine, per calcolare l'intercetta servono le medie delle quattro variabili. Dalla matrice dei dati si ottiene

$$\bar{x}_1 = 596,8, \quad \bar{x}_2 = 34,3,$$

$$\bar{x}_3 = 5, \quad \bar{x}_4 = 60,3,$$

e l'intercetta è dunque data da

$$\begin{aligned}\hat{c}_{1.234} &= 596,8 - (-5,480) \times 34,3 + \\ &\quad - (-90,435) \times 5 - 14,972 \times 60,3 \\ &= 334,127\end{aligned}$$

In definitiva, l'equazione dell'iperpiano ai minimi quadrati è data da

$$\hat{x}_1 = 334,127 - 5,480x_2 - 90,435x_3 + 14,972x_4.$$

A questo punto possiamo commentare i valori dei parametri dell'iperpiano:

- $\hat{\alpha}_{12.34} = -5,480$  ci dice che, secondo l'iperpiano, il fatturato annuale diminuisce di 5,480 milioni di euro se gli investimenti in attività promozionali aumentano di un milione di euro e il numero di concorrenti  $X_3$  e la popolazione di potenziali acquirenti  $X_4$  rimangono costanti.

L'interpretazione degli altri due coefficienti di regressione parziale  $\hat{\alpha}_{13.24}$  e  $\hat{\alpha}_{14.23}$  è analoga a quella di  $\hat{\alpha}_{12.34} = -5,480$ .

- L'intercetta  $\hat{c}_{1.234} = 334,127$  è il valore riprodotto dall'iperpiano quando le tre variabili esplicative sono tutte nulle.

b) L'indice di determinazione dell'iperpiano è dato da

$$\begin{aligned}
 I_{1.234}^2 &= \frac{1}{281382,36} \times [(-5,480) \times 8416,96 + \\
 &\quad + (-90,435) \times (-901,00) + \\
 &\quad + 14,972 \times 14767,86] \\
 &= \frac{256461,394}{281382,36} = 0,911
 \end{aligned}$$

L'iperpiano spiega il 91,1% della variabilità totale del fatturato  $X_1$ .

c) L'indice di determinazione del piano che spiega  $X_1$  in funzione di  $X_2$  e  $X_3$  è dato da

$$I_{1.23}^2 = 0,804$$

(esercizio!). Passando da questo piano all'iperpiano che include anche la variabile esplicativa  $X_4$  si ottiene

$$MVS = 0,911 - 0,804 = 0,107$$

e

$$MVR = \frac{0,911 - 0,804}{1 - 0,804} = 0,546.$$

Nel passaggio dal piano all'iperpiano viene quindi spiegato il 10,7% della variabilità di  $X_1$  e la devianza residua diminuisce del 54,6%.

- d) La matrice dei coefficienti di correlazione grezzi  $r_{ij} = \frac{\sigma_{ij}}{\sigma_i \times \sigma_j}$  è data da

$r_{ij}$	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	1,0000			
$X_2$	0,7846	1,0000		
$X_3$	-0,8952	-0,9017	1,0000	
$X_4$	0,9483	0,8471	-0,9174	1,0000

- e) La matrice con i coefficienti di correlazione parziale

$$r_{ij.1} = \frac{r_{ij} - r_{1i}r_{1j}}{\sqrt{(1 - r_{1i}^2) \times (1 - r_{1j}^2)}}$$

per  $i, j = 2, 3, 4$  (i coefficienti  $r_{ij.1}$  scontano l'effetto di variazioni di  $X_1$ ) è data da

$r_{ij.1}$	$X_2$	$X_3$	$X_4$
$X_2$	1,0000		
$X_3$	-0,7214	1,0000	
$X_4$	0,5237	-0,4841	1,0000

Si osservi che questa matrice non include la riga e la colonna riferite a  $X_1$  perché i corrispondenti coefficienti di correlazione che scontano l'effetto di variazioni di  $X_1$  non sono definiti (infatti, la varianza e quindi anche lo scarto quadratico medio dei residui della retta ai minimi quadrati che spiega  $X_1$  in funzione di  $X_1$  è nulla!!!).

Le altre matrici di coefficienti di correlazione parziale del primo ordine sono

$$\left[ \begin{array}{c|ccc} r_{ij.2} & X_1 & X_3 & X_4 \\ \hline X_1 & 1,0000 & & \\ X_3 & -0,7004 & 1,0000 & \\ X_4 & 0,8610 & -0,6685 & 1,0000 \end{array} \right]$$

$$\left[ \begin{array}{c|ccc} r_{ij.3} & X_1 & X_2 & X_4 \\ \hline X_1 & 1,0000 & & \\ X_2 & -0,1173 & 1,0000 & \\ X_4 & 0,7163 & 0,1155 & 1,0000 \end{array} \right]$$

$$\left[ \begin{array}{c|ccc} r_{ij.4} & X_1 & X_2 & X_3 \\ \hline X_1 & 1,0000 & & \\ X_2 & -0,1109 & 1,0000 & \\ X_3 & -0,1998 & -0,5890 & 1,0000 \end{array} \right]$$

- f) La matrice dei coefficienti di correlazione parziale del secondo ordine

$$\begin{aligned} r_{ij.12} &= \frac{r_{ij.1} - r_{2i.1}r_{2j.1}}{\sqrt{(1 - r_{2i.1}^2) \times (1 - r_{2j.1}^2)}} \\ &= \frac{r_{ij.2} - r_{1i.2}r_{1j.2}}{\sqrt{(1 - r_{1i.2}^2) \times (1 - r_{1j.2}^2)}} \end{aligned}$$

può essere ricavata sia dalla matrice con i coefficienti  $r_{ij.1}$  che da quella con i coefficienti  $r_{ij.2}$ . La matrice contiene solo un

coefficiente diverso da 1:

$$\left[ \begin{array}{c|cc} r_{ij.12} & X_3 & X_4 \\ \hline X_3 & 1,0000 & \\ X_4 & -0,1802 & 1,0000 \end{array} \right]$$

La matrice con gli  $r_{ij.13}$  può essere ricavata sia dalla matrice con gli  $r_{ij.1}$  che da quella con gli  $r_{ij.3}$ :

$$\left[ \begin{array}{c|cc} r_{ij.13} & X_2 & X_4 \\ \hline X_2 & 1,0000 & \\ X_4 & 0,2880 & 1,0000 \end{array} \right]$$

La matrice con gli  $r_{ij.14}$  può essere ricavata sia dalla matrice con gli  $r_{ij.1}$  che da quella con gli  $r_{ij.4}$ :

$$\left[ \begin{array}{c|cc} r_{ij.14} & X_2 & X_3 \\ \hline X_2 & 1,0000 & \\ X_3 & -0,6276 & 1,0000 \end{array} \right]$$

La matrice con gli  $r_{ij.23}$  può essere ricavata sia dalla matrice con gli  $r_{ij.2}$  che da quella con gli  $r_{ij.3}$ :

$$\left[ \begin{array}{c|cc} r_{ij.23} & X_1 & X_4 \\ \hline X_1 & 1,0000 & \\ X_4 & 0,7398 & 1,0000 \end{array} \right]$$

La matrice con gli  $r_{ij.24}$  può essere ricavata sia dalla matrice con gli  $r_{ij.2}$  che da quella con gli  $r_{ij.4}$ :

$$\left[ \begin{array}{c|cc} r_{ij.24} & X_1 & X_3 \\ \hline X_1 & 1,0000 & \\ X_3 & -0,3300 & 1,0000 \end{array} \right]$$

La matrice con gli  $r_{ij.34}$  può essere ricavata sia dalla matrice con gli  $r_{ij.3}$  che da quella con gli  $r_{ij.4}$ :

$$\left[ \begin{array}{c|cc} r_{ij.34} & X_1 & X_2 \\ \hline X_1 & 1,0000 & \\ X_2 & -0,2886 & 1,0000 \end{array} \right]$$

- g)  $r_{14.23}^2 = 0,7398^2 = 0,5473$  è uguale al miglioramento della devianza residua  $MVR$  calcolato al punto c). La differenza è dovuta solo agli arrotondamenti.
- h) In teoria,  $\hat{\alpha}_{14.23}$  e  $\hat{\alpha}_{41.23}$  dovrebbero essere dello stesso segno e la loro media geometrica dovrebbe coincidere con  $r_{14.23}$ :

$$r_{14.23} = \text{segno}(\hat{\alpha}_{14.23}) \sqrt{\hat{\alpha}_{14.23} \times \hat{\alpha}_{41.23}}.$$

Nella risposta al quesito a) abbiamo già visto che

$$\hat{\alpha}_{14.23} = 14,972.$$

Il valore di  $\hat{\alpha}_{41.23}$  è invece dato da

$$\hat{\alpha}_{41.23} = \frac{\begin{vmatrix} 14767,860 & 8416,960 & -901,000 \\ 502,910 & 409,010 & -34,600 \\ -51,100 & -34,600 & 3,600 \end{vmatrix}}{\begin{vmatrix} 281382,360 & 8416,960 & -901,000 \\ 8416,960 & 409,010 & -34,600 \\ -901,000 & -34,600 & 3,600 \end{vmatrix}} = \frac{554971,429}{15171029,64} = 0,0366.$$

Nel calcolo di  $\hat{\alpha}_{41.23}$  bisogna prestare attenzione al fatto che si tratta del coefficiente di regressione associato alla variabile **variabile esplicativa  $X_1$**  nell'iperpiano che spiega  $X_4$  in funzione delle altre tre variabili: la matrice al denominatore di  $\hat{\alpha}_{41.23}$  è dunque la matrice delle varianze e covarianze riferita alle variabili esplicative  $X_1$ ,  $X_2$  e  $X_3$  e la prima colonna della matrice al numeratore di  $\hat{\alpha}_{41.23}$  è data dalle covarianze delle tre variabili esplicative  $X_1$ ,  $X_2$  e  $X_3$  con la variabile dipendente  $X_4$ .

Come previsto dalla teoria precedentemente esposta,

$$\text{segno}(\hat{\alpha}_{14.23}) = \text{segno}(\hat{\alpha}_{41.23}),$$

e la media geometrica dei due coefficienti di regressione parziale  $\hat{\alpha}_{14.23} = 14,972$  e  $\hat{\alpha}_{41.23} = 0,0366$  è data da

$$r_{14.23} = \sqrt{14,972 \times 0,0366} = 0,7403.$$

La differenza rispetto al valore di  $r_{14.23}$  ottenuto al punto g) è dovuta solo agli arrotondamenti.



- i) Per definizione  $r_{14.23}$  è il coefficiente di correlazione (grezzo) tra i residui dei due piani che considerano rispettivamente  $X_1$  e  $X_4$  come variabili dipendenti, e  $X_2$  e  $X_3$  come variabili esplicative. Per ottenere i residui dei due piani bisogna innanzitutto determinare i parametri dei piani. Dopo alcuni calcoli (esercizio) si ottiene

$$\hat{x}_{1.23} = 2109,519 - 3,173x_2 - 280,777x_3$$

e

$$\hat{x}_{4.23} = 118,588 + 0,154x_2 - 12,714x_3.$$

I valori riprodotti da questi due piani ai minimi quadrati e i corrispondenti residui sono riportati nella seguente tabella:

mercato	$x_{1i}$	$\hat{x}_{1.23i}$	$z_{1.23i}$	$x_{4i}$	$\hat{x}_{4.23i}$	$z_{4.23i}$	$z_{1.23i}^2$	$z_{4.23i}^2$	$z_{1.23i} \times z_{4.23i}$
A	98	106,004	-8,004	20	31,438	-11,438	64,064	130,828	91,550
B	123	71,101	51,899	26	33,132	-7,132	2693,506	50,865	-370,144
C	235	386,781	-151,781	34	44,152	-10,152	23037,472	103,063	1540,881
D	360	386,781	-26,781	55	44,152	10,848	717,222	117,679	-290,520
E	465	291,591	173,409	61	48,772	12,228	30070,681	149,524	2120,445
F	567	597,752	-30,752	64	60,254	3,746	945,686	14,033	-115,197
G	599	594,579	4,421	58	60,408	-2,408	19,545	5,798	-10,646
H	698	594,579	103,421	76	60,408	15,592	10695,903	243,110	1612,540
I	780	1325,855	-545,855	83	103,94	-20,940	297957,681	438,484	11430,204
J	2043	1612,978	430,022	126	116,346	9,654	184918,920	93,200	4151,432
Tot	5968	5968,001	-0,001	603	603,002	-0,002	551120,680	1346,584	20160,545

Usando gli opportuni totali di colonna si ottiene

$$\begin{aligned} \text{corr}(Z_{1.23}, Z_{4.23}) &= \frac{20160,545}{\sqrt{551120,680 \times 1346,584}} \\ &= 0,740 \end{aligned}$$

e questo valore coincide con il valore di  $r_{14.23}$  ottenuto per altre vie ai punti f) e h) (la differenza è dovuta solo agli arrotondamenti).

□