

Capitolo V**RAPPRESENTAZIONE ANALITICA
DELLE VARIABILI STATISTICHE****1. Fasi della rappresentazione analitica**

Un modello simbolico o matematico rappresenta la *manifestazione esteriore di una teoria* formulata sulle relazioni intercorrenti tra le variabili che descrivono una determinata realtà fenomenica. Esso è costituito da elementi *variabili* nell'ambito dello stesso problema, cioè le variabili dipendenti e quelle indipendenti, e da elementi *stabili* nell'ambito dello stesso problema, cioè i parametri e le costanti caratteristiche. Quindi, se per studiare un determinato problema o un fenomeno particolare si vuol fare ricorso ad una schematizzazione formale del problema stesso, occorre attraversare tre fasi fondamentali di ricerca:

- anzitutto bisogna individuare quali sono le *variabili interagenti*, che definiscono il problema nei suoi vari aspetti;
- in secondo luogo è necessario interpretare il fenomeno osservato e cioè ipotizzare la *forma della interazione* tra le variabili;
- infine, sulla base dei dati osservati, si deve misurare il *valore dei parametri* specifici del problema considerato, senza di che riesce impossibile valutare le ripercussioni che certi comportamenti delle variabili esplicative determinano sulle variabili da spiegare.

Ora, mentre per la individuazione delle variabili interagenti e per la formulazione della forma della loro interazione si deve far ricorso alle conoscenze specifiche che già si hanno sul fenomeno studiato, per quanto riguarda il *problema della stima dei parametri* che caratterizzano ciascuna forma interpretativa, che rappresenta un tipico problema di natura statistica, occorre

esaminare i metodi disponibili per operare le stime suddette nei vari casi che si possono presentare.

Ovviamente, i criteri differiranno a seconda che si tratti di modelli *uniequazionali* o di modelli *pluriequazionali*, che si abbiano relazioni in due o più variabili o, infine, che si tratti di modelli *stocastici* oppure *normativi*.

In questo capitolo anzitutto si soffermerà l'attenzione sui metodi di uso più frequente, che sono impiegati per procedere alla *stima del valore dei parametri* nel caso dei *modelli stocastici uniequazionali*; in un secondo momento saranno anche illustrati gli accorgimenti da utilizzare nel caso in cui il modello prescelto per rappresentare il fenomeno allo studio sia un *modello stocastico pluriequazionale* formato da relazioni analitiche nelle quali figurino come variabili indipendenti certe entità che invece sono considerate come variabili dipendenti in altre relazioni funzionali dello stesso modello.

Invece, i criteri correntemente usati per valutare sia il *grado di precisione delle stime* ottenute per i parametri, sia *l'adattamento del modello* alla realtà osservata formeranno oggetto di trattazione nei capitoli successivi dedicati al problema della inferenza statistica.

Nel caso dei modelli uniequazionali, la forma di relazione più semplice che si può avere consiste nel *modello lineare in due variabili*. Se con il simbolo Y si indica la variabile dipendente, con il simbolo X la variabile indipendente e con a e b i parametri, si dice che Y è funzione lineare della X se per ogni valore di quest'ultima variabile la Y è data dalla relazione

$$Y = a + b \cdot X,$$

in cui il parametro a indica il valore che la Y assume quando la variabile X è uguale a zero, mentre il parametro b indica la propensione marginale a variare di Y per variazioni unitarie di X . Ad esempio, se Y è la domanda di un bene ed X è il prezzo del bene stesso, il parametro b misura la reazione della domanda quando il prezzo aumenta di una unità.

Generalizzando lo schema precedente al caso di più variabili, si dirà che Y è funzione lineare delle X_i , con $i = 1, 2, \dots, n$, se, per ogni combinazione di valori delle X_i , il livello di Y è determinato dalla relazione:

$$Y = b_0 \cdot X_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n.$$

Continuando nella generalizzazione, può darsi che la descrizione migliore di una relazione tra variabili non sia fornita da una relazione lineare, ma da una non lineare, oppure da una forma lineare per certe variabili e non lineare per altre. Ma, al di là di questi casi generali, data anche la complessità dei problemi di calcolo che sorgono per la stima dei parametri relativi a relazioni non lineari, è opportuno soffermarsi ora in particolare su un sottoinsieme di tali relazioni non lineari. Precisamente, su quel sottoinsieme che risulta formato da relazioni non lineari in due o più variabili, che, con opportune trasformazioni di variabile, possono essere ricondotte a *forme lineari nei parametri*.

Soffermandosi, ad esempio, sul caso dei modelli uniequazionali in due variabili, appartengono al sottoinsieme in questione le forme che seguono.

A. La funzione *esponenziale*. È definita dalla relazione o modello

$$X = A \cdot B^Y$$

che con una trasformazione semilogaritmica fornisce l'espressione

$$Y = a + b \cdot \log X$$

in cui si è posto $a = -\log A / \log B$ e $b = 1 / \log B$. In tale funzione, poichè è

$$\frac{dY}{dX} = \frac{b}{X}$$

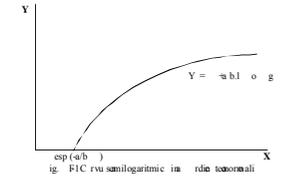
al crescere dei valori della variabile X la inclinazione della Y diminuisce e quando $Y=0$ è $\log X = -a/b$, cosicchè il punto di intersezione della funzione con l'asse delle X si incontra per $X = \exp(-a/b)$. L'inversa di questa funzione è data dalla espressione

$$X = e^{-\frac{a}{b}} \cdot e^{\frac{Y}{b}} = \exp(-a/b) \cdot \exp(Y/b)$$

che si può scrivere

$$X = A \cdot B^Y$$

ponendo $A = \text{esp}(-a/b)$ e $B = \text{esp}(1/b)$. Il diagramma seguente fornisce una rappresentazione della funzione esponenziale per $a < 0$ e $b > 0$.



Se invece si traccia il diagramma in termini delle variabili Y e $Z = \log X$, si ottiene il grafico della trasformata; cioè, la curva esponenziale precedente si trasforma in una retta.

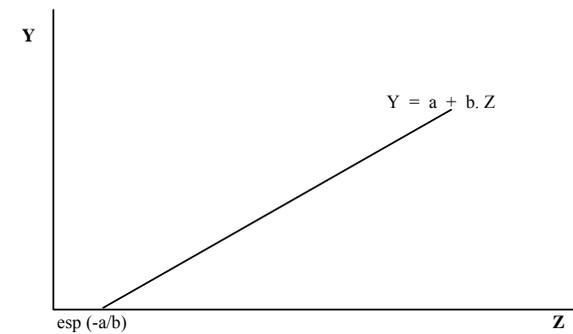


Fig. 2 - Curva semilogaritmica in coordinate semilogaritmiche

Va sottolineata la particolarità che, mentre nella espressione originaria i parametri A e B erano moltiplicati tra loro e quindi la relazione tra Y ed

X non era lineare nei parametri, per contro la trasformata semilogaritmica è lineare nei parametri.

La qual cosa è molto interessante per la maggiore facilità con cui si possono stimare i valori quantitativi dei parametri relativi ad una determinata situazione reale, quando essi figurano esplicitati in forma lineare nelle relazioni funzionali tra le variabili.

B. Altra forma notevole, sempre nel caso di modelli in due variabili, è la funzione doppio-logaritmica.

Si consideri la parabola generalizzata

$$Y = A \cdot X^b$$

che, con una trasformazione logaritmica, può essere riscritta nella forma

$$\log Y = a + b \cdot \log X$$

in cui si è posto $a = \log A$.

Anche in questo caso, la trasformazione consente di ottenere una relazione lineare nei parametri. Poiché è

$$\frac{dY}{dX} = A \cdot b \cdot X^{b-1} = b \cdot \frac{Y}{X}$$

si ha per la parabola generalizzata che, se è $b > 1$, l'inclinazione aumenta col crescere di X, mentre se è $0 < b < 1$, essa tende a diminuire col crescere della variabile X.

Inoltre, qualunque sia il valore di b, la variabile Y assumerà valori sempre più grandi e prossimi all'infinito al tendere all'infinito della variabile X. Se, anziché procedere alla rappresentazione diagrammatica in termini delle variabili originarie come indicato nella Fig. 3 che segue,

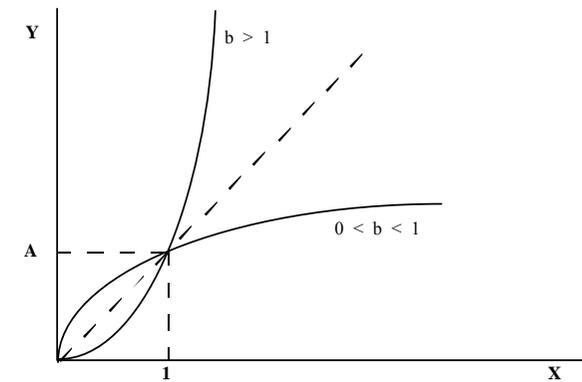


Fig. 3 - Curve della parabola generalizzata in coordinate normali

si opera quella in termini delle variabili trasformate, ponendo $H = \log Y$ ed anche $Z = \log X$, sarà

$$H = a + b \cdot Z,$$

cioè, le due curve si trasformeranno in due rette, come indicato nella Fig. 4.

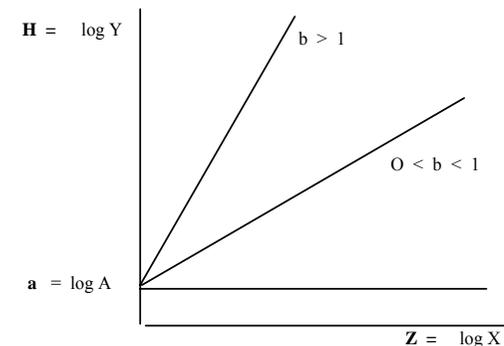


Fig. 4 - Curve della parabola generalizzata in coordinate logaritmiche

La forma doppio-logaritmica è diffusamente utilizzata dagli economisti, in quanto tale relazione corrisponde all'ipotesi di una *elasticità costante* tra le variabili Y ed X; inoltre, l'applicazione dei metodi di stima lineari ai valori dei logaritmi delle variabili permette di ottenere direttamente le stime di questa elasticità.

Vale la pena di ricordare che la *elasticità* di una funzione Y rispetto alla variabile indipendente X è definita come il *rapporto tra gli incrementi relativi* delle due variabili: cioè, sarà

$$\eta_{x/y} = \frac{\frac{dY}{Y}}{\frac{dX}{X}}$$

e, nel caso della funzione doppio-logaritmica, sarà, appunto,

$$\eta_{x/y} = \frac{dY}{dX} \cdot \frac{X}{Y} = A \cdot b \cdot X^{b-1} \cdot \frac{X}{Y} = b.$$

C. Un altro tipo di trasformazione molto usato è quella *reciproca*. Se si considera la forma *iperbolica*

$$Y = a + \frac{b}{X}$$

introducendo la trasformata $Z = 1 / X$, si ottiene la forma lineare

$$Y = a + b \cdot Z$$

Poiché la derivata della funzione rispetto ad X è

$$\frac{dY}{dX} = -\frac{b}{X^2}$$

se b è positivo, la inclinazione è sempre negativa e diminuisce in valore assoluto al crescere di X. Quando X tende a zero, la Y tende all'infinito e per X tendente ad infinito la Y tende ad a. La *trasformata reciproca* è quindi utile quando sia

ammissibile l'ipotesi di un livello asintotico dei valori della Y, l'asintoto essendo rappresentato appunto dal valore a.

In questo caso, la rappresentazione in termini delle variabili originarie fornisce il grafico riportato nella Fig. 5

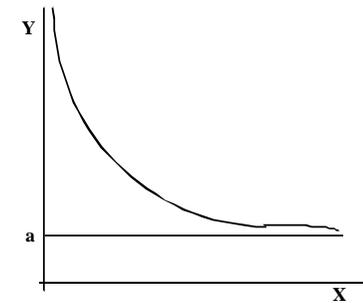


Fig. 5 - Curva dell'iperbole in coordinate normali

mentre, se si introduce la trasformata, quando b è positivo la rappresentazione diviene una retta discendente dai valori bassi di X verso i valori alti di X.

D. Un'altra trasformazione interessante è quella *logaritmico-reciproca*, per mezzo della quale la forma

$$Y = e^{a - \frac{b}{X}} = \text{esp}(a - b / X)$$

si trasforma nella relazione $H = a - b \cdot Z$, in cui $H = \log Y$ e $Z = 1/X$.

La curva logaritmico-reciproca passa per l'origine degli assi e, poiché la derivata prima è

$$\frac{dY}{dX} = \text{esp}\left(a - \frac{b}{X}\right) \cdot \frac{b}{X^2}$$

la inclinazione è positiva per b positivo. Inoltre, poiché la derivata seconda è

$$\frac{d^2Y}{dX^2} = \text{esp}\left(a - \frac{b}{X}\right) \cdot \left(\frac{b^2}{X^4} - \frac{2b}{X^3}\right)$$

vi è un punto di flesso in corrispondenza di $X = b/2$. A sinistra di questo punto l'inclinazione aumenta con X, mentre a destra diminuisce. Inoltre, al tendere di X all'infinito la Y tende al valore $\text{esp}(a)$. Rappresentata graficamente tale funzione da luogo alla curva descritta nella Fig. 6.

Curve di questo tipo sono state usate moltissimo nelle analisi statistiche delle abitudini di consumo, desunte dalle indagini sul comportamento dei consumatori, tramite la raccolta dei bilanci di famiglia. Infatti, in questo caso, l'ipotesi della esistenza di un livello asintotico o limite superiore della spesa, per certi beni può essere accettabile.

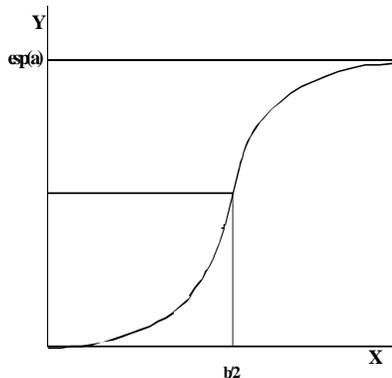


Fig 6 - Curva grafica tipo iccnaon teorali

Naturalmente, la stima dei parametri di relazioni non lineari potrebbe essere fatta direttamente sui dati originali, senza l'aiuto di trasformazioni preliminari, ma, in tal caso, sfortunatamente, non vi sarebbe la possibilità di utilizzare dei metodi e dei principi standard, come quello dei *minimi quadrati* e quello della *massima verosimiglianza*, i quali si prestano particolarmente ad essere usati, quando le relazioni tra variabili sono espresse o esprimibili tramite una preventiva trasformazione in forme lineari nei parametri.

Questo vale sia per le forme non lineari in due variabili, sia per quelle non lineari in più variabili. In quest'ultimo caso, infatti, gli accorgimenti che devono essere adottati per la linearizzazione delle forme sono di natura analoga a quelli considerati per il caso dei modelli in due variabili, naturalmente con qualche complicazione e difficoltà in più, per la maggiore complessità delle espressioni alle quali possono dar luogo.

Vale la pena di sottolineare che i metodi di rappresentazione analitica delle variabili statistiche illustrati in letteratura sono numerosi e soddisfano a certi requisiti particolari, come quello di adattare funzioni che passano *per* i punti osservati oppure *tra* i punti stessi o quello di soddisfare certe uguaglianze tra convenienti ed opportune operazioni fatte sulla distribuzione dei dati rilevati e le stesse operazioni ripetute sui valori teorici forniti dalla funzione interpolatrice.

In linea generale, per la determinazione dei valori numerici dei parametri che figurano nella funzione interpolatrice bisogna distinguere tre casi possibili: le variabili dipendenti ed indipendenti assumono valori privi di errori; solo i valori delle variabili dipendenti sono affetti da errori; tanto i valori delle variabili dipendenti, quanto quelli delle variabili indipendenti sono affetti da errori. A ciascun caso corrispondono metodi differenti, tra i quali se ne considereranno solo alcuni più importanti.

2. Metodo delle ordinate fisse

Per calcolare il valore dei parametri di un modello uniequazionale in più variabili, si possono usare metodi diversi, a seconda che la curva rappresentativa debba passare *per* tutti i punti osservati oppure *tra* i vari punti. Ma l'*interpolazione per punti* è possibile quando si ha ragione di ritenere che i valori osservati - e rappresentati nel diagramma cartesiano da punti di coordinate (x, y) - sono relativi a misure delle variabili, che possono ritenersi sufficientemente *esatte*. Infatti, se si adottassero curve passanti per punti-

immagine di misure affette da errore, la funzione che così si verrebbe ad individuare, sarebbe imprecisa, perché essa pure sarebbe affetta dagli stessi errori.

Pertanto, il metodo di determinazione dei parametri che stiamo considerando, essendo uno dei metodi di interpolazione *per* punti noti, si può utilizzare solo quando *non* ci sono errori nelle variabili. Esso consiste in quanto segue.

Supponiamo di avere s coppie di valori non affette da errori e di rappresentarle sul piano cartesiano; si avrà quindi una serie di s punti per i quali dovrà passare la funzione che si cerca. Ricordiamo che imporre ad una curva di passare per un punto equivale ad imporre alla sua equazione di soddisfare una condizione, per cui se nella funzione che si cerca figurano h parametri, occorrono h condizioni per determinarli.

Però, una volta fissati i punti, per questi possono passare una infinità di curve e, quindi, affinché il problema sia *determinato* occorre avere tante condizioni quanti sono i parametri che figurano nella funzione interpolatrice ed inoltre che tali condizioni siano *distinte* e *compatibili*, cioè le condizioni non siano l'una combinazione delle altre.

Se si deve interpolare con una parabola di grado $(s-1)$ con s parametri del tipo:

$$y = c_0 + c_1x + c_2x^2 + \dots + c_{s-1}x^{s-1}$$

il passaggio per i punti fissati si ottiene sostituendo ad x ed y dell'equazione generica i valori (x_i, y_i) delle coordinate dei punti, ottenendo il seguente sistema:

$$\begin{cases} y_1 = c_0 + c_1x_1 + c_2x_1^2 + \dots + c_{s-1}x_1^{s-1} \\ y_2 = c_0 + c_1x_2 + c_2x_2^2 + \dots + c_{s-1}x_2^{s-1} \\ \dots \dots \dots \dots \dots \dots \\ y_s = c_0 + c_1x_s + c_2x_s^2 + \dots + c_{s-1}x_s^{s-1} \end{cases}$$

in cui le incognite sono gli s parametri c_i . Tale sistema può risolversi con i normali mezzi di calcolo. e la soluzione generica sarà fornita dalla relazione:

$$c_i = \frac{\begin{vmatrix} 1 & x_1 & x_1^2 & \dots & y_1 & \dots & x_1^{s-1} \\ 1 & x_2 & x_2^2 & \dots & y_2 & \dots & x_2^{s-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_s & x_s^2 & \dots & y_s & \dots & x_s^{s-1} \end{vmatrix}}{\begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^1 & \dots & x_1^{s-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^1 & \dots & x_2^{s-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_s & x_s^2 & \dots & x_s^1 & \dots & x_s^{s-1} \end{vmatrix}}$$

Lo stesso si può dire se, anziché interpolare con una parabola di ordine (s-1) si interpola con un piano nello spazio a (s-1) dimensioni in cui alle singole potenze della variabile x sopra considerate, si sostituiscono altre (s-1) variabili z, t, k, h, ...tutte elevate alla prima potenza.

Vale la pena di osservare, comunque, che difficilmente conviene utilizzare forme così complesse, non tanto per le difficoltà del calcolo, quanto per il dubbio significato interpretativo della relazione trovata. Per cui, può riuscire molto più utile limitarsi a parabole di grado non elevato o a funzioni lineari in poche variabili, scegliendo soltanto alcuni punti per i quali far passare la curva interpolatrice, ma badando nel contempo che la curva stessa si discosti il meno possibile dagli altri punti.

Questo metodo è particolarmente usato con funzioni di primo grado quando si debba effettuare l'interpolazione tra due valori assegnati di una tavola numerica, come quelle dei tanti prontuari di matematica finanziaria e di statistica. In questo caso, si ottiene il valore di y in corrispondenza di un dato x non compreso nelle tavole, conoscendo i valori contigui, in base alla relazione:

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} (x - x_1).$$

Esempio V.1

L'equazione della logistica con asintoto inferiore nullo è:

$$y = \frac{m}{1 + n \cdot e^{-ax}}$$

e, ponendo e^{-a} = r, si ha

$$y = \frac{m}{1 + n \cdot r^x}$$

Per il calcolo dei parametri m, n, ed r si può far passare quest'ultima funzione per i tre punti (0, y₀), (1, y₁), (2, y₂), in modo da avere:

$$y_0 = \frac{m}{1 + n}; y_1 = \frac{m}{1 + n \cdot r}; y_2 = \frac{m}{1 + n \cdot r^2};$$

invertendo i membri di ciascuna equazione si ha:

$$\frac{1}{y_0} = \frac{1}{m} + \frac{n}{m}; \frac{1}{y_1} = \frac{1}{m} + \frac{n \cdot r}{m}; \frac{1}{y_2} = \frac{1}{m} + \frac{n \cdot r^2}{m},$$

per cui, sottraendo la seconda espressione dalla prima e la terza dalla seconda si ha:

$$\frac{1}{y_0} - \frac{1}{y_1} = \frac{n}{m} (1 - r); \frac{1}{y_1} - \frac{1}{y_2} = \frac{n}{m} \cdot r \cdot (1 - r)$$

e dividendo membro a membro la seconda per la prima equazione si ha:

$$\frac{y_0 (y_2 - y_1)}{y_2 (y_1 - y_0)} = r$$

che fornisce il valore di r. Calcolato questo, dalla

$$\frac{1}{y_0} - \frac{1}{y_1} = \frac{n}{m} (1 - r) \text{ si ricava } \frac{n}{m} = \frac{\frac{1}{y_0} - \frac{1}{y_1}}{(1 - r)}$$

che sostituito nella

$$\frac{1}{y_0} = \frac{1}{m} + \frac{n}{m}$$

consente di ricavare immediatamente m e poi n dalla

$$\frac{n}{m} = \frac{\frac{1}{y_0} - \frac{1}{y_1}}{(1 - r)}$$

ottenendo

$$n = \frac{y_2 (y_1 - y_0)^2}{y_0 (y_1^2 - y_0 y_2)}$$

Si osservi che, affinché i valori teorici di y siano reali occorre e basta che e^{-a} sia reale e positivo e quindi deve essere y₁ compreso tra y₀ e y₂ se i valori considerati sono positivi, come di fatto avviene nei casi pratici che interessano la statistica.

Inoltre, deve essere n>0, perché la logistica assuma per asintoto m una quantità finita e positiva.

Pertanto, affinché la logistica sia limitata, secondo l'ultima espressione scritta deve essere verificata la condizione y₁² > y₀y₂.

3. Metodo dei minimi quadrati

Più spesso accade che le misure statistiche a disposizione del ricercatore siano affette da errori di vario tipo derivanti da imperfezioni degli strumenti di misura, da distrazioni dell'operatore o dal fatto che si ha soltanto un campione dei dati possibili. In tali casi, per eliminare l'influenza di tali errori, occorre fare in modo che la curva non passi *per* i punti-immagine, ma passi *tra* tali punti. E, poiché, una volta fissato il tipo di funzione, vi sono svariate espressioni di tale funzione che possono passare tra i punti osservati - cioè, tante quante sono le possibili determinazioni dei parametri della funzione comprese in certi intervalli di variazione - sarà necessario anche *fissare la condizione di accostamento* che dovrà essere verificata, affinché, dal punto di vista della condizione fissata, sia scelta la funzione "migliore" tra tutte quelle possibili.

Uno dei metodi più utilizzati in statistica, nell'ipotesi che i valori della variabile dipendente siano affetti da errore e che quelli delle variabili indipendenti siano esatti è quello noto come *metodo dei minimi quadrati*.

Il *metodo dei minimi quadrati* pone la condizione di accostamento che *sia minima la somma dei quadrati delle differenze tra i k valori osservati della Y ed i k valori teorici forniti per la stessa variabile dalla relazione funzionale*, in corrispondenza dei diversi valori assunti dalle variabili indipendenti.

Alla base di questa condizione sta l'ipotesi che gli errori da cui sono affette le misure y_i della variabile Y siano accidentali e tendano a compensarsi, al crescere del numero dei casi osservati. Se si indica con y_i la generica determinazione che la variabile Y assume nella realtà, in corrispondenza della determinazione x_i della variabile X e se si indica con

$$y_i' = f(x_i; b_0, b_1, b_2, \dots, b_{s-1})$$

il valore teorico assunto dalla variabile Y nella relazione funzionale, quando la variabile X assume il valore x_i , la condizione dei minimi quadrati può esprimersi in termini formali come segue:

in cui D è il determinante formato dai coefficienti delle incognite del precedente sistema e cioè è

$$D = \begin{vmatrix} \Sigma f^2(x_i) & \Sigma f(x_i)\varphi(z_i) & \Sigma f(x_i)\psi(t_i) & \dots\dots\dots \\ \Sigma f(x_i)\varphi(z_i) & \Sigma \varphi^2(z_i) & \Sigma \varphi(x_i)\psi(t_i) & \dots\dots\dots \\ \Sigma f(x_i)\psi(t_i) & \Sigma \varphi(x_i)\psi(t_i) & \Sigma \psi^2(t_i) & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \end{vmatrix}$$

mentre A_{hj} è il cofattore o aggiunto dell'elemento che si trova alla intersezione della riga h con la colonna j.

Il sistema di equazioni che si è ora descritto considera come espressione generale la presenza di più funzioni di variabili diverse. Tale forma, essendo generale, è quindi valida per qualunque tipo di funzioni lineari nei parametri, in cui compaiano più variabili oppure la stessa variabile, ma con potenze successive.

Va inoltre rilevato che lo stesso metodo può essere applicato anche quando un dato valore della variabile Y corrispondente ad un dato valore della variabile X si presenta più di una volta, cioè con una frequenza nota: infatti, è possibile anche tener conto del fatto che certe combinazioni di valori delle variabili possono avere una frequenza superiore all'unità (*interpolazione ponderata*). In questo caso, quando è stato detto più sopra conserva tutta la sua validità con la sola complicazione che nelle formule già indicate figureranno anche le frequenze con cui si presentano i valori distinti della variabile Y. In base a quanto ora detto, se indichiamo con n_i la frequenza con cui si presenta l'associazione di valori (y_i, x_i, z_i, t_i) , con $i = 1, 2, \dots, k$, e ponendo

$$N = \sum_{i=1}^{i=k} n_i$$

la condizioni di minimo, nella sua espressione più generale, si scriverà:

$$\sum_{i=1}^{i=k} [y_i - b_0 \cdot f(x_i) - b_1 \cdot \varphi(z_i) - b_2 \cdot \Psi(t_i) - \dots\dots\dots]^2 \cdot n_i = \text{minimo}$$

e il sistema risolvente assumerà la forma:

$$\begin{cases} b_0 \cdot \Sigma f^2(x_i) \cdot n_i + b_1 \cdot \Sigma f(x_i) \cdot \varphi(z_i) \cdot n_i + b_2 \cdot \Sigma f(x_i) \cdot \psi(t_i) \cdot n_i + \dots\dots\dots = \Sigma y_i \cdot f(x_i) \cdot n_i \\ b_0 \cdot \Sigma f(x_i) \cdot \varphi(z_i) \cdot n_i + b_1 \cdot \Sigma \varphi^2(z_i) \cdot n_i + b_2 \cdot \Sigma \varphi(z_i) \cdot \psi(t_i) \cdot n_i + \dots\dots\dots = \Sigma y_i \cdot \varphi(z_i) \cdot n_i \\ b_0 \cdot \Sigma f(x_i) \cdot \psi(t_i) \cdot n_i + b_1 \cdot \Sigma \varphi(z_i) \cdot \psi(t_i) \cdot n_i + b_2 \cdot \Sigma \psi^2(t_i) \cdot n_i + \dots\dots\dots = \Sigma y_i \cdot \psi(t_i) \cdot n_i \\ \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots = \dots\dots\dots \end{cases}$$

La soluzione generica di questo sistema può essere scritta nella forma

$$b_j = \frac{1}{D} [A_{1j} \cdot \sum_{i=1}^{i=k} y_i \cdot f(x_i) \cdot n_i + A_{2j} \cdot \sum_{i=1}^{i=k} y_i \cdot \varphi(z_i) \cdot n_i + A_{3j} \cdot \sum_{i=1}^{i=k} y_i \cdot \Psi(t_i) \cdot n_i + \dots\dots]$$

in cui D è il determinante formato dai coefficienti delle incognite del precedente sistema e cioè è

$$D = \begin{vmatrix} \Sigma f^2(x_i) n_i & \Sigma f(x_i)\varphi(z_i) n_i & \Sigma f(x_i)\psi(t_i) n_i & \dots\dots\dots \\ \Sigma f(x_i)\varphi(z_i) n_i & \Sigma \varphi^2(z_i) n_i & \Sigma \varphi(x_i)\psi(t_i) n_i & \dots\dots\dots \\ \Sigma f(x_i)\psi(t_i) n_i & \Sigma \varphi(x_i)\psi(t_i) n_i & \Sigma \psi^2(t_i) n_i & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \end{vmatrix}$$

mentre A_{hj} è il cofattore o aggiunto dell'elemento che si trova alla intersezione della riga h con la colonna j.

Esistono vari criteri per mezzo dei quali si cerca di semplificare i calcoli necessari per pervenire alla stima del valore numerico dei parametri. Tali criteri sono specialmente utili quando si debbano utilizzare sistemi relativi a funzioni con molti parametri (*interpolazione con funzioni ortogonali*). Ma, per una più diffusa trattazione dei vari criteri, così come dei diversi metodi di interpolazione, si rimanda a pubblicazioni specializzate.

Quello che, invece, preme qui mettere in evidenza, è che il metodo finora illustrato è utile ogni qualvolta si sia in grado di utilizzare dei modelli uniequazionali che sono *lineari nei parametri* oppure che, adottando opportune trasformazioni delle variabili originarie, possono essere ricondotti a forme nelle quali i parametri risultano espressi in modo lineare. E' evidente che in quest'ultimo caso, la condizione di minimo non è rispettata nei riguardi dei valori delle variabili originarie, ma rispetto a quelli delle variabili trasformate.

Esempio V.2

Come esempio si consideri l'applicazione del metodo dei minimi quadrati ad una trasformata della $f(x)$, supponendo che la funzione contenente i parametri sia in forma non lineare, ma riducibile ad una forma lineare nei parametri mediante una opportuna trasformazione di variabile.

Consideriamo la funzione esponenziale

$$y = c_0 \cdot c_1^x$$

Con la trasformazione logaritmica si ottiene $\log y = \log c_0 + x \cdot \log c_1$, dalla quale si vede che $\log y$ è funzione lineare dei parametri $c'_0 = \log c_0$ e $c'_1 = \log c_1$.

Applicando il metodo dei minimi quadrati, la condizione di minimo risulterà essere:

$$\phi = \sum (\log y_i^* - \log y_i)^2 = \sum (c'_0 + c'_1 \cdot x_i - \log y_i)^2 = \text{minimo}$$

e non la condizione

$$\sum (y_i^* - y_i)^2 = \text{minimo}.$$

Poiché, però, un buon adattamento dei logaritmi non sempre assicura un altrettanto buon adattamento tra i valori originari y_i^* e y_i diviene indispensabile un controllo a posteriori di tale condizione.

Risolviendo il sistema delle due derivate

$$\frac{\delta \phi}{\delta c_0} = \sum 2 \cdot (c'_0 + c'_1 \cdot x_i - \log y_i) \cdot 1 = 0 \quad \text{e} \quad \frac{\delta \phi}{\delta c_1} = \sum 2 \cdot (c'_0 + c'_1 \cdot x_i - \log y_i) \cdot x_i = 0$$

rispetto ai parametri della funzione da minimizzare si ottiene la soluzione che segue:

$$\log c_0 = \frac{(\sum x^2)(\sum \log y) - (\sum x)(\sum x \cdot \log y)}{n(\sum x^2) - (\sum x)^2};$$

$$\log c_1 = \frac{n(\sum x \cdot \log y) - (\sum x)(\sum \log y)}{n(\sum x^2) - (\sum x)^2}.$$

4. Metodo della massima verosimiglianza

Anche questo è un metodo di che impone alla funzione interpolatrice di passare *tra* i valori osservati.

La distribuzione dei valori osservati, che si vuole rappresentare tramite il modello ipotizzato, può essere considerata come un campione possibile delle distribuzioni di ugual tipo che potrebbero ottenersi, se si avesse la possibilità di ripetere l'esperienza. Ad esempio, se si considera una serie di dati storici relativi agli investimenti in abitazioni effettuati in una certa area ed alla popolazione residente nell'area, tale serie potrebbe essere considerata come un campione, tratto da un universo di informazioni analoghe, relative ad un periodo storico più ampio di quello sotto osservazione. Mutando il campione, ovviamente, cambia la variabile statistica che esso rappresenta e quindi varia la stima dei parametri del modello, ciascuna delle quali è funzione di tutti i valori osservati.

La statistica moderna tende a studiare la distribuzione di queste ultime per poter ricavare quella che si può ritenere la più prossima al valore "vero" del parametro, intendendosi per "vero" quel valore che si otterrebbe per il parametro se si potessero considerare tutte le determinazioni dell'universo osservato, dal quale si considera tratto il campione di informazioni su cui invece si opera in concreto.

A tale scopo, se ci si fonda sulla teoria degli errori di osservazione e se la distribuzione campionaria di ogni stima c_j è normale, come vero valore del parametro γ_j si può assumere il suo valore medio $M(c_j)$. Se, invece, la distribuzione campionaria non è normale, allora viene meno la ragione teorica di scelta di $M(c_j)$ come valore vero, anche se, in pratica, si continua, generalmente, ad assumere $M(c_j)$ come il valore più prossimo a quello riguardante la massa o universo delle determinazioni.

Da quanto precede, si comprende l'importanza che assume, nella statistica moderna, il tener presente, nella scelta dei metodi di calcolo dei parametri, la possibilità di ottenere stime che, variando il campione e facendo crescere le sue dimensioni ossia il numero dei casi o determinazioni, si

distribuiscono il più possibile secondo la legge di distribuzione degli errori accidentali o di Gauss e, per di più, con la minima varianza.

Le stime che soddisfano queste due condizioni sono dette stime "efficienti". Il Fisher ed altri autori, in una serie di lavori in materia probabilistica, hanno cercato di esaminare gran parte dei vari metodi di calcolo dei parametri, dal punto di vista della possibilità di dare stime efficienti.

Tra questi, quello senza dubbio migliore è proprio il *metodo della massima verosimiglianza*, il quale si fonda essenzialmente su considerazioni di carattere probabilistico. Senza addentrarci in tale campo, che richiede una preparazione specifica, si può giungere egualmente alla *individuazione della funzione di verosimiglianza minimizzando la media geometrica ponderata degli scarti geometrici tra valori osservati e valori teorici* (y_i / y'_i) , anziché la somma dei quadrati degli scarti aritmetici $(y_i - y'_i)$, come nel metodo dei minimi quadrati, assumendo come pesi i valori y_i .

Detta condizione può scriversi come segue:

$$\left\{ \prod_{i=1}^{i=k} \left[\frac{y_i}{f(x_i; c_0, c_1, c_2, \dots, c_{s-1})} \right]^{y_i} \right\}^{\frac{1}{\sum y_i}} = \text{minimo}$$

dalla quale, prendendo i valori logaritmici, si ottiene

$$\frac{1}{\sum y_i} \left[\sum_{i=1}^{i=k} y_i \cdot \log y_i - \sum_{i=1}^{i=k} y_i \cdot \log f(x_i; c_0, c_1, c_2, \dots, c_{s-1}) \right] = \text{minimo}$$

In questa relazione la funzione $f(x_i; c_0, c_1, c_2, \dots, c_{s-1})$ rappresenta il modello o i valori teorici della variabile Y, le cui determinazioni osservate sono le y_i . Ma, poiché i parametri incogniti figurano soltanto nel secondo termine della relazione suddetta, il minimo della espressione si otterrà quando sarà

$$\sum_{i=1}^{i=k} y_i \cdot \log f(x_i; c_0, c_1, c_2, \dots, c_{s-1}) = \text{massimo}$$

Per la determinazione del massimo di quest'ultima funzione in più incognite, cioè i parametri del modello, dovranno essere annullate le derivate

parziali prime rispetto a $c_0, c_1, c_2, \dots, c_{s-1}$ e cioè dovrà essere soddisfatto il sistema

$$\begin{cases} \sum_{i=1}^{i=k} \frac{y_i}{f(x_i)} \frac{\partial f(x_i)}{\partial c_0} = 0 \\ \sum_{i=1}^{i=k} \frac{y_i}{f(x_i)} \frac{\partial f(x_i)}{\partial c_1} = 0 \\ \sum_{i=1}^{i=k} \frac{y_i}{f(x_i)} \frac{\partial f(x_i)}{\partial c_2} = 0 \\ \dots\dots\dots \\ \sum_{i=1}^{i=k} \frac{y_i}{f(x_i)} \frac{\partial f(x_i)}{\partial c_{s-1}} = 0 \end{cases}$$

dove, per semplicità, si è scritto $f(x_i)$ in luogo di $f(x_i; c_0, c_1, c_2, \dots, c_{s-1})$.

I valori $c_0, c_1, c_2, \dots, c_{s-1}$ che annullano le derivate parziali prime del sistema suddetto soddisfano anche le condizioni di minimo poste all'inizio. Va tuttavia osservato, che non sempre è possibile ricavare facilmente i parametri incogniti, perché questo metodo presenta talvolta lo stesso difetto del metodo dei minimi quadrati, cioè, quello di fornire equazioni non risolvibili per via algebrica.

Esempio V.3

Si voglia rappresentare una distribuzione statistica con una curva normale di equazione:

$$y = \frac{1}{c_0 \sqrt{2\pi}} e^{-\frac{(x-c_1)^2}{2c_0^2}}$$

nella quale, come è noto, c_0 è lo scarto quadratico medio e c_1 è la media aritmetica della popolazione di origine, entrambi incogniti.

La funzione da massimizzare ha l'espressione che segue:

$$\log L = \sum y_i \log \left(\frac{1}{c_0 \sqrt{2\pi}} e^{-\frac{(x-c_1)^2}{2c_0^2}} \right) = -\sum y_i \log(c_0 \sqrt{2\pi}) + \sum y_i \log \left(e^{-\frac{(x-c_1)^2}{2c_0^2}} \right) =$$

$$= -\sum y_i \log c_0 - \sum y_i \log \sqrt{2\pi} - \sum y_i \frac{(x_i - c_1)^2}{2c_0^2}$$

Uguagliando a zero le derivate parziali rispetto a c_0 e c_1 si ha:

$$\begin{cases} \frac{\partial \log L}{\partial c_0} = -\frac{\sum y_i}{c_0} + \frac{\sum y_i (x_i - c_1)^2}{c_0^3} = 0 \\ \frac{\partial \log L}{\partial c_1} = \frac{\sum y_i (x_i - c_1)}{c_0^2} = 0 \end{cases}$$

La prima di queste due relazioni fornisce

$$c_0^2 \sum y_i = \sum y_i (x_i - c_1)^2$$

e la seconda

$$c_1 = \frac{\sum y_i x_i}{\sum y_i} = \bar{x}$$

Quindi, sostituendo si ha, infine:

$$c_0^2 = \frac{\sum y_i (x_i - \bar{x})^2}{\sum y_i} = \sigma_x^2$$

5. Metodo delle somme

Il metodo delle somme costituisce l'applicazione più semplice della condizione generale per la determinazione dei parametri di una funzione che consiste nell'imporre alla funzione interpolatrice che vi sia *eguaglianza tra convenienti ed opportune operazioni fatte sui valori rilevati e le stesse operazioni ripetute sui valori teorici ottenuti con la funzione.*

In questo metodo, la condizione generale che si impone di rispettare impone che la *somma dei valori effettivi rilevati* per certi valori della variabile dipendente in corrispondenza di certi valori della variabile indipendente deve essere *uguale alla somma dei valori teorici* della variabile dipendente in corrispondenza degli stessi valori della variabile indipendente.

Si supponga che in una distribuzione i dati non siano espressi singolarmente, ma ad esempio, in corrispondenza dei valori x_1, x_2, x_3 sia data la somma $S_1=y_1+ y_2+ y_3$; in corrispondenza dei valori x_4, x_5, x_6, x_7 sia data la

somma $S_2=y_4+ y_5+ y_6+ y_7$; e così via. In questo caso è possibile esprimere una relazione tra i valori effettivi e quelli teorici del seguente tipo:

$$f(x_1; c_0, c_1, \dots, c_h) + f(x_2; c_0, c_1, \dots, c_h) + f(x_3; c_0, c_1, \dots, c_h) = S_1$$

e così per le altre somme date, in modo da avere tante equazioni quanti sono i parametri da determinare. Quando il numero dei parametri è inferiore al numero delle condizioni da soddisfare, si presenta l'arbitrio di scegliere, tra le condizioni date, quelle su cui operare. Tale scelta dipenderà ovviamente dalla conoscenza del fenomeno che ha l'operatore, tenuto conto del fatto che a volte può essere vantaggioso tralasciare quelle condizioni o quei valori che si ritengono poco sicuri.

A questo metodo si fa ricorso anche quando siano dati i valori singoli, in quanto, essendo possibile il loro raggruppamento, generalmente si formano *tanti gruppi quanti sono i parametri da determinare.* La scelta dei gruppi si fa in modo che vengano compensati, almeno in parte, gli errori accidentali. Come criterio generale, si cerca di fare *gruppi che siano all'incirca egualmente numerosi.*

I valori dei parametri che si ottengono variano generalmente cambiando i raggruppamenti, ma se questi sono scelti bene la differenza tra i valori delle stime spostano di poco la curva geometrica nel piano.

Il metodo delle somme è largamente usato, per la sua semplicità, quando si ha la sensazione che i valori rilevati siano alquanto rudimentali ed imprecisi e che gli errori si compensino all'interno dei singoli gruppi considerati.

Esempio V.4

Un'azienda ha incassato tra il 1985 ed il 1996 le somme seguenti, in miliardi di lire:

1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
2	4	7	9	12	15	17	20	21	25	27	30

Vogliamo interpolare la precedente serie storica col metodo delle somme mediante una retta di equazione $y=ax+b$. Introdotta le ascisse di lavoro -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, mediante la traslazione $z=x-1990$, raggruppiamo i valori in due classi 1985-1990 e 1991-1996.

Per determinare i parametri a e b basta scrivere le due relazioni:

$$\begin{aligned} 2+4+7+9+12+15 &= 5a+b-4a+b-3a+b-2a+b-a+b+b \\ 17+20+21+25+27+30 &= a+b+2a+b+3a+b+4a+b+5a+b+6a+b \end{aligned}$$

da cui, semplificando e risolvendo si ottiene $a=2,53$ e $b=14,49$. Pertanto l'equazione della retta interpolatrice è:

$$y=2,53z+14,49=2,53(x-1990)+14,49=2,53x-5020,21.$$

6. Metodo delle aree o di Cantelli

Questo metodo è indicato quando i valori osservati non si riferiscono a valori singoli, ma ad intervalli di valori della variabile indipendente e precisamente alla variabile statistica divisa in intervalli la cui rappresentazione grafica costituisce un istogramma.

Si fissa una funzione $y = f(x; c_0, c_1, \dots, c_s)$, poi in base a questa si scrive una serie di eguaglianze simili a quelle del metodo delle somme, solo che i due membri dell'eguaglianza anziché esprimere delle ordinate esprimono delle aree.

Tali relazioni che legano l'area racchiusa in determinati intervalli e i valori osservati in ciascun intervallo sono del tipo:

$$\int_{x_i}^{x_{i+1}} f(x; c_0, c_1, \dots, c_{s-1}) dx = y_i, \text{ con } i = 0, 1, 2, \dots, (s-1)$$

nelle quali si presuppone, per la concreta utilizzazione del metodo, che l'integrazione si sappia esprimere in termini finiti così che il sistema delle relazioni del tipo suddetto diventi un sistema nei parametri incogniti c_i .

Se il numero delle classi supera quello dei parametri, bisognerà scegliere tante classi, tra quelle date, quanti sono i parametri, oppure fare delle classi più ampie, riunendo più classi date in un'unica classe. Sull'arbitrarietà di tale scelta valgono le stesse considerazioni fatte a proposito del metodo delle somme.

Il metodo delle aree, come anche quello delle somme, vengono spesso adoperati per la loro semplicità, per decomporre classi ampie (ad esempio poliennali) in classi più ristrette (ad esempio, annuali).

Esempio V.5

Si abbia la distribuzione della produzione di acciaio (milioni di tonnellate) per classi di capacità produttiva (migliaia di tonnellate) degli stabilimenti descritta dalla tavola seguente:

Capacità produttiva $x_i \text{ --- } x_{i+1}$	Produzione y_i
10 --- 15	4,21
15 --- 20	4,00
20 --- 25	4,03
25 --- 30	3,89

Come sempre in questi casi conviene effettuare un cambiamento di variabile ponendo:

$$z = \frac{x - 20}{5}$$

La variabile Z è quindi una trasformata affine della variabile X. Le nuove ascisse di lavoro sono, perciò, -2, -1, 0, 1, 2.

Vogliamo interpolare la distribuzione con una parabola di secondo grado, sotto la condizione che l'area delimitata dalle rette parallele all'asse delle ordinate per (-2,0), (-1, 0) e dalla curva interpolatrice sia y_0 , quella delimitata dalla curva interpolatrice e dalle rette per (-1, 0) e (0, 0) sia y_1 , e quella delimitata dalla curva interpolatrice e dalle rette per (0, 0) e (2, 0) sia y_2+y_3 . Rappresentando la distribuzione data con la parabola di secondo grado si possono scrivere tre relazioni lineari in c_0, c_1 e c_2 , che permettono di ricavare questi parametri incogniti. Si ha:

$$\begin{cases} \int_{-2}^{-1} (c_0 + c_1 z + c_2 z^2) dz = 4,21 \\ \int_{-1}^0 (c_0 + c_1 z + c_2 z^2) dz = 4,00 \\ \int_0^2 (c_0 + c_1 z + c_2 z^2) dz = 4,03 + 3,89 = 7,92 \end{cases}$$

Risolviendo gli integrali si ha:

$$\begin{cases} \left[c_0 z + \frac{c_1 z^2}{2} + \frac{c_2 z^3}{3} \right]_{-2}^{-1} = 4,21 \\ \left[c_0 z + \frac{c_1 z^2}{2} + \frac{c_2 z^3}{3} \right]_{-1}^0 = 4,00 \\ \left[c_0 z + \frac{c_1 z^2}{2} + \frac{c_2 z^3}{3} \right]_0^2 = 7,92 \end{cases}$$

e semplificando si ricava:

$$\begin{cases} 6.c_0 - 9.c_1 + 14.c_2 = 25,26 \\ 6.c_0 - 3.c_1 + 2.c_2 = 24,00 \\ 6.c_0 + 6.c_1 + 8.c_2 = 23,76 \end{cases}$$

da cui, trattandosi di un sistema lineare, si ottiene la soluzione:

$$\begin{cases} c_0 = 3,94 \\ c_1 = -0,0725 \\ c_2 = 0,06875 \end{cases}$$

7. Metodo dei momenti

Questo metodo è particolarmente adatto quando la funzione scelta per la rappresentazione analitica non sia lineare nei parametri, oppure quando vengono meno i motivi che fanno preferire il metodo dei minimi quadrati.

Come condizione generale, si pone una *eguaglianza tra i momenti teorici ed i momenti empirici* della funzione. Come si è visto si dice *momento empirico di ordine r e origine 0* la quantità

$$m_r = \sum_{i=1}^{i=s} x_i^r y_i$$

contenente le x_i e le y_i note. Se le y_i sono frequenze assolute, si pone $1/N$ davanti alla sommatoria. Casi particolari sono il momento empirico di ordine zero che è uguale a $\sum y_i$ e il momento empirico di ordine uno, che è uguale a $\sum x_i y_i$; quest'ultimo coincide con la media ponderata dei valori x_i .

Si dice *momento teorico di ordine r e origine 0* la quantità

$$\mu_r = \int_{\alpha}^{\beta} x^r f(x) dx,$$

ove α -- β è l'intervallo di definizione della variabile indipendente X e $y = f(x)$ è la funzione di densità contenente le incognite c_0, c_1, \dots, c_{s-1} .

Le relazioni su cui si fonda il metodo dei momenti nascono dall'eguaglianza $\mu_r = m_r$, tra momenti teorici e momenti empirici, facendo assumere ad r successivamente i valori $0, 1, 2, \dots, (s-1)$ ossia fino ad un ordine tale che il numero delle equazioni coincida col numero dei parametri da determinare, i quali scaturiscono dalla risoluzione dell'integrale.

Questa eguaglianza non è corretta in quanto pone a confronto un momento calcolato con la sommatoria e un momento calcolato con l'integrale. A questo inconveniente si rimedia o correggendo μ_r con l'uso delle correzioni di Sheppard, oppure sostituendo ai μ_r le quantità

$$m_r^* = \sum_{i=0}^{i=s-1} x_i^r y_i^*$$

in cui

$$y_i^* = \int_{x_{i-d/2}}^{x_{i+d/2}} f(x; c_0, c_1, \dots, c_{s-1}) dx$$

dove d è l'ampiezza della classe x_i -- x_{i+1} .

L'eguaglianza tra momenti teorici e momenti empirici porta quindi ad un sistema del tipo

$$\begin{cases} \sum y_i^* = \sum y_i \\ \sum x_i y_i^* = \sum x_i y_i \\ \sum x_i^2 y_i^* = \sum x_i^2 y_i \\ \dots = \dots \\ \sum x_i^{s-1} y_i^* = \sum x_i^{s-1} y_i \end{cases}$$

in cui i parametri incogniti figurano in y_i^* .

Per esplicitare questi parametri incogniti occorre allora risolvere gli integrali in termini finiti. Il metodo dei momenti non è subordinato alla

limitazione di avere una $f(x)$ lineare nei coefficienti, perché basta che dalla suddetta integrazione risulti un sistema nei parametri incogniti che si sappia risolvere.

Se la funzione interpolatrice è un polinomio completo, il metodo dei momenti dà gli stessi risultati del metodo dei minimi quadrati. Ciò non è vero se il polinomio non è completo.

Il metodo dei momenti è adatto ad interpolare la curva normale come anche le curve del Pearson, in quanto per esse si riesce ad esprimere l'integrale della $y = f(x)$ in termini piuttosto comodi per determinare i parametri ricercati. Va però osservato che i momenti empirici di ordine elevato (ad esempio a partire da m_4) sono molto sensibili alla grandezza degli errori contenuti nei dati osservati, per cui in pratica è preferibile evitare di usare questo metodo quando i parametri da calcolare siano più di quattro.

Esempio V.6

Come applicazione consideriamo la distribuzione normale definita da $-\infty$ a $+\infty$:

$$y = \frac{c_0}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

la quale contiene due parametri c_0 e σ e pertanto dovrebbe bastare conoscere i primi due momenti della funzione. Ma in questa funzione $\mu_0=c_0$ e $\mu_1=0$ e dato che il primo momento non contiene alcun parametro della funzione, occorre utilizzare il secondo momento che è $\mu_2=c_0\sigma^2$. Il sistema che si forma uguagliando i momenti teorici a quelli empirici è quindi:

$$\begin{cases} c_0 = \sum y_i \\ c_0\sigma^2 = \sum x_i^2 y_i \end{cases}$$

dal quale si ricavano immediatamente c_0 e σ in funzione degli elementi noti x_i e y_i .

8. Accorgimenti per i sistemi di equazioni simultanee

Nella generalità dei casi le relazioni tra più variabili sono piuttosto complesse e possono essere espresse in maniera più idonea attraverso

l'uso di sistemi di più equazioni, ciascuna contenente parecchie variabili indipendenti e dipendenti, le quali possono anche figurare in più di una equazione del sistema

generale. E' questo dunque un caso ben più generale di quello considerato finora delle equazioni di regressione singole, con una variabile dipendente ed uno o più variabili indipendenti.

Nel caso di un sistema di equazioni, si potrebbe risolvere il problema della stima dei parametri anche trattando le equazioni *una alla volta* e stimando i parametri che vi figurano con le usuali tecniche della regressione multipla. Tuttavia, alcune importanti considerazioni sconsigliano una applicazione così meccanica di questa tecnica, dato che si corre il rischio di avere dei coefficienti privi di significato in certi casi, o che forniscono stime non accettabili del valore "vero" dei parametri in altri casi. Occorre ricordare, infatti, che per il momento si sta considerando l'ipotesi che le variabili indipendenti di ciascuna equazione siano misurate correttamente e che solo le misure della variabile dipendente siano affette da errori. Nel caso dei modelli pluriequazionali, sorgono così due distinti tipi di problemi e cioè il *problema dell'identificazione delle relazioni* e quello della *stima dei parametri*.

Va subito precisato, infatti, che, quando la variabile dipendente di una equazione figura come variabile esplicativa o indipendente in un'altra equazione o in più altre equazioni dello stesso sistema, ci si trova in condizioni alquanto diverse da quelle finora esaminate per i modelli uniequazionali.

Quando si devono trattare dei sistemi di equazioni si possono presentare i due casi seguenti:

- *quando in equazioni diverse di uno stesso insieme di relazioni si presentano tutte o soltanto alcune delle variabili esplicative o indipendenti, ma nessuna delle variabili endogene al sistema, tranne quelle che ciascuna relazione cerca di spiegare, non sorge alcun problema di identificazione o di stima simultanea dei parametri e le equazioni possono essere risolte una per una applicando gli usuali procedimenti di stima per equazioni singole;*

- *quando, per contro, in un sistema di relazioni alcune o tutte le variabili che debbono essere spiegate con l'analisi di regressione (cioè le variabili dipendenti) appaiono anche come variabili esplicative (cioè come variabili indipendenti) in alcune altre equazioni del sistema, allora la distinzione tra variabili dipendenti e variabili indipendenti che si ha nel caso di una sola*

equazione non è più valida per il sistema nel suo complesso. In tale caso, invece, occorre fare una

distinzione tra *variabili congiuntamente dipendenti*, per ciascuna delle quali va ricercata l'influenza determinata su ogni altra variabile, e *variabili predeterminate*, le quali esercitano una influenza sulle variabili congiuntamente dipendenti, senza essere a loro volta influenzate da queste ultime.

Il fatto che una variabile sia considerata endogena (dipendente) o esogena (indipendente) dipende esclusivamente dallo schema teorico o modello che si è adottato. Ad esempio, in un modello economico potranno essere considerate invariabilmente esogene le variabili non economiche come le condizioni climatiche, mentre altre variabili economiche possono essere trattate come endogene in un modello ed esogene in un altro modello.

In un sistema di equazioni simultanee figurano tante equazioni quante sono le variabili endogene o dipendenti ed in esso può figurare un numero qualsiasi di variabili esogene o indipendenti o predeterminate, che possono essere numericamente di più o di meno di quelle endogene.

Dato che ciascuna variabile endogena deve comparire almeno una volta, si può scrivere formalmente ciascuna di esse al primo membro di una equazione del sistema, con coefficiente pari all'unità. Se si suppone di misurare tutte le variabili come deviazioni o scarti dalla loro media, allora con h variabili endogene o dipendenti y_1, y_2, \dots, y_h e k variabili esogene o indipendenti x_1, x_2, \dots, x_k , il sistema di equazioni simultanee può essere scritto esprimendo ciascuna variabile endogena come funzione delle altre variabili endogene e delle variabili esogene selezionate come importanti per il modello, ottenendo un sistema di h equazioni in $(h + k)$ variabili e $h(h - 1 + k)$ parametri connessi tra loro nella forma che segue:

$$\begin{cases} y_1 = \beta_{12} \cdot y_2 + \beta_{13} \cdot y_3 + \dots + \beta_{1h} \cdot y_h + \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \dots + \gamma_{1k} \cdot x_k + \varepsilon_1 \\ y_2 = \beta_{21} \cdot y_1 + \beta_{23} \cdot y_3 + \dots + \beta_{2h} \cdot y_h + \gamma_{21} \cdot x_1 + \gamma_{22} \cdot x_2 + \dots + \gamma_{2k} \cdot x_k + \varepsilon_2 \\ \dots \\ y_h = \beta_{h1} \cdot y_1 + \beta_{h2} \cdot y_2 + \dots + \beta_{h,h-1} \cdot y_{h-1} + \gamma_{h1} \cdot x_1 + \gamma_{h2} \cdot x_2 + \dots + \gamma_{hk} \cdot x_k + \varepsilon_h \end{cases}$$

in cui le variabili ε_i (per $i = 1, 2, \dots, h$) rappresentano le componenti stocastiche delle relazioni.

Queste equazioni sono chiamate *equazioni strutturali* del modello o sistema di equazioni. Esse mostrano le relazioni che si suppone esistano tra le variabili *sulla base di una determinata teoria* che può essere nota oppure ipotizzata e da verificare.

I coefficienti di regressione sono chiamati *coefficienti strutturali* del modello ed il loro numero, come si è detto, è pari ad $h \cdot (h - 1 + k)$.

Il sistema delle equazioni strutturali può essere riscritto anche come sistema di equazioni lineari delle h variabili dipendenti y_1, y_2, \dots, y_h espresse come funzioni lineari delle sole k variabili indipendenti x_1, x_2, \dots, x_k . In termini matriciali il sistema, infatti, potrebbe risciversi come segue:

$$\beta \cdot Y = \gamma \cdot X + \varepsilon$$

con β e γ matrici di ordine $h \times h$ la prima e $h \times k$ la seconda, mentre Y, X ed ε sono tre vettori colonna di ordine $h \times 1$ il primo ed il terzo ed $k \times 1$ il secondo.

Implicitamente, si suppone quindi che *nessuna di queste ultime equazioni possa essere espressa come funzione lineare delle altre*, perché in tal caso una equazione potrebbe essere ridondante o il sistema potrebbe essere contraddittorio. *Questa espressione equivale alla condizione che il determinante della matrice formata con i coefficienti delle variabili dipendenti sia diverso da zero ($\beta \neq 0$).*

Il sistema delle equazioni strutturali potrebbe essere risolto per le variabili y_1, y_2, \dots, y_h in modo che ciascuna di esse sia espressa solamente come funzione lineare delle variabili indipendenti o predeterminate, nel qual caso l'insieme delle equazioni risultanti può essere scritto come segue:

$$\begin{cases} y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \dots + \delta_{1k} \cdot x_k + \eta_1 \\ y_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + \dots + \delta_{2k} \cdot x_k + \eta_2 \\ \dots \\ y_h = \delta_{h1} \cdot x_1 + \delta_{h2} \cdot x_2 + \dots + \delta_{hk} \cdot x_k + \eta_h \end{cases}$$

in cui gli errori η_i sono funzioni lineari degli errori ε_i e, sia i coefficienti δ_{ij} sia gli errori η_i sono funzioni, generalmente non lineari, dei coefficienti β_{ij} e γ_{it} .

In forma matriciale avremo:

$$Y = \beta^{-1} \cdot (\gamma \cdot X + \varepsilon) = \delta \cdot X + \eta.$$

Quest'ultima trasformazione del sistema iniziale è chiamata *forma ridotta* del sistema di equazioni strutturali e gli h.k coefficienti δ_{ij} sono chiamati *coefficienti della forma ridotta*.

Dunque, in questo caso si possono far regredire le y_i singolarmente prese sulle x_1, x_2, \dots, x_k , considerate come insieme, ottenendo così le stime d_{ij} dei coefficienti della forma ridotta con il metodo dei minimi quadrati ordinari. Infatti, in ogni equazione figurerebbe una sola variabile dipendente (affetta da errori) e tutte o alcune variabili indipendenti (non affette da errori).

Per contro, *quando il metodo dei minimi quadrati ordinari viene applicato a ciascuna singola equazione della forma strutturale, in generale può succedere che i risultati ottenuti possono non produrre affatto le stime dei coefficienti strutturali*, ma soltanto stime distorte che spesso sono dotate anche di ampi margini di incertezza.

Ciò accade per il fatto che qualche variabile endogena figura come variabile indipendente in qualche equazione, mentre i suoi valori sono affetti da errore e pertanto danno luogo a stime dei parametri anch'esse affette da errore. Ciò che non avviene con le variabili indipendenti, i cui valori sono per ipotesi misurati correttamente e non affetti da errori di osservazione e di misura.

D'altra parte, *non sempre il problema della stima dei coefficienti strutturali può essere risolto con il metodo dei minimi quadrati "indiretti", operando, cioè, sulla forma ridotta del sistema*, dato che il modo in cui si può passare dall'una all'altra forma delle relazioni dipende dal modo in cui può essere risolto il problema della identificazione delle relazioni.

9. Stima dei parametri per equazioni sottoidentificate

Per esporre in maniera piana e chiara il problema della identificazione, consideriamo il caso in cui sia $h = 2$ e k qualunque, il che equivale a considerare un sistema di due equazioni con due variabili endogene, ma con un numero qualsiasi di variabili indipendenti. Il sistema strutturale assume allora la forma:

$$\begin{cases} y_1 = & \beta_{12} \cdot y_2 + \gamma_{11} \cdot x_1 + \dots + \gamma_{1k} \cdot x_k + \varepsilon_1 \\ y_2 = \beta_{21} \cdot y_1 & + \gamma_{21} \cdot x_1 + \dots + \gamma_{2k} \cdot x_k + \varepsilon_2 \end{cases}$$

Poiché y_1 e y_2 sono *congiuntamente dipendenti*, non vi è alcuna ragione ovvia per scrivere y_2 piuttosto che y_1 al primo membro della seconda equazione. Se si adotta il semplice accorgimento di trasferire il termine y_2 a secondo membro, il termine $\beta_{21} \cdot y_1$ al primo membro e di dividere per $-\beta_{21}$, la seconda equazione si può scrivere:

$$y_1 = \frac{y_2}{\beta_{21}} - \frac{\gamma_{21} \cdot x_1}{\beta_{21}} - \dots - \frac{\gamma_{2k} \cdot x_k}{\beta_{21}} - \frac{\varepsilon_2}{\beta_{21}}$$

che è della stessa forma della prima equazione, ma con differenti coefficienti. Nel caso qui considerato può essere calcolata, però, soltanto una regressione di y_1 su y_2, x_1, \dots, x_k e non si ha modo di dire se i coefficienti di regressione possono essere interpretati come $\beta_{12}, \gamma_{11}, \dots, \gamma_{1k}$ oppure come $1/\beta_{21}, -\gamma_{21}/\beta_{21}, \dots, -\gamma_{2k}/\beta_{21}$.

Per rendere più complicate le cose, inoltre, le equazioni possono essere moltiplicate per fattori arbitrari λ_1 e λ_2 , sommate e divise sia per

$$\lambda_1 - \lambda_2 \beta_{12} \quad \text{sia per} \quad \lambda_2 - \lambda_1 \beta_{21}.$$

Con questa procedura, dunque, si può costruire un numero illimitato di nuove equazioni della stessa forma di quelle strutturali.

In questa situazione si parla di *sottoidentificazione* e si dice che *ambidue* le equazioni sono sottoidentificate. Ma si può anche verificare il caso che una equazione sia sottoidentificata e l'altra non lo sia.

Le equazioni della forma ridotta non sono mai sottoidentificate.

Esse nel caso esaminato assumono la forma già indicata in precedenza, cioè:

$$\begin{cases} y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \dots + \delta_{1k} \cdot x_k + \eta_1 \\ y_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + \dots + \delta_{2k} \cdot x_k + \eta_2 \end{cases}$$

Queste due equazioni sono ben distinte e le stime dei minimi quadrati d_{ij} relative ai coefficienti δ_{ij} hanno un significato ben definito.

Se vi fosse un solo modo nel quale combinare queste equazioni per ottenere una equazione come la prima o come la seconda delle equazioni strutturali, allora le stime dei coefficienti strutturali potrebbero derivarsi dalle stime dei coefficienti della forma ridotta. Ma vi sono molti modi di combinarle ed il metodo dei minimi quadrati indiretti fornisce, dunque, risultati non accettabili. E' anche chiaro che dai $2k$ coefficienti della forma ridotta non possono ricavarsi le stime dei $2(k+1)$ coefficienti strutturali.

A meno che non si introducano nuove ipotesi sulle componenti stocastiche ϵ_1 ed ϵ_2 , vi è soltanto un modo per superare tale difficoltà e cioè quello di *introdurre delle ipotesi a priori su alcuni coefficienti delle equazioni strutturali.*

Se una data variabile non svolge un ruolo importante in una relazione, una semplificazione potrebbe ottenersi adottando ad esempio una delle misure seguenti:

- il suo coefficiente può porsi uguale a zero, per cui la variabile sparisce dalla relazione;
- come alternativa, si può supporre che alcuni coefficienti siano uguali in valore;
- oppure alcuni coefficienti possono essere assoggettati a qualche altra restrizione da prefissare.

Nel modello considerato, ad esempio, la prima equazione strutturale è identificata se è $\beta_{12} = 0$, in quanto essa viene a coincidere con la sua forma ridotta e quindi si può procedere alla stima diretta dei coefficienti strutturali, adottando il metodo dei minimi quadrati ordinari. La equazione è identificata anche se uno qualunque dei $\gamma_{1j} = 0$ e $\gamma_{2j} \neq 0$, poiché in questo caso nessuna equazione simile può essere ottenuta come combinazione lineare delle due equazioni date. Infatti, una qualsiasi combinazione del genere dovrebbe avere coefficiente non nullo per x_j , mentre è per ipotesi $\gamma_{1j} = 0$.

In questo caso le stime dei coefficienti strutturali possono essere derivate dalle stime dei coefficienti della forma ridotta attraverso opportune combinazioni lineari delle equazioni della forma ridotta.

Si può quindi affermare che se una equazione è identificata, le stime dei suoi coefficienti strutturali possono essere ottenute da quelle dei coefficienti della forma ridotta, con opportune combinazioni lineari delle equazioni della forma ridotta.

Infatti, esprimendole in termini delle stime dei coefficienti ottenute con il metodo dei minimi quadrati, le equazioni della forma ridotta del sistema considerato diventano

$$\begin{cases} y_1 = d_{11} \cdot x_1 + d_{12} \cdot x_2 + \dots + d_{1k} \cdot x_k + v_1 \\ y_2 = d_{21} \cdot x_1 + d_{22} \cdot x_2 + \dots + d_{2k} \cdot x_k + v_2 \end{cases}$$

Se nel sistema delle equazioni strutturali è $\gamma_{1j} = 0$, si può cercare una combinazione lineare delle equazioni del sistema ridotto o ora trascritto tale che il coefficiente di y_1 sia 1 e quello di x_j sia 0. Se consideriamo la combinazione

$$\lambda_1 + \lambda_2 \cdot y_2$$

si avrà

$$\lambda_1 = 1 ; \quad \lambda_1 \cdot \delta_{1j} + \lambda_2 \cdot \delta_{2j} = 0 ; \quad \lambda_2 = - \delta_{1j} / \delta_{2j}.$$

L'equazione risultante allora fornisce le stime b_{12} di β_{12} e c_{1i} di γ_{1i} ($j \neq i$); per esempio

$$b_{12} = \frac{d_{1j}}{d_{2j}}; \quad c_{1i} = d_{1i} - \frac{d_{1j} \cdot d_{2i}}{d_{2j}}.$$

Similmente, se, per esempio, è $\gamma_{11} = \gamma_{12}$, ma è $\gamma_{21} \neq \gamma_{22}$, la prima equazione del sistema strutturale è identificata. Le stime dei suoi coefficienti si possono ottenere quindi dalla forma ridotta moltiplicando le equazioni rispettivamente per 1 e per λ cosicché, se è $\gamma_{11} = \gamma_{12}$, sarà

$$d_{11} + \lambda \cdot d_{21} = d_{12} + \lambda \cdot d_{22}$$

e quindi

$$\lambda = - (d_{11} - d_{12}) / (d_{21} - d_{22}).$$

per cui nell'equazione risultante i coefficienti di x_1 e x_2 saranno gli stessi.

Se una restrizione simile sui parametri si applica alla seconda equazione strutturale, entrambe le equazioni risultano identificate e si parla allora di sistema di equazioni identificate. Per esempio, con $k \geq 2$, $\gamma_{1k} = 0$ e $\gamma_{2k} = 0$ si otterrà un sistema identificato dal momento che $\gamma_{11} \neq 0$ e $\gamma_{2k} \neq 0$. Se non vi sono altre restrizioni sui parametri, si avranno tanti parametri strutturali quanti sono quelli della forma ridotta e tutti i parametri strutturali potranno essere derivati da quelli della forma ridotta.

Questo metodo di stima dei coefficienti strutturali è noto, come si è detto, con il nome di metodo dei *minimi quadrati indiretti* o come *metodo della forma ridotta*.

Nel sistema strutturale un caso particolare è quello in cui è $\beta_{12} = 0$, mentre $\beta_{21} \neq 0$ ed inoltre tutti i $\gamma_{ij} \neq 0$. In questo caso la prima equazione è identificata, ma la seconda non lo è. Nella seconda equazione la variabile y_1 può essere considerata come una delle variabili indipendenti insieme alle variabili x_1, x_2, \dots, x_k . La variabile y_1 non dipende da y_2 , ma secondo la prima equazione esiste una relazione tra le variabili $y_1, x_1, x_2, \dots, x_k$ che sono le variabili esplicative della seconda equazione. *La seconda equazione è sottoidentificata poiché vi è multicollinearità tra le variabili e pertanto non è possibile stimare i singoli coefficienti*, cioè le stime che si potrebbero ottenere non sarebbero

significative

perchè la presenza di multicollinearità tra le variabili esplicative determina la incertezza dei parametri dato il loro elevato errore standard. Ciò mostra il legame che intercorre tra la multicollinearità e la sottoidentificazione; cioè, è la *esistenza di troppe relazioni tra le stesse variabili che rende impossibile la stima*.

I concetti di identificazione e di sottoidentificazione possono essere generalizzati facilmente al caso di sistemi formati da più di due equazioni. Tenuto conto della struttura del modello, ogni equazione può essere identificata o meno, ma se non lo è, l'identificazione può essere ottenuta, con le usuali ipotesi, soltanto tramite l'omissione di alcune variabili oppure aggiungendo altre restrizioni sui parametri.

Si potrebbe pensare a tutta prima che l'omissione di una variabile diversa in ogni equazione sia sufficiente a rendere identificato l'intero sistema, ma ciò non è vero. *In un sistema di h equazioni, ve ne sono $(h-1)$, oltre a quella allo studio, che possono essere combinate linearmente per dare una equazione con $(h-2)$ coefficienti nulli o con altre restrizioni sui parametri. Se la prima equazione ha meno di $(h-1)$ coefficienti nulli (o restrizioni), può essere ottenuta un'equazione di forma simile tramite una combinazione lineare e quindi l'equazione non è identificata. Una condizione necessaria per l'identificazione è perciò l'esistenza di almeno $(h-1)$ restrizioni sui parametri dell'equazione in esame.* Poiché il coefficiente di una variabile dipendente è reso uguale all'unità, ciò implica che per poter effettuare la stima in ciascuna equazione figurino non più di k parametri diversi.

Il caso più semplice è quello in cui l'insieme dei vincoli o restrizioni posti origina il necessario numero di coefficienti nulli. In tal caso, infatti, la condizione ulteriore che consente di completare il sistema delle condizioni necessaria e sufficiente può essere formulata come segue: si individua l'insieme delle variabili escluse dalla equazione in esame e si considerano i coefficienti con cui esse appaiono nelle rimanenti equazioni, con i quali si può formare almeno un determinante non nullo di ordine $(h-1)$, cioè con $(h-1)$ righe e colonne; se non è possibile trovare un determinante non nullo di ordine $(h-1)$, allora vuole dire che si può costruire una equazione in cui le variabili sotto esame hanno coefficienti nulli e quindi viene a cadere il fondamento del problema

dell'identificazione. Si può pertanto dire che *condizione sufficiente per la identificabilità di una equazione è che sia possibile individuare un determinante non nullo di ordine $(h-1)$ nella matrice*

formata con i coefficienti che le variabili escluse dalla equazione in esame hanno nelle rimanenti equazioni.

Nel caso generale di restrizioni sui parametri non vi sono regole così semplici ed è necessario usare molta attenzione per accertare la identificabilità delle equazioni strutturali.

E' anche possibile dedurre le proprietà della identificazione dalla forma ridotta. *Se un'equazione è identificata, i parametri strutturali possono essere stimati con l'ausilio della forma ridotta; se essa è sottoidentificata ciò non è possibile.* Con h variabili dipendenti e k variabili indipendenti, un sistema di h equazioni, nessuna delle quali contenga più di k parametri, è di solito completamente identificato.

10. Stima dei parametri di equazioni esattamente identificate

In questo paragrafo, si supporrà che tutte le equazioni che figurano nell'insieme studiato siano esattamente identificate, dato che se tale ipotesi non è soddisfatta è praticamente impossibile una corretta stima delle equazioni strutturali, anche se le equazioni della forma ridotta possono essere ancora studiate.

I parametri delle equazioni strutturali identificate possono essere stimati direttamente con il metodo dei minimi quadrati applicato alle singole equazioni (SELS = single equation least squares) e da essi possono poi essere dedotte le stime dei coefficienti della forma ridotta. Oppure si può procedere prima alla stima dei coefficienti della forma ridotta utilizzando il metodo dei minimi quadrati indiretti (ILS = indirect least squares) e poi, in qualche caso, si possono derivare le stime dei coefficienti strutturali. I due insiemi di stime in generale differiranno l'uno dall'altro, come può essere dimostrato dal seguente semplice esempio:

$$\begin{cases} y_1 = \gamma \cdot x + \varepsilon_1 \\ y_2 = \beta \cdot y_1 + \varepsilon_2 \end{cases}$$

in cui la prima equazione coincide con la sua forma ridotta, mentre la forma ridotta della seconda equazione è

$$y_2 = \beta \cdot \gamma \cdot x + (\beta \cdot \varepsilon_1 + \varepsilon_2)$$

Le stime SELS forniscono il risultato

$$\begin{cases} b_{(sels)} = \frac{\sum y_1 \cdot y_2}{\sum y_1^2} \\ c_{(sels)} = \frac{\sum x \cdot y_1}{\sum x^2} \end{cases}$$

Allora, se d è la stima di $\delta = \beta \cdot \gamma$ sarà $d = b \cdot c$, cioè

$$d_{(sels)} = \frac{\sum x \cdot y_1 \cdot \sum y_1 \cdot y_2}{\sum x^2 \cdot \sum y_1^2}$$

Con i minimi quadrati indiretti invece si ottengono le stime

$$\begin{cases} c_{(ILS)} = \frac{\sum x \cdot y_1}{\sum x^2} = c_{(SELS)} \\ d_{(ILS)} = \frac{\sum x \cdot y_2}{\sum x^2} \neq d_{(SELS)} \\ b_{(ILS)} = d_{(ILS)} \cdot c_{(ILS)} = \frac{\sum x \cdot y_2}{\sum x \cdot y_1} \neq b_{(SELS)} \end{cases}$$

Ora, se possono essere applicati due metodi di stima diversi, che conducono a risultati diversi ci si può chiedere quale sia preferibile. *Si dimostra che il metodo di stima per equazioni singole (SELS) generalmente conduce a risultati che sono distorti e non consistenti ed è per questa ragione che spesso si evita di usare questo metodo.*

Per dimostrare che il SELS fornisce stime distorte e non consistenti, consideriamo la versione identificata del sistema di equazioni

strutturali

che si è assunto come esempio all'inizio del paragrafo precedente. Tenendo presente che abbiamo $\gamma_{21} = \gamma_{1k} = 0$, $\gamma_{11} \neq 0$ e $\gamma_{2k} \neq 0$, esso in forma esplicita si può scrivere come segue :

$$\begin{cases} y_1 = & \beta_{12} \cdot y_2 + \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \dots + \gamma_{1, k-1} \cdot x_{k-1} & + \varepsilon_1 \\ y_2 = & \beta_{21} \cdot y_1 & + \gamma_{22} \cdot x_2 + \dots + \gamma_{2, k-1} \cdot x_{k-1} + \gamma_{2k} \cdot x_k & + \varepsilon_2 \end{cases}$$

e in forma implicita

$$\begin{cases} y_1 - \beta_{12} \cdot y_2 = & \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \dots + \gamma_{1, k-1} \cdot x_{k-1} & + \varepsilon_1 \\ -\beta_{21} \cdot y_1 + y_2 = & \gamma_{22} \cdot x_2 + \dots + \gamma_{2, k-1} \cdot x_{k-1} + \gamma_{2k} \cdot x_k & + \varepsilon_2 \end{cases}$$

e consideriamo altresì la sua forma ridotta. Applicando il metodo dei minimi quadrati alla seconda equazione strutturale e ponendo

$$D = \begin{vmatrix} \sum y_1^2 & \sum x_2 \cdot y_1 & \dots & \sum x_k \cdot y_1 \\ \sum y_1 \cdot x_2 & \sum x_2^2 & \dots & \sum x_k \cdot x_2 \\ \dots & \dots & \dots & \dots \\ \sum y_1 \cdot x_k & \sum x_2 \cdot x_k & \dots & \sum x_k^2 \end{vmatrix}$$

e A_{11} , A_{21} , ..., A_{k1} come cofattori di $\sum y_1^2$, $\sum y_1 \cdot x_2$, ..., $\sum y_1 \cdot x_k$ rispettivamente si ha, per esempio,

$$b_{21} = \frac{(A_{11} \cdot \sum y_1 \cdot y_2 + A_{21} \cdot \sum x_2 \cdot y_2 + \dots + A_{k1} \cdot \sum x_k \cdot y_2)}{D}$$

Se in tale relazione si sostituisce ad y_2 la sua espressione che figura nel sistema strutturale e se si ricorda che per le note proprietà dei determinanti si ha:

$$A_{1h} \cdot \sum x_{ji} \cdot x_{1i} + A_{2h} \cdot \sum x_{ji} \cdot x_{2i} + \dots + A_{kh} \cdot \sum x_{ji} \cdot x_{ki} = \begin{cases} D & \text{per } h = j \\ 0 & \text{per } h \neq j \end{cases}$$

l'espressione scritta per b_{21} fatte le dovute semplificazioni, diventa

$$b_{21} = \beta_{21} + \frac{(A_{11} \cdot \sum y_1 \cdot \varepsilon_2 + A_{21} \cdot \sum x_2 \cdot \varepsilon_2 + \dots + A_{k1} \cdot \sum x_k \cdot \varepsilon_2)}{D}$$

A questo punto, ricordiamo che è:

$$y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \dots + \delta_{1k} \cdot x_k + \eta_1$$

con

$$\eta_1 = \frac{\varepsilon_1 + \beta_{12} \cdot \varepsilon_2}{1 - \beta_{12} \cdot \beta_{21}}$$

Quindi, introducendo le costanti C_1, C_2, \dots, C_k e C che pure dipendono da y_1 (ma non è necessario che sia scritto esplicitamente), sostituendo ad y_1 la sua forma ridotta, si può scrivere

$$b_{21} = \beta_{21} + C_1 \cdot \sum x_1 \cdot \varepsilon_2 + C_2 \cdot \sum x_2 \cdot \varepsilon_2 + \dots + C_k \cdot \sum x_k \cdot \varepsilon_2 + C \cdot \frac{\sum \varepsilon_1 \cdot \varepsilon_2 + \beta_{12} \cdot \sum \varepsilon_2^2}{1 - \beta_{12} \cdot \beta_{21}}$$

per cui risulta evidente che

$$M(b_{21}) \neq \beta_{21}$$

perché è $\sum \varepsilon_2^2 \neq 0$.

Ciò vuol dire che b_{21} è una stima non corretta e la differenza non tende ad annullarsi con l'aumentare del numero delle osservazioni. Quindi b_{21} non è una stima consistente e lo stesso si può dire anche per le stime degli altri coefficienti. Si può allora concludere affermando che *le stime dei minimi quadrati diretti per relazioni identificate non sono corrette, ma distorte e non consistenti.*

Il metodo dei minimi quadrati indiretti fornisce stime consistenti e non distorte per i coefficienti della forma ridotta, sebbene essi non siano generalmente e completamente efficienti, a causa dell'usuale errore di specificazione. Le stime dei coefficienti strutturali a cui il metodo conduce sono generalmente distorte, ma sono invariabilmente consistenti. Pertanto tale metodo è di solito preferito al procedimento di stima per equazioni singole.

Esiste, tuttavia, un caso in cui la posizione è diversa ed è quello del *sistema ricorrente*. Ad esempio, nel caso di due equazioni ciò accade se nella forma strutturale è $\beta_{12} = 0$ (oppure $\beta_{21} = 0$) e, siccome ambedue le equazioni debbono essere identificate, uno dei coefficienti γ_{2j} deve essere uguale a zero. In tal caso, infatti, si ha

$$\begin{cases} y_1 = & + \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \dots + \gamma_{1, k-1} \cdot x_{k-1} + \gamma_{1k} \cdot x_k + \varepsilon_1 \\ y_2 = \beta_{21} \cdot y_1 & + \gamma_{22} \cdot x_2 + \dots + \gamma_{2, k-1} \cdot x_{k-1} + \gamma_{2k} \cdot x_k + \varepsilon_2 \end{cases}$$

ed il sistema mette bene in evidenza che non vi è una reale interdipendenza tra y_1 e y_2 in quanto tra le stesse esiste soltanto una relazione unidirezionale e cioè di dipendenza di y_2 da y_1 ; inoltre, non dovrebbe esservi multicollinearità per il fatto che la y_1 è spiegata, tra le altre variabili, anche da x_1 che non figura nella seconda relazione. Anche in questo caso si può scrivere:

$$b_{21} = \beta_{21} + \frac{(A_{11} \cdot \sum y_1 \cdot \varepsilon_2 + A_{21} \cdot \sum x_2 \cdot \varepsilon_2 + \dots + A_{k1} \cdot \sum x_k \cdot \varepsilon_2)}{D}$$

ma, se si usa la prima equazione del sistema strutturale, si ottiene

$$b_{21} = \beta_{21} + C_1 \cdot \sum x_1 \cdot \varepsilon_2 + C_2 \cdot \sum x_2 \cdot \varepsilon_2 + \dots + C_k \cdot \sum x_k \cdot \varepsilon_2 + C \cdot \sum \varepsilon_1 \cdot \varepsilon_2$$

dalla quale, dato che ε_1 è indipendente da ε_2 per ipotesi e quindi $\sum \varepsilon_1 \cdot \varepsilon_2 = 0$, risulta verificata la relazione

$$M(b_{21}) = \beta_{21}$$

Dunque, nel caso di questo sistema ricorrente, b_{21} è una stima non distorta di β_{21} e non è difficile vedere che essa è anche una stima consistente. La stessa affermazione vale ovviamente anche per le stime degli altri coefficienti. Allora possiamo dire che: *in un modello ricorrente ambedue i metodi, quello per equazioni singole e quello della forma ridotta, conducono a stime consistenti; mentre il primo fornisce stime non distorte per i parametri strutturali, il*

Si consideri il seguente modello, nel quale è $\gamma_{1k} = \gamma_{21} = \gamma_{22} = 0$, di modo che la prima equazione ha k parametri, mentre la seconda ne ha (k-1):

$$\begin{cases} y_1 = \beta_{12} \cdot y_2 + \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \gamma_{13} \cdot x_3 + \dots + \gamma_{1, k-1} \cdot x_{k-1} + \varepsilon_1 \\ y_2 = \beta_{21} \cdot y_1 + \gamma_{23} \cdot x_3 + \dots + \gamma_{2, k-1} \cdot x_{k-1} + \gamma_{2k} \cdot x_k + \varepsilon_2 \end{cases}$$

Se si stimano i coefficienti della forma ridotta, si ottengono nuovamente le equazioni tipo quelle già trovate in precedenza e pari a

$$\begin{cases} y_1 = d_{11} \cdot x_1 + d_{12} \cdot x_2 + \dots + d_{1k} \cdot x_k + v_1 \\ y_2 = d_{21} \cdot x_1 + d_{22} \cdot x_2 + \dots + d_{2k} \cdot x_k + v_2 \end{cases}$$

ed è facile ottenere una combinazione lineare con coefficiente nullo per la variabile x_k , dalla quale possano derivarsi le stime per i coefficienti della prima equazione strutturale. Ma, in generale, sarà impossibile costruire una equazione con coefficienti nulli sia per x_1 sia per x_2 ed il metodo della forma ridotta non potrà più essere usato per ricavare le misure dei coefficienti strutturali.

Si potrebbe, per esempio, ignorare la condizione

$$\lambda \cdot d_{12} + d_{22} = 0$$

e specificare semplicemente λ in modo che

$$\lambda \cdot d_{11} + d_{21} = 0$$

ottenendo così stime dei coefficienti strutturali basate sul rapporto d_{12}/d_{11} . Ma, alternativamente, può essere utilizzato il rapporto d_{22}/d_{21} e i due insiemi di stime potrebbero essere discordanti. La qual cosa lascia insoddisfatti. In tal caso si dice che l'equazione è *sovraidentificata*. Questo non implica che i coefficienti strutturali non possano essere stimati, ma semplicemente che debbono essere usati metodi diversi da quello della forma ridotta, il quale non tiene conto di alcune restrizioni imposte ai parametri.

La sovraidentificazione non costituisce un problema serio per un sistema ricorrente, poiché in questo caso è possibile usare, ed anzi è raccomandato, il metodo dei minimi quadrati per equazioni singole, che non fa sorgere alcun problema di stima.

Con un sistema non ricorrente la soluzione non è invece così ovvia. Per tale caso sono disponibili metodi di stima della massima verosimiglianza, nonostante che essi presentino alcune difficoltà di calcolo. Il metodo più semplice che d'altronde sembra essere anche una logica generalizzazione del metodo dei minimi quadrati indiretti, è quello noto come *metodo dei minimi quadrati a due stadi*.

L'idea base di questo metodo è la seguente. Per una equazione sovraidentificata si ottengono i valori teorici della variabile endogena (o delle variabili endogene, se si hanno più di due equazioni) che si trova al secondo membro con il metodo della forma ridotta. Quindi, nella equazione strutturale originaria, i valori effettivi della (o delle) variabile endogena che figura a secondo membro si sostituiscono con i valori teorici calcolati per mezzo della forma ridotta. Infine, si applica il metodo dei minimi quadrati ordinari al nuovo insieme di variabili. Allora, nel modello già incontrato, la variabile y_1 è sostituita nella seconda equazione dalla variabile y_{1C} che è pari a

$$y_{1C} = d_{11} \cdot x_1 + d_{12} \cdot x_2 + \dots + d_{1k} \cdot x_k$$

L'effetto di questo procedimento è quello di eliminare nella misura del possibile il termine erratico v_1 e con esso la fonte della distorsione asintotica. Le stime che ne risultano sono stime *consistenti*.

In pratica non è necessario calcolare tutti i valori teorici di y_{1C} o delle variabili endogene di pari natura, dato che si può far ricorso a metodi di calcolo che consentono di ridurre tali lungaggini.

Se si applica il metodo dei minimi quadrati a due stadi ad una equazione esattamente identificata, il risultato che si ottiene è identico a quello

fornito dai minimi quadrati indiretti. Per esempio, se si sostituisce a y_2 nella prima equazione del sistema strutturale la regressione

$$y_{2C} = d_{21} \cdot x_1 + d_{22} \cdot x_2 + \dots + d_{2k} \cdot x_k$$

l'equazione diventa

$$y_1 = \beta_{12} \cdot y_{2C} + \gamma_{11} \cdot x_1 + \gamma_{12} \cdot x_2 + \gamma_{13} \cdot x_3 + \dots + \gamma_{1,k-1} \cdot x_{k-1} + \varepsilon_1 =$$

$$= (\beta_{12} \cdot d_{21} + \gamma_{11}) x_1 + \dots + (\beta_{12} \cdot d_{2,k-1} + \gamma_{1,k-1}) x_{k-1} + \beta_{12} \cdot d_{2k} \cdot x_k + \varepsilon_1$$

Siccome entrambe le formulazioni comportano una regressione multipla su k variabili, la stima è equivalente e le stime dei coefficienti nella seconda espressione sono eguali ai coefficienti della forma ridotta. Allora

$$d_{1j} = b_{12}d_{2j} + c_{1j} \quad e \quad d_{1k} = b_{12}d_{2k} \quad (\text{per } j= 1, 2, \dots, k-1)$$

da cui segue che

$$b_{12} = \frac{d_{1k}}{d_{2k}} \quad e \quad c_{1j} = \left(d_{1j} - \frac{d_{2j} \cdot d_{1k}}{d_{2k}} \right) \quad (\text{per } j= 1, 2, \dots, k-1)$$

Ma i minimi quadrati indiretti danno come risultato

$$y_1 = \frac{d_{1k}}{d_{2k}} \cdot y_2 + \left(d_{11} - \frac{d_{21} \cdot d_{1k}}{d_{2k}} \right) \cdot x_1 + \dots + \left(d_{1,k-1} - \frac{d_{2,k-1} \cdot d_{1k}}{d_{2k}} \right) \cdot x_k$$

e quindi i due metodi in questo caso sono equivalenti.

Con un sistema di equazioni ricorrenti, a condizione che tutte le equazioni siano identificate, non sorge alcun reale problema di stima, poiché, sotto le ipotesi usuali, la stima per equazioni singole è non solo possibile, ma ottimale. La stima coi minimi quadrati indiretti potrebbe non avere in questo caso vantaggi teorici e pratici.

Con un sistema non ricorrente, inoltre, non vi è alcuna reale difficoltà in connessione con le equazioni che sono esattamente identificate, vale a

dire con le equazioni i cui parametri possono essere derivati in un modo univoco da quelli della forma ridotta.

Il metodo dei minimi quadrati indiretti, che produce stime consistenti e non distorte per i coefficienti della forma ridotta e stime consistenti per i coefficienti strutturali, è generalmente preferito alla stima per equazioni singole che non fornisce stime consistenti. Ma nonostante il metodo non possa dare coefficienti con la minima varianza campionaria, la sua semplicità offre molti vantaggi pratici.

Il principale problema di stima sorge in connessione con le equazioni sovraidentificate, che possono essere tutte o parte dell'insieme intero. Come nel caso precedente, le stime dei minimi quadrati per equazioni singole sono inficiate dal difetto della non consistenza, ma i minimi quadrati indiretti non consentono in questo caso il passaggio alle stime dei coefficienti strutturali e nonostante i coefficienti della forma ridotta possano essere usati per scopi previsivi, essi non sono compatibili con le ipotesi implicite nel modello originario.

Il metodo della massima verosimiglianza ad informazioni complete, che talvolta è usato per stimare i parametri, è basato su una idea anch'essa molto semplice e le stime che fornisce hanno proprietà ottimali sotto le ipotesi di normalità della distribuzione degli errori.

Tuttavia, esso, quando si ha un gran numero di equazioni e di vincoli sui parametri, può condurre a calcoli complicati anche con l'ipotesi di errori indipendenti nelle diverse equazioni strutturali ed in pratica è usato raramente.

Una sua variante è rappresentata dal *metodo della massima verosimiglianza ad informazioni incomplete*, il quale differisce da quello ad informazioni complete per il fatto che *sono ignorate le restrizioni poste su qualche parametro e che influiscono sul sistema di equazioni come insieme*, ma non su una singola equazione. Di conseguenza, il sistema diviene di più facile manipolazione in termini matematici, sebbene comparativamente siano anche

richiesti dei calcoli abbastanza lunghi, che includono qualche iterazione o qualche soluzione di equazioni di grado più elevato. Le stime che si ottengono hanno la proprietà formale di produrre quello che è chiamato *rapporto di minima varianza*. Le varianze in questione sono i residui della regressione sulle variabili indipendenti nell'equazione

e su tutte le variabili predeterminate rispettivamente; e le stime allora minimizzano il miglioramento dell'adattamento che potrebbe essere determinato dall'aggiunta delle variabili escluse dal modello.

Il metodo dei minimi quadrati a due stadi e quello della massima verosimiglianza ad informazioni limitate possono considerarsi come applicazioni particolari del principio delle variabili strumentali. La teoria dell'uso delle variabili strumentali generalmente è sviluppata nell'ambito della teoria della regressione semplice, in connessione con la distorsione originata dagli errori nelle variabili, ma lo stesso problema può presentarsi anche nella regressione multipla. *Con un sistema di equazioni simultanee, infatti, la fonte della distorsione risiede nel fatto che le variabili dipendenti, che sono affette da errore, sono usate nello stesso modo delle variabili esogene nel calcolare le somme dei quadrati e dei prodotti misti.*

La distorsione si evita se le variabili dipendenti sono eliminate come variabili dipendenti e sono sostituite da variabili strumentali predeterminate: solo allora, infatti, si ottengono stime consistenti.

Quanto ai vantaggi ed agli svantaggi dei vari metodi è da osservare come il problema sia di per se complicato, perché possono esservi errori di specificazione, ma anche errori di stima. Risultati teorici validi per una data forma matematica del modello, cessano di essere tali se vengono introdotte delle restrizioni sui parametri che siano scorrette. Questa è una eventualità piuttosto seria in economia ed in generale nel campo delle scienze sociali dato che spesso è dubbia e non nota la esatta natura delle relazioni tra le variabili.

Una posizione più estremista su questa materia fu presa da Lin, il quale sostiene che mentre la maggior parte dei modelli econometrici che sono stati costruiti sono sovraidentificati, in realtà le equazioni possono essere sottoidentificate se correttamente specificate, tenendo conto del gran numero di fattori che esercitano un'influenza sulle variabili economiche.

Se si accetta questa argomentazione, la conclusione logica è che si possono costruire e stimare forme ridotte con un gran numero di variabili

indipendenti e che si possono usare a fini di previsione i coefficienti ottenuti con i minimi quadrati indiretti.

Una difficoltà pratica che contrasta con questo approccio è che molto verosimilmente potrebbero incontrarsi multicollinearità abbastanza gravi tra le variabili esplicative. Questo potrebbe non essere un problema per la previsione, fintanto che le variabili esplicative mantengono più o meno le relazioni l'una con l'altra, ma se queste relazioni possono variare, le previsioni allora possono anche essere invalidate.

Una posizione meno estremista potrebbe essere quella di *trasformare una struttura sovraidentificata in un'altra esattamente identificata, aggiungendo il numero richiesto di variabili o lasciando cadere alcune restrizioni sui parametri*. Ciò potrebbe essere sempre possibile e potrebbe contribuire così a risolvere il problema della stima, tuttavia, il problema della multicollinearità rimarrebbe.

Avverso a questo punto di vista, si può osservare che la teoria economica può aiutare nel ridurre il numero dei parametri da stimare a quelli veramente essenziali, dando quindi relazioni su cui possa riporsi più fiducia. In particolare, l'uso di modelli ricorrenti che riflettono la dipendenza casuale corretta delle variabili economiche l'una dall'altra è stato difeso da Wold e Juréen in un loro ormai classico studio. Se si accetta questo punto di vista, allora con equazioni formate propriamente è giustificata la stima per equazioni singole. Un vantaggio di questo criterio è che spesso sarà possibile lavorare con un piccolo o comunque modesto numero di variabili che sono effettivamente usate per la stima in ciascuna equazione.

Non vi è dubbio che l'approccio con equazioni simultanee ha dato un potente contributo a tutta la teoria della modellistica, in quanto ha puntualizzato la necessità di una appropriata formulazione delle relazioni ed anche quella di considerare, in connessione con lo studio di una relazione, altre relazioni che hanno influenza sulle variabili comprese nella prima.

Non si nega, tuttavia, che in molti casi è sufficiente ignorare queste altre relazioni e che l'applicazione della tecnica dei minimi quadrati per equazioni singole continua a mantenere un posto di rilievo nella risoluzione dei problemi di stima che sorgono nel campo delle scienze sociali.