

1 IL MODELLO LINEARE GENERALE

Indice del capitolo

1.1	Serie storiche, dati sezionali e longitudinali.....	2
	<i>Dati longitudinali.....</i>	3
1.2	Il criterio dei minimi quadrati.....	4
1.3	I minimi quadrati nel modello lineare semplice.....	7
	<i>Le stime dei minimi quadrati nel modello lineare semplice.....</i>	7
1.4	I minimi quadrati nel modello lineare multiplo.....	11
	<i>La condizione necessaria per i minimi quadrati e le equazioni normali.....</i>	13
	<i>L'ortogonalità dei residui rispetto alle variabili esplicative.....</i>	14
	<i>Un esempio: il modello lineare semplice in termini matriciali.....</i>	15
	<i>La condizione sufficiente per i minimi quadrati.....</i>	16
1.5	La scomposizione della devianza ed il coefficiente di determinazione.....	18
	<i>Il coefficiente di determinazione in termini matriciali.....</i>	19
	<i>Il coefficiente di determinazione corretto.....</i>	20
	<i>Il coefficiente di determinazione per il modello con le variabili scarto.....</i>	21
1.6	I residui come enti aleatori: le ipotesi deboli.....	24
	<i>Lo stimatore dei minimi quadrati per il modello lineare semplice.....</i>	27
	<i>Lo stimatore dei minimi quadrati per il modello lineare multiplo.....</i>	29
1.7	La stima della varianza dei residui.....	32
	<i>La distorsione della varianza campionaria.....</i>	32
1.8	Il teorema di Gauss-Markov e gli stimatori BLU.....	35
1.9	La matrice di correlazione degli stimatori dei parametri di regressione.....	37
1.10	La stima dei minimi quadrati di una funzione delle importazioni.....	39
1.11	Il criterio dei minimi quadrati vincolati.....	44
	<i>La stima dei minimi quadrati vincolati.....</i>	45
	<i>Le stime dei minimi quadrati vincolati per una funzione delle importazioni..</i>	48
	<i>Il vincolo di omogeneità di grado zero sui prezzi.....</i>	50
	<i>Il doppio vincolo dell'uguaglianza delle elasticità e dell'omogeneità sui prezzi</i>	50
1.12	Riferimenti bibliografici.....	53

1.1 Serie storiche, dati sezionali e longitudinali

Fin dall'inizio è stata presa in considerazione la semplice funzione del consumo di derivazione keynesiana (I-2.1.1) nella quale consumo e reddito, legati da una relazione lineare, possono essere riferiti ad istanti differenti di tempo, $t = 1, 2, \dots, n$, oppure ad unità di consumo e di reddito (ad esempio *famiglie*), $i = 1, 2, \dots, N$, considerate allo stesso tempo t . Si possiede, allora, nel primo caso un *campione* di osservazioni che formano *serie storiche*

$$c_t = \alpha + \beta y_t \quad t = 1, 2, \dots, n \quad (1.1.1)$$

mentre nel secondo le osservazioni compongono *dati sezionali*¹.

$$c_i = \alpha + \beta y_i \quad i = 1, 2, \dots, N \quad (1.1.2)$$

Un campione temporale di *ampiezza* n può essere costruito mediante indagini che si protraggono nel tempo, oppure tramite una disaggregazione temporale (ad esempio trimestralizzazione o mensilizzazione di dati annuali), mentre un campione sezionale di ampiezza N può essere estratto da un'inchiesta puntuale nel tempo, ad esempio un'indagine sulla spesa di un gruppo di famiglie oppure un censimento.

I modelli (1.1.1) e (1.1.2) sono analoghi e differiscono unicamente nel modo con cui i dati sono stati reperiti. Naturalmente esistono modelli i cui dati sono contemporaneamente *sezionali* e *temporali*, come nell'esempio seguente

$$c_{it} = \alpha_i + \beta_i y_{it} \quad t = 1, 2, \dots, n; i = 1, 2, \dots, N \quad (1.1.3)$$

rappresentativo di una funzione del consumo nella quale ciascuna famiglia i possiede una propria funzione definita dai parametri α_i e β_i , considerati costanti nel *periodo di osservazione* campionario, cioè per $t = 1, 2, \dots, n$.

Se poniamo

$$c_t = \sum_{i=1}^N c_{it}, \quad \alpha = \sum_{i=1}^N \alpha_i, \quad y_t = \sum_{i=1}^N y_{it}$$

e nell'ipotesi che tutte le propensioni marginali al consumo siano uguali, $\beta_1 = \beta_2 = \dots = \beta_N = \beta$, le equazioni (1.1.3) possono essere sommate membro a membro in modo da dare

$$c_t = \alpha + \beta y_t \quad t = 1, 2, \dots, n$$

costituendo questa l'aggregazione sezionale delle (1.1.3).

¹ Le serie storiche o temporali vengono dette in lingua inglese *time series* mentre i dati sezionali sono detti *cross-section data*.

Un altro modo di aggregare le equazioni (1.1.3) è quello che si basa sulla conoscenza della distribuzione del reddito. Se la quota di reddito y_t posseduta dalla i -esima famiglia in ogni tempo è λ_i , con il vincolo

$$\sum_{i=1}^N \lambda_i = 1$$

si ha che

$$y_{it} = \lambda_i y_t \dots \quad t=1,2,\dots,n \ ; \ i=1,2,\dots,N \quad (1.1.4)$$

per cui, sostituendo le (1.1.4) nelle (1.1.3) e tenendo conto del vincolo, si ottiene sommando membro a membro

$$c_t = \alpha + \beta^0 y_t$$

dove $\beta^0 = \sum_{i=1}^N \lambda_i \beta_i$, di nuovo del tipo (1.1.1) ma con un'altra aggregazione sezionale.

Dati longitudinali

Se il campione di famiglie considerato nella (1.1.3) rimane costante negli n tempi, i dati ad esso relativi, $\{c_{it}\}$ e $\{y_{it}\}$ sono chiamati *longitudinali*, alludendo al fatto che un campione di più individui viene seguito lungo il tempo². Per il trattamento dei dati longitudinali si usano procedure econometriche specifiche che non saranno trattate nel presente modulo.

² In lingua inglese i dati longitudinali vengono chiamati *panel data* (dal termine *panel*, che indica un gruppo di individui).

1.2 Il criterio dei minimi quadrati

Generalmente i valori dei parametri di un'equazione non sono conosciuti ed occorre *stimarli* a partire da un campione di osservazioni. Volendo, ad esempio, determinare i valori di α e β nella funzione del consumo (1.1.1), è necessario disporre di un campione costituito dagli n consumi $[c_1 \ c_2 \ \dots \ c_n]$ e dai corrispondenti redditi $[y_1 \ y_2 \ \dots \ y_n]$, che possono essere temporali o sezionali a seconda delle circostanze. In generale, nel prosieguo, supporremo che i dati siano temporali, essendo relativamente facile utilizzare nel caso sezionale le tecniche sviluppate partendo da osservazioni temporali.

Il campione costituito dalle c_t e dalle y_t può essere riportato in un grafico del tipo illustrato nella figura 1.1, detto *diagramma di dispersione* delle coppie di valori (c_t, y_t) , dal quale risulta evidente l'ovvia circostanza che in generale non esiste una retta che passi esattamente per tutti i punti individuati da questi dati, e che pertanto non esistono valori di α e β che soddisfino perfettamente la (1.1.1) per ognuna delle coppie campionarie.

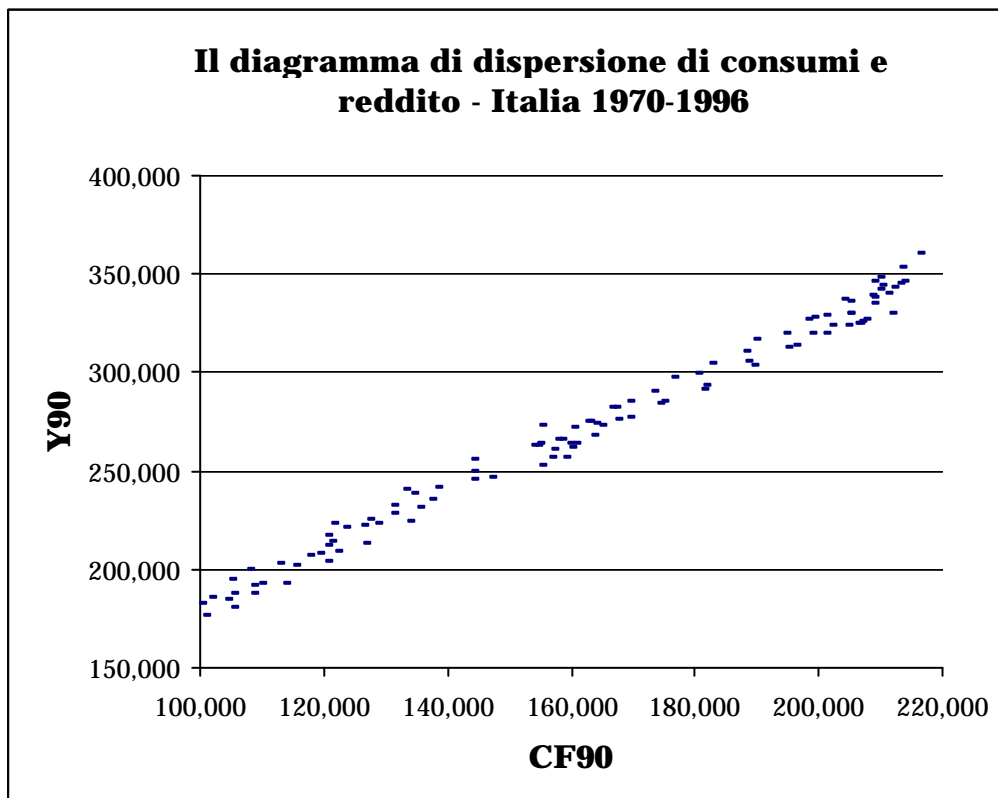


Figura 1.1 - Il diagramma di dispersione di consumi e reddito - dati trimestrali italiani 1970:1-1996:3. Le serie utilizzate per costruire il grafico sono quelle rappresentate nella figura I-2.2.

Si può, invece, trovare una retta del tipo (1.1.1), e quindi una coppia di valori α e β , tale che attraversi la “nuvola” di punti (c_t, y_t) nel diagramma di figura 1.1 con una distanza u_t da questi punti, misurata secondo le ordinate, che obbedisca ad un particolare criterio di ottimo, scelto soggettivamente. Se il criterio fa determinare i valori $\hat{\alpha}$ e $\hat{\beta}$, l’equazione della retta è

$$\hat{c}_t = \hat{\alpha} + \hat{\beta}y_t$$

definente i valori *teorici* \hat{c}_t per il consumo, in funzione delle y_t campionarie e di $\hat{\alpha}$ e $\hat{\beta}$; le differenze tra i valori *osservati* e quelli *teorici* per le c_t costituiscono i *residui* \hat{u}_t

$$\hat{u}_t = c_t - \hat{c}_t = c_t - \hat{\alpha} - \hat{\beta}y_t \quad t = 1, 2, \dots, n$$

Ad ogni coppia $\hat{\alpha}$, $\hat{\beta}$ corrisponde una serie storica di residui $\{\hat{u}_t\}$ diversa, per cui finché ad α e a β non si danno valori determinati è possibile considerare una serie di residui $\{u_t\}$ non specificata ma che si suppone generata dal modello

$$u_t = c_t - \alpha - \beta y_t \quad t = 1, 2, \dots, n \quad (1.2.1)$$

Si è detto che il criterio con il quale si determinano α e β è soggettivo; spesso si usa il *criterio dei minimi quadrati*, sviluppato indipendentemente dai matematici K. F. Gauss e A. M. Legendre tra la fine del diciottesimo e gli inizi del diciannovesimo secolo, mediante il quale i valori $\hat{\alpha}$ e $\hat{\beta}$ sono calcolati minimizzando la somma dei quadrati dei residui

$$S(\alpha, \beta) = \sum_{t=1}^n (c_t - \alpha - \beta y_t)^2 = \sum_{t=1}^n u_t^2 \quad (1.2.2)$$

detta *devianza* dei residui³.

Questo criterio si basa sulla ricerca della retta che attraversa la nuvola di punti (c_t, y_t) in modo tale da minimizzare la somma dei quadrati delle distanze tra se stessa e i punti, con tali distanze prese rispetto all’asse delle ordinate. Al posto di questo criterio se ne possono scegliere altri, ad esempio quello basato sulla minimizzazione della somma dei valori assoluti

$$S'(\alpha, \beta) = \sum_{t=1}^n |c_t - \alpha - \beta y_t| = \sum_{t=1}^n |u_t| \quad (1.2.3)$$

mediante il quale, ovviamente, si ricavano valori diversi, per α e β , dai precedenti $\hat{\alpha}$ e $\hat{\beta}$. Generalmente, con criteri differenti si ottengono *stime* diverse per α e β . A prescindere dal criterio utilizzato, i parametri α e β sono supposti valere identici

³ In lingua inglese: *Residual Sum of Squares*, denotata con l’acronimo RSS.

per ogni coppia e cioè per ogni t . Ipotizziamo, in altre parole, che la “struttura dell’economia”, nella fattispecie l’equazione del consumo, rimanga inalterata nel *periodo campionario* ovvero che il campione sia *omogeneo*.

Osservazione 1.1- In ambedue i criteri sopra indicati i valori negativi e quelli positivi delle u_t sono trattati alla stessa stregua (simmetricamente). Nel criterio definito dalla (1.2.2) i residui intervengono a comporre la devianza in modo non lineare, in quanto il loro contributo ad essa è proporzionale al loro quadrato. In termini geometrici questo equivale a dire che la retta dei minimi quadrati del tipo (1.1.1) è determinata in modo che sia più sensibile ai punti più “esterni” nella nuvola (c_t, y_t) che non a quelli più “interni” (e vicini alla stessa retta).

Osservazione 1.2 - Le distanze sono prese parallelamente all’asse delle ordinate ma potrebbero essere parimenti considerate parallelamente all’asse delle ascisse, oppure anche ortogonali alla retta che definisce i valori teorici \hat{c}_t .

1.3 I minimi quadrati nel modello lineare semplice

Il modello (1.2.1) può essere scritto nella forma un poco più generale

$$u_t = y_t - \beta_1 - \beta_2 x_t \quad t = 1, 2, \dots, n \quad (1.3.1)$$

dove y_t è la variabile endogena (il consumo) e x_t un'esplicativa (il reddito). In effetti la x_t non può essere considerata generalmente un'esogena poiché talvolta rappresenta un'endogena in altre equazioni; ad esempio, la x_t che costituisce un'esogena nell'equazione singola (I-2.1.1) corrisponde all'endogena nella seconda equazione del sistema (I-2.3.1); è più conveniente, pertanto, chiamare la x_t nella (1.3.1) variabile *esplicativa* (di y). In questo caso più generale i valori teorici per le y_t sono dati da

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t \quad t = 1, 2, \dots, n \quad (1.3.2)$$

mentre le differenze tra i valori osservati y_t e quelli teorici \hat{y}_t costituiscono i valori dei residui

$$\hat{u}_t = y_t - \hat{y}_t = y_t - \hat{\beta}_1 + \hat{\beta}_2 x_t \quad t = 1, 2, \dots, n \quad (1.3.3)$$

Se le variabili esplicative sono k , il modello (1.3.1) viene generalizzato nell'altro

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t \quad \forall t \quad (1.3.4)$$

che non necessariamente corrisponde ad una generica equazione della forma ridotta (I-3.4.1); in linea di principio le x_{it} sono variabili esplicative di y_t per cui la (1.3.4) può essere sia un'equazione qualsiasi del sistema (I-3.4.1), in forma ridotta, sia una del sistema (I-3.2.1), in forma strutturale, risolta rispetto ad una variabile endogena. Il modello (1.3.4) è detto *lineare multiplo*. La (1.3.2) è a sua volta un caso particolare della seguente

$$u_t = g(y_t, x_{1t}, x_{2t}, \dots, x_{kt}; a_1, a_2, \dots, a_h) \quad t = 1, 2, \dots, n \quad (1.3.5)$$

dove le x_{it} sono le variabili esplicative, le a_j sono i parametri del modello, variabili nello spazio parametrico A , e g è una funzione che può essere lineare, come nella formulazione (1.3.4), oppure non lineare.

Le stime dei minimi quadrati nel modello lineare semplice

Le stime dei parametri a_i possono essere determinate con il criterio dei minimi quadrati, per mezzo del quale si calcolano i valori \hat{a}_j per i quali si ottiene il minimo della devianza

$$\min_{a_1, \dots, a_h} \sum_{t=1}^n u_t^2 = \min_{a_1, \dots, a_h} \sum_{t=1}^n [g(y_t, x_{1t}, x_{2t}, \dots, x_{kt}; a_1, a_2, \dots, a_h)]^2 = \min_{a_1, \dots, a_h} S(a_1, a_2, \dots, a_h) \quad (1.3.6)$$

Le condizioni necessarie affinché valga la (1.3.6) impongono che siano soddisfatte le equazioni seguenti

$$\frac{\partial}{\partial a_i} S(a_1, a_2, \dots, a_h) = \frac{\partial}{\partial a_i} \sum_{t=1}^n g(y_t, x_{1t}, x_{2t}, \dots, x_{ht}; a_1, a_2, \dots, a_h) = 0$$

per $i = 1, 2, \dots, h$, dette *equazioni normali*. Queste possono essere lineari oppure non lineari, a seconda delle (1.3.5). Nel primo caso si perviene ai minimi quadrati (ordinari) lineari (OLS: *Ordinary Least Squares*, in lingua inglese); nel secondo caso ai minimi quadrati non lineari (NLLS: *Non Linear Least Squares* in inglese).

Nel caso lineare della (1.3.1) il criterio dei minimi quadrati (1.3.6) comporta la determinazione del minimo seguente

$$\min_{\beta_1, \beta_2} \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)^2 = \min_{\beta_1, \beta_2} S(\beta_1, \beta_2)$$

per il quale occorre trovare le derivate prime di $S(\beta_1, \beta_2)$ ed uguagliarle a zero, ottenendosi le due equazioni normali

$$\begin{cases} \frac{\partial S}{\partial \beta_1} = 2 \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)(-1) = 0 \\ \frac{\partial S}{\partial \beta_2} = 2 \sum_{t=1}^n (y_t - \beta_1 - \beta_2 x_t)(-x_t) = 0 \end{cases}$$

cioè

$$\begin{cases} \sum_{t=1}^n y_t = n\beta_1 + \beta_2 \sum_{t=1}^n x_t \\ \sum_{t=1}^n x_t y_t = \beta_1 \sum_{t=1}^n x_t + \beta_2 \sum_{t=1}^n x_t^2 \end{cases} \quad (1.3.7)$$

Se si pone

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t, \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t, \quad m_{xx} = \frac{1}{n} \sum_{t=1}^n x_t^2, \quad m_{xy} = \frac{1}{n} \sum_{t=1}^n x_t y_t \quad (1.3.8)$$

dalla prima delle (1.3.7) si ricava, dividendo per n ,

$$\bar{y} = \beta_1 + \beta_2 \bar{x} \quad (1.3.9)$$

e dalla seconda, sostituendo il valore di β_1 dato dalla (1.3.9),

$$\sum_{t=1}^n x_t y_t = (\bar{y} - \beta_2 \bar{x}) \sum_{t=1}^n x_t + \beta_2 \sum_{t=1}^n x_t^2$$

cioè

$$m_{xy} = \bar{y}\bar{x} + \beta_2(m_{xx} - \bar{x}^2)$$

dalle quali si ottiene la *stima dei minimi quadrati* di β_2

$$\hat{\beta}_2 = \frac{m_{xy} - \bar{y}\bar{x}}{m_{xx} - \bar{x}^2} \quad m_{xx} \neq \bar{x}^2 \quad (1.3.10)$$

e, sostituendo nella (1.3.9), quella di β_1

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad (1.3.11)$$

Le stime $\hat{\beta}_1$ e $\hat{\beta}_2$ costituiscono effettivamente un punto di minimo per $S(\beta_1, \beta_2)$ in quanto sono soddisfatte anche le condizioni sufficienti, date dalle

$$\frac{\partial^2 S}{\partial \beta_1^2} > 0, \quad \frac{\partial^2 S}{\partial \beta_2^2} > 0, \quad \frac{\partial^2 S}{\partial \beta_1^2} \cdot \frac{\partial^2 S}{\partial \beta_2^2} - \left(\frac{\partial^2 S}{\partial \beta_1 \partial \beta_2} \right)^2 > 0;$$

infatti

$$\frac{\partial^2 S}{\partial \beta_1^2} = 2n > 0, \quad \frac{\partial^2 S}{\partial \beta_2^2} = 2 \sum_{t=1}^n x_t^2 > 0, \quad \frac{\partial^2 S}{\partial \beta_1 \partial \beta_2} = 2 \sum_{t=1}^n x_t$$

dalle quali segue la terza condizione

$$2n \times 2 \sum_{t=1}^n x_t^2 - \left(2 \sum_{t=1}^n x_t \right)^2 = 4n^2 (m_{xx} - \bar{x}^2) = 4n \sum_{t=1}^n (x_t - \bar{x})^2 > 0$$

Osservazione 1.3 - Se ad \bar{x} e \bar{y} si dà l'interpretazione (statistica) di valori medi campionari, e a m_{xx} e m_{xy} quella di momenti secondi campionari, la stima $\hat{\beta}_2$ è data dal rapporto della covarianza sulla varianza, anch'esse campionarie.

Osservazione 1.4 - Dalla (1.3.9) segue che la retta di regressione

$$y_t = \beta_1 + \beta_2 x_t \quad (1.3.12)$$

passa sempre nel punto (\bar{y}, \bar{x}) quali che siano i valori di β_1 e β_2 che soddisfano le equazioni (1.3.7).

Osservazione 1.5 - Si noti che in corrispondenza del punto di ottimo le equazioni normali possono essere scritte come segue

$$\begin{cases} \sum_{t=1}^n (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = \sum_{t=1}^n \hat{u}_t = 0 \\ \sum_{t=1}^n (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) x_t = \sum_{t=1}^n \hat{u}_t x_t = 0 \end{cases} \quad (1.3.13)$$

Di conseguenza, i residui stimati nel modello (1.3.1) sommano a zero (e quindi hanno media campionaria nulla).

Osservazione 1.6 - Ancora dall'Osservazione 1.4 e dalla (1.3.2) segue che

$$\frac{1}{n} \sum_{t=1}^n \hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{n} \sum_{t=1}^n x_t = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} = \bar{y}$$

cioè \bar{y} è il valore medio campionario non soltanto delle y_t osservate ma anche delle \hat{y}_t teoriche.

1.4 I minimi quadrati nel modello lineare multiplo

Ci poniamo ora il problema di determinare le stime dei minimi quadrati dei parametri del modello (1.3.4) a partire da un campione omogeneo (nel senso illustrato nel paragrafo 1.2) di osservazioni, di ampiezza n , relativo alla y_t

$$\mathbf{y} = [y_1 \quad y_2 \quad \dots \quad y_n]'$$

e alle variabili esplicative x_{it} , che riportiamo in forma matriciale⁴

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix} \quad (1.4.1)$$

Nella \mathbf{X} , di ordine $n \times k$, contrariamente a quanto è consuetudinario nell'algebra delle matrici, il primo indice di ogni elemento indica la colonna (e definisce la variabile) ed il secondo la riga (e denota l'elemento del campione o del tempo).

Anche nel caso del modello lineare multiplo (1.3.4), detto anche *di regressione lineare multipla*, i residui u_t servono a bilanciare i due membri dell'equazione una volta che siano definiti i parametri $\mathbf{b} = [\beta_1, \beta_2, \dots, \beta_k]'$. Similmente, ad una stima dei parametri $\hat{\mathbf{b}} = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]'$ corrisponde una stima \hat{u}_t per il residuo in ogni tempo e un particolare valore teorico \hat{y}_t

$$\hat{u}_t = y_t - \hat{y}_t = y_t - (\hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \dots + \hat{\beta}_k x_{kt}) = y_t - \hat{\mathbf{b}}' \mathbf{x}_t \quad (1.4.2)$$

dove \mathbf{x}_t è la t -esima riga della matrice \mathbf{X} .

Osservazione 1.7 - Così, \hat{u}_t è dato dalla differenza tra il valore osservato e quello teorico di y_t .

Il termine additivo u_t misura tutto quanto non è spiegato dalle variabili esplicative x_{it} e per questo motivo è chiamato *residuale*; esso è costituito tra l'altro dalla possibile aggregazione di:

- variabili che non sono state inserite tra le esplicative (*omesse*) e che invece spiegherebbero parte di y_t ,
- impulsi accidentali prodotti dal sistema economico su y_t , validi soltanto per alcune t e non in modo sistematico per tutto il campione,

⁴ I concetti di algebra matriciale necessari alla comprensione dei passaggi successivi sono esposti nel capitolo XIX-1.

- elementi caratteristici di y_t , ad esempio le stagionalità, che non si riesce a spiegare per mezzo delle x_{it} ,
- errori nella misurazione della y_t ,
- elementi di disturbo dovuti al fatto che la specificazione della (1.3.4) è lineare, mentre avrebbe dovuto essere non lineare rispetto ad alcune delle variabili esplicative.

Osservazione 1.8 - Da questa caratterizzazione segue che non ha senso considerare u_t come un errore, anche se in tale modo sovente viene chiamato a seguito delle prime utilizzazioni del modello (1.3.4) in demografia e nelle scienze fisiche. Questa denominazione, in econometria, è chiaramente un errore.

Molti sono i criteri che si possono utilizzare per ottenere delle stime dei parametri \mathbf{b} a partire da un campione \mathbf{y} e \mathbf{X} ; tra questi, di nuovo quello che minimizza la somma dei valori assoluti dei residui, del tipo (1.2.3), e quello dei minimi quadrati (1.3.6), che ora diventa

$$\min_{\mathbf{b}} S(\mathbf{b}) = \min_{\mathbf{b}} \sum_{t=1}^n (y_t - \mathbf{b}' \mathbf{x}_t)^2 \quad (1.4.3)$$

Per ottenere questo minimo è utile riscrivere il modello (1.3.4), per tutte le t , in forma matriciale

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ \dots \\ \dots \\ u_n \end{bmatrix}$$

cioè

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u} \quad (1.4.4)$$

dove $\mathbf{u} = [u_1 \ u_2 \ \dots \ u_n]'$. Allora la devianza $S(\mathbf{b})$ è

$$S(\mathbf{b}) = \sum_{t=1}^n u_t^2 = \mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) \quad (1.4.5)$$

La condizione necessaria per i minimi quadrati e le equazioni normali

Facendo uso delle definizioni precedenti è semplice determinare le stime dei minimi quadrati di b con la minimizzazione della devianza (1.4.5) che può essere scritta nella forma

$$\begin{aligned} S(b) &= (\mathbf{y}-\mathbf{X}b)'(\mathbf{y}-\mathbf{X}b) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}b - b'\mathbf{X}'\mathbf{y} + b'\mathbf{X}'\mathbf{X}b = \\ &= \mathbf{y}'\mathbf{y} - 2b'\mathbf{X}'\mathbf{y} + b'\mathbf{X}'\mathbf{X}b \end{aligned} \quad (1.4.6)$$

dove si è sfruttato il fatto che $b'\mathbf{X}'\mathbf{y}$ è uno scalare e quindi uguale al suo trasposto. La condizione necessaria affinché $S(b)$ sia minimizzata consiste nell'uguaglianza a zero delle sue derivate parziali rispetto a b , che possono essere calcolate applicando le regole di derivazione vettoriale esposte nel paragrafo XIX-1.9 nel modo seguente

$$\frac{\partial S(b)}{\partial b'} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}b = \mathbf{0} \quad (1.4.7)$$

essendo nulle le derivate parziali della $\mathbf{y}'\mathbf{y}$, costante rispetto ai b . Dalla (1.4.7) si ottengono le *equazioni normali* dei minimi quadrati

$$\mathbf{X}'\mathbf{X}b = \mathbf{X}'\mathbf{y} \quad (1.4.8)$$

Raccogliendo i termini le (1.4.8) possono essere espresse come $\mathbf{X}'(\mathbf{y}-\mathbf{X}b) = \mathbf{X}'\mathbf{u} = \mathbf{0}$, per cui di fatto le condizioni del primo ordine per la soluzione del problema dei minimi quadrati, cioè le equazioni normali, equivalgono a una condizione di ortogonalità fra le colonne di \mathbf{X} , cioè le variabili esplicative del modello, e i residui \mathbf{u} .⁵

Se la matrice $\mathbf{X}'\mathbf{X}$ è invertibile, cioè se⁶

$$\det(\mathbf{X}'\mathbf{X}) \neq 0 \quad (1.4.9)$$

si ottiene la stima dei minimi quadrati ordinari

$$\hat{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (1.4.10)$$

Più oltre, in questo paragrafo, determineremo la condizione sufficiente per il minimo.

In conclusione, l'unica ipotesi che è necessario fare per calcolare la *stima dei minimi quadrati* \hat{b} mediante la formula (1.4.10), oltre al requisito dell'omogeneità del campione (\mathbf{y}, \mathbf{X}) , è che valga la (1.4.9).

⁵ Per la nozione di ortogonalità fra vettori si veda il par. XIX-1.1.

⁶ Nel paragrafo XIX-1.8 si mostra che il determinante di $\mathbf{X}'\mathbf{X}$ è nonnegativo per una \mathbf{X} qualsiasi.

Osservazione 1.9 – Poiché il rango massimo di $\mathbf{X}'\mathbf{X}$ è k , per i teoremi XIX-1.4 e XIX-1.3 la condizione (1.4.9) è equivalente alla seguente

$$n > k, \quad r(\mathbf{X}) = k \quad (1.4.11)$$

L'ortogonalità dei residui rispetto alle variabili esplicative

Poiché le equazioni normali (1.4.8) sono soddisfatte da $\hat{\mathbf{b}}$ definita dalla (1.4.10), vale la relazione

$$\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y} \quad (1.4.12)$$

che viene utilizzata per dimostrare l'ortogonalità delle stime $\hat{\mathbf{u}} = [\hat{u}_1 \quad \hat{u}_2 \quad \dots \quad \hat{u}_n]'$ dei residui rispetto alle variabili esplicative⁷. Coerentemente con la (1.4.2) si ha

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} \quad (1.4.13)$$

costituito dalla differenza tra il vettore \mathbf{y} dei valori osservati e quello $\hat{\mathbf{y}}$ dei valori teorici

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} \quad (1.4.14)$$

formati dalla componente sistematica delle y_i con i parametri $\hat{\beta}_i$ stimati tramite il criterio dei minimi quadrati. L'ortogonalità di $\hat{\mathbf{u}}$ rispetto alle variabili esplicative consiste, dunque, nel fatto che è

$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{0} \quad (1.4.15)$$

per la (1.4.8). Trasponendo la (1.4.15) si ottiene, ovviamente,

$$\hat{\mathbf{u}}'\mathbf{X} = \mathbf{0}' \quad (1.4.16)$$

Le (1.4.15) o (1.4.16) generalizzano le (1.3.13).

L'ortogonalità mostrata dalla (1.4.15) può essere interpretata nel senso che $\hat{\mathbf{u}}$ è la proiezione ortogonale di \mathbf{y} sullo spazio ortogonale a quello generato dalle colonne di \mathbf{X} . D'altro canto $\hat{\mathbf{y}}$ consiste nella proiezione ortogonale di \mathbf{y} sullo spazio generato dalle colonne di \mathbf{X} , per cui $\hat{\mathbf{u}}$ e $\hat{\mathbf{y}}$ sono ortogonali. Infatti

$$\hat{\mathbf{y}}'\hat{\mathbf{u}} = \hat{\mathbf{b}}'\mathbf{X}'\hat{\mathbf{u}} = 0$$

avendo sfruttato le (1.4.14) e (1.4.15).

Le relazioni (1.4.15) e (1.4.16) permettono di determinare la seguente scomposizione del prodotto scalare $\mathbf{y}'\mathbf{y}$ che otteniamo in virtù delle definizioni (1.4.13) e (1.4.14)

⁷ Talvolta la stima \hat{u}_i è indicata con il simbolo e_i ed il vettore $\hat{\mathbf{u}}$ con \mathbf{e} .

$$\mathbf{y}'\mathbf{y} = (\mathbf{X}\hat{\mathbf{b}} + \hat{\mathbf{u}})'(\mathbf{X}\hat{\mathbf{b}} + \hat{\mathbf{u}}) = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + \hat{\mathbf{b}}'\mathbf{X}'\hat{\mathbf{u}} + \hat{\mathbf{u}}'\mathbf{X}\hat{\mathbf{b}} + \hat{\mathbf{u}}'\hat{\mathbf{u}} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{u}}'\hat{\mathbf{u}} \quad (1.4.17)$$

e che lega i valori osservati \mathbf{y} con i teorici e i residui stimati.

Se l'equazione (1.3.4) contiene l'intercetta, una delle colonne di \mathbf{X} , ad esempio l'ultima, è costituita da unità, per cui anche l'ultima riga di \mathbf{X}' è formata da 1, e dalla (1.4.15) deriva che

$$\sum_{t=1}^n \hat{u}_t = 0 \quad (1.4.18)$$

cioè il valor medio campionario dei residui stimati è nullo, se viene calcolata l'intercetta. La (1.4.18) generalizza la prima delle (1.3.13). In questo caso, considerando la (1.4.2), si ha anche che il valor medio campionario delle y_t è uguale al valor medio campionario delle \hat{y}_t stimate

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t = \frac{1}{n} \sum_{t=1}^n \hat{y}_t \quad (1.4.19)$$

Osservazione 1.10 - L'equazione i -esima del sistema (1.4.15) può essere scritta nella forma

$$\sum_{t=1}^n x_{it}(y_t - \mathbf{x}'_t \hat{\mathbf{b}}) = \sum_{t=1}^n x_{it} \hat{u}_t = 0 \quad (1.4.20)$$

dove \mathbf{x}'_t è il vettore delle k variabili esplicative al tempo t . Se in virtù della presenza dell'intercetta è, ad esempio, $x_{kt}=1$, per ogni t , si ottiene nuovamente la (1.4.18).

Un esempio: il modello lineare semplice in termini matriciali

Trattiamo il caso del modello lineare semplice (1.3.1) in termini matriciali. La (1.4.4) è in forma esplicita

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ \dots \\ u_n \end{bmatrix}$$

per cui la matrice $\mathbf{X}'\mathbf{X}$ è

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{t=1}^n x_t \\ \sum_{t=1}^n x_t & \sum_{t=1}^n x_t^2 \end{bmatrix}$$

con determinante $\det(\mathbf{X}'\mathbf{X}) = n \sum_{t=1}^n x_t^2 - \left(\sum_{t=1}^n x_t \right)^2$ e aggiunta

$$\text{agg}(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} \sum_{t=1}^n x_t^2 & - \sum_{t=1}^n x_t \\ - \sum_{t=1}^n x_t & n \end{bmatrix}$$

Si ha, allora, facendo uso delle posizioni (1.3.8)

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \frac{1}{m_{xx} - \bar{x}^2} \begin{bmatrix} m_{xx} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \bar{y} \\ m_{xy} \end{bmatrix} = \frac{1}{m_{xx} - \bar{x}^2} \begin{bmatrix} m_{xx}\bar{y} - \bar{x}m_{xy} \\ m_{xy} - \bar{x}\bar{y} \end{bmatrix}$$

stime uguali alle (1.3.11) e (1.3.10), rispettivamente.

La condizione sufficiente per i minimi quadrati

Determiniamo ora le condizioni sufficienti affinché la devianza (1.4.6) possenga un minimo nel punto estremante $\hat{\mathbf{b}}$ trovato tramite la condizione necessaria (1.4.10). Se consideriamo il differenziale del secondo ordine della funzione $S(\mathbf{b})$

$$d^2S(\mathbf{b}) = d\mathbf{b}'\mathbf{H}d\mathbf{b} \tag{1.4.21}$$

dove $d\mathbf{b} = (db_1 \ db_2 \ \dots \ db_k)$ ed \mathbf{H} è la matrice quadrata simmetrica di ordine k , detta *hessiana*, il cui elemento generico (i, j) è

$$\frac{\partial^2 S(\mathbf{b})}{\partial \beta_i \partial \beta_j}$$

$d^2S(\mathbf{b})$ è positivo o negativo a seconda che \mathbf{H} sia definita positiva o negativa. Così la condizione sufficiente affinché $\hat{\mathbf{b}}$ corrisponda ad un minimo (la positività del differenziale $d^2S(\mathbf{b})$) è che la matrice

$$\mathbf{H} = \frac{\partial^2 S(\mathbf{b})}{\partial \beta_i \partial \beta_j} = 2\mathbf{X}'\mathbf{X} \tag{1.4.22}$$

ottenuta derivando la (1.4.7) rispetto a \mathbf{b} sia definita positiva. Ma per ipotesi (1.4.11) e il teorema XIX-1.8 la $\mathbf{X}'\mathbf{X}$ è definita positiva e quindi $\hat{\mathbf{b}}$ è un punto di minimo.

Dunque la condizione (1.4.9) oppure l'equivalente (1.4.11) è necessaria e sufficiente perché $\hat{\mathbf{b}}$ sia un minimo per $S(\mathbf{b})$.

Osservazione 1.11 - In analogia a quanto asserito sopra, la condizione sufficiente affinché $\hat{\mathbf{b}}$ corrisponda ad un massimo (la negatività del differenziale $d^2S(\mathbf{b})$) è che la matrice \mathbf{H} sia definita negativa.

Le condizioni sufficienti possono essere dimostrate anche senza ricorrere al differenziale del secondo ordine (1.4.21). A questo scopo scriviamo la devianza $S(\mathbf{b})$ nel modo seguente

$$\begin{aligned} S(\mathbf{b}) &= \sum_{i=1}^n u_i^2 = \mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \\ &= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} + \mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} + \mathbf{X}\hat{\mathbf{b}} - \mathbf{X}\mathbf{b}) = \\ &= [(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - \mathbf{X}(\mathbf{b} - \hat{\mathbf{b}})]'[(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - \mathbf{X}(\mathbf{b} - \hat{\mathbf{b}})] = \\ &= (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) + (\mathbf{b} - \hat{\mathbf{b}})' \mathbf{X}'\mathbf{X}(\mathbf{b} - \hat{\mathbf{b}}) \end{aligned} \quad (1.4.23)$$

dove si è sfruttato il fatto che i prodotti incrociati sono nulli per la proprietà di ortogonalità esposta nella (1.4.16)⁸

$$(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})' \mathbf{X}(\mathbf{b} - \hat{\mathbf{b}}) = \hat{\mathbf{u}}' \mathbf{X}(\mathbf{b} - \hat{\mathbf{b}}) = 0$$

da cui anche, trasponendo,

$$0 = (\mathbf{b} - \hat{\mathbf{b}})' \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

Sfruttando il corollario XIX-1.1 del par. XIX-1.8, dato che $\det(\mathbf{X}'\mathbf{X}) \neq 0$ la *forma quadratica* in $(\mathbf{b} - \hat{\mathbf{b}})$ data dalla $(\mathbf{b} - \hat{\mathbf{b}})' \mathbf{X}'\mathbf{X}(\mathbf{b} - \hat{\mathbf{b}})$ nella (1.4.23) sarà maggiore di zero se $(\mathbf{b} - \hat{\mathbf{b}}) \neq \mathbf{0}$. Allora, poiché $\hat{\mathbf{b}}$ è un vettore di costanti, il minimo di $S(\mathbf{b})$ viene ottenuto se si prende $\mathbf{b} = \hat{\mathbf{b}}$ dato dalla (1.4.10), che quindi costituisce il vettore delle stime dei minimi quadrati dei parametri del modello (1.4.4).

Osservazione 1.12 - La condizione sufficiente consistente nella definitezza positiva della matrice $\mathbf{X}'\mathbf{X}$ corrisponde al fatto che $(\mathbf{b} - \hat{\mathbf{b}})' \mathbf{X}'\mathbf{X}(\mathbf{b} - \hat{\mathbf{b}}) > 0$ per un qualsiasi $\mathbf{b} - \hat{\mathbf{b}} \neq \mathbf{0}$, cioè alla disuguaglianza utilizzata per mostrare che la devianza $S(\mathbf{b})$ è minima nella (1.4.23) quando $\mathbf{b} = \hat{\mathbf{b}}$.

⁸ Nelle due uguaglianze seguenti sono utilizzate le proprietà dell'operazione di trasposizione di una matrice, per le quali si veda il paragrafo XIX-1.3.

1.5 La scomposizione della devianza ed il coefficiente di determinazione

Esiste una serie di indici che permettono di misurare la capacità del modello lineare (1.3.4) di adattarsi ai dati del campione. Per definire il primo di questi indici supponiamo, ovviamente senza perdere in generalità, che il modello contenga l'intercetta (che, stimata, può anche valere zero) e scomponiamo la devianza delle y_t nel seguente modo

$$\begin{aligned} \sum_{t=1}^n (y_t - \bar{y})^2 &= \sum_{t=1}^n (y_t - \hat{y}_t + \hat{y}_t - \bar{y})^2 = \\ &= \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 + 2 \sum_{t=1}^n (y_t - \hat{y}_t)(\hat{y}_t - \bar{y}) \end{aligned} \quad (1.5.1)$$

dove $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ come nelle (1.3.8). Il termine misto è nullo poiché

$$\begin{aligned} \sum_{t=1}^n (y_t - \hat{y}_t)(\hat{y}_t - \bar{y}) &= \sum_{t=1}^n \hat{u}_t \hat{y}_t - \bar{y} \sum_{t=1}^n \hat{u}_t = \sum_{t=1}^n \hat{u}_t \left(\sum_{i=1}^k \hat{\beta}_i x_{it} \right) - \bar{y} \sum_{t=1}^n \hat{u}_t = \\ &= \sum_{i=1}^k \hat{\beta}_i \sum_{t=1}^n x_{it} \hat{u}_t - \bar{y} \sum_{t=1}^n \hat{u}_t = 0 \end{aligned}$$

avendo applicato le (1.4.18) e (1.4.20), per cui vale la scomposizione della *devianza (totale)* di y_t nella *devianza di regressione* ed in quella *residua*⁹, essendo \bar{y} per la (1.4.19) il valor medio campionario sia di y_t che di \hat{y}_t ,

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t (y_t - \hat{y}_t)^2 \quad (1.5.2)$$

Dev. totale Dev. di regress. Dev. residua

Se dividiamo i due membri della (1.5.2) e per la devianza totale otteniamo

$$1 = (\text{Dev. di regressione})/(\text{Dev. totale}) + (\text{Dev. residua})/(\text{Dev. totale})$$

per mezzo della quale definiamo il *coefficiente di determinazione*

$$R^2 = \frac{\text{Dev. di regressione}}{\text{Dev. totale}} = 1 - \frac{\text{Dev. residua}}{\text{Dev. totale}} \quad (1.5.3)$$

pari al quadrato del *coefficiente di correlazione multipla* tra y_t e l'insieme delle variabili esplicative.

⁹ In lingua inglese: Dev. totale = *Total Sum of Squares* (TSS); Dev. di regressione = *Explained Sum of Squares* (ESS); Dev. residua = *Residual Sum of Squares* (RSS).

Quando tutta la variabilità totale è spiegata da quella di regressione si ha che l'andamento del modello è perfetto, la devianza residua è nulla ed $R^2 = 1$; nel caso opposto la parte sistematica del modello non spiega niente e la variabilità totale coincide con quella residua, per cui $R^2 = 0$. In generale dunque, si ha

$$0 \leq R^2 \leq 1 \quad (1.5.4)$$

Il coefficiente di determinazione in termini matriciali

Dal punto di vista computazionale è talvolta utile calcolare il coefficiente R^2 per mezzo dei dati campionari e delle stime $\hat{\mathbf{b}}$; per arrivare a questo, osserviamo che la devianza totale è

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}'\mathbf{y} + \sum_{i=1}^n \bar{y}^2 - 2\bar{y}\sum_{i=1}^n y_i = \mathbf{y}'\mathbf{y} - n\bar{y}^2 \quad (1.5.5)$$

e quella residua

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\mathbf{b}} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \hat{\mathbf{b}}'(\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}) = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} \end{aligned} \quad (1.5.6)$$

dove si è nuovamente sfruttato il fatto che $\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}$ è uno scalare e quindi uguale al suo trasposto, e si è utilizzata la (1.4.12). Sostituendo queste relazioni nella (1.5.3) si ottiene

$$R^2 = 1 - \frac{\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} = \frac{\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} \quad (1.5.7)$$

in funzione del campione, di $\hat{\mathbf{b}}$ e di \bar{y} . Il coefficiente di determinazione (1.5.7) è detto *centrato*. Se si elimina $n\bar{y}^2$ si ottiene il coefficiente di determinazione *non centrato*:

$$R_u^2 = \frac{\hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}}{\mathbf{y}'\mathbf{y}} = \frac{\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y}} \quad (1.5.8)$$

dove il pedice u sta per *uncentered*, che significa appunto, in lingua inglese, non centrato.

Si noti che la (1.5.8) deriva dalla scomposizione (1.4.17) della *somma dei quadrati* della variabile dipendente, mentre la (1.5.7) deriva dalla scomposizione (1.5.2) della *devianza*, cioè della somma dei quadrati *degli scarti* della variabile dipendente dalla propria media campionaria. In altre parole, la differenza fra i due coefficienti risiede nel fatto che la (1.5.7) confronta due devianze, mentre la (1.5.8) due somme di quadrati. Ne consegue che il coefficiente non centrato (1.5.8) è

sempre positivo, mentre quello espresso dalla (1.5.7) può essere negativo se dal modello viene omessa l'intercetta, perché in tal caso l'ortogonalità (1.4.16) non implica più il risultato (1.4.18) e quindi il termine misto nella scomposizione (1.5.1) non si annulla.

Il coefficiente non centrato R_u^2 riveste particolare importanza nella costruzione di test diagnostici sul modello, come si vedrà in seguito nei moduli III e VIII. Viceversa, per quanto attiene alla valutazione della *bontà di adattamento*¹⁰ del modello ai dati, e quindi alla scelta delle variabili da considerare nel modello stesso, si fa riferimento al coefficiente centrato R^2 o alla sua versione corretta, che passiamo a descrivere.

Il coefficiente di determinazione corretto

Se si dividono per n le due devianze nella (1.5.3) si ottiene

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2} = 1 - \frac{\frac{1}{n} \sum_{t=1}^n \hat{u}_t^2}{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2} \quad (1.5.9)$$

che mostra chiaramente come l' R^2 misuri la proporzione di varianza totale spiegata dal modello di regressione. Tuttavia nella (1.5.9) si utilizzano gli stimatori costituiti dalle varianze campionarie, che sono distorti. Se a tali stimatori distorti si sostituiscono quelli non distorti si ottiene un coefficiente di determinazione leggermente diverso dall'(1.5.9), detto *corretto* rispetto ai gradi di libertà,¹¹

$$R_c^2 = 1 - \frac{\frac{1}{n-k} \sum_{t=1}^n \hat{u}_t^2}{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2} \quad (1.5.10)$$

Siamo così passati dal rapporto fra somme di quadrati (1.5.8) e dal rapporto fra devianze (1.5.7) al rapporto fra varianze (1.5.10), nell'ultimo dei quali si tiene esplicito conto del numero di variabili esplicative k .

Se, dato un modello, gli si aggiunge una variabile esplicativa qualsiasi, assolutamente non significativa, cioè non legata da alcuna effettiva relazione con la variabile dipendente, l' R^2 comunque aumenterà. Al limite, inserendo nel modello n variabili esplicative (cioè tante quante sono le osservazioni disponibili) si otterrà un

¹⁰ In lingua inglese: *goodness of fit*.

¹¹ Il coefficiente corretto è stato introdotto da Theil [1961].

adattamento perfetto ai dati ($R^2 = 1$), in conseguenza del fatto che una nuvola di n punti può essere interpolata esattamente da un iperpiano a n dimensioni. L' R_c^2 invece diminuisce, poiché a parità di devianze è $R_c^2 < R^2$ come si può ricavare comparando la (1.5.9) con la (1.5.10). In questa maniera il confronto tra due modelli con un diverso numero di variabili esplicative, effettuato ricercando quale dei due possiede un coefficiente di determinazione maggiore, diventa più significativo in quanto al modello con k più grande si attribuisce uno svantaggio, funzione appunto della sua maggiore dimensione. Talvolta R_c^2 è indicato mediante una soprallineatura: \overline{R}^2 .

La relazione esistente tra R^2 ed R_c^2 è presto trovata

$$R_c^2 = 1 - \frac{n-1}{n-k}(1-R^2) = \frac{1-k}{n-k} + \frac{n-1}{n-k}R^2 \quad (1.5.11)$$

La (1.5.11) mostra che quando k si avvicina molto a n il coefficiente corretto \overline{R}^2 diventa negativo tendendo a meno infinito.

Si noti che nonostante questa penalizzazione possa apparire molto severa, in realtà è possibile dimostrare che anche il coefficiente R_c^2 può aumentare (anche se non aumenta necessariamente) quando al modello vengono aggiunte variabili irrilevanti.¹² Di conseguenza le misure di bontà dell'interpolazione, anche se costituiscono un utile indicatore sintetico della bontà complessiva del modello, non possono essere considerate come unica guida nella strategia di specificazione econometrica.

Il coefficiente di determinazione per il modello con le variabili scarto

Il coefficiente di determinazione definito dalla (1.5.3), o, in modo equivalente, dalla (1.5.7), può essere calcolato anche attraverso le variabili espresse come scarti dai rispettivi valori medi. Se nell'equazione (1.3.4) si pone $x_{kt} \equiv 1$, per cui il coefficiente β_k diventa l'intercetta della nostra funzione di regressione, e si stimano i parametri coi minimi quadrati, sommando i due membri dell'equazione per tutti i tempi e dividendo per n si ottiene

$$\bar{y} = \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \dots + \hat{\beta}_{k-1} \bar{x}_{k-1} + \hat{\beta}_k + \frac{1}{n} \sum_{t=1}^n \hat{u}_t \quad (1.5.12)$$

dove $\hat{\beta}_k$ è la stima del termine noto e le soprallineature denotano i valori medi campionari delle variabili; considerando poi che per la (1.4.18) la somma dei residui

¹² Una descrizione più rigorosa di questo fenomeno verrà fornita nel paragrafo 2.3 dopo aver introdotto i test di significatività dei coefficienti.

stimati è nulla, facendo la differenza tra la (1.3.4) stimata e la (1.5.12) si ottiene il modello degli scarti rispetto ai valori medi

$$y_t - \bar{y} = \hat{\beta}_1(x_{1t} - \bar{x}_1) + \hat{\beta}_2(x_{2t} - \bar{x}_2) + \dots + \hat{\beta}_{k-1}(x_{k-1t} - \bar{x}_{k-1}) + \hat{u}_t \quad (1.5.13)$$

nella quale il termine noto non è più presente. Il coefficiente di determinazione per la (1.5.13) è allora calcolato semplicemente applicando la (1.5.7) all'equazione (1.5.13): questo può essere effettuato partizionando la matrice \mathbf{X} originale in modo da isolare l'ultima colonna formata da tutti uno ed indicata con \mathbf{i} ¹³

$$\mathbf{X} = [\mathbf{X}_1 | \mathbf{i}]$$

considerando il nuovo vettore dei parametri $\mathbf{b}_1 = [\beta_1 \ \beta_2 \ \dots \ \beta_k]'$, ed effettuando le seguenti sostituzioni, derivate dall'uso del modello (1.5.12):

al posto di	si considera
y_t	$y_t - \bar{y} = y_t - \mathbf{i}'\mathbf{y} / n$
\mathbf{y}	$\mathbf{y} - \mathbf{i}\bar{y} = \mathbf{y} - \mathbf{ii}'\mathbf{y} / n = \mathbf{C}\mathbf{y}$
$\mathbf{y}'\mathbf{y}$	$\mathbf{y}'\mathbf{C}\mathbf{y}$
\mathbf{x}_i	$\mathbf{x}_i - \mathbf{i}\bar{x}_i = \mathbf{x}_i - \mathbf{ii}'\mathbf{x}_i / n = \mathbf{C}\mathbf{x}_i$
\mathbf{X}	$\mathbf{C}\mathbf{X}_1$
\mathbf{X}'	$\mathbf{X}_1'\mathbf{C}$
\bar{y}	0

dove si è utilizzata la matrice di centraggio \mathbf{C} definita dalla XIX-(1.10.5) come

$$\mathbf{C} = \mathbf{I}_n - \frac{1}{n} \mathbf{ii}'$$

e le sue proprietà di simmetria e idempotenza dimostrate dalla (XIX-1.10.6). Si ha allora

$$R^2 = \frac{\hat{\mathbf{b}}_1' \mathbf{X}_1' \mathbf{C} \mathbf{y}}{\mathbf{y}' \mathbf{C} \mathbf{y}} \quad (1.5.14)$$

dove si è ancora fatto uso dell'idempotenza di \mathbf{C} e $\hat{\mathbf{b}}_1 = [\hat{\beta}_1 \ \hat{\beta}_2 \ \dots \ \hat{\beta}_{k-1}]'$.

Osservazione 1.13 - Dalle (1.5.12) e (1.5.13) si trae che i residui stimati con il modello degli scarti sono uguali a quelli stimati con il modello (1.3.4) quando quest'ultimo contenga l'intercetta. In altre parole, l'introduzione dell'intercetta vale a depurare i residui del modello

¹³ La barra verticale nella (1.5.13) indica semplicemente che la \mathbf{X} è partizionata in \mathbf{X}_1 e \mathbf{i} .

dall'influsso delle medie delle variabili, qualora queste siano diverse da zero (come sono, invece, per costruzione se il modello è costruito con variabili scarto). D'altra parte, abbiamo visto che la presenza dell'intercetta assicura il risultato (1.4.18). Segue che anche le devianza residue dei due modelli sono le stesse.

Per l'osservazione 1.13 il coefficiente di determinazione centrato è anche esprimibile tramite l'espressione seguente

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{C}\mathbf{y}}$$

dalla quale è possibile derivare il corrispondente coefficiente di determinazione centrato corretto per mezzo della correzione rispetto ai gradi di libertà analoga a quella effettuata nella (1.5.10).

Scarsa significatività economica si ha quando nell'equazione di regressione (1.3.4) sussiste una forte tendenza sia nella endogena che in una o più variabili esplicative. In questo caso un R^2 alto può significare semplicemente un buon adattamento della tendenza di y_t a quella delle esplicative, e non necessariamente una buona capacità esplicativa della componente sistematica del modello a prescindere dall'andamento tendenziale.

1.6 I residui come enti aleatori: le ipotesi deboli

Finora i residui u_t sono stati considerati come scarti tra i valori osservati e quelli teorici di una variabile y_t , per ogni tempo t : in sostanza, come entità deterministiche senza collegamenti intertemporali. È ora opportuno estendere questa connotazione, supponendo che tali residui costituiscano realizzazioni di variabili aleatorie \tilde{u}_t ,¹⁴ $t = 1, 2, \dots, n$ dotate di proprietà stocastiche (ipotesi) che mutano a seconda del grado di approfondimento con cui si vuole studiare il modello (1.3.4) oppure della diversa conformazione dei dati campionari.

Supponendo sempre che l'equazione (1.3.4) rimanga inalterata nel periodo campionario, l'insieme più semplice di ipotesi stocastiche che possono essere formulate rispetto ad essa è dato da

$$\begin{aligned}
 (i) \quad & x_{it} \quad \text{valori noti } \forall i, t \\
 (ii) \quad & E(\tilde{u}_t) = 0 \quad \forall t \\
 (iii) \quad & E(\tilde{u}_t \tilde{u}_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}
 \end{aligned} \tag{1.6.1}$$

La prima ipotesi indica che le variabili esplicative x_i sono conosciute. In particolare, quindi, essa comporta che le x_{it} , a differenza delle y_t , siano misurate senza errori.

La seconda ipotesi non è affatto restrittiva in quanto se fosse $E(\tilde{u}_t) = \mu \neq 0, \forall t$, ci si potrebbe sempre ricondurre al caso di valor medio nullo semplicemente aggiungendo μ al termine noto dell'equazione (1.3.4). L'osservazione 1.13 ci ricorda che l'introduzione dell'intercetta garantisce che i residui stimati abbiano media campionaria nulla, proprietà che è appunto il corrispettivo campionario della seconda delle (1.6.1).

La terza ipotesi delle (1.6.1) è, viceversa, restrittiva in quanto presuppone sia che i residui siano non correlati tra di loro quando sono associati a tempi diversi sia che abbiano tutti la stessa varianza σ^2 . Ambedue queste sottoipotesi sono raramente verificate nella realtà, ma sono molto utili nell'introduzione didattica della (1.3.4) in ambiente stocastico.

Le ipotesi (1.6.1)-(ii) e (iii) vengono talora sintetizzate dicendo che il residuo della (1.3.4) è un *rumore bianco*, dove per rumore bianco si intende appunto una

¹⁴ Indichiamo con una tilde una variabile aleatoria. Tale simbolo è utilizzato soltanto quando la variabile è considerata in un contesto dichiaratamente stocastico (ad esempio sotto il simbolo di valor medio E). In contesti più generali (ad esempio in un modello) è solitamente omissivo.

successione temporale di variabili aleatorie incorrelate con valor medio nullo e varianza costante.

Si noti anche che le prime due delle (1.6.1) implicano che sia $E(\tilde{u}_t x_{it}) = 0 \quad \forall i, t$, poiché $E(\tilde{u}_t x_{it}) = x_{it} E(\tilde{u}_t) = 0$.¹⁵

L'ipotesi che alcune variabili aleatorie abbiano la stessa varianza è detta di *omoschedasticità*¹⁶, mentre quella alternativa di varianze diverse è chiamata di *eteroschedasticità*.

Poiché le ipotesi (1.6.1) non presuppongono alcuna forma per la distribuzione delle \tilde{u}_t , sono dette *deboli*; nel caso contrario, che esamineremo nel prossimo capitolo, di assunzione di una specifica distribuzione, si ipotizzeranno *ipotesi forti* per \tilde{u}_t .

L'immersione del modello (1.3.4) nell'ambiente stocastico produce come risultato che \tilde{y}_t deve essere considerato come una variabile aleatoria. Si è detto, infatti, in precedenza, che le caratteristiche del membro a destra nell'equazione (1.3.4) devono essere rispecchiate in quello a sinistra: se la parte di destra è stocastica (a causa di \tilde{u}_t), così deve essere quella a sinistra, per cui l'equazione (1.3.4) diventa

$$\tilde{y}_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \tilde{u}_t \quad (1.6.2)$$

che indica chiaramente come la variabile endogena \tilde{y}_t sia rappresentata da un modello scisso in una *componente sistematica* data dalla combinazione lineare $\sum_{i=1}^k \beta_i x_{it}$ delle esplicative, ed in una *componente aleatoria* formata dal residuo \tilde{u}_t .

La prima componente è detta sistematica in quanto rappresenta la struttura di y_t in funzione dei parametri considerati, invariabili nel tempo in virtù dell'omogeneità (nel senso illustrato nel paragrafo 1.2) del campione, e delle esplicative, supposte note per la prima delle (1.6.1). La componente sistematica quindi non contiene alcun elemento aleatorio.

Questa considerazione è importante anche perché mette in luce che le ipotesi stocastiche (1.6.1), che per motivi didattici e storici vengono spesso espresse in termini dei residui non osservabili \tilde{u}_t , in effetti possono essere viste come ipotesi

¹⁵ Si tratta di una conseguenza della proprietà di linearità dell'operatore di valor medio o speranza matematica, definita dalla XXI-1.3.6

¹⁶ Dai termini greci ομοιοσ, uguale, e σκεδασσ, dispersione. Una definizione più rigorosa di omoschedasticità richiede l'impiego delle distribuzioni di probabilità condizionate ed è fornita nel par. XXI-2.4

sulle variabili osservabili y_t , considerate come realizzazioni di una variabile aleatoria \tilde{y}_t . In particolare, dalle (1.6.1) scaturiscono le seguenti ipotesi per \tilde{y}_t

$$E(\tilde{y}_t) = E\left(\sum_j \beta_j x_{jt} + \tilde{u}_t\right) = \sum_j \beta_j x_{jt}$$

$$V(\tilde{y}_t) = E[(\tilde{y}_t - E(\tilde{y}_t))^2] = E(\tilde{u}_t^2) = \sigma^2$$

$$\text{Cov}(\tilde{y}_t, \tilde{y}_s) = E[(\tilde{y}_t - E(\tilde{y}_t))(\tilde{y}_s - E(\tilde{y}_s))] = E(\tilde{u}_t \tilde{u}_s) = 0 \quad \forall t \neq s$$

In altre parole, la struttura di covarianza ipotizzata per la \tilde{u}_t si applica ugualmente alla \tilde{y}_t , il che, visto che le due variabili aleatorie differiscono per una costante additiva (la parte sistematica del modello), discende immediatamente da note proprietà della varianza.¹⁷ Questo risultato non ha un mero interesse teorico, ma ha anche una rilevante importanza pratica perché ci permette di valutare immediatamente, almeno in modo informale, la plausibilità delle ipotesi (1.6.1).

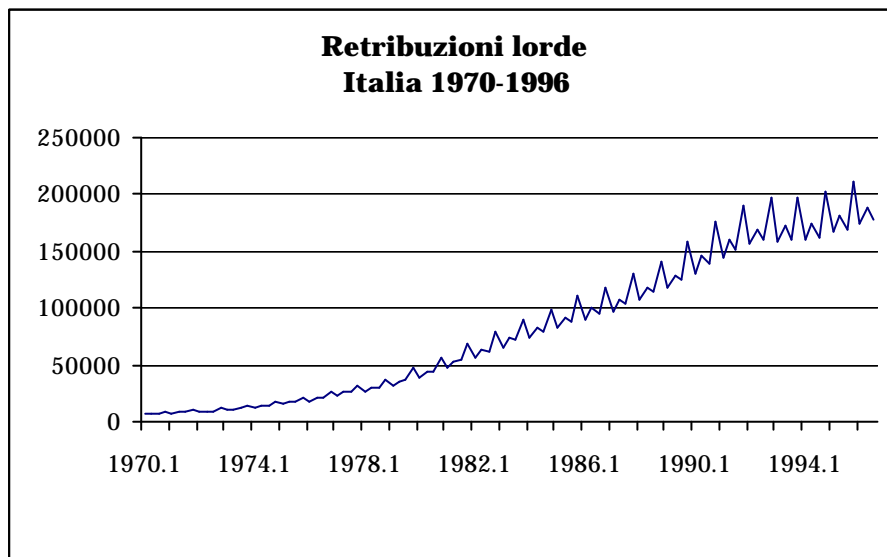


Figura 1.2 Le retribuzioni lorde complessive in Italia, dati trimestrali grezzi dal 1970 al 1996 (fonte ISTAT).

Ad esempio, se la nostra variabile dipendente avesse un andamento analogo a quello della variabile rappresentata nella figura 1.2 (le retribuzioni lorde in Italia dal 1970 al 1996) l'ipotesi di costanza delle varianze nel tempo sarebbe poco

¹⁷ Una costante additiva può essere eliminata dall'operatore di varianza: $V(a + \tilde{x}) = V(\tilde{x})$, dato che la traslazione determinata dall'aggiunta della costante influenza la locazione (e quindi la media) ma non la dispersione (e quindi la varianza) della \tilde{x} .

plausibile, visto che la variabile in questione manifesta variabilità crescente col passare del tempo. Quindi se fra le esplicative non figurano variabili caratterizzate anch'esse da una simile variabilità crescente, questa si scaricherà sul residuo dell'equazione, determinando la violazione delle ipotesi stocastiche deboli.

Lo stimatore dei minimi quadrati per il modello lineare semplice

Un secondo risultato dell'approccio stocastico riguarda le stime dei minimi quadrati dei parametri che, nel caso della regressione semplice, possono essere ancora calcolate mediante le (1.3.10) e (1.3.11) ma che se sono interpretate in termini aleatori, in funzione delle y definite dalla

$$\tilde{y}_t = \beta_1 + \beta_2 x_t + \tilde{u}_t \quad (1.6.3)$$

diventano¹⁸

$$\begin{aligned} \hat{\beta}_2 &= \frac{1}{m_{xx} - \bar{x}^2} (m_{xy} - \bar{y}\bar{x}) = \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) y_t = \\ &= \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) (\beta_1 + \beta_2 x_t + \tilde{u}_t) = \\ &= \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) (\beta_1 + \beta_2 x_t) + \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) \tilde{u}_t = \\ &= \beta_2 + \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) \tilde{u}_t \end{aligned} \quad (1.6.4)$$

$$\begin{aligned} \hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x} = \frac{1}{n} \sum_{t=1}^n (\beta_1 + \beta_2 x_t + \tilde{u}_t) - \hat{\beta}_2 \bar{x} = \\ &= \beta_1 + \beta_2 \bar{x} - \hat{\beta}_2 \bar{x} + \frac{1}{n} \sum_{t=1}^n \tilde{u}_t = \beta_1 + \bar{x} (\beta_2 - \hat{\beta}_2) + \frac{1}{n} \sum_{t=1}^n \tilde{u}_t \end{aligned} \quad (1.6.5)$$

Le $\hat{\beta}_1$ e $\hat{\beta}_2$ in ambito stocastico sono dunque *stimatori* per β_1 e β_2 , rispettivamente. I loro valori medi sono, in virtù delle ipotesi (stocastiche) deboli,

$$E(\hat{\beta}_2) = E \left[\beta_2 + \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) \tilde{u}_t \right] = \beta_2 \quad (1.6.6)$$

¹⁸ Le $\hat{\beta}_1$ e $\hat{\beta}_2$ seguenti sono variabili aleatorie e, seguendo la nostra convenzione, dovrebbero essere indicate con una tilde, sovrapposta al cappello. Per semplicità di notazione omettiamo la tilde, per cui $\hat{\beta}_1$ e $\hat{\beta}_2$ possono indicare, in funzione del contesto, sia le stime (1.3.10) e (1.3.11) sia le variabili aleatorie (stimatori) (1.6.4) e (1.6.5). Così nel seguito per quanto riguarda le generiche $\hat{\beta}_j$.

$$E(\hat{\beta}_1) = \beta_1 + \bar{x}E(\beta_2 - \hat{\beta}_2) + \frac{1}{n} \sum_{t=1}^n E(\tilde{u}_t) = \beta_1 \quad (1.6.7)$$

per cui gli stimatori $\hat{\beta}_1$ e $\hat{\beta}_2$ sono *non distorti*.¹⁹

Calcoliamo ora le loro varianze:

$$\begin{aligned} Var(\hat{\beta}_2) &= E[(\hat{\beta}_2 - \beta_2)^2] = E\left[\frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) \tilde{u}_t\right]^2 = \\ &= \frac{\sigma^2 \sum_{t=1}^n (x_t - \bar{x})^2}{n^2 (m_{xx} - \bar{x}^2)^2} = \frac{\sigma^2}{n} \frac{1}{m_{xx} - \bar{x}^2} \end{aligned} \quad (1.6.8)$$

avendo utilizzato la relazione

$$\sum_{t=1}^n (x_t - \bar{x})^2 = n(m_{xx} - \bar{x}^2)$$

Inoltre

$$\begin{aligned} Var(\hat{\beta}_1) &= E[(\hat{\beta}_1 - \beta_1)^2] = E\left\{\left[\bar{x}(\beta_2 - \hat{\beta}_2) + \frac{1}{n} \sum_{t=1}^n \tilde{u}_t\right]^2\right\} = \\ &= E\left\{\bar{x}^2(\beta_2 - \hat{\beta}_2)^2 + \frac{1}{n^2} \left(\sum_{t=1}^n \tilde{u}_t\right)^2 + \frac{2}{n} \bar{x}(\beta_2 - \hat{\beta}_2) \sum_{t=1}^n \tilde{u}_t\right\} = \\ &= \bar{x}^2 E[(\beta_2 - \hat{\beta}_2)^2] + \frac{1}{n^2} E\left[\left(\sum_{t=1}^n \tilde{u}_t\right)^2\right] + \frac{2\bar{x}}{n} E(\beta_2 - \hat{\beta}_2) \times E\left(\sum_{t=1}^n \tilde{u}_t\right) = \\ &= \frac{\sigma^2}{n} \frac{\bar{x}^2}{m_{xx} - \bar{x}^2} + \frac{\sigma^2}{n} = \frac{\sigma^2}{n} \left[1 + \frac{\bar{x}^2}{m_{xx} - \bar{x}^2}\right] \end{aligned} \quad (1.6.9)$$

dove nel quarto passaggio si è utilizzata la non correlazione dei due fattori $(\beta_2 - \hat{\beta}_2)$ e $\sum_{t=1}^n \tilde{u}_t$. Infatti la loro covarianza è nulla

¹⁹ Uno stimatore $\tilde{\theta}$ del parametro θ è detto *non distorto* se $E(\tilde{\theta}) = \theta$. La non distorsione è una buona proprietà di uno stimatore se, come avviene di solito, la distribuzione di probabilità di $\tilde{\theta}$ è concentrata intorno al suo valor medio. In questo caso le stime, realizzazioni dello stimatore, hanno alta probabilità di trovarsi vicino al valore θ del parametro.

$$\begin{aligned}
 \text{Cov}[(\beta_2 - \hat{\beta}_2), \sum_{t=1}^n \tilde{u}_t] &= E[(\beta_2 - \hat{\beta}_2) \cdot \sum_{t=1}^n \tilde{u}_t] = E\left\{ \left[\frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x}) \tilde{u}_t \right] \sum_{t=1}^n \tilde{u}_t \right\} = \\
 &= \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \left\{ E\left[\left(\sum_{t=1}^n x_t \tilde{u}_t \right) \sum_{t=1}^n \tilde{u}_t \right] - \bar{x} E\left[\sum_{t=1}^n \tilde{u}_t \right] \right\} = \\
 &= \frac{1}{m_{xx} - \bar{x}^2} \frac{1}{n} \left\{ \sigma^2 \sum_{t=1}^n x_t - n\sigma^2 \bar{x} \right\}
 \end{aligned}$$

dove nel quarto passaggio è stata utilizzata la non correlazione delle \tilde{u}_t per indici diversi.

Lo stimatore dei minimi quadrati per il modello lineare multiplo

Se le equazioni (1.3.4) sono sintetizzate nella forma matriciale (1.4.4) è conveniente riscrivere le ipotesi (1.6.1) nel modo seguente

$$\begin{cases} \mathbf{X} & \text{matrice di costanti} \\ E(\tilde{\mathbf{u}}) &= \mathbf{0} \\ \text{Cov}(\tilde{\mathbf{u}}) &= E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}') = \sigma^2 \mathbf{I}_n \end{cases} \quad (1.6.10)$$

con $\tilde{\mathbf{u}} = [\tilde{u}_1 \quad \tilde{u}_2 \quad \dots \quad \tilde{u}_n]'$ e dove la $E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}')$ indica la *matrice di dispersione* (o di *covarianza*, o di *varianze e covarianze*) del vettore di residui $\tilde{\mathbf{u}}$, essendo così formata

$$E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}') = \begin{bmatrix} E(\tilde{u}_1^2) & E(\tilde{u}_1\tilde{u}_2) & \dots & E(\tilde{u}_1\tilde{u}_n) \\ E(\tilde{u}_2\tilde{u}_1) & E(\tilde{u}_2^2) & \dots & E(\tilde{u}_2\tilde{u}_n) \\ \dots & \dots & \dots & \dots \\ E(\tilde{u}_n\tilde{u}_1) & E(\tilde{u}_n\tilde{u}_2) & \dots & E(\tilde{u}_n^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n \quad (1.6.11)$$

Osservazione 1.14 - Gli elementi della diagonale principale della matrice di dispersione costituiscono le varianze delle variabili aleatorie del vettore $\tilde{\mathbf{u}}$; gli elementi fuori di tale diagonale ne formano le covarianze. La scrittura delle varianze e delle covarianze nella (1.6.11) tiene conto del fatto che i valori medi delle \tilde{u}_t sono nulli.

Osservazione 1.15 - La matrice di dispersione (1.6.11) può essere indicata mediante il suo elemento generico

$$\{E(\tilde{u}_i\tilde{u}_j)\} \quad (1.6.12)$$

e poiché $E(\tilde{u}_i \tilde{u}_j) = E(\tilde{u}_j \tilde{u}_i)$, essa è simmetrica. Questa proprietà di simmetria vale per qualsiasi matrice di dispersione (si veda la XXI-2.3.6).

Le prime due ipotesi implicano che sia $E(\mathbf{X}'\tilde{\mathbf{u}}) = \mathbf{X}'E(\tilde{\mathbf{u}}) = \mathbf{0}$. In altre parole, le ipotesi deboli implicano che le colonne della matrice \mathbf{X} siano ortogonali²⁰ rispetto ai residui $\tilde{\mathbf{u}}$. Si noti che questa condizione equivale di fatto a imporre sui momenti della popolazione una condizione che abbiamo già visto essere verificata dai momenti campionari (si veda la (1.4.15)).

In termini stocastici il modello matriciale (1.4.4) è scritto

$$\tilde{\mathbf{y}} = \mathbf{X}\mathbf{b} + \tilde{\mathbf{u}} \quad (1.6.13)$$

dove $\tilde{\mathbf{y}} = [\tilde{y}_1 \quad \tilde{y}_2 \quad \dots \quad \tilde{y}_n]'$, con valor medio vettoriale

$$E(\tilde{\mathbf{y}}) = \mathbf{X}\mathbf{b} \quad (1.6.14)$$

e matrice di dispersione

$$Cov(\tilde{\mathbf{y}}) = Cov(\tilde{\mathbf{y}} - \mathbf{X}\mathbf{b}) = Cov(\tilde{\mathbf{u}}) = \sigma^2 \mathbf{I}_n \quad (1.6.15)$$

La stima dei minimi quadrati $\hat{\mathbf{b}}$ può ancora essere calcolata mediante la (1.4.10) ma se è interpretata in termini aleatori, in funzione della (1.6.13), diventa

$$\begin{aligned} \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\tilde{\mathbf{y}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\mathbf{b} + \tilde{\mathbf{u}}) = \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\tilde{\mathbf{u}} = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\tilde{\mathbf{u}} \end{aligned} \quad (1.6.16)$$

e rappresenta quindi uno *stimatore* (quello dei minimi quadrati) di \mathbf{b} .

Osservazione 1.16 - Evidenziamo che la validità dello stimatore (1.6.16) è subordinata all'assunzione congiunta delle ipotesi deboli (1.6.10) e delle altre (1.4.11).

Osservazione 1.17 - Rimarchiamo la differenza (di interpretazione) tra la stima $\hat{\mathbf{b}}$ (1.4.10), funzione delle variabili osservate (\mathbf{y} , \mathbf{X}), e lo stimatore $\hat{\mathbf{b}}$ (1.6.16), funzione del vettore aleatorio $\tilde{\mathbf{y}}$ oltre che delle \mathbf{X} .

Dalla (1.6.16) si ricava il valor medio vettoriale dello stimatore $\hat{\mathbf{b}}$ ²¹

$$E(\hat{\mathbf{b}}) = \mathbf{b} + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\tilde{\mathbf{u}}] = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E[\tilde{\mathbf{u}}] = \mathbf{b} \quad (1.6.17)$$

che indica che lo stimatore dei minimi quadrati è *non distorto* (o *corretto*).

²⁰ Ci riferiamo qui alla nozione di ortogonalità in senso stocastico definita dalla XXI-2.3.11).

²¹ I passaggi seguenti sfruttano la proprietà di linearità dell'operatore E discussa nel paragrafo XXI-2.3 (si veda la XXI-(2.3.13)).

Sempre tramite la (1.6.16) si calcola facilmente la matrice di dispersione di $\hat{\mathbf{b}}$

$$\begin{aligned} Cov(\hat{\mathbf{b}}) &= E[(\hat{\mathbf{b}} - \mathbf{b})(\hat{\mathbf{b}} - \mathbf{b})'] = E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\tilde{\mathbf{u}}\tilde{\mathbf{u}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\sigma^2 \mathbf{I}_n) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (1.6.18)$$

dove si è impiegata la terza delle ipotesi (1.6.10).

Osservazione 1.18 - La matrice di dispersione (1.6.18), tenendo anche conto dell'Osservazione 1.14 è formata nel modo seguente

$$\begin{aligned} Cov(\hat{\mathbf{b}}) &= E[(\hat{\mathbf{b}} - \mathbf{b})(\hat{\mathbf{b}} - \mathbf{b})'] = \\ &= \begin{bmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & Var(\hat{\beta}_k) \end{bmatrix} = \\ &= \begin{bmatrix} E[(\hat{\beta}_1 - \beta_1)^2] & E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)] & \dots & E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k)] \\ E[(\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1)] & E[(\hat{\beta}_2 - \beta_2)^2] & \dots & E[(\hat{\beta}_2 - \beta_2)(\hat{\beta}_k - \beta_k)] \\ \dots & \dots & \dots & \dots \\ E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_1 - \beta_1)] & E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_2 - \beta_2)] & \dots & E[(\hat{\beta}_k - \beta_k)^2] \end{bmatrix} \end{aligned}$$

dove si è sfruttato il risultato (1.6.17) per la notazione dei valori medi $E(\hat{\beta}_i)$.

1.7 La stima della varianza dei residui

Dalla (1.6.18) si nota che se la varianza dei residui è conosciuta, lo è anche la matrice di dispersione di $\hat{\mathbf{b}}$; altrimenti σ^2 deve essere stimata tramite uno stimatore che generalmente ha una delle due forme

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / n \quad (1.7.1)$$

$$\bar{\sigma}^2 = \frac{1}{n-k} \sum_{t=1}^n \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / (n-k) \quad (1.7.2)$$

che discendono in maniera “naturale” dalla definizione di varianza. Il primo di questi stimatori costituisce la *varianza campionaria* ed è *distorto*; il secondo non lo è, come si dimostra nel seguente teorema 1.1 che utilizza alcune proprietà della *traccia* di una matrice quadrata.

La radice quadrata della stima della varianza dei residui (1.7.2) è chiamata *errore standard dell'equazione* (o *della regressione*)²².

Per la piena comprensione delle (1.7.1) e (1.7.2) considerate come stimatori è necessario considerare che anche il vettore $\hat{\mathbf{u}}$ è aleatorio quando è espresso in funzione dello stimatore $\hat{\mathbf{b}}$ tramite la (1.4.13). Dunque anche le \hat{u}_t posseggono la doppia fisionomia di stime e di stimatori dei residui.

La distorsione della varianza campionaria

Vale, dunque, il seguente

Teorema 1.1 - Lo stimatore (1.7.2) è non distorto.

Infatti, in virtù delle (1.4.13) e (1.4.10) si ha

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\mathbf{b} + \mathbf{u}) = \\ &= \mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}\mathbf{b} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{u} = \mathbf{M}\mathbf{u} \end{aligned} \quad (1.7.3)$$

dove si è posto

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \quad (1.7.4)$$

matrice quadrata di ordine n . La \mathbf{M} è simmetrica, come il lettore verifica facilmente, e idempotente, cioè tale che $\mathbf{M}\mathbf{M} = \mathbf{M}$; infatti

²² In inglese: *Standard Error of the Equation* (SEE) oppure *Standard Error of the Regression* (SER).

$$\begin{aligned}
 \mathbf{M}\mathbf{M} &= [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \\
 &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \\
 &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M}
 \end{aligned} \tag{1.7.5}$$

(si veda anche il paragrafo XIX-1.11)

Allora, adoperando questa proprietà della \mathbf{M} e la (1.7.3) si ha

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u} \tag{1.7.6}$$

per cui, sfruttando i risultati (1.10.1)-(1.10.3) del modulo XIX,

$$\begin{aligned}
 E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) &= E(\tilde{\mathbf{u}}'\mathbf{M}\tilde{\mathbf{u}}) = E[\text{tr}(\tilde{\mathbf{u}}'\mathbf{M}\tilde{\mathbf{u}})] = E[\text{tr}(\mathbf{M}\tilde{\mathbf{u}}\tilde{\mathbf{u}}')] = \sigma^2 \text{tr}\mathbf{M} = \\
 &= \sigma^2 \text{tr}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \sigma^2 \{ \text{tr}\mathbf{I}_n - \text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \} = \\
 &= \sigma^2 \{ n - \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \} = \sigma^2 (n - \text{tr}\mathbf{I}_k) = \sigma^2 (n - k)
 \end{aligned} \tag{1.7.7}$$

essendo $\mathbf{X}'\mathbf{X}$ una matrice quadrata di ordine k . Dunque

$$E(\bar{\sigma}^2) = E(\hat{\mathbf{u}}'\hat{\mathbf{u}})/(n - k) = \sigma^2$$

e lo stimatore (1.7.2) è non distorto.

Il denominatore $n-k$ nella (1.7.2) definisce il *numero dei gradi di libertà*²³ del modello lineare (1.3.4).

Osservazione 1.19 - È utile studiare la dimostrazione del teorema 1.1 in primo luogo perché costituisce una semplice esercitazione di calcolo matriciale; inoltre è un esempio di come in molte dimostrazioni di statistica matematica si usi la tecnica di eseguire dapprima dei passaggi matriciali [la (1.7.3)] e di effettuare poi operazioni stocastiche che utilizzano le espressioni trovate [la (1.7.7)]. Questa stessa tecnica è stata usata più sopra per mostrare la correttezza dello stimatore $\hat{\mathbf{b}}$. Nel teorema 1.1 si definisce inoltre la matrice \mathbf{M} e si sfrutta l'operatore traccia, la proprietà di simmetria e quella definita dalla (1.7.5), chiamata *idempotenza*; sono definizioni e proprietà che ritroveremo di frequente nel seguito. Si noti, infine che $\hat{\mathbf{u}}$ dato dalla (1.7.3) è una *trasformazione lineare* di \mathbf{u} , tramite \mathbf{M} .

Per mezzo del teorema 1.1 si calcola facilmente la distorsione della varianza campionaria (1.7.1); infatti

²³ In lingua inglese: *number of Degrees of Freedom* (DF).

$$Dist(\hat{\sigma}^2) = \sigma^2 - E(\hat{\sigma}^2) = \sigma^2 - E\left(\frac{n-k}{n}\bar{\sigma}^2\right) = \frac{k}{n}\sigma^2 \quad (1.7.8)$$

che tende a zero per $n \rightarrow \infty$; lo stimatore varianza campionaria è detto allora *asintoticamente non distorto* e non si differenzia molto da quello non distorto (1.7.2) se n è grande rispetto a k .

Come già accennato, una stima della matrice di dispersione (1.6.18) di \hat{b} è ottenuta sostituendo al σ^2 una sua stima, ad esempio $\hat{\sigma}^2$ o $\bar{\sigma}^2$ date dalle (1.7.1) e (1.7.2), dalle quali si nota che le stime delle varianze e delle covarianze dei parametri di regressione sono, *ceteris paribus*, tanto meno disperse quanto più grandi sono i valori n o $n-k$. Di fondamentale importanza, quindi, per avere stime precise (con varianze piccole) e poco correlate tra loro (con covarianze piccole) è che n o la differenza $n-k$ siano sufficientemente grandi.

Questa indicazione, tuttavia, è di carattere statistico. Da un punto di vista economico, invece, l'ingrandimento di n , cioè dell'ampiezza del campione, può comportare la violazione del principio di omogeneità della struttura dell'economia del periodo campionario, necessaria affinché la componente sistematica del modello (1.6.10) possa rappresentare tale struttura in modo adeguato. Per determinare n (o $n-k$) è allora necessario trovare un compromesso tra un valore abbastanza grande per avere stime delle varianze e covarianze dei parametri β_i precise, ed uno sufficientemente piccolo in modo tale che la struttura dell'economia non si modifichi troppo nel periodo campionario.

1.8 Il teorema di Gauss-Markov e gli stimatori BLU

Si è detto nel paragrafo 1.6 che la non distorsione è una buona proprietà per gli stimatori; tra quelli lineari rispetto alle y_i lo stimatore $\hat{\mathbf{b}}$ dei minimi quadrati non soltanto gode di questa proprietà ma possiede variabilità minima nel senso del teorema di Gauss-Markov che enunciamo e dimostriamo nel prosieguo, facendo uso dell'espressione matriciale della varianza di una combinazione lineare con pesi $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_k]'$ delle variabili raccolte nel vettore aleatorio $\tilde{\mathbf{z}} = [\tilde{z}_1 \ \tilde{z}_2 \ \dots \ \tilde{z}_k]'$

$$Var(\mathbf{c}'\tilde{\mathbf{z}}) = Var\left(\sum_{i=1}^k c_i z_i\right) = \sum_{i=1}^k \sum_{j=1}^k c_i c_j Cov(\tilde{z}_i \tilde{z}_j) = \mathbf{c}'Cov(\tilde{\mathbf{z}})\mathbf{c} \quad (1.8.1)$$

La (1.8.1) discende dal teorema sui momenti di una combinazione lineare di variabili aleatorie enunciato dalla XXI-(2.3.13).

Vale dunque il

Teorema 1.2 (di Gauss-Markov) - Nella classe degli stimatori lineari rispetto alle \tilde{y}_i e non distorti, se $\hat{\mathbf{b}}$ è lo stimatore dei minimi quadrati definito dalla (1.6.16) e $\tilde{\mathbf{b}}$ è un qualsiasi altro stimatore, si ha

$$Var(\mathbf{c}'\hat{\mathbf{b}}) \leq Var(\mathbf{c}'\tilde{\mathbf{b}}) \quad (1.8.2)$$

dove $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_k]'$ è un qualsiasi vettore di costanti reali non tutte nulle.

Poiché $\tilde{\mathbf{b}}$ è uno stimatore lineare rispetto alle \tilde{y}_i la sua combinazione lineare $\mathbf{c}'\tilde{\mathbf{b}}$ può essere espressa come funzione lineare delle \tilde{y}_i mediante i pesi $\mathbf{h} = [h_1 \ h_2 \ \dots \ h_n]'$

$$\mathbf{c}'\tilde{\mathbf{b}} = \mathbf{h}'\tilde{\mathbf{y}} \quad (1.8.3)$$

per cui si ha, sfruttando la (1.6.14),

$$\mathbf{h}'\mathbf{X}\mathbf{b} = \mathbf{h}'E(\tilde{\mathbf{y}}) = E(\mathbf{h}'\tilde{\mathbf{y}}) = E(\mathbf{c}'\tilde{\mathbf{b}}) = \mathbf{c}'\mathbf{b} \quad (1.8.4)$$

dove nell'ultimo passaggio è stata sfruttata la non distorsione di $\tilde{\mathbf{b}}$. Segue che $\mathbf{h}'\mathbf{X} = \mathbf{c}'$, per cui, in virtù delle (1.8.2), (1.6.18) e (1.8.4)

$$Var(\mathbf{c}'\hat{\mathbf{b}}) = \mathbf{c}'Cov(\hat{\mathbf{b}})\mathbf{c} = \sigma^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c} = \sigma^2 \mathbf{h}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{h}$$

D'altro canto dalla (1.6.15), ed ancora considerando la (1.8.2), si ha

$$Var(\mathbf{c}'\tilde{\mathbf{b}}) = Var(\mathbf{h}'\tilde{\mathbf{y}}) = \mathbf{h}'Cov(\tilde{\mathbf{y}})\mathbf{h} = \sigma^2 \mathbf{h}'\mathbf{h}$$

per cui la tesi è dimostrata se si dimostra che

$$\begin{aligned} 0 \leq Var(\mathbf{c}'\tilde{\mathbf{b}}) - Var(\mathbf{c}'\hat{\mathbf{b}}) &= \sigma^2 [\mathbf{h}'\mathbf{h} - \mathbf{h}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{h}] = \\ &= \sigma^2 \mathbf{h}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{h} = \sigma^2 \mathbf{h}'\mathbf{M}\mathbf{h} \end{aligned} \quad (1.8.5)$$

dove \mathbf{M} è la matrice quadrata (1.7.4). Ma la matrice \mathbf{M} , essendo simmetrica e idempotente, è semidefinita positiva per il teorema XIX-1.12 e quindi vale la (1.8.5), essendo σ^2 sempre non negativo.

Gli stimatori a varianza minima nel senso del teorema di Gauss-Markov sono detti *ottimi*; sinteticamente essi sono chiamati BLU, dalle iniziali dei termini inglesi *Best* (ottimi), *Linear* (lineari), *Unbiased* (non distorti).

L'uso della stima dei minimi quadrati è stato in precedenza giustificato sulla base dell'interpretazione del relativo criterio, fornita nel paragrafo 1.2. Da un punto di vista stocastico, l'uso dello stimatore (e quindi della stima) dei minimi quadrati è motivato proprio dal fatto di essere BLU.

Osservazione 1.20 - Il teorema di Gauss-Markov può essere dimostrato anche sotto ipotesi meno restrittive di quelle sopra utilizzate, in altre parole le (1.4.8) e (1.6.10). Ad esempio, se si suppone che il rango di \mathbf{X} sia inferiore a k non esiste lo stimatore unico (1.6.16) dei minimi quadrati, ma il teorema di Gauss-Markov può ancora essere dimostrato in virtù delle sole equazioni normali (1.4.8) che, ovviamente, continuano a sussistere.

1.9 La matrice di correlazione degli stimatori dei parametri di regressione

Nelle applicazioni, piuttosto della matrice di dispersione (1.6.18), è più utile considerare la matrice delle correlazioni di $\hat{\mathbf{b}}$, che indica come gli stimatori dei parametri di regressione siano correlati tra di loro. Questo perché le covarianze, così come le varianze, risentono dell'unità di misura delle variabili ed è quindi difficile valutare se una covarianza sia rilevante o trascurabile. I coefficienti di correlazione sono invece normalizzati fra -1 e 1 per cui l'individuazione di quelli più elevati (cioè più prossimi in valore assoluto all'unità) è immediata.

Tale matrice è data dalla

$$\text{Corr}(\hat{\mathbf{b}}) = \sigma^{-1} \mathbf{D}^{-1} \text{Cov}(\hat{\mathbf{b}}) \mathbf{D}^{-1} \sigma^{-1} = \sigma^{-1} \mathbf{D}^{-1} [\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}] \mathbf{D}^{-1} \sigma^{-1} = \mathbf{D}^{-1} [(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{D}^{-1} \quad (1.9.1)$$

dove \mathbf{D} è la matrice diagonale i cui elementi nonnulli a_{ii} sono le radici quadrate aritmetiche degli elementi diagonali di $(\mathbf{X}'\mathbf{X})^{-1}$.

Esempio 1.9 - Supponiamo che sia, con $k=2$,

$$\text{Cov}(\hat{\mathbf{b}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{bmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{bmatrix}$$

per cui

$$\sigma \mathbf{D} = \sigma \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$$

Poiché l'inversa di una matrice diagonale \mathbf{A} è una matrice che ha per elementi diagonali gli inversi degli elementi diagonali di \mathbf{A} , si ha

$$\begin{aligned} \text{Corr}(\hat{\mathbf{b}}) &= \sigma^{-1} \begin{bmatrix} 1/a_{11} & 0 \\ 0 & 1/a_{22} \end{bmatrix} \sigma^2 \begin{bmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{bmatrix} \begin{bmatrix} 1/a_{11} & 0 \\ 0 & 1/a_{22} \end{bmatrix} \sigma^{-1} = \\ &= \begin{bmatrix} 1/a_{11} & 0 \\ 0 & 1/a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12}^2/a_{22} \\ a_{21}^2/a_{11} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & a_{12}^2/a_{11}a_{22} \\ a_{21}^2/a_{11}a_{22} & 1 \end{bmatrix} \end{aligned} \quad (1.9.2)$$

matrice che contiene, al di fuori della diagonale principale, i coefficienti di correlazione tra le coppie di stimatori dei minimi quadrati dei parametri \mathbf{b} .

Osservazione 1.21 - Il coefficiente di correlazione lineare tra le due variabili aleatorie $\hat{\beta}_1$ e $\hat{\beta}_2$ è dato da

$$\rho(\hat{\beta}_1, \hat{\beta}_2) = \frac{\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{[\text{Var}(\hat{\beta}_1) \cdot \text{Var}(\hat{\beta}_2)]^{1/2}} = \frac{\sigma^2 a_{12}^2}{[\sigma^2 a_{11}^2 \cdot \sigma^2 a_{22}^2]^{1/2}} = a_{12}^2 / a_{11} a_{22} \quad (1.9.3)$$

che si ritrova nei due elementi della (1.9.2) fuori della diagonale principale, uguali poiché la matrice di dispersione è simmetrica. Si noti, altresì, come la struttura del coefficiente di correlazione (1.9.3) sia simile a quella della matrice di correlazione (1.9.1), con la matrice delle varianze e delle covarianze “divisa” per le radici quadrate aritmetiche delle varianze degli stimatori dei parametri disposte lungo la diagonale della matrice $\sigma\mathbf{D}$.

Osservazione 1.22 - Generalizzando il caso definito dalla (1.9.3) si ottiene che il coefficiente di correlazione tra due stimatori $\hat{\beta}_i$ e $\hat{\beta}_j$ è dato dall'elemento a_{ij}^2 della matrice $(\mathbf{X}'\mathbf{X})^{-1}$ diviso per il prodotto $a_{ii}a_{jj}$ dei due elementi diagonali di posto i e j della matrice \mathbf{D} , come anche indicato nella (1.9.1).

In generale, se la correlazione fra $\hat{\beta}_i$ e $\hat{\beta}_j$ è elevata ciò significa che a scostamenti di $\hat{\beta}_i$ dalla propria media β_i saranno associati scostamenti (positivi o negativi, a seconda del segno di $\rho(\hat{\beta}_i, \hat{\beta}_j)$) di $\hat{\beta}_j$ dalla propria media β_j . Dato che, come abbiamo appena ricordato, le medie delle stime dei parametri coincidono con i valori veri degli stessi (per la proprietà di non distorsione), correlazioni elevate implicano che errori nella stima di un parametro tenderanno ad essere associati a errori nella stima degli altri parametri. In questo senso quindi l'incorrelazione delle stime contribuisce alla robustezza del modello, poiché garantisce che errori accidentali compiuti nella stima di uno dei β_i non si propagano alle stime degli altri coefficienti.

1.10 La stima dei minimi quadrati di una funzione delle importazioni

Stimiamo con il criterio dei minimi quadrati ordinari la seguente funzione delle importazioni per l'Italia

$$\ln y_t = \beta_1 + \beta_2 \ln x_{1t} + \beta_3 \ln x_{2t} + \beta_4 \ln x_{3t} + \beta_5 \ln x_{4t} + u_t \quad (1.10.1)$$

dove

y_t = importazioni di beni e servizi,

x_{1t} = consumi finali interni delle famiglie + consumi collettivi,

x_{2t} = investimenti fissi lordi + esportazioni di beni e servizi + variazioni delle scorte,

x_{3t} = deflatore implicito delle importazioni,

x_{4t} = deflatore implicito del prodotto interno lordo

per cui $\beta_2 > 0$, $\beta_3 > 0$, $\beta_4 < 0$, $\beta_5 > 0$.

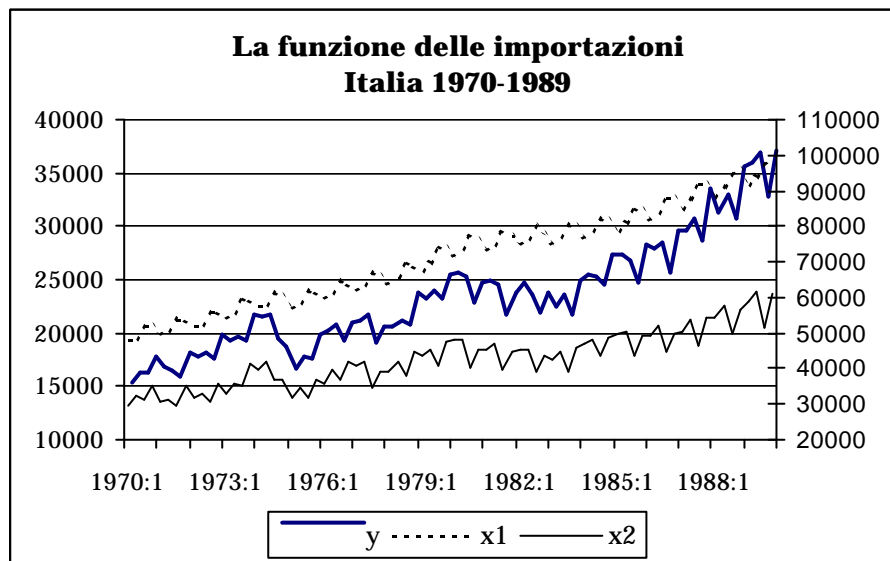


Figura 1.3 – La funzione delle importazioni in Italia, dati trimestrali grezzi a prezzi 1980 sul campione 1970:1-1989:4; y sono le importazioni di beni e servizi (scala di sinistra), x_1 è la somma dei consumi finali interni delle famiglie e collettivi (scala di destra) e x_2 è la somma di investimenti fissi lordi, esportazioni di beni e servizi e variazioni delle scorte (scala di destra); si veda la (1.10.1).

La (1.10.1) differisce dalle funzioni standard delle importazioni in quanto:

- (i) considera separatamente due componenti della domanda aggregata, e

- (ii) considera separatamente i prezzi dei beni importati e di quelli prodotti internamente, cioè non impone l'ipotesi che la domanda di beni importati dipenda dal loro prezzo relativo.

La prima caratteristica è motivata dal fatto che si può ipotizzare che la domanda di beni importati reagisca con diverse elasticità alla domanda per consumi e a quella per investimenti. La seconda è motivata dal fatto che inserire nella funzione di domanda i prezzi relativi significa di fatto imporre un vincolo sui parametri dell'equazione. Vedremo più avanti, nel corso di questo capitolo, come può essere formulato questo vincolo, mentre nel prossimo capitolo studieremo i metodi statistici che permettono di verificare se esso sia o meno respinto dai dati campionari.

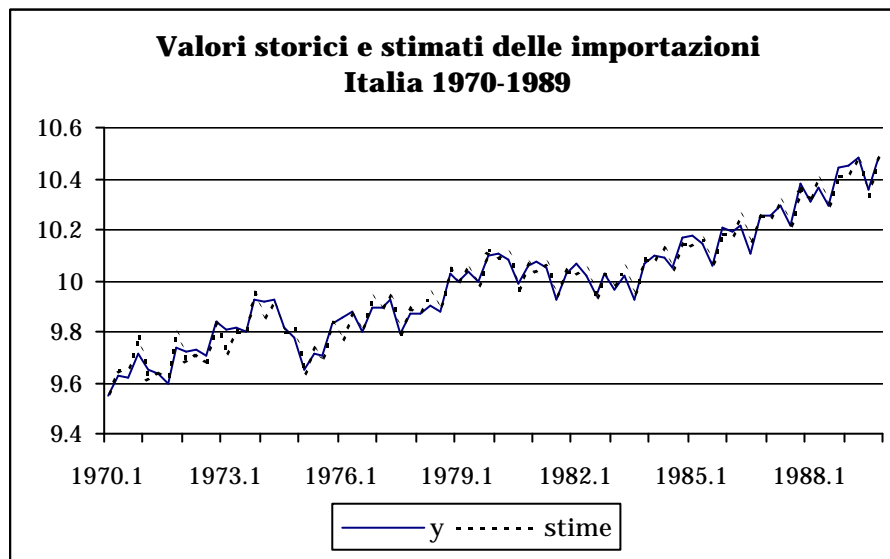


Figura 1.4 – I valori storici delle importazioni e quelli stimati con l'equazione (1.10.2).

Per stimare la (1.10.1) utilizziamo dati trimestrali nel periodo campionario che va dal primo trimestre del 1970 al quarto del 1989 (1970:1 – 1989:4), grezzi (cioè non depurati delle stagionalità), a prezzi 1980 e di fonte ISTAT (1989).

Le serie y_t , x_{1t} e x_{2t} sono rappresentate nella figura 1.3.²⁴ Si vede facilmente come la serie x_{1t} dei consumi abbia un andamento più livellato di quelle delle altre componenti del prodotto. È questo un fatto stilizzato che abbiamo già incontrato nel capitolo I-2 (si vedano ad esempio le figure I-2.1 e I-2.4 e i relativi commenti) e che

²⁴ Le serie storiche impiegate in questo esempio possono essere scaricate dalla home page della *Traccia* nel sito Internet <http://econometria.net>.

ha motivato diverse teorie economiche, fra le quali l'ipotesi del reddito relativo di Duesenberry (esposta nel paragrafo I-2.2) e la teoria del ciclo di vita di Ando-Modigliani (nota dai testi di macroeconomia). Nel presente contesto ciò comporta che i consumi da soli possono spiegare in parte la tendenza della variabile dipendente, ma non le sue oscillazioni cicliche, per tener conto delle quali diventa determinante il contributo della serie x_{2t} .²⁵

L'equazione stimata è:

$$\ln y_t = -6.023 + 0.572 \ln x_{1t} + 0.900 \ln x_{2t} - 0.164 \ln x_{3t} + 0.104 \ln x_{4t} + \hat{u}_t \quad (1.10.2)$$

$$R^2 = 0.984, R_c^2 = 0.983, \text{RSS} = 0.0633, \text{SEE} = 0.029$$

dove \hat{u}_t è la serie storica dei residui stimati, definiti dalla (1.4.2) come differenza fra la serie dei valori storici e di quelli stimati della variabile dipendente (si veda la figura 1.4) e rappresentati nella figura 1.5; RSS è la devianza (minima) dei residui stimati, cioè la $S(\hat{\mathbf{b}}) = \hat{\mathbf{u}}' \hat{\mathbf{u}}$; SEE è l'errore standard dell'equazione, cioè la radice quadrata della (1.7.2), determinato con un numero di gradi di libertà pari a $n-k = 75$.

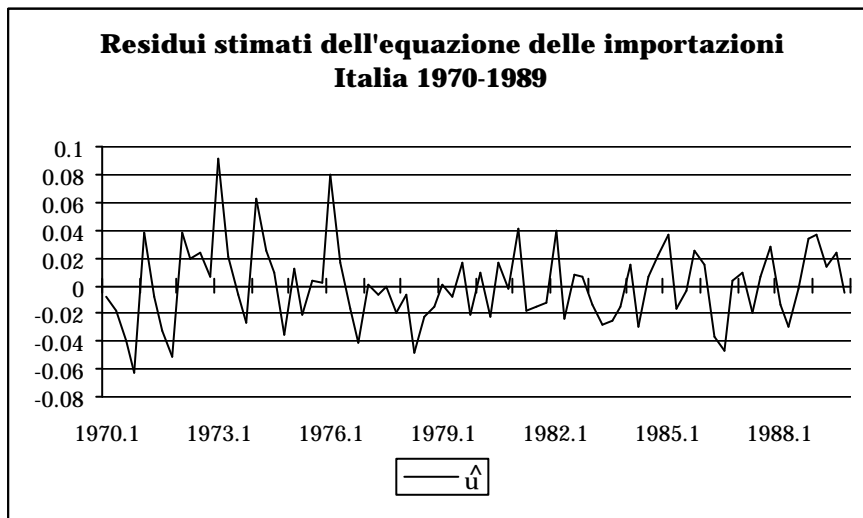


Figura 1.5 - Il grafico dei residui dell'equazione (1.10.2).

La figura 1.4 mostra che l'accostamento fra valori storici e stimati è buono. Di conseguenza i due coefficienti di determinazione, calcolati mediante le variabili scarto (cioè centrati e quindi non influenzati dal termine noto), posseggono valori

²⁵ Un'osservazione più attenta mostra che anche le ciclicità stagionali di importazioni e consumi differiscono, presentando una diversa conformazione di picchi.

alti. Questo risultato va però considerato con cautela perché potrebbe dipendere dalla presenza di tendenza nelle variabili dell'equazione.

Un buon accostamento fra valori storici e stimati significa residui “piccoli”. Il grafico dei residui 1.4 mette però in luce che i residui, pur essendo relativamente piccoli, potrebbero non rispettare le ipotesi stocastiche deboli (1.6.10). Si vede infatti distintamente che fino a tutto il 1977 i residui presentano una variabilità maggiore rispetto a quella mostrata nel periodo successivo (dal 1978 al 1989). Si noti che il grafico 1.4 rappresenta i residui *stimati* mentre le ipotesi stocastiche si riferiscono ai residui *teorici*. Tuttavia, nella misura in cui i residui stimati possono essere interpretati come una buona approssimazione di quelli teorici, la figura 1.5 suggerisce che nella (1.10.2) viene a cadere l'ipotesi di omoschedasticità implicita nella terza delle (1.6.10). Un'analisi più accurata mostra che in certi periodi i residui tendono a seguire un ciclo stagionale (cioè di periodo quattro). Questo fatto contrasta con l'ipotesi di non autocorrelazione (anch'essa implicata dalla terza delle (1.6.10)), dato che in presenza di simili cicli ogni valore dei residui stimati è positivamente correlato con quelli precedenti e seguenti a distanza di multipli interi di quattro trimestri.²⁶

L'inversa della matrice dei momenti è data dalla

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1126.194 & & & & & & \\ -89.272 & 8.727 & & & & & \\ -11.938 & -0.786 & 1.945 & & & & \\ -3.425 & 0.189 & 0.129 & 0.442 & & & \\ 27.674 & -2.151 & -0.344 & -0.525 & 1.151 & & \end{bmatrix} \quad (1.10.3)$$

nella quale sono omessi gli elementi simmetrici, e dalla quale è semplice determinare la matrice di correlazione degli stimatori dei parametri di regressione definita dalla (1.9.1): la matrice diagonale \mathbf{D} è data da

$$\mathbf{D} = < 33.559 \quad 2.954 \quad 1.395 \quad 0.665 \quad 1.073 >$$

per cui la matrice di correlazione è

²⁶ Ribadiamo che il grafico 1.4 rappresenta i residui stimati, non quelli teorici (non osservabili). Le condizioni che giustificano l'uso dei residui stimati per analizzare le proprietà dei residui teorici verranno descritte in modo più puntuale nei capitoli successivi.

$$Corr(\hat{\mathbf{b}}) = \begin{bmatrix} 1 & & & & \\ -0.900 & 1 & & & \\ -0.255 & -0.191 & 1 & & \\ -0.153 & 0.096 & 0.139 & 1 & \\ 0.769 & -0.679 & -0.230 & -0.736 & 1 \end{bmatrix} \quad (1.10.4)$$

nella quale si nota come i coefficienti di correlazione più alti in valore assoluto riguardino la coppia $(\hat{\beta}_1, \hat{\beta}_2)$, con correlazione -0.9 , e quelle formate dal parametro $\hat{\beta}_5$ con $\hat{\beta}_1$, $\hat{\beta}_4$, e $\hat{\beta}_2$, pari rispettivamente a 0.77 , -0.74 e -0.68 ; gli altri valori della matrice sono bassi.

1.11 Il criterio dei minimi quadrati vincolati

In molte equazioni econometriche è necessario stimare i parametri subordinatamente al fatto che essi soddisfino una o più relazioni lineari. La funzione di produzione del tipo Cobb-Douglas (I-2.4.5), ad esempio, che logaritmizzata e completata con il residuo stocastico diventa

$$\ln x = \ln \gamma + \alpha \ln l + \beta \ln k + u \quad (1.11.1)$$

deve essere stimata subordinatamente al vincolo tra i parametri

$$\alpha + \beta = 1 \quad (1.11.2)$$

che corrisponde all'ipotesi di omogeneità di grado uno della x rispetto alle variabili esplicative,²⁷ ipotesi che dal punto di vista economico esprime quella di rendimenti di scala costanti.

Un secondo esempio riguarda la funzione delle importazioni (1.10.1) nella quale si può imporre il vincolo che le elasticità delle importazioni rispetto ai consumi ed agli investimenti siano uguali

$$\beta_2 = \beta_3 \quad (1.11.3)$$

oppure si può imporre la condizione di omogeneità di grado zero di y rispetto ai due deflatori, che vale²⁸

$$\beta_4 + \beta_5 = 0 \quad (1.11.4)$$

Dal punto di vista economico quest'ultima ipotesi corrisponde alla nota ipotesi di teoria della domanda secondo la quale le funzioni di domanda dipendono dai prezzi relativi dei beni.

Un semplice modo di trattare questi problemi di stime vincolate riposa nell'inserimento del vincolo nella stessa equazione da stimare: nel caso della (1.11.1) si ha

$$\ln x = \ln \gamma + \alpha \ln l + (1-\alpha) \ln k + u$$

²⁷ Una funzione $y = f(x_1, x_2)$ di due variabili x_1 e x_2 è omogenea di grado r rispetto a queste se accade che $f(\lambda x_1, \lambda x_2) = \lambda^r f(x_1, x_2)$ per ogni punto (x_1, x_2) e ogni valore della costante reale λ . La generalizzazione al caso di più variabili esplicative è immediata. Per la (1.11.1) si ha

$$f(\lambda l, \lambda k) = \gamma \lambda^{\alpha+\beta} l^\alpha k^\beta$$

che comporta la condizione di omogeneità di grado uno se $\alpha+\beta = 1$.

²⁸ Infatti in questo caso, ricordando quanto riportato nella nota precedente,

$$f(x_1, x_2, \lambda x_3, \lambda x_4) = (\beta_4 + \beta_5) \ln \lambda + f(x_1, x_2, x_3, x_4)$$

che nella stima diventa

$$y = \beta_1 + \beta_2 z + u \quad (1.11.5)$$

con $y = \ln(x/k)$, $z = \ln(I/k)$, $\beta_1 = \ln \gamma$, $\beta_2 = \alpha$.

Nel caso della funzione delle importazioni, d'altro canto, la (1.10.1) subordinata al vincolo (1.11.3) diventa

$$\ln y = \beta_1 + \beta_2 \ln(x_1 x_2) + \beta_4 \ln x_3 + \beta_5 \ln x_4 + u \quad (1.11.6)$$

che può essere stimata, ad esempio, con il criterio dei minimi quadrati ordinari sostituendo alle due variabili esplicative distinte x_1 ed x_2 l'altra variabile consistente nel loro prodotto.

Infine, se nella (1.10.1) si vogliono inserire contemporaneamente i vincoli (1.11.3) e (1.11.4) si ottiene

$$\ln y = \beta_1 + \beta_2 \ln(x_1 x_2) + \beta_5 \ln \frac{x_3}{x_4} + u \quad (1.11.7)$$

che possiede due variabili esplicative soltanto, la $z_1 = (x_1 x_2)$ e la $z_2 = (x_3/x_4)$, che rappresenta il prezzo relativo delle importazioni, calcolato come rapporto tra il deflatore implicito delle importazioni di beni e servizi ed il deflatore implicito del PIL.

La stima dei minimi quadrati vincolati

In generale l'equazione (1.3.4) può essere stimata con il criterio dei minimi quadrati subordinatamente al soddisfacimento di q vincoli lineari sui k parametri, che possono essere scritti nella forma matriciale seguente

$$\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{q1} & r_{q2} & \dots & r_{qk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_k \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_q \end{bmatrix} \quad (1.11.8)$$

resa più compatta nell'altra

$$\mathbf{Rb} = \mathbf{r} \quad (1.11.9)$$

dove \mathbf{R} è una matrice di ordine $q \times k$, con $q \leq k$, ed \mathbf{r} è un vettore di dimensione q .

Si può inoltre supporre che il rango di \mathbf{R} sia q poiché altrimenti una o più righe di \mathbf{R} sarebbero combinazioni lineari delle altre e le equazioni (1.11.8) non sarebbero

linearmente indipendenti (cioè ci sarebbero almeno due vincoli equivalenti), come si può dimostrare nell'algebra delle matrici.

Il vincolo (1.11.2) per il modello (1.11.1) rappresenta una sola delle equazioni del sistema (1.11.9) che diventa

$$[0 \quad 1 \quad 1] \begin{bmatrix} \ln \gamma \\ \alpha \\ \beta \end{bmatrix} = 1 \quad (1.11.10)$$

con \mathbf{R} matrice formata dal solo vettore riga $[0 \ 1 \ 1]$ ed \mathbf{r} vettore costituito dal solo scalare 1. Anche il vincolo (1.11.3) per il modello (1.10.1) rappresenta una sola delle equazioni (1.11.8), che è

$$[0 \quad 1 \quad -1 \quad 0 \quad 0] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = 0 \quad (1.11.11)$$

con \mathbf{R} matrice formata dal solo vettore riga $[0 \ 1 \ -1 \ 0 \ 0]$ ed \mathbf{r} vettore costituito dal solo scalare 0.

Se le due condizioni (1.11.3) e (1.11.4) sono imposte simultaneamente sulla (1.10.1), il sistema (1.11.8) diventa

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (1.11.12)$$

con \mathbf{R} matrice 2×5 ed \mathbf{r} vettore bidimensionale.

Nei tre esempi precedenti la stima dei parametri delle equazioni subordinata al valore dei vincoli è stata trasformata in una stima non vincolata di altre equazioni i cui parametri tenevano conto dei vincoli stessi. È tuttavia possibile stimare con il criterio dei minimi quadrati direttamente le equazioni originali, del tipo lineare (1.3.4), sotto i q vincoli lineari rappresentati dal sistema (1.11.8); dal punto di vista matematico si tratta di effettuare la minimizzazione (1.4.3) sotto la condizione (1.11.8). A questo scopo si costruisce la funzione lagrangiana

$$S(\mathbf{b}, \mathbf{m}) = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) - \mathbf{m}'(\mathbf{Rb} - \mathbf{r}) = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) - \mathbf{b}'\mathbf{R}'\mathbf{m} + \mathbf{m}'\mathbf{r} \quad (1.11.13)$$

ottenuta combinando linearmente la somma dei quadrati (1.4.5) e il primo membro del vincolo (1.11.9) con il vettore \mathbf{r} portato a sinistra, tramite il vettore $\mathbf{m} = [m_1, m_2, \dots, m_q]'$ dei pesi, detti *moltiplicatori di Lagrange*. La minimizzazione vincolata può avvenire ora con la procedura dei minimi liberi applicata alla (1.11.13). Seguendo quanto esposto nel paragrafo 1.4, notiamo che le condizioni del primo ordine per minimizzare la (1.11.13) impongono che le derivate parziali prime siano nulle, per cui, ricordando la (1.4.7)

$$\begin{cases} \frac{\partial S(\mathbf{b}, \mathbf{m})}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{R}'\mathbf{m} = \mathbf{0} \\ \frac{\partial S(\mathbf{b}, \mathbf{m})}{\partial \mathbf{m}} = -\mathbf{R}\mathbf{b} + \mathbf{r} = \mathbf{0} \end{cases}$$

Dalla prima si ottiene

$$\mathbf{b} = \frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{m} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\mathbf{b}} + \frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{m} \quad (1.11.14)$$

dove si è fatto uso della stima $\hat{\mathbf{b}}$, e dalla seconda, sostituendovi la (1.11.14),

$$\mathbf{R}\hat{\mathbf{b}} + \frac{1}{2}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{m} = \mathbf{r}$$

cioè

$$\frac{1}{2}\mathbf{U}\mathbf{m} = \mathbf{r} - \mathbf{R}\hat{\mathbf{b}} \quad (1.11.15)$$

dove si è posto

$$\mathbf{U} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \quad (1.11.16)$$

Sfruttando ora il teorema XIX-1.9, se $(\mathbf{X}'\mathbf{X})^{-1}$ è una matrice definita positiva - e lo è perché esiste $(\mathbf{X}'\mathbf{X})^{-1}$ - è definita positiva e quindi non singolare anche la matrice \mathbf{U} data dalla (1.11.16), con \mathbf{R} di ordine $q \times k$, $q \leq k$, purché $r(\mathbf{R}) = q$, per cui si ottiene dalla (1.11.15)

$$\mathbf{m} = 2\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}}) \quad (1.11.17)$$

Sostituendo \mathbf{m} nella (1.11.14) si ottiene *la stima dei minimi quadrati vincolati*, che chiamiamo \mathbf{b}_0

$$\mathbf{b}_0 = \hat{\mathbf{b}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}}) \quad (1.11.18)$$

Lo stimatore dei minimi quadrati vincolati è non distorto; infatti,

$$E(\tilde{\mathbf{b}}_0) = E(\hat{\mathbf{b}}) + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}})] = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\mathbf{b}) = \mathbf{b} \quad (1.11.19)$$

in virtù della non distorsione di $\hat{\mathbf{b}}$ e della sussistenza del vincolo (1.11.9). Inoltre la matrice di dispersione di $\tilde{\mathbf{b}}_0$ è, per la (1.11.19),

$$\begin{aligned} \text{Cov}(\tilde{\mathbf{b}}_0) &= E[(\tilde{\mathbf{b}}_0 - \mathbf{b})(\tilde{\mathbf{b}}_0 - \mathbf{b})'] = \\ &= E\left\{[\mathbf{I}_k - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}](\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{u}}\tilde{\mathbf{u}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{I}_k - \mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}]\right\} = \quad (1.11.20) \\ &= \sigma^2[\mathbf{I}_k - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}](\mathbf{X}'\mathbf{X})^{-1}[\mathbf{I}_k - \mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= \sigma^2[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}] = \sigma^2[\mathbf{I}_k - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}](\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Le stime dei minimi quadrati vincolati per una funzione delle importazioni

Stimiamo con il criterio dei minimi quadrati l'equazione delle importazioni (1.10.1) sotto il vincolo (1.11.3) delle elasticità uguali. Volendo utilizzare la stima (1.11.18) si osserva che la matrice \mathbf{R} vale

$$\mathbf{R} = [0 \ 1 \ -1 \ 0 \ 0]$$

come anche indicato nella (1.11.11) per cui

$$\mathbf{R}\hat{\mathbf{b}} = [0 \ 1 \ -1 \ 0 \ 0] \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} = -0.328$$

dove $\hat{\mathbf{b}}$ è la stima dei minimi quadrati ordinari riportata nella (1.10.2). Inoltre, facendo uso della matrice (1.10.3) si ha

$$[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} = \mathbf{U}^{-1} = 0.0817 \quad (1.11.21)$$

per cui

$$\mathbf{R}'\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}}) = [0 \ 0.0268 \ -0.0268 \ 0 \ 0]'$$

essendo $\mathbf{r} = \mathbf{0}$. Infine, si calcola, in virtù della (1.11.18),

$$\mathbf{b}_0 = \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} + (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} 0.0000 \\ 0.0268 \\ -0.0268 \\ 0.0000 \\ 0.0000 \end{bmatrix} = \begin{bmatrix} -8.095 \\ 0.827 \\ 0.827 \\ -0.162 \\ 0.056 \end{bmatrix}$$

che si può ritrovare stimando la (1.10.1) con il vincolo inserito e con i minimi quadrati ordinari

$$\ln y_t = -8.095 + 0.827(\ln x_{1t} + \ln x_{2t}) - 0.162 \ln x_{3t} + 0.056 \ln x_{4t} + \hat{u}_t \quad (1.11.22)$$

$$R^2 = 0.982, R_c^2 = 0.981, \text{RSS} = 0.0721, \text{SEE} = 0.030$$

La devianza residua (RSS) nell'equazione vincolata è ovviamente maggiore di quella dell'equazione stimata senza vincoli ma l'errore standard (SEE) nelle due equazioni è molto simile poiché nell'equazione (1.11.22) esso è calcolato con un numero maggiore di grado di libertà: $n-4 = 76$.

Per determinare la matrice di correlazione degli stimatori dei parametri vincolati possiamo far uso della (1.11.20), per calcolare la quale notiamo che \mathbf{U}^{-1} è lo scalare dato dalla (1.11.21), che la matrice $\mathbf{R}'\mathbf{R}$ è quella calcolata nell'esempio (XIX-1.6) e che

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\mathbf{U}^{-1} \mathbf{R} = \begin{bmatrix} 0 & -6.316 & 6.316 & 0 & 0 \\ 0 & 0.777 & -0.777 & 0 & 0 \\ 0 & -0.223 & 0.223 & 0 & 0 \\ 0 & 0.005 & 0.005 & 0 & 0 \\ 0 & -0.148 & 0.148 & 0 & 0 \end{bmatrix}$$

per cui

$$[\mathbf{I}_5 - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\mathbf{U}^{-1} \mathbf{R}](\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 637.76 & & & & & \\ -29.19 & 1.34 & & & & \\ -29.19 & 1.34 & 1.34 & & & \\ -3.05 & 0.14 & 0.14 & 0.44 & & \\ 16.26 & -0.75 & -0.75 & -0.52 & 0.88 & \end{bmatrix}$$

nella quale sono omissi gli elementi simmetrici.

La matrice diagonale \mathbf{D} vale, in questo caso,

$$\mathbf{D} = \langle 25.254 \quad 1.156 \quad 1.156 \quad 0.665 \quad 0.940 \rangle$$

per cui la matrice di correlazione è

$$\text{Corr}(\tilde{\mathbf{b}}_0) = \begin{bmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ -1 & 1 & 1 & & & \\ -0.182 & 0.182 & 0.182 & 1 & & \\ 0.685 & -0.690 & -0.690 & -0.832 & 1 & \end{bmatrix}$$

In questa si verifica che, poiché $\beta_{02} = \beta_{03}$,

$$\rho_{21} = \rho_{31}, \quad \rho_{32} = 1, \quad \rho_{42} = \rho_{43}, \quad \rho_{52} = \rho_{53}$$

Il vincolo di omogeneità di grado zero sui prezzi

Determiniamo ora le stime dei minimi quadrati dei parametri dell'equazione (1.10.1) sotto il vincolo (1.11.4) dell'omogeneità di grado zero rispetto ai due deflatori impliciti delle importazioni e del PIL. L'inversa della matrice dei momenti è ancora fornita dalla (1.10.3), mentre la matrice del vincolo \mathbf{R} vale

$$\mathbf{R} = [0 \ 0 \ 0 \ 1 \ 1]$$

per cui

$$\mathbf{R}\hat{\mathbf{b}} = [0 \ 0 \ 0 \ 1 \ 1] \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} = -0.060$$

$$[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} = \mathbf{U}^{-1} = 1.8416$$

$$\mathbf{R}\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}}) = [0 \ 0 \ 0 \ 0.1105 \ 0.1105]'$$

Infine si calcola

$$\mathbf{b}_0 = \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} + (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.1105 \\ 0.1105 \end{bmatrix} = \begin{bmatrix} -3.344 \\ 0.355 \\ 0.876 \\ -0.173 \\ 0.173 \end{bmatrix}$$

che può essere ritrovata stimando la (1.10.1) con il vincolo inserito e con i minimi quadrati ordinari

$$\ln y_t = -3.344 + 0.355 \ln x_{1t} + 0.876 \ln x_{2t} - 0.173 (\ln x_{3t} - \ln x_{4t}) + \hat{u}_t \quad (1.11.23)$$

$$R^2 = 0.982, \quad R_c^2 = 0.982, \quad \text{RSS} = 0.0699, \quad \text{SEE} = 0.030$$

In questo caso il vincolo produce una devianza residua leggermente inferiore a quella dell'equazione (1.11.22), ma sempre maggiore di quella dell'equazione non vincolata (1.10.2).

Il doppio vincolo dell'uguaglianza delle elasticità e dell'omogeneità sui prezzi

Infine, calcoliamo le stime dei minimi quadrati dei parametri della (1.10.1) sotto i due vincoli delle elasticità uguali (1.11.3) e dell'omogeneità di grado zero rispetto ai

due deflatori (1.11.4). L'inversa della matrice dei momenti è la stessa (1.10.3), mentre la matrice \mathbf{R} dei due vincoli, data dalla (1.11.12), vale

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

per cui

$$\mathbf{R}\hat{\mathbf{b}} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} = \begin{bmatrix} -0.328 \\ -0.060 \end{bmatrix}$$

$$[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} = \begin{bmatrix} 12.244 & -1.747 \\ -1.747 & 0.543 \end{bmatrix}^{-1} = \begin{bmatrix} 0.1510 & 0.4857 \\ 0.4857 & 3.4044 \end{bmatrix}$$

$$\mathbf{R}'\mathbf{U}^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{b}}) = [0 \quad 0.079 \quad -0.079 \quad 0.363 \quad 0.363]'$$

Infine si calcola

$$\mathbf{b}_0 = \begin{bmatrix} -6.023 \\ 0.572 \\ 0.900 \\ -0.164 \\ 0.104 \end{bmatrix} + (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} 0.0000 \\ 0.0787 \\ -0.0787 \\ 0.3636 \\ 0.3636 \end{bmatrix} = \begin{bmatrix} -3.292 \\ 0.607 \\ 0.607 \\ -0.189 \\ 0.189 \end{bmatrix}$$

i cui elementi possono essere ritrovati stimando la (1.11.7) con il doppio vincolo inserito e con i minimi quadrati ordinari

$$\ln y_t = -3.292 + 0.607(\ln x_{1t} \cdot \ln x_{2t}) - 0.190(\ln x_{3t} / \ln x_{4t}) + \hat{u}_t \quad (1.11.24)$$

$$R^2 = 0.972, R_c^2 = 0.972, \text{RSS} = 0.1109, \text{SEE} = 0.037$$

dove si nota che la devianza dei residui è ben più alta di quella calcolata nel modello senza vincoli: è del 75.2% maggiore.

Per determinare la matrice di correlazione degli stimatori dei parametri vincolati facciamo ancora una volta uso della (1.11.20), per calcolare la quale notiamo che

$$\mathbf{U} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = \begin{bmatrix} 12.244 & -1.748 \\ -1.748 & 0.548 \end{bmatrix}$$

$$\mathbf{U}^{-1} = \begin{bmatrix} 0.151 & 0.485 \\ 0.485 & 3.399 \end{bmatrix}$$

$$\mathbf{I}_5 - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\mathbf{U}^{-1} \mathbf{R} = \begin{bmatrix} 1 & -0.093 & 0.093 & -44.898 & -44.898 \\ 0 & 0.516 & 0.484 & 2.054 & 2.054 \\ 0 & 0.516 & 0.484 & 2.054 & 2.054 \\ 0 & 0.031 & 0.031 & 1.252 & 0.252 \\ 0 & 0.031 & -0.031 & 1.252 & 0.252 \end{bmatrix}$$

$$[\mathbf{I}_5 - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\mathbf{U}^{-1} \mathbf{R}](\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 44.629 & & & & \\ -2.044 & 0.094 & & & \\ -2.044 & 0.094 & 0.094 & & \\ 0.276 & -0.009 & -0.009 & 0.423 & \\ -0.276 & 0.009 & 0.009 & -0.423 & 0.423 \end{bmatrix}$$

dalla quale si trae la matrice diagonale \mathbf{D}

$$\mathbf{D} = \langle 6.680 \quad 0.307 \quad 0.307 \quad 0.650 \quad 0.650 \rangle$$

e quindi la matrice di correlazione

$$\text{Corr}(\tilde{\mathbf{b}}_0) = \begin{bmatrix} 1 & & & & \\ -0.997 & 1 & & & \\ -0.997 & 1 & 1 & & \\ 0.064 & -0.045 & -0.045 & 1 & \\ -0.064 & 0.045 & 0.045 & -1 & 1 \end{bmatrix}$$

In questa si verifica che, poiché $\beta_{02} = \beta_{03}$,

$$\rho_{21} = \rho_{31}, \quad \rho_{32} = 1, \quad \rho_{42} = \rho_{43}, \quad \rho_{52} = \rho_{53}$$

e che, poiché $\beta_{04} = -\beta_{05}$,

$$\rho_{54} = 1, \quad \rho_{41} = -\rho_{51}$$

dove ρ_{ij} indica l'elemento di posto (i,j) nella matrice di correlazione di $\tilde{\mathbf{b}}_0$.

1.12 Riferimenti bibliografici

Amemiya, T. [1980], “Selection of regressors”, *International Economic Review*, **21**, pp. 331-354.

Theil, H. [1961], *Economic Forecasts and Policy*, II edizione, Amsterdam: North Holland.