# Minimi quadrati vincolati e test F

## Impostazione del problema

Spesso, i modelli econometrici che stimiamo hanno dei parametri che sono passibili di interpretazione diretta nella teoria economica. Consideriamo ad esempio una funzione di produzione Cobb-Douglas $Q = AK^{\alpha_1}L^{\alpha_2}$ . È noto dalla teoria microeconomica (o almeno, dovrebbe) che la Cobb-Douglas ha rendimenti di scala costanti se e solo se $\alpha + \alpha_2 = 1$. Scrivendo la funzione in logaritmi si ha

$$q = a + \alpha_1 k + \alpha_2 l$$

Supponiamo di condurre un esperimento in cui facciamo variare a nostro piacimento $k$ e $l$, e osserviamo i cambiamenti in $q$. Se le ipotesi classiche sono rispettate, è naturale pensare di stimare il vettore di parametri

$$\boldsymbol{\beta} = \begin{pmatrix} a \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

con i minimi quadrati.

Se però sapessimo — o congetturassimo — che la funzione è a rendimenti di scala costanti, vorremmo che la nostra stima di $\boldsymbol{\beta}$ incorporasse l'informazione $\alpha_1 + \alpha_2 = 1$. Ovviamente, non c'è alcuna garanzia che $\widehat{\boldsymbol{\beta}}$ rispetti questa condizione. In altri termini, cerchiamo uno stimatore $\widetilde{\boldsymbol{\beta}}$, che rispetti per costruzione il vincolo

$$R\widetilde{\boldsymbol{\beta}} = d$$

dove $R = [0\ 1\ 1]$ e $d = 1$.

Più in generale, vogliamo uno stimatore del modello $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ che soddisfi *a priori* un insieme di $l$ restrizioni che possiamo scrivere come $R\boldsymbol{\beta} = d$.

## Soluzione

Per trovarlo, minimizziamo la somma dei quadrati dei residui sotto vincolo. Definendo i residui come $\mathbf{e}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ il Lagrangiano sarà

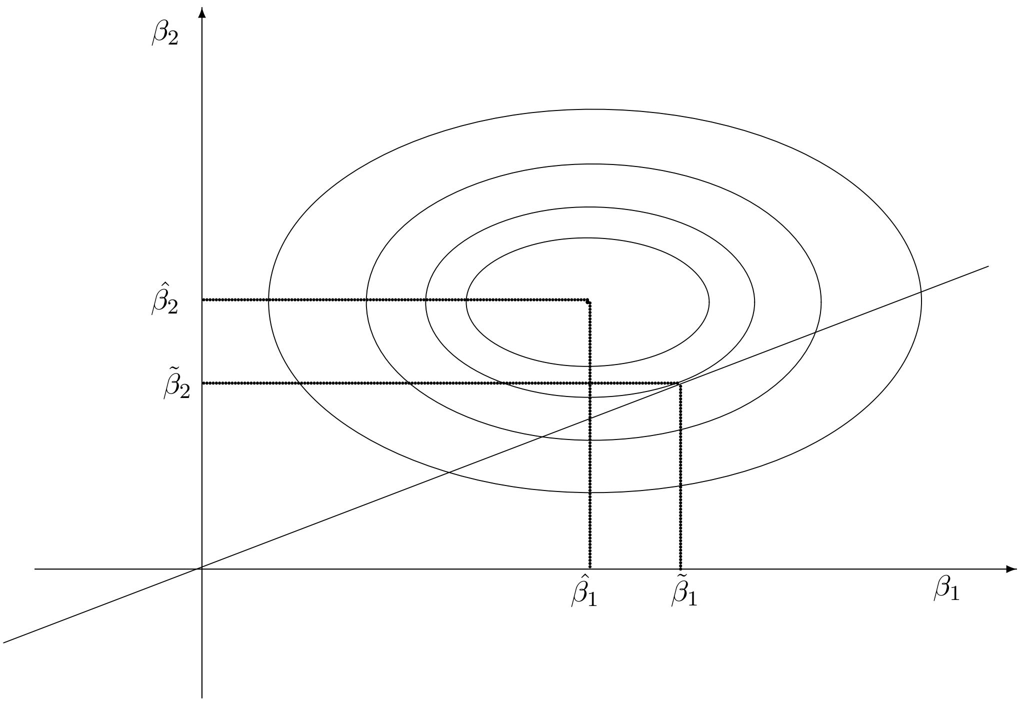
$$L = \frac{1}{2}\mathbf{e}'\mathbf{e} + \lambda'(R\boldsymbol{\beta} - d).$$

Poiché la derivata di $\mathbf{e}$ rispetto a $\boldsymbol{\beta}$ è $-\mathbf{X}$, la condizione di primo ordine può essere scritta

$$\mathbf{X}'\tilde{\mathbf{e}} = R'\lambda, \tag{1}$$

dove indichiamo con $\widetilde{\boldsymbol{\beta}}$ il vettore che rende vera la (1) e con $\tilde{\mathbf{e}}$ il vettore $\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$.

Figura 1: Esempio: vettore di due parametri




*Le ellissi sono le curve di livello della funzione* $\mathbf{e}'\mathbf{e}$. *Il vincolo è* $\beta_1 = 3\beta_2$. *Il numero di parametri k è uguale a 2 e il numero di vincoli l è pari a 1. Il punto di minimo non vincolato è* $\widehat{\beta}_1, \widehat{\beta}_2$*; Il punto di minimo vincolato è* $\tilde{\beta}_1, \tilde{\beta}_2$.

L'equazione (1) può essere riscritta in modo tale da rendere evidenti le relazioni che esistono fra il problema di minimo vincolato (e la sua soluzione) e il problema di minimo libero (e la sua soluzione, che è ovviamente lo stimatore OLS). In particolare, possiamo considerare le implicazioni della (1)

1. nello spazio dei parametri ($\mathbb{R}^k$)
2. nello spazio dei vincoli ($\mathbb{R}^l$)
3. nello spazio delle osservazioni ($\mathbb{R}^T$)
4. nello spazio della funzione obiettivo ($\mathbb{R}$).

Cominciamo coi parametri: premoltiplicando la (1) per $(\mathbf{X}'\mathbf{X})^{-1}$ si ottiene

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} R' \lambda$$

da cui si può ricavare una relazione interessante fra lo stimatore vincolato e quello libero:

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} R' \lambda \tag{2}$$

Lo stimatore vincolato, quindi, è uguale a quello libero a cui viene aggiunto un "fattore di correzione" proporzionale a $\lambda$.

La seconda cosa che si può dire riguarda lo spazio dei vincoli, e quindi il valore di $\lambda$: premoltiplicando la (2) per $R$ si ha che

$$\lambda = \left[ R(\mathbf{X}'\mathbf{X})^{-1} R' \right]^{-1} (R\widehat{\boldsymbol{\beta}} - d) \tag{3}$$

perché $R\widetilde{\boldsymbol{\beta}} = d$ per costruzione.

Dovrebbe essere chiaro dalla (3) che, se lo stimatore non vincolato rispetta già di per sé il vincolo ($R\widehat{\boldsymbol{\beta}} = d$), allora $\lambda = 0$ e quindi lo stimatore vincolato coincide con quello libero. In questo senso, si può dire che il vettore $\lambda$ ci dà una misura di quanto lo stimatore vincolato è diverso da quello libero; sarò più preciso fra poco. La formula che si trova di solito nei libri di testo la si ottiene combinando le equazioni (2) e (3):

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} R' \left[ R(\mathbf{X}'\mathbf{X})^{-1} R' \right]^{-1} (R\widehat{\boldsymbol{\beta}} - d) \tag{4}$$

Possiamo esaminare cosa succede nello spazio delle osservazioni premoltiplicando la (2) per $\mathbf{X}$:

$$\mathbf{X}\widetilde{\boldsymbol{\beta}} = \tilde{\mathbf{y}} = \hat{\mathbf{y}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} R' \lambda$$

da cui discende

$$\tilde{\mathbf{e}} = \hat{\mathbf{e}} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} R' \lambda$$

Consideriamo ora lo spazio della funzione obiettivo: la somma dei quadrati dei residui vincolati (cioè il minimo vincolato) $\tilde{\mathbf{e}}'\tilde{\mathbf{e}}$ può essere scritta nel seguente modo:

$$\tilde{\mathbf{e}}'\tilde{\mathbf{e}} = \hat{\mathbf{e}}'\hat{\mathbf{e}} + \lambda' R(\mathbf{X}'\mathbf{X})^{-1}R'\lambda \tag{5}$$

dove abbiamo sfruttato il fatto che, per costruzione, $\mathbf{X}'\hat{\mathbf{e}} = 0$. Ora, la (5) ci dice una cosa importante: la differenza che c'è fra il minimo vincolato e il minimo libero (che è evidentemente sempre positiva) può essere scritta come una forma quadratica in $\lambda$.

Mettendo assieme la equazioni (3) e (5) si arriva alla seguente uguaglianza:

$$\begin{gathered} \tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}} = \\ \lambda' R(\mathbf{X}'\mathbf{X})^{-1}R'\lambda = \\ (R\widehat{\boldsymbol{\beta}} - d)' \left[R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1} (R\widehat{\boldsymbol{\beta}} - d) \end{gathered} \tag{6}$$

L'espressione (6) è molto interessante, perché ci dice che la stessa quantità può essere interpretata in tre modi diversi ed equivalenti:

1. $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}$ è la differenza che c'è fra la funzione obiettivo vincolata e non. Maggiore è questa differenza, maggiore è la perdita di capacità che il modello vincolato ha di accostarsi ai dati empiricamente osservati;

2. $\lambda' R(\mathbf{X}'\mathbf{X})^{-1}R'\lambda$ è una forma quadratica che vale 0 solo se $\lambda = 0$. poiché abbiamo già visto che $\lambda = 0$ solo se le stime vincolate coincidono con quelle libere, questa grandezza varia sostanzialmente con la distanza fra il vettore $\widehat{\boldsymbol{\beta}}$ ed il vettore $\boldsymbol{\beta}$; poiché $\lambda$ è il vettore dei moltiplicatori di Lagrange del problema di minimo vincolato è possibile — come è noto — darne una lettura in termini di prezzo ombra: l'$i$-esimo elemento del vettore $\lambda$ ci dice quanto migliora la funzione obiettivo ad una variazione 'piccola' del vincolo corrispondente[1];

3. la grandezza $(R\widehat{\boldsymbol{\beta}} - d)' \left[R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1} (R\widehat{\boldsymbol{\beta}} - d)$ è una forma quadratica (definita positiva) in $(R\widehat{\boldsymbol{\beta}} - d)$, ossia in un vettore che è pari a 0 solo se lo stimatore libero rispetta già di per sé il vincolo.

## Prova di ipotesi

Sotto le ipotesi classiche è possibile utilizzare i risultati esposti fin qui a fini di prova di ipotesi: in questo caso il nostro vincolo $(R\boldsymbol{\beta} - d)$ rappresenta l'ipotesi nulla che intendiamo sottoporre a verifica.

Sulla base delle considerazioni fatte in precedenza ci si potrebbe attendere che una statistica test ragionevole possa essere basata sulle diverse capacità di adattamento ai dati dei modelli libero e vincolato. In effetti, è proprio così.

[1] Formalmente, si può dimostrare che $\lambda$ è il vettore di derivate parziali di $\tilde{\mathbf{e}}'\tilde{\mathbf{e}}/2$ rispetto a $d$.

Consideriamo la differenza $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}$. Come abbiamo visto prima, questa grandezza può essere scritta anche come

$$(R\widehat{\boldsymbol{\beta}} - d)' \left[R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1} (R\widehat{\boldsymbol{\beta}} - d).$$

Se valgono le ipotesi classiche, sappiamo che $\widehat{\boldsymbol{\beta}} \sim N\left[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right]$. Di conseguenza, date le proprietà della normale,

$$(R\widehat{\boldsymbol{\beta}} - d) \sim N\left[R\boldsymbol{\beta} - d, \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'\right]$$

Se effettivamente il vincolo vale (e quindi $R\boldsymbol{\beta} = d$), allora il valore atteso di $(R\widehat{\boldsymbol{\beta}} - d)$ è 0 e, più in generale, $(R\widehat{\boldsymbol{\beta}} - d) \sim N\left[0, \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'\right]$, cosicché l'espressione

$$(R\widehat{\boldsymbol{\beta}} - d)' \left[R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1} (R\widehat{\boldsymbol{\beta}} - d)$$

è proporzionale (a meno di un fattore $\frac{1}{\sigma^2}$) ad una variabile casuale $\chi^2$ con $l$ gradi di libertà (tanti quante sono le righe di $R$).

Questa quantità può essere quindi presa come base per un test sull'ipotesi $H_0 : R\boldsymbol{\beta} - d$. Infatti, se quest'ipotesi è vera, allora la grandezza nella (6) — divisa per $\sigma^2$ — si distribuisce come una $\chi^2$, e valori "grandi" possono verificarsi

- o in casi eccezionali
- o quando l'ipotesi nulla è falsa.

Sulla base di questa considerazione, possiamo decidere di rifiutare $H_0$ quando $\frac{\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2}$ assume valori "grandi" (relativamente alle tavole del $\chi^2$). C'è solo un problema: normalmente $\sigma^2$ non è una grandezza nota. Abbiamo però una sua stima

$$s^2 = \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T - k}$$

che, come sappiamo, è corretta.

Per derivare una statistica test con una distribuzione trattabile, si consideri che la grandezza

$$\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2} = (T - k)\frac{s^2}{\sigma^2}$$

si distribuisce anch'essa come una $\chi^2$ con $T - k$ gradi di libertà. Se riuscissimo a dimostrare che $s^2$ è indipendente da $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}$, allora potremmo sfruttare le proprietà[2] delle distribuzioni $\chi^2$ ed $F$.

[2] Se

$$\begin{cases} A \sim \chi^2_n \\ B \sim \chi^2_d \\ A \perp\!\!\!\perp B \end{cases}$$

allora $\frac{A/n}{B/d}$ si distribuisce come una $F$ con gradi di libertà pari a $n$ e $d$.

Consideriamo il rapporto

$$\frac{\frac{\tilde{\mathbf{e}}'\tilde{\mathbf{e}}-\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2}}{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2}} \cdot \frac{T-k}{l}, \tag{7}$$

che può essere visto come il rapporto fra due variabili casuali $\chi^2$ divise per i rispettivi gradi di libertà. Da quanto detto sopra, se riuscissimo a provare che il denominatore ed il numeratore sono indipendenti, avremmo che la statistica

$$F = \frac{\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}} \cdot \frac{T-k}{l}$$

(ottenuta semplicemente semplificando il fattore $\sigma^2$ dall'equazione (7)) si distribuisce come una F di Snedecor. La dimostrazione, tuttavia, è immediata, in quanto abbiamo già visto che $s^2$ è indipendente da $\widehat{\beta}$. Poiché, in forza della (6), si ha che $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}}$ può essere scritta come una forma quadratica in $\widehat{\beta}$ dove la matrice "in mezzo" è nonstocastica, se ne può dedurre anche il risultato che ci serve[3].

[3]Può essere un simpatico esercizio dimostrare che il test $F$ può anche essere scritto come

$$F = \frac{\tilde{\mathbf{e}}'\mathbf{P_X}\tilde{\mathbf{e}}}{\tilde{\mathbf{e}}'\mathbf{M_X}\tilde{\mathbf{e}}} \cdot \frac{T-k}{l}$$

e dimostrare l'indipendenza di numeratore e denominatore per questa via.