2 LE IPOTESI FORTI SUI RESIDUI

Indice del capitolo

2.1 Un riassunto delle ipotesi imposte al modello lineare

È opportuno, a questo punto, riassumere le ipotesi di vario tipo sinora fatte in relazione al modello lineare. Se questo è il semplice (1.3.1) le ipotesi sono:

1) il campione è omogeneo ed i parametri α e β sono invariabili nel tempo;
2) i valori di x_t sono noti, cioè non aleatori;
3) $m_{xx} - \bar{x}^2 \neq 0$
4) $E(\tilde{u}_t) = 0 \quad , \quad E(\tilde{u}_t \cdot \tilde{u}_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases} \quad \forall t,s$

(2.1.1)

Se, viceversa, il modello lineare è quello multiplo (1.3.4), in termini matriciali le ipotesi sono:

i) il campione ($\mathbf{y}$, $\mathbf{X}$) è omogeneo; in altre parole i parametri del vettore $\boldsymbol{\beta}$ sono considerati invariabili nel tempo;
ii) $\mathbf{X}$ è una matrice di costanti;
iii) $det(\mathbf{X}'\mathbf{X}) \neq 0$
iv) $E(\tilde{\mathbf{u}}) = \mathbf{0} \quad , \quad E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}') = \sigma^2 \mathbf{I}_n$

(2.1.2)

Con la i) si suppone che la struttura dell'economia rimanga invariata nel periodo campionario e che quindi sia possibile considerare valide per tutte le $t = 1, 2,\ldots, n$, le equazioni (1.3.1) e (1.3.4). La ii) è un'ipotesi semplificatrice, che in seguito elimineremo, che limita gli elementi stocastici del modello al residuo ed alla variabile endogena. La iii) è necessaria per poter determinare la stima dei minimi quadrati ordinari, sia nel caso non stocastico che in quello stocastico.

Infine, le ipotesi deboli iv) sono utilizzate per determinare alcune caratteristiche degli stimatori: la non distorsione e l'efficienza, nonché le matrici di dispersione e di correlazione di quelli dei minimi quadrati e la distorsione della varianza $\hat{\sigma}^2$ campionaria dei residui.

Si noti che l'ipotesi (3), che abbiamo specificato nel caso semplice per simmetria con l'ipotesi (iii) formulata nel caso multiplo, in effetti è automaticamente verificata purché le x_t non siano costanti.

2.2 Residui distribuiti normalmente

La funzione di densità congiunta dei residui

Le ipotesi stocastiche precedenti, tuttavia, non permettono di effettuare un'inferenza statistica completa sul modello lineare; ad esempio, non sono sufficienti per determinare intervalli di confidenza o per fare verifiche di ipotesi. Aggiungiamo, allora, l'*ipotesi forte* che i residui siano *distribuiti normalmente* con media nulla e varianza costante

$$\tilde{u}_t \sim N(0, \sigma^2) \qquad \forall t \tag{2.2.1}$$

cioè che la loro funzione di densità di probabilità sia del tipo XXI-(1.4.19)

$$f(u_t) = (2\pi\sigma^2)^{-1/2} \exp\left\{-u_t^2 / 2\sigma^2\right\} \tag{2.2.2}$$

Lavorando in termini matriciali si prende in considerazione l'intero vettore delle **u** ed è quindi utile disporre della loro funzione di densità congiunta. Questa si costruisce facilmente ricordando che per l'ultima delle (2.1.1) o delle (2.1.2) le u_t sono incorrelate. Come è noto, se due o più variabili aleatorie distribuite in modo normale sono incorrelate, allora esse sono anche indipendenti.[1] Ma è altresì noto che se due o più variabili sono stocasticamente indipendenti la loro densità congiunta è uguale al prodotto delle rispettive densità marginali.[2] La densità congiunta delle u_t si ottiene quindi moltiplicando tra di loro n densità del tipo (2.2.2)

$$f(u_1, u_2, \ldots, u_n) = \prod_{t=1}^{n} f(u_t) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\sum\nolimits_{t=1}^{n} u_t^2 / 2\sigma^2\right\} \tag{2.2.3}$$

e può essere espressa in termini matriciali come

$$f(\mathbf{u}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\mathbf{u}'\mathbf{u} / 2\sigma^2\right\} \tag{2.2.4}$$

La (2.2.4) è una funzione di densità congiunta di tipo normale multivariato con valor medio **0** e matrice di dispersione $\sigma^2\mathbf{I}$ la cui espressione nel caso generale è data dalla XXI-(2.5.5), che riportiamo per comodità

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1} (\mathbf{x}-\mu)\right\} \tag{2.2.5}$$

dove **x** è un vettore aleatorio con n elementi, di valor medio μ e matrice di dispersione Σ.

---

[1] Si veda il teorema XXI-2.4.

[2] Si veda il paragrafo XXI-2.2 e in particolare la XXI-(2.2.16).

Osservazione 2.1 – La (2.2.5) si riduce al caso particolare (2.2.4) se si considera che per le (2.1.1)-(2.1.2) le **u** hanno media nulla (e quindi $\boldsymbol{\mu} = \mathbf{0}$), e inoltre sono incorrelate e omoschedastiche, per cui $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ e quindi $det(\boldsymbol{\Sigma}) = \sigma^{2n}$ e $\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\mathbf{I}$. Effettuando queste sostituzioni nella (2.2.5) con **u** al posto di **x** si ottiene la (2.2.4). In simboli la (2.2.4) viene espressa asserendo che

$$\tilde{\mathbf{u}} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n) \tag{2.2.6}$$

Osservazione 2.2 - Nelle ipotesi precedenti (forti) i residui sono Indipendenti ed Identicamente Distribuiti; queste loro qualità sono indicate con l'acronimo IID.

La funzione di densità dello stimatore dei minimi quadrati

La (2.2.4) o (2.2.6) ci permette di ricavare immediatamente la distribuzione di probabilità dello stimatore $\hat{\boldsymbol{\beta}}$. Abbiamo visto nella (1.6.16) che esso può essere espresso in funzione del vettore aleatorio $\tilde{\mathbf{u}}$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{u}} \tag{2.2.7}$$

e quindi basterà applicare il teorema XXI-2.3 sulle trasformazioni lineari di variabili normalmente distribuite[3] per ottenere

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}] \tag{2.2.8}$$

La funzione di densità dello stimatore è quindi

$$f(\hat{\boldsymbol{\beta}}) = (2\pi\sigma^2)^{-k/2}(det\mathbf{X}'\mathbf{X})^{-1/2}\exp\left\{-(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})/2\sigma^2\right\} \tag{2.2.9}$$

dove si è fatto uso del teorema XIX-1.5 applicando il quale si ricava che

$$det(\sigma^{-2}\mathbf{X}'\mathbf{X}) = det(\sigma^{-2}\mathbf{I}_k)\cdot det(\mathbf{X}'\mathbf{X}).$$

Sempre applicando il teorema XXI-2.3 si trova immediatamente la distribuzione dei residui stimati $\hat{\mathbf{u}}$. Questi, infatti, per la (1.7.3) sono una funzione lineare del vettore aleatorio $\tilde{\mathbf{u}}$ che per la (2.2.6) è distribuito in modo multinormale. Avremo allora

$$\hat{\mathbf{u}} = \mathbf{M}\tilde{\mathbf{u}} \tag{2.2.10}$$

da cui ricaviamo la distribuzione dei residui stimati

[3] Il teorema asserisce che se $\tilde{\mathbf{x}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, allora $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. Nel caso della (2.2.7) abbiamo le posizioni seguenti: $\tilde{\mathbf{x}} = \tilde{\mathbf{u}}$; $\mathbf{b} = \boldsymbol{\beta}$; $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$; $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, applicando le quali si ottiene il risultato (2.2.8).

$$\hat{\mathbf{u}} \sim N(\mathbf{0}, \sigma^2\mathbf{M}) \tag{2.2.11}$$

dato che in questo caso

$$Cov(\mathbf{M}\tilde{\mathbf{u}}) = E(\mathbf{M}\tilde{\mathbf{u}}\,\tilde{\mathbf{u}}'\mathbf{M}') = \mathbf{M}E(\tilde{\mathbf{u}}\,\tilde{\mathbf{u}}')\mathbf{M}' = \sigma^2\,\mathbf{M}\mathbf{M}' = \sigma^2\,\mathbf{M}$$

Vedremo in seguito che, sfruttando il teorema del limite centrale, se si elimina l'ipotesi di normalità dei residui, mantenendo le altre, la (2.2.8) continua a valere, seppure asintoticamente.

Intervalli di confidenza per i parametri del modello lineare semplice con σ^2 *noto*

Se poniamo

$$\frac{1}{n}\left(1+\frac{\bar{x}^2}{m_{xx}-\bar{x}^2}\right) = a_{11}^2$$

$$\frac{1}{n\left(m_{xx}-\bar{x}^2\right)} = a_{22}^2$$

ricordando le (1.6.9) e (1.6.8) le distribuzioni degli stimatori dei parametri di regressione β_1 e β_2 sono

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 a_{11}^2) \qquad \hat{\beta}_2 \sim N(\beta_2, \sigma^2 a_{22}^2) \tag{2.2.12}$$

Effettuando le trasformazioni lineari

$$\frac{\hat{\beta}_1-\beta_1}{\sigma a_{11}} \quad \text{e} \quad \frac{\hat{\beta}_2-\beta_2}{\sigma a_{22}} \tag{2.2.13}$$

possiamo standardizzare gli stimatori, che così trasformati hanno distribuzione normale standard (si veda la XXI-(1.6.4)).[4]

Sia per $\hat{\beta}_1$ che per $\hat{\beta}_2$ allora, si può calcolare l'*intervallo di confidenza* più corto (simmetrico rispetto al valor medio) *al livello* α, supponendo σ^2 noto, partendo dalla condizione

$$P(z' \le \tilde{z} < z'') = 1-\alpha \tag{2.2.14}$$

dove $\tilde{z} \sim N(0, 1)$ e $z',\ z''$ sono i quantili di probabilità $\alpha/2$ e $1-\alpha/2$ rispettivamente, forniti dalle tavole della distribuzione normale standardizzata. L'intervallo è, dunque,

---

[4] Nel derivare le (2.2.12) abbiamo applicato il risultato esposto dalla XXI-(2.5.11) secondo cui le distribuzioni marginali di una distribuzione normale multivariata sono esse stesse normali. Si veda il paragrafo XXI-2.5.

$$z' \le \frac{\hat{\beta}_1 - \beta_1}{\sigma a_{11}} < z''$$

cioè

$$\hat{\beta}_1 - \sigma a_{11} z'' < \beta_1 \le \hat{\beta}_1 - \sigma a_{11} z' \qquad (2.2.15)$$

nel caso di β_1 e

$$z' \le \frac{\hat{\beta}_2 - \beta_2}{\sigma a_{22}} < z''$$

cioè

$$\hat{\beta}_2 - \sigma a_{22} z'' < \beta_2 \le \hat{\beta}_2 - \sigma a_{22} z' \qquad (2.2.16)$$

nel caso di β_2. Al livello di significatività α=0.05 i due quantili sono z'=–1.96 e z''=1.96 (si veda la figura XXI-1.1).

Gli intervalli di confidenza (2.2.15) e (2.2.16) sono aleatori poiché i loro estremi sono funzioni dello stimatore. Il livello di confidenza α, che definisce la probabilità che β_i risulti al di fuori dell'intervallo, è determinato soggettivamente: se è scelto troppo piccolo l'intervallo risulta grande e quindi poco utile ad indicare la dispersione dello stimatore; se è troppo grande, la probabilità che l'intervallo contenga β_1 o β_2 diminuisce, e di nuovo l'informazione fornita dall'intervallo di confidenza diventa scarsa.

> *Osservazione* 2.3 - Considerando quanto argomentato sopra si trae che $1-\alpha$ è la probabilità che l'intervallo di confidenza (2.2.15) contenga il parametro β_1 e che il (2.2.16) contenga β_2. Ad esempio, se poniamo $\alpha = 0.05$ le (2.2.15)-(2.2.16) definiranno gli intervalli di confidenza al 95% per i due parametri del modello di regressione semplice.

Stima intervallare nel modello lineare multiplo con σ^2 noto

Più in generale, passando al caso del modello lineare multiplo, se $\sigma^2 a_{ii}^2$ indica la varianza di $\hat{\beta}_i$, dove a_{ii}^2 è l'elemento *i*-esimo della diagonale principale di $(\mathbf{X}'\mathbf{X})^{-1}$, si ha che

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii}^2) \qquad (2.2.17)$$

per cui la trasformazione lineare che standardizza $\hat{\beta}_i$ è

$$\frac{\hat{\beta}_i - \beta_i}{\sigma \cdot a_{ii}} \qquad (2.2.18)$$

e l'intervallo di confidenza all'(1-α)×100% diventa

$$\hat{\beta}_i - \sigma \cdot a_{ii} \cdot z'' < \beta_i \le \hat{\beta}_i - \sigma \cdot a_{ii} \cdot z' \qquad (2.2.19)$$

Osservazione 2.4 – Riassumiamo per comodità del lettore la catena di implicazioni che permette di pervenire alla (2.2.19) a partire dalle ipotesi forti sui residui

$$\tilde{\mathbf{u}} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \Rightarrow \hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}] \Rightarrow \hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii}^2) \Rightarrow (\hat{\beta}_i - \beta_i)/\sigma a_{ii} \sim N(0,1)$$

Verifica di ipotesi lineari per i parametri del modello con σ^2 noto

Come è noto, esiste una dualità fra stima intervallare e verifica delle ipotesi (si veda il capitolo XXIII-2), per cui la trasformazione (2.2.18) può essere utilizzata anche come statistica per verificare l'*ipotesi nulla* che il parametro β_i sia uguale ad un prefissato valore r_i, sempre supponendo σ^2 noto

$$\begin{cases} H_0 : \beta_i = r_i \\ H_1 : \beta_i \ne r_i \end{cases} \qquad (2.2.20)$$

Nel *sistema di ipotesi* (2.2.20) l'ipotesi nulla viene contrapposta all'*ipotesi alternativa* H_1: $\beta_i \ne r_i$.

Per controllare l'ipotesi nulla è sufficiente sostituire $\beta_i = r_i$ nella (2.2.20) e verificare se il valore trovato

$$z = \frac{\hat{\beta}_i - r_i}{\sigma a_{ii}} \qquad (2.2.21)$$

è compreso nell'*intervallo di accettazione* di H_0 formato da $[z', z'')$ *oppure* in quello *di rifiuto*, composto dalle due semirette

$$z < z' \quad \text{e} \quad z \ge z''$$

In forma poco rigorosa ma espressiva si può asserire che l'ipotesi nulla $\beta_i - r_i = 0$ è accettata se la differenza $\hat{\beta}_i - r_i$ si ritrova vicino allo zero (appartiene all'intervallo di accettazione); è rifiutata se tale differenza è lontana dallo zero (appartiene all'intervallo di rifiuto). La divisione per lo scarto quadratico medio serve a standardizzare la distanza fra $\hat{\beta}_i$ e il valore ipotizzato, depurandola dell'unità di misura e permettendone il confronto con i quantili della normale standard.

Fra le applicazioni più frequenti di questo test vi è la verifica dell'ipotesi di nullità del parametro β_i. In questo caso l'ipotesi nulla è che il parametro sia uguale a zero e l'alternativa che sia diverso da zero

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \ne 0 \end{cases} \qquad (2.2.22)$$

Questo sistema di ipotesi, per quanto semplice, ha una notevole rilevanza ai fini della specificazione del modello, perché se l'ipotesi nulla non viene respinta ciò significa che l'i-esima variabile viene moltiplicata per una costante nulla, cioè, semplicemente, scompare dal modello. In altri termini, il *mancato* rifiuto dell'ipotesi nulla nella (2.2.22) implica che la i-esima variabile vada eliminata dal modello.[5]

In caso di verifica della (2.2.22) la statistica (2.2.21) diventa semplicemente

$$z = \frac{\hat{\beta}_i}{\sigma a_{ii}} \tag{2.2.23}$$

cioè il rapporto fra il coefficiente stimato e il suo scarto quadratico medio.

Ad esempio, se si conduce il test al livello di significatività $\alpha = 0.05$ la regione di accettazione va da -1.96 a $+1.96$, per cui se, poniamo, è $\hat{\beta}_i = 53428.4$ e $\sigma a_{ii} = 113245.8$ la statistica z è pari a 0.47 e quindi cade nella regione di accettazione dell'ipotesi H_0 formulata nel sistema (2.2.22); di conseguenza la i-esima variabile andrà eliminata dal modello. Se, viceversa, è $\hat{\beta}_i = 0.04$ e $\sigma a_{ii} = 0.001$ la statistica z è pari a 40, cioè cade nella regione di rifiuto dell'ipotesi nulla e quindi la i-esima variabile contribuisce significativamente a spiegare l'andamento della y_t.

L'esempio precedente vale a ricordare che dal punto di vista statistico non esistono coefficienti "grandi" o "piccoli", ma solo coefficienti "significativi" o "non significativi", intendendo per significativi quelli che risultano significativamente diversi da zero nel test definito dal sistema di ipotesi (2.2.22).

Verifica di ipotesi riguardanti più parametri nel modello lineare multiplo con σ^2 noto

Abbiamo visto che siccome il vettore $\hat{\beta}$ è normalmente distribuito secondo la (2.2.8) le proprietà della normale multivariata assicurano che sono distribuiti in modo

---

[5] Il ricercatore però non deve adattarsi passivamente all'esito del test. In alcuni casi, infatti, può darsi che un parametro, per quanto stimato in modo impreciso, sia essenziale nel determinare le proprietà complessive del modello, e quindi si può decidere di mantenerlo nella specificazione per ragioni di ordine economico anche se esso risulta non significativo al convenzionale livello di significatività del 5%. In altri casi possono essere motivi di ordine statistico a imporre il mantenimento di un parametro non significativo. Ad esempio, la presenza dell'intercetta è essenziale per garantire la proprietà (1.4.18), la quale, a sua volta, garantisce che sia possibile scomporre la devianza come nella (1.5.2). Se questa scomposizione non è possibile, il coefficiente di determinazione può assumere valori negativi. Di conseguenza si suggerisce di inserire sempre un'intercetta, anche qualora essa dovesse risultare non significativa, a meno che le variabili non siano già espresse come scarto dalle proprie medie, come nel modello (1.5.13).

normale anche tutti i $\hat{\beta}_i$ come nella (2.2.17), il che permette di costruire statistiche del tipo (2.2.21) per la verifica di sistemi di ipotesi come il (2.2.20).

In effetti, le proprietà della normale garantiscono che *qualsiasi* combinazione lineare degli elementi di $\hat{\boldsymbol{\beta}}$ sia distribuita in modo normale. In particolare, applicando la XXI-(2.5.8) abbiamo che dato un vettore $\mathbf{c}$ di ordine k la combinazione lineare $\mathbf{c}'\hat{\boldsymbol{\beta}}$ ha la seguente distribuzione

$$\mathbf{c}'\hat{\boldsymbol{\beta}} \sim N(\mathbf{c}'\boldsymbol{\beta}, \sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}) \tag{2.2.24}$$

Si noti che il risultato (2.2.17) è in effetti un caso particolare della (2.2.24) nella quale il vettore $\mathbf{c}$ sia un *vettore di selezione* che contiene tutti elementi nulli tranne l'i-esimo, nel qual caso

$$\mathbf{c}'\hat{\boldsymbol{\beta}} = [\,0\;0\;\ldots\;1\;\ldots\;0\,]\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix} = \hat{\beta}_i\;;\quad \mathbf{c}'\boldsymbol{\beta} = \beta_i; \qquad \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = a_{ii}^2$$

se l'unico elemento unitario del vettore $\mathbf{c}$ è in i-esima posizione, e quindi ci si riconduce alla (2.2.17).[6]

La (2.2.24) può essere standardizzata nel modo seguente

$$\frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - \mathbf{c}'\boldsymbol{\beta}}{\sigma\sqrt{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \sim N(0,1) \tag{2.2.25}$$

e questo risultato può essere usato per costruire statistiche relative a test di ipotesi su più parametri, del tipo

$$\begin{cases} H_0 : \mathbf{c}'\boldsymbol{\beta} = r \\ H_1 : \mathbf{c}'\boldsymbol{\beta} \neq r \end{cases} \tag{2.2.26}$$

nel qual caso la statistica si ottiene sostituendo al valore vero della combinazione lineare quello ipotizzato dalla (2.2.26). La distribuzione della statistica sotto la nulla è quindi

$$z = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - r}{\sigma\sqrt{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \sim N(0,1) \tag{2.2.27}$$

---

[6] Il lettore può verificare direttamente che $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$ è un vettore colonna di k elementi contenente la k-esima colonna di $(\mathbf{X}'\mathbf{X})^{-1}$, per cui premoltiplicando $\mathbf{c}'$ si ottiene il risultato esposto.

Ad esempio, supponiamo di voler verificare nel modello

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \tag{2.2.28}$$

l'ipotesi che sia $\beta_1+\beta_2 = 0$ (analoga alla (1.11.4)). Questa ipotesi può essere espressa nella forma (2.2.26) considerando il vettore $\mathbf{c}' = [1\ 1\ 0]$ e lo scalare $r = 0$. Se la varianza del modello è pari a $\sigma^2 = 4$, la matrice $(\mathbf{X}'\mathbf{X})^{-1}$ è così costituita

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 5 \end{bmatrix}$$

e i coefficienti stimati sono

$$\hat{\beta} = \begin{bmatrix} 1 \\ -0.7 \\ 3 \end{bmatrix}$$

il numeratore della statistica è

$$\mathbf{c}'\hat{\beta} = [1\ 1\ 0] \begin{bmatrix} 1 \\ -0.7 \\ 3 \end{bmatrix} = 0.3$$

mentre il denominatore è

$$\sigma\sqrt{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} = 2\sqrt{7} = 5.29$$

per cui la z definita dalla (2.2.27) è pari a 0.3/5.29 = 0.05 e l'ipotesi nulla non viene rifiutata dai dati.

2.3 Intervalli di confidenza e verifica di ipotesi lineari semplici per i parametri del modello con σ^2 ignoto

Stima intervallare e verifica di ipotesi nel modello lineare semplice

I procedimenti inferenziali delineati nel paragrafo precedente si basano sull'impiego degli stimatori standardizzati (2.2.13) o (2.2.21) nel cui calcolo si utilizzano i valori dei parametri veri β_i ($i = 1, 2$) e σ^2. Come abbiamo visto, in realtà non è necessario conoscere i valori veri dei β_i; ad esempio, nell'effettuare il test di ipotesi (2.2.20) ciò che a noi interessa è la distribuzione dello stimatore sotto l'ipotesi nulla, e quindi al posto del valore vero β_i del coefficiente di regressione usiamo il valore r_i postulato dall'ipotesi nulla. Viceversa, il valore di σ^2 è essenziale, perché senza di esso non possiamo effettuare la standardizzazione. Se σ^2 è sconosciuta, può essere stimata per mezzo dello stimatore non distorto (1.7.2), ma la distribuzione degli stimatori dei parametri di regressione standardizzati non è più normale.

In teoria della probabilità si dimostra che nel modello di regressione semplice vale il seguente risultato distribuzionale[7]

$$\frac{\sum_t \hat{u}_t^2}{\sigma^2} \sim \chi^2_{n-2} \tag{2.3.1}$$

Se ora nella prima delle (2.2.13) sostituiamo a σ il valore ottenuto tramite la radice quadrata dello stimatore non distorto

$$\overline{\sigma} = \sqrt{\frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-2}} \tag{2.3.2}$$

otteniamo

$$\frac{\hat{\beta}_1 - \beta_1}{\overline{\sigma} a_{11}} = \frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sigma a_{11}}}{\sqrt{\dfrac{\sum_t \hat{u}_t^2}{\sigma^2 (n-2)}}} \tag{2.3.3}$$

---

7 Questo risultato è relativamente semplice da dimostrare nel modello di regressione multipla usando l'algebra matriciale (vedi oltre in questo stesso paragrafo). Viceversa, se non si ricorre ai metodi matriciali la sua dimostrazione è inutilmente complicata e viene quindi omessa.

da cui si vede che in effetti il coefficiente di regressione "standardizzato" mediante la stima (2.3.2) corrisponde al rapporto fra una normale standard (la prima delle (2.2.13)) e la radice quadrata del χ^2_{n-2} nella (2.3.1) diviso per i propri gradi di libertà.

Si noti che le equazioni normali (1.3.13) implicano che sia

$$\sum_t \hat{y}_t \hat{u}_t = \hat{\beta}_1 \sum_t \hat{u}_t + \hat{\beta}_2 \sum_t x_t \hat{u}_t = 0 \tag{2.3.4}$$

per cui le stime dei residui sono incorrelate rispetto a quelle della parte sistematica, e in particolare, quindi, rispetto alle stime dei $\hat{\beta}_i$. Dato che sia i residui stimati che gli stimatori dei coefficienti sono distribuiti normalmente, la loro incorrelazione implica che essi siano indipendenti. Nella (2.3.3) quindi il numeratore e il denominatore sono stocasticamente indipendenti. Applicando il risultato XXI-(1.6.10), secondo il quale se $z \sim N(0,1)$ e $x \sim \chi^2_n$ sono due variabili aleatorie stocasticamente indipendenti, allora il rapporto fra la prima e la radice quadrata della seconda divisa per i propri gradi di libertà si distribuisce come una t di Student centrale con n gradi di libertà, si ha quindi che

$$\frac{\hat{\beta}_1 - \beta_1}{\bar{\sigma} a_{11}} \sim t_{n\text{-}2} \tag{2.3.5}$$

e analogamente si avrà

$$\frac{\hat{\beta}_2 - \beta_2}{\bar{\sigma} a_{22}} \sim t_{n\text{-}2} \tag{2.3.6}$$

Il denominatore delle (2.3.5) è l'*errore standard* [8] di $\hat{\beta}_1$ e quello della (2.3.6) l'errore standard di $\hat{\beta}_2$.

Per verificare, allora, l'ipotesi nulla

$$H_0\text{: } \beta_1 = r_1$$

oppure l'altra

$$H_0\text{: } \beta_2 = r_2$$

basta verificare che $(\hat{\beta}_1 - r_1)/\bar{\sigma} \cdot a_{11}$ oppure che $(\hat{\beta}_2 - r_2)/\bar{\sigma} \cdot a_{22}$ siano compresi nell'intervallo $[t'_{n-2}, t''_{n-2})$, dove t'_{n-2} e t''_{n-2} sono dati dalle tavole dei quantili della distribuzione della t di Student per n–2 gradi di libertà, e generalmente si trovano per livelli di significatività pari al 10%, 5% e 1%. Se vi sono compresi si è spinti ad accettare le ipotesi nulle; altrimenti si è indotti ad accettare le ipotesi alternative

[8] In lingua inglese: *Standard Error*, da cui l'acronimo SE.

$$H_1\text{: } \beta_1 \neq r_1 \qquad \text{e} \qquad H_1\text{: } \beta_2 \neq r_2$$

rispettivamente.

Usando la stessa logica è possibile costruire stime intervallari come le (2.2.15), utilizzando al posto dei quantili $z,' z''$ della normale standard quelli t'_{n-2}, t''_{n-2} della t di Student centrale con n-2 gradi di libertà.

La distribuzione della t di Student è più schiacciata della normale, alla quale si avvicina progressivamente all'aumentare dei gradi di libertà. Questa convergenza è illustrata dalla figura XXI-1.4.

Dato che le "code" della distribuzione della t sono più alte, i quantili, a parità di area, sono tanto più esterni rispetto a quelli della normale quanto minore il numero di gradi di libertà. Ad esempio, per $\alpha = 0.05$ i due quantili t'_{n-k} e t''_{n-k} valgono ±2.571, ±2.086 e ±1.980 per i tre numeri dei gradi di libertà $n-k = 5, 20, 120$, rispettivamente; i relativi quantili di una normale standard valgono, come è noto, ±1.960 e quindi ai fini pratici una t_{120} è praticamente equivalente a una normale standard. Questo significa che gli intervalli di confidenza e le regioni di accettazione definiti usando la distribuzione t sono maggiori di quelli costruiti usando la normale. Questo risultato ha un fondamento intuitivo, dato che quando il parametro σ^2 è ignoto l'incertezza relativa al modello è maggiore, e sono quindi più ampi anche i margini di incertezza nelle stime (gli intervalli di confidenza).

Stima intervallare nel modello lineare multiplo

Gli stessi risultati si estendono al modello lineare multiplo, dove l'uso del calcolo matriciale ci consente di provare facilmente il risultato corrispondente a quello esposto dalla (2.3.1).

Si è visto nel paragrafo 1.7 che la matrice $\mathbf{M} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ è simmetrica, idempotente e con traccia pari ad $n-k$, per cui il suo rango è $n-k$ per il teorema XIX-1.13 e l'applicazione del teorema XXI-2.6 e del suo corollario XXI-2.1 ci assicura che la variabile aleatoria

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} / \sigma^2 = \mathbf{u}'\mathbf{M}\mathbf{u} / \sigma^2 \sim \chi^2_{n-k} \tag{2.3.7}$$

Ora, dalla (2.2.13) segue che la variabile aleatoria

$$(\hat{\beta}_i - \beta_i) / \sigma a_{ii} \sim N(0,1)$$

per cui il suo rapporto alla radice quadrata della variabile χ^2_{n-k} definita dalla (2.36), divisa per i propri gradi di libertà,

$$\frac{(\hat{\beta}_i - \beta_i)/\sigma a_{ii}}{[\hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2(n-k)]^{1/2}} = \frac{\hat{\beta}_i - \beta_i}{a_{ii}\overline{\sigma}} \tag{2.3.8}$$

possiede distribuzione della t di Student con $n–k$ gradi di libertà se $\hat{\beta}_i$ - β_i e $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ sono variabili aleatorie indipendenti (si veda la XXI-(1.6.10)).

Ora, per la (1.7.6) $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ è una forma quadratica nel vettore normale multivariato $\mathbf{u}$

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}\mathbf{u} \tag{2.3.9}$$

mentre per la (1.6.16) $\hat{\boldsymbol{\beta}}$ -$\boldsymbol{\beta}$ è una trasformazione lineare del medesimo vettore

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \tag{2.3.10}$$

dove $\mathbf{M}$ è la matrice definita nella (1.7.4). Per il teorema XXI-2.8, l'indipendenza fra le (2.3.9) e (2.3.10) è dimostrata purché sia $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0}$, come si verifica immediatamente sfruttando la definizione della $\mathbf{M}$.

Il denominatore della (2.3.8)

$$\overline{s}_i = a_{ii}[\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)]^{1/2} \tag{2.3.11}$$

è l'*errore standard* della stima $\hat{\beta}_i$

L'intervallo di confidenza più corto (simmetrico rispetto all'origine) al livello α per questa stima è allora facilmente determinabile partendo dalla condizione

$$P\left(t'_{n-k} \le \frac{\hat{\beta}_i - \beta_i}{\overline{s}_i} < t''_{n-k}\right) = 1 - \alpha$$

dove t'_{n-k} e t''_{n-k} sono i quantili con probabilità $\alpha/2$ e $1-\alpha/2$, rispettivamente, della distribuzione della t di Student con $n–k$ gradi di libertà. L'intervallo è dunque

$$\hat{\beta}_i - \overline{s}_i \cdot t''_{n-k} < \beta_i \le \hat{\beta}_i - \overline{s}_i \cdot t'_{n-k} \tag{2.3.12}$$

Stima intervallare della varianza dei residui

Il risultato (2.3.7) ci permette di calcolare un intervallo di confidenza al livello α per la varianza dei residui partendo dalla condizione

$$P\left({\chi^2_{n-k}}' \le \hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2 < {\chi^2_{n-k}}''\right) = 1 - \alpha$$

dove ${\chi^2_{n-k}}'$ e ${\chi^2_{n-k}}''$ sono i quantili con probabilità $\alpha/2$ e $1-\alpha/2$ della distribuzione del chi quadrato con $n–k$ gradi di libertà. L'intervallo è dunque

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} / \chi^2_{n-k}{}'' < \sigma^2 \le \hat{\mathbf{u}}'\hat{\mathbf{u}} / \chi^2_{n-k}{}' \tag{2.3.13}$$

Per α =0.05 e per i tre numeri di gradi di libertà $n{-}k$ =10, 20, 30, i quantili $\chi^2_{n-k}{}'$ valgono 3.51, 9.95 e 17.21, ed i quantili $\chi^2_{n-k}{}''$ 21.72, 35.22 e 47.96 rispettivamente (si noti che dato che la distribuzione del χ^2 non è simmetrica,[9] i quantili sinistro e destro per il test bilaterale differiscono in valore assoluto). Come vedremo al termine di questo paragrafo, con la stessa logica si può costruire un test per l'ipotesi lineare $H_0 : \sigma^2 = r$.

Verifica di ipotesi lineari semplici nel modello lineare multiplo

L'ipotesi nulla (2.2.20) può essere verificata contro l'alternativa H_1: $\beta_i \neq r_i$, nel caso di σ^2 sconosciuto, direttamente utilizzando l'intervallo di confidenza (2.3.12): se questo contiene il punto $\beta_i = r_i$ si accetta l'ipotesi nulla, altrimenti la si rifiuta e si accetta l'alternativa.

Una seconda maniera di procedere è analoga al test di ipotesi presentata nel paragrafo 2.2 e consiste nel verificare se la statistica

$$t_i = \frac{\hat{\beta}_i - r_i}{\bar{s}_i} \tag{2.3.14}$$

è compresa nell'intervallo di accettazione $[t'_{n-k}, t''_{n-k})$ oppure in quello di rifiuto composto dalle due semirette

$$t < t_{n-k}' \qquad \text{e} \qquad t \ge t_{n-k}''$$

dove la t di Student è considerata per $n{-}k$ gradi di libertà.

Questo test è eseguibile in un'ulteriore forma. Si calcola la probabilità che la variabile aleatoria $\tilde{t}$ assuma un valore assoluto maggiore od uguale al valore trovato $|t_i|$ e la si confronta con $\alpha/2$: se è minore si rifiuta l'ipotesi nulla. Tale probabilità è detta *livello di significatività marginale*[10] e un suo valore vicino allo zero implica che la t_i data dalla (2.3.14) molto difficilmente possa essere considerata statisticamente nulla.

Naturalmente il test può essere esteso coinvolgendo nel medesimo vincolo più parametri, come accade nella (2.2.26), nel qual caso la statistica (2.2.27) va modificata nel modo seguente

---

[9] La distribuzione del chi quadrato è illustrata nella figura XXI-1.2.

[10] In lingua inglese: *marginal* o *empirical significance level* o anche, più spesso, *p-value*.

$$t = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - r}{\bar{\sigma}\sqrt{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \sim t_{n-k} \tag{2.3.15}$$

Un'applicazione: il test di nullità sui parametri della funzione delle importazioni

Riprendiamo ora il sistema di ipotesi (2.2.22) con il quale si verifica l'ipotesi di nullità dell'*i*-esimo coefficiente

$$H_0\text{: } \beta_i = r_i = 0 \qquad i = 1, 2, \ldots, k$$

In questo caso la statistica (2.3.14) si riduce semplicemente al rapporto fra il coefficiente stimato e il relativo scarto quadratico medio

$$t_i = \frac{\hat{\beta}_i}{\bar{s}_i} \tag{2.3.16}$$

In altre parole, la statistica consiste nel coefficiente "standardizzato", o, come anche si usa dire, con una terminologia forse più precisa, ma ugualmente cattiva dal punto di vista linguistico, "studentizzato" in modo da renderlo adimensionale.[11]

> *Osservazione* 2.5 - Il valore (2.3.14) o il suo caso particolare (2.3.16), per ovvii motivi, sono talvolta chiamati rapporto t.[12] I programmi di calcolo econometrico associano automaticamente a ogni coefficiente il rapporto (2.3.16), che viene definito per antonomasia la "t di Student" del coefficiente.

La consultazione delle t di Student fornisce utili indizi circa la validità del modello stimato. Vediamo ora come si interpretano concretamente questi rapporti , che usualmente vengono riportati sotto le stime, analizzando le quattro versioni stimate dell'equazione delle importazioni, cioè le (2.3.17), (2.3.18), (2.3.19) e (2.3.20), che riportiamo per comodità del lettore corredando i coefficienti con le relative t di Student:

$$\begin{array}{llllll} \ln \hat{y}_t = & -6.023 & + 0.572 \ln x_{1t} & + 0.900 \ln x_{2t} & - 0.164 \ln x_{3t} & + 0.104 \ln x_{4t} \\ & (-6.1) & (6.6) & (22.2) & (-8.5) & (3.3) \end{array} \tag{2.3.17}$$

$$n = 80,\ R^2 = 0.984,\ R^2_c = 0.983,\ RSS = 0.063,\ SEE = 0.029$$

---

[11] Si parla di standardizzazione quando una variabile aleatoria viene divisa per il proprio scarto quadratico medio. Nel caso in questione lo scarto quadratico medio non è noto e si usa una sua stima; dato che il rapporto che ne risulta è distribuito come una t di Student, viene talora impiegato il termine "studentizzazione".

[12] In lingua inglese: *t-ratio*.

$$\ln \hat{y}_t = \underset{(-10.4)}{-8.095} + \underset{(23.2)}{0.827} (\ln x_{1t} + \ln x_{2t}) - \underset{(-7.9)}{0.162} \ln x_{3t} + \underset{(3.3)}{0.056} \ln x_{4t} \qquad (2.3.18)$$

$$n = 80,\ R^2 = 0.982,\ R^2_c = 0.981,\ RSS = 0.071,\ SEE = 0.030$$

$$\ln \hat{y}_t = \underset{(-16.4)}{-3.344} + \underset{(9.1)}{0.355} \ln x_{1t} + \underset{(21.1)}{0.876} \ln x_{2t} - \underset{(-8.7)}{0.173} (\ln x_{3t} - \ln x_{4t}) \qquad (2.3.19)$$

$$n = 80,\ R^2 = 0.982,\ R^2_c = 0.982,\ RSS = 0.069,\ SEE = 0.030$$

$$\ln \hat{y}_t = \underset{(-13.0)}{-3.292} + \underset{(52.3)}{0.607} (\ln x_{1t} + \ln x_{2t}) - \underset{(-7.6)}{0.190} (\ln x_{3t} - \ln x_{4t}) \qquad (2.3.20)$$

$$n = 80,\ R^2 = 0.972,\ R^2_c = 0.972,\ RSS = 0.111,\ SEE = 0.037$$

I valori delle t di Student, riportati fra parentesi sotto i coefficienti, vanno confrontati con i quantili ricavati dalle tavole statistiche della distribuzione, che dipendono dai gradi di libertà del modello. Ricordiamo che al divergere dei gradi di libertà la t di Student converge a una normale standardizzata, con un processo di convergenza molto rapido per valori piccoli dei gradi di libertà (si veda la figura XXI-1.4). In pratica, ciò significa che la forma della distribuzione (e quindi il valore dei relativi quantili) cambia molto meno passando da 60 a 120 gradi di libertà che non passando da 2 a 3 gradi di libertà. Di conseguenza le tavole statistiche riportano i quantili per incrementi unitari (1, 2, 3,…) dei gradi di libertà solo fino a un certo valore (ad es., 30), dopo il quale la distribuzione viene tabulata per incrementi via via più ampi.

Questo vuol dire che se il nostro modello ha un numero elevato di gradi di libertà (indicativamente, superiore a cinquanta), probabilmente non troveremo tabulata la distribuzione esattamente corrispondente. Ad esempio, nelle quattro equazioni considerate i gradi di libertà variano da 75 (nella 1.9.5) a 77 (nella 1.11.4), ma la tavola della t di Student riportata da Johnston [1984, p. 548] tabula la distribuzione solo per 60 o 120 gradi di libertà. In questi casi possiamo utilizzare una zona di accettazione approssimata costruita scegliendo i quantili riferiti alla distribuzione con il numero di gradi di libertà più vicino a quello del nostro modello, ad esempio, 60. Leggendo sulla tavola i quantili della t_{60} che isolano un'area pari ad $\alpha/2$ costruiamo per simmetria gli intervalli di accettazione, che risultano pari a

$[-2.660, +2.660)$ per $\alpha = 0.01$

[−2.000, +2.000) per α =0.05

[1.671, +1.671) per α =0.10

Questi intervalli sono approssimazioni dell'intervallo esatto (cioè di quello relativo a una distribuzione con 75, 76 o 77 g.d.l.). Dato che la distribuzione diviene progressivamente più concentrata al divergere dei gradi di libertà,[13] se i valori prescelti sono inferiori a quelli effettivi (come nel nostro caso, essendo 60<75) l'intervallo approssimato è più ampio di quello esatto, per cui il test tende a far accettare la nulla più spesso del dovuto (è quindi meno *potente*, perché tende a far rifiutare di meno la nulla sia quando è vera che quando è falsa): in pratica, si corre il rischio di considerare non significativo un parametro che invece lo è. Se invece scegliessimo un valore dei g.d.l. superiore a quello effettivo, ad esempio costruendo le regioni di accettazione per 120 g.d.l., si avrebbero intervalli più concentrati

[-2.617, +2.617) per α =0.01

[−1.980, +1.980) per α =0.05

[−1.658, +1.658) per α =0.10

nel qual caso l'approssimazione porterebbe a un test che fa respingere la nulla troppo spesso, e quindi ha livello di significatività effettivo superiore a quello nominale (la probabilità di respingere la nulla quando è vera è superiore al valore α prescelto).

Come si vede, però, per g.d.l. superiori a 50 l'errore di approssimazione riguarda al più la seconda cifra decimale ed è quindi praticamente irrilevante nella maggior parte dei casi, compreso quello della funzione delle importazioni sopraindicata. Infatti, nelle quattro equazioni riportate la t di Student più piccola in valore assoluto è quella di $\ln x_{4t}$ nella (2.3.18) pari a 3.3, e quindi cade al di fuori dell'intervallo di accettazione più ampio fra quelli considerati. In altri termini, l'ipotesi che in queste equazioni vi sia un coefficiente statisticamente non significativo viene respinta.

Si noti anche che, come è intuitivo, l'intervallo di accettazione si restringe all'aumentare di α. Questo perché α è la probabilità di rifiutare l'ipotesi nulla

[13] Questa proprietà ha un significato intuitivo: i gradi di libertà aumentano con la numerosità campionaria; quanto maggiore è l'informazione campionaria di cui disponiamo, tanto minore sarà l'incertezza relativa ai risultati delle stime, e quindi tanto più concentrate saranno le relative distribuzioni di probabilità. Naturalmente questo ragionamento si basa sul presupposto che all'aumentare della numerosità campionaria la struttura dei dati non cambi, cioè che i dati siano perfettamente omogenei. Questo presupposto in economia deve essere attentamente verificato. Alcuni strumenti statistici per effettuare questa verifica verranno introdotti nel successivo capitolo 3.

quando è vera, e la probabilità di rifiutare la nulla è tanto maggiore quanto più piccola è la regione di accettazione del test. La scelta di α dipende da diversi fattori, fra i quali le convinzioni *a priori* del ricercatore o il fatto che il test venga condotto isolatamente ovvero in una procedura sequenziale di verifica sui parametri del modello.

> *Osservazione* 2.6 - Il livello di significatività più usuale per questo test, detto della t di Student, come per gli altri che seguiranno, è $\alpha = 0.05$. I quantili ad esso associati per 60 gradi di libertà, ± 2.0, vengono spesso utilizzati per effettuare il test in modo approssimativo, senza consultare le tavole della distribuzione relativa. Come abbiamo appena constatato, l'errore di approssimazione è trascurabile nella maggior parte dei casi.

Spesso, nel riportare le equazioni econometriche, al posto della t di Student si usa mettere tra parentesi tonde gli errori standard $\bar{s}_i$ delle stime; in questo caso il valore t da adoperare nel test della t di Student è calcolato semplicemente mediante il rapporto $\hat{\beta}_i / \bar{s}_i$. Generalmente, se si è interessati al solo test della t, con $r_i = 0$, è più utile avere tra parentesi tonde direttamente i valori t_i; se invece si desidera effettuare test con $r_i \neq 0$ è più conveniente avere tra parentesi gli errori standard (se non li si hanno, vengono immediatamente determinati tramite il rapporto $\hat{\beta}_i / t_i$).

Verifica di ipotesi lineari semplici per la varianza dei residui

L'ipotesi nulla

$$H_0\colon \sigma^2 = r \tag{2.3.21}$$

può essere verificata contro l'alternativa $H_1\colon \sigma^2 \neq r$ utilizzando l'intervallo di confidenza (2.3.13): se questo contiene il punto $\sigma^2 = r$ si accetta l'ipotesi nulla, altrimenti la si rifiuta e si accetta l'alternativa.

Si può, altro canto, procedere con un test di ipotesi che sfrutti direttamente il risultato (2.3.7), costruendo il valore

$$\chi^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} / r \tag{2.3.22}$$

e verificando se è compreso nell'intervallo di accettazione per H_0 formato da $[\chi^{2\,\prime}_{n-k}, \chi^{2\,\prime\prime}_{n-k})$ oppure in quello di rifiuto composto dalle due parti

$$0 \leq \chi^2 < \chi^{2\,\prime}_{n-k} \qquad \text{e} \qquad \chi^2 \geq \chi^{2\,\prime\prime}_{n-k} \tag{2.3.23}$$

dove i due quantili hanno probabilità $\alpha/2$ e $1-\alpha/2$ rispettivamente.

L'intervallo di confidenza (2.3.13) per $\alpha = 0.05$ nel caso dell'equazione (2.3.17) è facilmente determinabile, se si considera che dalla tavola della distribuzione del χ^2 si ottiene, per 75 gradi di libertà,

$$\chi^{2\,\prime}_{75} = 52.96 \quad , \quad \chi^{2\,\prime\prime}_{75} = 100.82$$

Si ha, inoltre, che $\hat{\mathbf{u}}\hat{\mathbf{u}} = RSS = 0.0633$, per cui l'intervallo è

$$[0.00063 \quad , \quad 0.00119)$$

2.4 Verifica di ipotesi lineari multiple

Nel paragrafo 1.11 sono state determinate le stime dei minimi quadrati dei parametri $\boldsymbol{\beta}$ del modello lineare (1.3.4) in modo che esse soddisfacessero identicamente ai vincoli rappresentati dal sistema di q equazioni lineari (1.11.9); si ottenevano così le stima dei minimi quadrati vincolati $\boldsymbol{\beta}_0$, tali che $\mathbf{R}\boldsymbol{\beta}_0 = \mathbf{r}$. Ora modifichiamo questo argomento supponendo di stimare prima $\boldsymbol{\beta}$ *senza* vincoli, ad esempio con il criterio dei minimi quadrati, e di verificare *dopo* se questi vincoli sono rispettati, cioè se è valida l'ipotesi nulla, formata dalle q relazioni lineari

$$H_0\text{: } \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \tag{2.4.1}$$

rispetto all'alternativa

$$H_1\text{: } \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \tag{2.4.2}$$

dove $\mathbf{R}$ è una matrice di ordine $q{\times}k$ ed $\mathbf{r}$ è un vettore di q elementi. Dunque il problema di *stima vincolata* del paragrafo 1.11 viene modificato in uno di *verifica delle ipotesi lineari* (2.4.1): $\boldsymbol{\beta}_0$ è una stima che necessariamente soddisfa ai vincoli (1.11.9), mentre $\hat{\boldsymbol{\beta}}$ è una stima che può verificare o meno l'ipotesi nulla (2.4.1) costituita dagli stessi vincoli.

Il sistema di ipotesi (2.4.1)-(2.4.2) comprende come caso particolare quello in cui la matrice $\mathbf{R}$ ha $q = 1$ righe, nel qual caso siamo ricondotti alla (2.2.26), dove in effetti si è posto $\mathbf{R} = \mathbf{c}'$. Di conseguenza, l'ipotesi nulla (2.2.22)

$$H_0\text{: } \beta_i = 0$$

che è un caso particolare della (2.2.26) e che viene verificata mediante il test della z con la statistica (2.2.21) nel caso in cui σ^2 sia noto e con quello della t di Student dato dalla statistica (2.3.14) nel caso in cui σ^2 sia ignoto, può anche essere vista come caso particolare della (2.4.1) per $q = 1$ e con

$$\mathbf{R} = [0\ 0\ \ldots\ \underset{\substack{\uparrow \\ \text{elemento } i\text{-esimo}}}{1}\ \ldots 0] \qquad\qquad \mathbf{r} = 0 \tag{2.4.4}$$

La stessa ipotesi nulla (2.2.20) diventa la (2.4.1) per q=1, con $\mathbf{R}$ ancora data dalla prima della (2.4.4) ed $\mathbf{r} = r_i$, mentre l'ipotesi di uguaglianza di due parametri, ad esempio la (1.11.3), può essere scritta nel modo

$$H_0\text{: } \beta_2 - \beta_3 = 0 \tag{2.4.5}$$

cioè, nella forma generale (2.4.1) con q=1,

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

Nella discussione riferita al modello (2.2.28) abbiamo anche illustrato come impostare il test per un'ipotesi di omogeneità del tipo (1.11.4).

Tuttavia è chiaro che in tutti questi casi, nei quali abbiamo a che fare con *un unico vincolo* (eventualmente riguardante più parametri) disponiamo già di un risultato distribuzionale, il (2.2.27), opportunamente esteso dal (2.3.15) al caso di σ^2 ignoto, che consente di verificare l'ipotesi di interesse mediante una statistica distribuita rispettivamente come una normale o come una t di Student. La formulazione (2.4.1) diventa quindi veramente utile nel momento in cui abbiamo a che fare con l'imposizione simultanea di *più di un vincolo*. Questo va tenuto presente anche se, nel prosieguo, alcuni nostri esempi saranno riferiti, per non appesantire i calcoli, al caso in cui $q = 1$ (cioè al caso in cui si verifica un unico vincolo).[14]

Per verificare l'ipotesi (2.4.1) si considera la trasformata lineare $\mathbf{R}\hat{\beta}$. Ricordando che $\hat{\beta}$ ha la distribuzione multinormale specificata dalla (2.2.8), il teorema XXI-2.3 ci permette di ricavare la distribuzione di $\mathbf{R}\hat{\beta}$, che sarà anch'essa normale multivariata

$$\mathbf{R}\hat{\beta} \sim N[\mathbf{R}\beta, \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']$$

ovvero, ponendo $\mathbf{U} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ come nella (1.11.16),

$$\mathbf{R}\hat{\beta} \sim N[\mathbf{R}\beta, \sigma^2\mathbf{U}] \tag{2.4.6}$$

da cui segue immediatamente che

$$\mathbf{R}\hat{\beta} - \mathbf{R}\beta \sim N[\mathbf{0}, \sigma^2\mathbf{U}] \tag{2.4.7}$$

Ora, per quanto esposto nel paragrafo 1.11, la matrice di dispersione $\sigma^2\mathbf{U}$ è definita positiva ed è quindi possibile applicare il teorema XIX-1.14. Allora, la forma quadratica

$$(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)'\mathbf{U}^{-1}(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)/\sigma^2 \sim \chi^2_q \tag{2.4.8}$$

e divisa per l'altra forma quadratica (2.3.7), rapportate ambedue ai rispettivi gradi di libertà, conduce alla

$$\frac{(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)'\mathbf{U}^{-1}(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)/\sigma^2 q}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2(n-k)} = \frac{(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)'\mathbf{U}^{-1}(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \cdot \frac{(n-k)}{q} \sim F_{q,n-k} \tag{2.4.9}$$

[14] Alcuni programmi di calcolo econometrico, fra i quali Easy Reg 2000 Basic, utilizzano sempre per la verifica di ipotesi lineari la formulazione generale (2.4.1) e quindi non effettuano il test (2.3.15) in chiave di t di Student.

avendo sfruttato il risultato di teoria delle probabilità secondo il quale il rapporto di due variabili aleatorie distribuite come χ^2 si distribuisce come una F di Fisher se esse sono indipendenti e divise ciascuna per il proprio numero dei gradi di libertà; questi due numeri costituiscono, poi, la coppia dei gradi di libertà della F di Fisher (si veda la XXI-(1.6.11)).

Rimane quindi da dimostrare l'indipendenza del numeratore della (2.4.9) rispetto al denominatore. A questo scopo utilizziamo il teorema XIX-2.7, considerando che sia il numeratore che il denominatore della (2.4.9) sono forme quadratiche aleatorie nel vettore $\tilde{\mathbf{u}}$ associate a matrici idempotenti. Abbiamo infatti per il numeratore

$$(\hat{\beta}-\beta)'\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\hat{\beta}-\beta) = \tilde{\mathbf{u}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{u}} \tag{2.4.10}$$

dove si è fatto uso della (1.6.16) mentre per il denominatore abbiamo già visto che vale la

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \tilde{\mathbf{u}}'\mathbf{M}\tilde{\mathbf{u}} \tag{2.4.11}$$

dove si sono utilizzate le (1.7.6) e (1.7.4). L'idempotenza della matrice in (2.4.11) è stata mostrata con la (1.7.5) e la dimostrazione dell'idempotenza della matrice in (2.4.10) è lasciata al lettore. Poiché $\mathbf{MX} = \mathbf{0}$ è anche

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0}$$

e quindi applicando il teorema XIX-2.7 si dimostra l'indipendenza delle due forme quadratiche a numeratore e a denominatore della (2.4.9).

Se è vera l'ipotesi nulla (2.4.1) il rapporto

$$\frac{(\mathbf{R}\hat{\beta}-\mathbf{R}\beta)'\mathbf{U}^{-1}(\mathbf{R}\hat{\beta}-\mathbf{R}\beta)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}\cdot\frac{(n-k)}{q} \sim F \tag{2.4.12}$$

si colloca vicino allo zero, perché la distanza fra $\mathbf{R}\hat{\beta}$ e $\mathbf{R}\beta$ è trascurabile e quindi l'ipotesi nulla stessa è verificata al livello α se il valore F cade nell'intervallo di accettazione $[0\ ,\ F_{q,n-k}')$ dove l'estremo destro $F_{q,n-k}'$ è il quantile $1-\alpha$ di probabilità della distribuzione della F di Fisher con q ed $n-k$ gradi di libertà. Se il valore F cade fuori dell'intervallo, si rifiuta H_0 e si accetta l'ipotesi alternativa (2.4.2). La funzione di densità del tipo F è illustrata nella figura XXI-1.5.

Anche in questo caso il test può essere condotto utilizzando il livello di significatività marginale o *p-value* definito dalla probabilità $P(\tilde{F} \geq F)$ dove F è il valore dato dalla (2.4.12). Se questa probabilità è inferiore ad α si è spinti a rifiutare l'ipotesi nulla (2.4.1); se è maggiore la si accetta.

Esempio 2.1 - Se occorre verificare l'ipotesi nulla (2.4.3) con le posizioni (2.4.4) si ha che q=1 e

$$\mathbf{R}\hat{\beta} - \mathbf{r} = \hat{\beta}_i \qquad \text{e} \qquad \mathbf{U} = a_{ii}^2$$

dove a_{ii}^2 è l'elemento i-esimo della diagonale principale di $(\mathbf{X}'\mathbf{X})^{-1}$. Allora il rapporto (2.4.12) diventa

$$F = \frac{\hat{\beta}_i^2}{a_{ii}^2 \cdot \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)} = \frac{\hat{\beta}_i^2}{\bar{s}_{ii}^2} \tag{2.4.13}$$

che si distribuisce come il quadrato della variabile aleatoria t di Student con $n-k$ gradi di libertà; questo è allora uguale ad una F di Fisher con 1 e $n-k$ gradi di libertà, per cui ritroviamo il risultato XXI-(1.6.11) secondo cui il test della F di Fisher con un grado di libertà al numeratore è equivalente al test della t di Student (elevata al quadrato) illustrato nel paragrafo 2.3.

Due applicazioni della verifica di ipotesi lineari

Verifichiamo nella (1.10.1) l'ipotesi che le due elasticità β_2 e β_3 siano uguali, cioè l'ipotesi (1.11.3) che può esser scritta nella forma (2.4.5). Come abbiamo già notato, in questo caso,visto che il vincolo sottoposto a verifica è uno solo, potremmo utilizzare il test della t con la statistica data dalla (2.3.15). Volendo utilizzare il test (2.4.12) riprendiamo dal paragrafo 1.11 i seguenti dati

$$\mathbf{R}\hat{\beta} = -0.328 \qquad \text{e} \qquad \mathbf{U}^{-1} = 0.0817$$

mentre dalla stima (2.3.17) sappiamo che $\hat{\mathbf{u}}'\hat{\mathbf{u}} = RSS = 0.0633$, $n-k = 75$ ed inoltre è $\mathbf{r} = 0$ (uno scalare) e $q = 1$. Inserendo questi valori nella (2.4.12) si ottiene $F = 10.414$ che è più grande del quantile della distribuzione della F di Fisher con 1 e 75 gradi di libertà, che vale 3.97 per $\alpha = 0.05$. Questo test, quindi, ci induce a rifiutare l'ipotesi di uguaglianza delle due elasticità.

Verifichiamo ora che nella (1.10.1) contemporaneamente le due elasticità β_2 e β_3 siano uguali e che sussista il vincolo di omogeneità di grado zero sui prezzi. In questo caso abbiamo a che fare con la verifica contemporanea di due vincoli, e quindi è possibile procedere solo con la statistica (2.4.12). Formalmente, il sistema di ipotesi sottoposto a verifica è

$$\begin{cases} H_0 : \beta_2 = \beta_3; \quad \beta_4 = -\beta_5 \\ H_1 : (\beta_2 \neq \beta_3) \cup (\beta_4 \neq -\beta_5) \end{cases}$$

dove nell'alternativa si è fatto uso del segno di unione o somma logica $\cup$,[15] ad indicare che l'ipotesi nulla è respinta anche se è solo $\beta_2 \neq \beta_3$ o solo $\beta_4 \neq -\beta_5$ (*a fortiori*, l'ipotesi nulla sarà respinta se è simultaneamente $\beta_2 \neq \beta_3$ e $\beta_4 \neq -\beta_5$).

Facendo ancora uso del test con statistica (2.4.12) riprendiamo dal paragrafo 1.11 i valori seguenti

$$\mathbf{R}\hat{\boldsymbol{\beta}} = \begin{bmatrix} -0.328 \\ -0.060 \end{bmatrix} \qquad \text{e} \qquad \mathbf{U}^{-1} = \begin{bmatrix} 0.1510 & 0.4857 \\ 0.4857 & 3.4044 \end{bmatrix}$$

mentre dalla (2.3.17) sappiamo che $\hat{\mathbf{u}}'\hat{\mathbf{u}} = 0.0633$, $n-k = 75$ ed inoltre è $\mathbf{r} = [0\ 0]'$ e $q = 2$. Inserendo questi dati nella (2.4.12) otteniamo $F = 28.210$ che è sensibilmente più grande del quantile della distribuzione della F di Fisher con 2 e 75 gradi di libertà che vale 3.12 per $\alpha = 0.05$. Questo test, quindi, ci induce a rifiutare l'ipotesi doppia formulata sopra.

Verifica della bontà di adattamento complessiva di un modello

In ogni modello (1.3.4) occorre, tra le altre, verificare l'ipotesi che i parametri relativi a tutte le variabili esplicative siano contemporaneamente nulli, definendo, quindi, un test sulla bontà di adattamento dell'intero modello ai dati. L'ipotesi nulla è

$$H_0\colon \beta_1 = \beta_2 = \ldots = \beta_k = 0 \tag{2.4.14}$$

formata da $q = k$ relazioni lineari, per cui la (2.4.1) diventa

$$\boldsymbol{\beta} = \mathbf{0} \tag{2.4.15}$$

con $\mathbf{R} = \mathbf{I}$ ed $\mathbf{r} = \mathbf{0}$. La (2.4.9) è

$$\frac{\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}/k}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}/k}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)} \sim F_{k,n-k} \tag{2.4.16}$$

avendo fatto uso delle equazioni normali (1.4.11).

Dato che la (2.4.14) annulla tutti i coefficienti, essa equivale a ipotizzare, di fatto, che la componente sistematica della (1.3.4), cioè il valore atteso della y_t, sia nulla. In generale, fatto salvo il caso in cui la y_t sia espressa come scarto dalla propria media, questa ipotesi è eccessivamente restrittiva. Di conseguenza la bontà dell'adattamento viene verificata prescindendo dall'intercetta, che supponiamo essere β_k, cioè sostituendo l'ipotesi multipla (2.4.14) con la seguente

$$H_0\colon \beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0 \tag{2.4.17}$$

[15] Si veda il par. XXI-1.2.

che equivale all'ipotesi che la componente sistematica sia costante, e quindi, in particolare, che la variabile dipendente non venga influenzata da variazioni nel livello delle esplicative. La statistica F data dalla (2.4.16) deve essere modificata di conseguenza. È conveniente a questo scopo definire in termini matriciali l'ipotesi (2.4.17), ponendo

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

matrice di ordine $(k\text{-}1)\times k$, ed $\mathbf{r} = \mathbf{0}$; così la matrice $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ è data dalla sottomatrice quadrata di ordine k-1 ricavata dalla $(\mathbf{X}'\mathbf{X})^{-1}$ escludendo l'ultima riga e l'ultima colonna. Poiché si può effettuare la partizione $\mathbf{X} = [\ \mathbf{X}_1\ \mathbf{i}\ ]$, si ha

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{i} \\ \mathbf{i}'\mathbf{X}_1 & n \end{bmatrix}$$

per cui, tenendo conto della XIX-(1.4.9)

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = \left(\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{i}\frac{1}{n}\mathbf{i}'\mathbf{X}_1 \right)^{-1} = \left[\mathbf{X}_1'\left(\mathbf{I} - \mathbf{ii}'/n\right)\mathbf{X}_1\right]^{-1} = \left[\mathbf{X}_1'\mathbf{C}\mathbf{X}_1\right]^{-1}$$

dove $\mathbf{C}$ è la matrice di centraggio definita dalla XIX-(1.10.5) per cui la statistica (2.4.12) diventa

$$\frac{\hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{C}\mathbf{X}_1\hat{\boldsymbol{\beta}}_1}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \cdot \frac{n-k}{k-1} \sim F_{k-1,n-k} \tag{2.4.18}$$

> *Osservazione* 2.8 – La verifica dell'ipotesi che i parametri di un modello, esclusa l'intercetta, siano globalmente nulli, viene detta per antonomasia *test della F* di Fisher, anche le statistiche dei test per la verifica di molte altre ipotesi possiedono la stessa distribuzione (in altre parole, il (2.4.18) non è l'*unico* test che utilizza la F di Fisher).

Poiché, ora, per l'Osservazione 1.13 si ha che

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \left(1 - R^2\right) \times \text{Devianza totale}$$

ed è

$$\hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{C}\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 = R^2 \times \text{Devianza totale}$$

segue che il rapporto della (2.4.18) che indichiamo con F può essere espresso in funzione di R^2

$$F = \frac{R^2(n-k)}{(1-R^2)(k-1)} \tag{2.4.19}$$

Applicando la (2.4.18) (o la (2.4.19)) alle equazioni (2.3.17), (2.3.18), (2.3.19) e (2.3.20) otteniamo rispettivamente

eq. (2.3.17): $F_{4,75} = 1193.5$ (2.49) [0.00]

eq. (2.3.18): $F_{3,76} = 1413.1$ (2.73) [0.00]

eq. (2.3.19): $F_{3,76} = 1457.4$ (2.73) [0.00]

eq. (2.3.20): $F_{2,77} = 1382.9$ (3.11) [0.00]

dove abbiamo indicato accanto a ogni statistica, fra parentesi tonde il rispettivo valore soglia, e fra parentesi quadre il *p-value*. In tutti e quattro i casi la statistica è molto più grande del corrispondente quantile della distribuzione della F di Fisher e pertanto si può asserire che i quattro modelli si adattano complessivamente bene ai dati. Questo risultato conferma quelli dei test univariati della t di Student. D'altra parte, per la struttura stessa dell'ipotesi (2.4.17), si vede che affinché essa sia respinta basta che anche uno solo dei primi k-1 coefficienti sia significativamente diverso da zero. Questo significa che se anche uno solo dei coefficienti diversi dall'intercetta ha una t di Student significativa, ci dovremo attendere che sia significativa (cioè che cada nella regione critica del test) anche la F. Nel caso in esame, in tutte e quattro le equazioni tutti i coefficienti presi singolarmente sono molto significativi al test della t, per cui l'esito del test della F non fa altro che confermare questi risultati.

D'altra parte, questo non deve portare a pensare che l'informazione fornita dal test congiunto sia ridondante, dato che l'analisi dei valori dei test della t sui singoli coefficienti non ci permette di stabilire quale modello sia complessivamente superiore in termini di significatività. Ad esempio, dalle statistiche F emerge che l'equazione (2.3.19) è superiore alla (2.3.18), perché a parità di gradi di libertà la seconda ha una statistica inferiore alla prima. A questo risultato non si può pervenire consultando le singole t dei coefficienti, tra l'altro anche perché queste si riferiscono a variabili definite in modo diverso.

> *Osservazione* 2.9 – La verifica nelle (2.3.17)-(2.3.20) della (2.4.19) porta a un'uguaglianza approssimata, a causa degli arrotondamenti effettuati nel riportare i coefficienti di determinazione. Ciò non pregiudica l'esito del test, dato che le statistiche sono sensibilmente superiori ai valori soglia.

2.5 Formulazioni alternative dei test di ipotesi lineari multiple

Quando sono in vigore le q restrizioni lineari (1.11.9) o (2.4.1) i residui $\mathbf{u}_0$ del modello (1.3.4) a esse vincolato sono

$$\mathbf{u}_0 = \mathbf{y} - \mathbf{y}_0 = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0 = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}_0 = \hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) \tag{2.5.1}$$

dove la stima dei minimi quadrati vincolata $\boldsymbol{\beta}_0$ è fornita tramite lo stimatore (1.11.18), $\hat{\boldsymbol{\beta}}$ è quella dei minimi quadrati ordinari e $\hat{\mathbf{u}}$ è la stima dei residui ottenuta con questo criterio. Per mezzo della (2.5.1) si trae la devianza residua

$$\mathbf{u}_0'\mathbf{u}_0 = [\hat{\mathbf{u}}' - (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})'\mathbf{X}'][\hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})] = \hat{\mathbf{u}}'\hat{\mathbf{u}} + (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) \tag{2.5.2}$$

dove si è fatto uso delle relazioni di ortogonalità (1.4.15) e (1.4.16).

Sostituendo la (1.11.18) nella (2.5.2) si ottiene

$$\begin{aligned} \mathbf{u}_0'\mathbf{u}_0 - \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) = \\ &= (\mathbf{r}-\mathbf{R}\hat{\boldsymbol{\beta}})'\mathbf{U}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\,\mathbf{U}^{-1}(\mathbf{r}-\mathbf{R}\hat{\boldsymbol{\beta}}) = (\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})'\mathbf{U}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r}) \end{aligned} \tag{2.5.3}$$

per cui il test della F di Fisher (2.4.9) può esser scritto anche nelle due forme equivalenti

$$\frac{\mathbf{u}_0'\mathbf{u}_0 - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}\frac{(n-k)}{q} = \frac{(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}\frac{(n-k)}{q} = F \tag{2.5.4}$$

La scelta, per poter effettuare la verifica dell'ipotesi multipla (2.5.1), tra la forma (2.4.12) ed una delle (2.5.4) dipende dalle quantità che si hanno a disposizione. Se, oltre alla devianza $\hat{\mathbf{u}}'\hat{\mathbf{u}}$, necessaria in ciascuno dei tre casi, si dispone della devianza vincolata $\mathbf{u}_0'\mathbf{u}_0$, conviene usare la prima delle forme (2.5.4); se si dispone di $\hat{\boldsymbol{\beta}}$ è conveniente usare la (2.4.12), mentre se si hanno a disposizione le due stime $\boldsymbol{\beta}_0$ e $\hat{\boldsymbol{\beta}}$ è utile fare uso della seconda delle (2.5.4).

In pratica, dato che tutti i software econometrici disponibili calcolano automaticamente la devianza di una regressione, mentre non tutti dispongono di funzioni di calcolo matriciale, la scelta del ricercatore si orienterà più spesso verso la prima delle (2.5.4).

La verifica dell'ipotesi di omogeneità nei prezzi

Verifichiamo l'ipotesi di omogeneità di grado zero sui prezzi (1.11.4) facendo uso della prima delle (2.5.4). Dalla (2.3.17) sappiamo che $\hat{\mathbf{u}}'\hat{\mathbf{u}} = 0.0633$ mentre dalla (2.3.19) che $\mathbf{u}_0'\mathbf{u}_0 = 0.0699$; inoltre è n-k = 75 e q = 1, per cui si ottiene

$$\frac{0.0699-0.0633}{0.0633}\frac{75}{1}=7.820 \tag{2.5.5}$$

che è più grande del quantile della distribuzione della F di Fisher con 1 e 75 gradi di libertà che vale 3.97 per $\alpha = 0.05$. Se abbassiamo il livello di significatività del test al valore $\alpha = 0.01$ il quantile è pari a 6.99, anche in questo caso più basso del rapporto (2.5.5).