Appunti di Econometria

ARGOMENTO [2]: ESTENSIONI DEL MODELLO LINEARE

Tommaso Nannicini – Università Bocconi

Ottobre 2010

1 Scelta della forma funzionale

Abbiamo visto che abbandonare l'assunzione di normalità degli errori (A6) non crea troppi problemi ai nostri stimatori OLS, visto che possiamo continuare a usare le stesse procedure inferenziali discusse sotto l'ipotesi di normalità facendo appello alle proprietà asintotiche degli stimatori (cioè, potendo disporre di campioni grandi). Anche l'abbandono dell'assunzione di linearità (A1) non pone sfide insormontabili agli stimatori OLS. In molti casi, è sufficiente aggiustare la forma funzionale del modello statistico per catturare le non-linearità. Seguono alcuni esempi importanti.

Modello log-log

$$y_i = e^{\beta_1} x_i^{\beta_2} e^{\epsilon_i} \Rightarrow \log y_i = \beta_1 + \beta_2 \log x_i + \epsilon_i.$$

In questo caso, l'effetto marginale non è più costante come nel modello interamente lineare già analizzato (dove $dy/dx = \beta_2$ per ogni x). In questo caso, è l'**elasticità** ad essere costante, dato che possiamo interpretare il coefficiente β_2 appunto come l'elasticità di y rispetto ad x:

$$\beta_2 = \frac{d \log y_i}{d \log x_i} = \frac{(d \log y_i / dy_i) dy_i}{(d \log x_i / dx_i) dx_i} = \frac{dy_i / y_i}{dx_i / x_i} = E_{yx}.$$

Sulla base di questo risultato, possiamo interpretare β_2 come la variazione percentuale di y associata a una variazione di x dell'1% (approssimativamente per variazioni discrete). Il rapporto tra il coefficiente e il semplice effetto marginale (dy/dx) è $\beta_2 = (dy/dx)(y/x)$ e ovviamente dipende dai valori di y ed x.

Modello log-lin

$$\log y_i = \beta_1 + \beta_2 x_i + \epsilon_i,$$

dove la variabile dipendente è in forma logaritmica e il regressore no. In questo caso il coefficiente cattura la variazione relativa di y rispetto a una variazione di x:

$$\beta_2 = \frac{d \log y_i}{dx_i} = \frac{(d \log y_i / dy_i) dy_i}{dx_i} = \frac{dy_i / y_i}{dx_i}.$$

Sempre approssimativamente per variazioni discrete, possiamo quindi interpretare β_2 come la variazione percentuale di y associata a una variazione di x di una unità ($\Delta x = 1$). Questa approssimazione è ottima per valori bassi di β_2, mentre comincia a fare un po' acqua per valori grandi. Può essere allora utile ricordare anche il valore esatto della variazione percentuale di y associata a $\Delta x = 1$: $\Delta\% y = e^{\beta_2} - 1$.

Modello lin-log

$$y_i = \beta_1 + \beta_2 \log x_i + \epsilon_i,$$

dove il regressore è in forma logaritmica e la variabile dipendente no. In questo caso il coefficiente cattura la variazione di y rispetto a una variazione relativa di x:

$$\beta_2 = \frac{dy_i}{d\log x_i} = \frac{dy_i}{(d\log x_i/dx_i)dx_i} = \frac{dy_i}{dx_i/x_i}.$$

Sempre approssimativamente per variazioni discrete, abbiamo che un aumento dell'1% di x è associato a una variazione di y pari a $0.01\beta_2$.

Specificazione quadratica

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \epsilon_i,$$

dove il regressore x appare sia linearmente sia al quadrato. Anche in questo caso, l'effetto marginale smette di essere costante e dipende dai valori di x:

$$\frac{dy_i}{dx_i} = \beta_2 + 2\beta_3 x_i.$$

Per esempio, può essere utile calcolare l'effetto marginale della x sulla y per valori particolari del regressore (la media o altro). Se i segni di β_2 e β_3 sono diversi, la relazione (non-lineare) tra la y e la x cambia direzione a un certo punto, come in una parabola. Ricordando l'equazione di una **parabola** con asse verticale sul piano ($y = ax^2 + bx + c$) e la formula per l'ascissa del suo vertice ($x_v = -b/2a$), possiamo facilmente vedere che il punto in cui la relazione tra y e x cambia direzione è dato da: $x^* = -\beta_2/2\beta_3$.

Modello con interazioni

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 x_i z_i + \epsilon_i,$$

dove i regressori x e z sono interagiti tra loro. In questo caso, il coefficiente β_2 cattura l'effetto marginale della x sulla y quando $z = 0$. In generale, l'effetto marginale dipende dall'altro regressore z: $dy/dx = \beta_2 + \beta_4 z_i$. Lo stesso vale per z, il cui effetto dipende dall'altro regressore x: $dy/dz = \beta_3 + \beta_4 x_i$.

Test RESET sulla corretta specificazione della forma funzionale

A questo punto, nasce spontanea la domanda: come scegliere la forma funzionale corretta? Come possiamo sapere se la semplice forma lineare è appropriata o meno? Una possibilità è quella di partire dall'ipotesi nulla che la forma lineare è corretta (A1) e usare un procedura di test per valutare la plausibilità di questa ipotesi. È quello che fa il cosiddetto **test di Ramsey** o **test RESET** (*REgression Specification Error Test*). Si stimi il modello multivariato (che diventerà il nostro modello ristretto)

$$M_r \rightarrow y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

e si calcolino i valori fittati $\hat{y}$. I termini $\hat{y}^m$ (per m interi e maggiori di 1) rappresentano funzioni non-lineari delle x. Possiamo quindi aggiungerli al modello ristretto e stimare un modello completo con generiche non-linearità incluse tra i regressori (ci fermiamo ad $m = 4$ ma altre scelte sono altrettanto plausibili):

$$M_c \to y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + \delta_3 \hat{y}^4 + \epsilon.$$

L'ipotesi nulla di linearità è quindi data da: $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$. Possiamo usare il solito test F per decidere se rifiutare H_0. Per valori alti della statistica F osservata, il test RESET ci dice che dobbiamo rifiutare l'ipotesi di linearità e che ci sono problemi di specificazione funzionale nel modello ristretto. L'esatta forma della non-linearità da includere nel modello, però, non è conosciuta e va ricercata nei dati o nel problema economico oggetto di studio.

2 Variabili dummy e stabilità dei parametri

In molte applicazioni, è estremamente utile l'uso di variabili **dummy** o **binarie** (cioè, variabili che possono prendere soltanto 0 o 1 come valori). Per esempio, possiamo usarle per testare l'ipotesi che la relazione catturata dal nostro modello di regressione sia stabile tra gruppi diversi di osservazioni (osservazioni appartenenti a periodi di tempo diversi; oppure osservazioni appartenenti a individui di sesso diverso; e così via). Si consideri la variabile binaria D_i, con $D_i = 1$ se l'osservazione appartiene al gruppo A e $D_i = 0$ se l'osservazione appartiene al gruppo B. Se stimiamo il modello

$$y_i = \alpha_0 + \beta_0 x_i + D_i(\alpha_1 + \beta_1 x_i) + \epsilon_i,$$

stiamo di fatto stimando due diverse rette di regressione:

- una per il gruppo B ($D_i = 0$), $\hat{y}_i = \hat{\alpha}_0 + \hat{\beta}_0 x_i$,
- ed una per il gruppo A ($D_i = 1$), $\hat{y}_i = (\hat{\alpha}_0 + \hat{\alpha}_1) + (\hat{\beta}_0 + \hat{\beta}_1)x_i$.

Il coefficiente della dummy additiva ($\hat{\alpha}_1$) cattura l'effetto dell'appartenenza al gruppo A (rispetto a B) sulla variabile y, mentre il coefficiente della dummy moltiplicativa ($\hat{\beta}_1$) cattura l'effetto differenziale di x su y per il gruppo A (rispetto al gruppo B). In altre parole, $\hat{\alpha}_1$ è la differenza tra l'intercetta della retta di regressione stimata per A e l'intercetta della retta per B, mentre $\hat{\beta}_1$ è la differenza tra la pendenza della retta di regressione stimata per A e la pendenza della retta per B. Per testare la stabilità dei parametri tra i gruppi A e B, possiamo quindi eseguire un test F sull'ipotesi nulla $H_0 : \hat{\alpha}_1 = \hat{\beta}_1 = 0$ (**Chow test**). Possiamo anche guardare ai singoli test t sui due coefficienti per capire se l'eventuale mancanza di stabilità proviene da un effetto sui livelli o da un effetto differenziale dell'altro regressore (x).

In notazione matriciale, abbiamo il modello:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{n_a} \\ y_{n_a+1} \\ \vdots \\ y_{n_a+n_b} \end{pmatrix} = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n_a} & 0 & 0 \\ 0 & 0 & 1 & x_{n_a+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{n_a+n_b} \end{pmatrix} \begin{pmatrix} \alpha_0 + \alpha_1 \\ \beta_0 + \beta_1 \\ \alpha_0 \\ \beta_0 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{n_a+n_b} \end{pmatrix},$$

dove n_a è la numerosità del gruppo A e n_b la numerosità del gruppo B, e le osservazioni sono state ordinate sulla base della loro appartenenza ai due gruppi. Il modello può essere riscritto sinteticamente come:

$$\begin{pmatrix} \mathbf{y_a} \\ \mathbf{y_b} \end{pmatrix} = \begin{pmatrix} X_a & 0 \\ 0 & X_b \end{pmatrix} \begin{pmatrix} \beta_\mathbf{a} \\ \beta_\mathbf{b} \end{pmatrix} + \epsilon.$$

Applicando gli stimatori OLS su tutto il modello, si vede facilmente che alla fine: $\hat{\beta}_\mathbf{a} = (X'X)^{-1}X'\mathbf{y_a}$ e $\hat{\beta}_\mathbf{b} = (X'X)^{-1}X'\mathbf{y_b}$. In altre parole, separare i due campioni e applicare OLS in ognuno di essi darebbe lo stesso risultato. Tenendoli insieme, però, possiamo testare la stabilità del modello tra i due gruppi applicando un test F su restrizioni lineari multiple. In questo caso, dobbiamo testare due ipotesi congiunte della forma $R\beta = \mathbf{q}$, cioè:

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_0 + \alpha_1 \\ \beta_0 + \beta_1 \\ \alpha_0 \\ \beta_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Il test di Chow, quindi, non è altro che un test F su queste ipotesi:

$$\frac{(R\hat{\beta} - \mathbf{q})'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - \mathbf{q})}{\mathbf{e}'\mathbf{e}} \cdot \frac{n_a + n_b - 4}{2} = F_{oss}.$$

Se la F_{oss} è maggiore del valore critico che ci siamo scelti sulla base del livello di significatività attribuito al test, il test di Chow rigetta l'ipotesi che la relazione lineare sia identica nei due gruppi.

3 Variabili omesse e variabili irrilevanti

Vediamo meglio, adesso, che implicazioni ha la scelta dei regressori x_j. Quante variabili dobbiamo includere nel nostro modello? E quali? Un risultato importante ci dice che non è la stessa cosa omettere variabili rilevanti rispetto a includere variabili irrilevanti: i due problemi sono molto diversi in termini di conseguenze sugli stimatori OLS.

Esclusione di variabili rilevanti

Assumiamo che il vero modello sia

$$\mathbf{y} = X_1\beta_\mathbf{1} + X_2\beta_\mathbf{2} + \epsilon,$$

mentre noi stimiamo il modello $\mathbf{y} = X_1\beta_\mathbf{1} + \epsilon$ (omettendo quindi i regressori rilevanti X_2). Lo stimatore OLS sarà:

$$\hat{\beta}_\mathbf{1} = (X_1'X_1)^{-1}X_1'\mathbf{y} = \beta_\mathbf{1} + (X_1'X_1)^{-1}X_1'X_2\beta_\mathbf{2} + (X_1'X_1)^{-1}X_1'\epsilon.$$

Da cui:

$$E(\hat{\beta}_\mathbf{1}) = \beta_\mathbf{1} + (X_1'X_1)^{-1}X_1'X_2\beta_\mathbf{2}. \tag{1}$$

A causa delle variabili omesse X_2, lo stimatore OLS non è più corretto. O meglio: il **termine di distorsione** è nullo soltanto se le variabili omesse non hanno nessun effetto sulla variabile dipendente ($\beta_\mathbf{2} = \mathbf{0}$), oppure se le variabili omesse sono ortogonali rispetto agli altri regressori ($(X_1'X_1)^{-1}X_1'X_2 = \mathbf{0}$). Nel caso sia violata la prima condizione, si può dire che le variabili omesse hanno un *outcome effect*; nel caso sia violata la seconda, si può dire che le variabili omesse hanno un *selection effect*: questi due effetti congiuntamente creano la distorsione degli stimatori OLS. Si noti che l'omissione di variabili rilevanti equivale a una violazione della nostra assunzione A4, visto che adesso i termini di errore contengono anche variabili che sono correlate con i regressori. Questa violazione è grave: gli OLS smettono di essere corretti e consistenti! Non basta una fermata ai box per aggiustare gli stimatori dei minimi quadrati, come abbiamo fatto abbandonando A6 (normalità) o A1 (linearità). Se è l'assunzione A4 ad essere violata, dobbiamo proprio cambiare macchina (cioè, stimatori): è quello che faremo introducendo le cosiddette **variabili strumentali**. Per il momento, ci siamo limitati a dimostrare la gravità del problema per gli stimatori dei minimi quadrati.

Notando che il termine $(X_1'X_1)^{-1}X_1'X_2$ non è altro che una matrice in cui ogni colonna contiene i coefficienti OLS della regressione della rispettiva colonna di X_2 su X_1, nel caso bivariato possiamo scrivere:

$$E(\hat{\beta}) = \beta + \frac{Cov(x,z)}{Var(x)}\delta, \tag{2}$$

dove z è la variabile omessa, δ il coefficiente di z nel vero modello, e β il coefficiente del regressore x. Questa semplice formula ci può permettere di prevedere almeno il segno della distorsione degli stimatori dei minimi quadrati in casi particolari (ammesso di avere una qualche idea sulla natura della variabile omessa).

Inclusione di variabili irrilevanti

Se l'omissione di variabili irrilevanti equivale a imporre sul modello una restrizione che non è vera ($\boldsymbol{\beta_2} = \mathbf{0}$), l'inclusione di variabili irrilevanti equivale a non imporre una restrizione che sarebbe vera. Già questo ci fa intuire che il problema è minore. È facile vedere, infatti, che se il vero modello è $\mathbf{y} = X_1\boldsymbol{\beta_1} + \epsilon$, ma noi stimiamo $\mathbf{y} = X_1\boldsymbol{\beta_1} + X_2\boldsymbol{\beta_2} + \epsilon$, allora avremo che: $E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta_1}$ e $E(\hat{\boldsymbol{\beta}}_2) = \mathbf{0}$. Quindi, gli stimatori OLS sono ancora corretti. Si tenga presente, tuttavia, che nessun pasto è gratis: a furia di aggiungere regressori irrilevanti al nostro modello, riduciamo i gradi di libertà e l'accuratezza delle stime (inflazionando la varianza). Inoltre, possiamo facilmente incorrere nel problema di *over-controlling* da un punto di vista teorico o nel problema di multicollinearità imperfetta.

4 Eteroschedasticità

Vediamo adesso che cosa succede ai nostri stimatori OLS se rimuoviamo l'assunzione di omoschedasticità (A5). Sappiamo che gli stimatori restano corretti e consistenti (visto che per dimostrare queste proprietà facciamo appello solo alle assunzioni da A1 ad A4). Ci aspettiamo, invece, che la presenza di eteroschedasticità crei problemi alla stima della varianza degli stimatori e alle procedure inferenziali. Partiamo dal **caso bivariato**. Sulla base dell'equazione (13) nell'argomento 1 di questi appunti, possiamo scrivere la varianza dello stimatore della pendenza come:

$$Var(\hat{\beta}_2) = \frac{Var(\sum(x_i - \bar{x})\epsilon_i)}{(\sum(x_i - \bar{x})^2)^2} = \frac{\sum(x_i - \bar{x})^2\sigma_i^2}{(\sum(x_i - \bar{x})^2)^2} \tag{3}$$

dato che questa volta $\sigma_i^2 \neq \sigma^2$ (eteroschedasticità). Quindi, le procedure inferenziali sviluppate in precedenza sulla base della varianza $\frac{\sigma^2}{\sum(x_i - \bar{x})^2}$ non sono più corrette. Gli stimatori OLS non sono efficienti e i test non danno i risultati previsti.

Lo stesso vale, ovviamente, nel **caso multivariato**. Se assumiamo che $E(\epsilon\epsilon') = \sigma^2\Omega \neq \sigma^2 I$ (eteroschedasticità), la formula per la matrice di varianza e covarianza degli stimatori diventa:

$$Var(\hat{\beta}) = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}, \tag{4}$$

che di nuovo è diversa da quella derivata sotto condizione di omoschedasticità ($\sigma^2(X'X)^{-1}$, si veda l'argomento 1 di questi appunti). La stima della varianza, la dimostrazione del teorema di Gauss-Markov e le procedure inferenziali derivate in precedenza non sono valide.

Test di Breusch-Pagan

Nella letteratura econometrica, sono stati proposti vari test per individuare la presenza di eteroschedasticità. Vediamone uno: il **test di Breusch-Pagan**. Partendo dal solito modello multivariato

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon,$$

vogliamo testare l'ipotesi nulla $H_0 : Var(\epsilon|\mathbf{x}) = E(\epsilon^2|\mathbf{x}) = E(\epsilon^2) = \sigma^2$ (assumendo che valgano le altre assunzioni del modello di regressione lineare, tranne A5). In altre parole, assumendo una relazione lineare, vogliamo stimare

$$\epsilon^2 = \delta_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v$$

e testare l'ipotesi nulla $H_0 : \delta_2 = \cdots = \delta_k = 0$. Poiché non osserviamo i veri termini di errore, stimiamo la seconda regressione usando come variabile dipendente i quadrati dei residui della prima regressione:

$$e^2 = \delta_1 + \delta_2 x_2 + \cdots + \delta_k x_k + \eta.$$

E testiamo H_0 con il test F:

$$\frac{R_e^2/(K-1)}{(1-R_e^2)/(N-K)} \sim F_{K-1,N-K}.$$

Se la F osservata è maggiore del valore critico prescelto, abbiamo problemi di eteroschedasticità.

Una prima soluzione: lo stimatore Huber-White della varianza

Una prima soluzione al problema è quella di trovare uno stimatore dell'equazione (3) per il caso bivariato e dell'equazione (4) per il caso multivariato che sia robusta alla presenza di eteroschedasticità. È quello che fa il cosiddetto **stimatore Huber-White** della varianza (e quindi dello *standard error*) degli stimatori. In pratica, la stima dell'equazione (3) diventa:

$$\hat{Var}(\hat{\beta}_2) = \frac{\sum (x_i - \bar{x})^2 e_i^2}{(\sum (x_i - \bar{x})^2)^2}. \tag{5}$$

La stima dell'equazione (4) diventa:

$$\hat{Var}(\hat{\beta}) = (X'X)^{-1} \sum e_i^2 \mathbf{x_i}' \mathbf{x_i} (X'X)^{-1}, \tag{6}$$

dove $\mathbf{x_i}$ è il vettore delle k variabili osservate per l'osservazione i. Lo stimatore Huber-White è chiamato a volte *sandwich estimator*, visto che la parte rilevante della stima sta nel mezzo al "panino" con i due estremi $(X'X)^{-1}$. Una volta applicato questo metodo per stimare la matrice di varianza e covarianza, anche in presenza di eteroschedasticità, si può procedere nel solito modo sia per individuare lo *standard error* di ogni coefficiente stimato e quindi il test t, sia per individuare il test F su ipotesi multiple. L'unico "problema" è che questi risultati valgono solo in grandi campioni perché si basano su proprietà asintotiche.

Una seconda soluzione: i minimi quadrati pesati (WLS)

Una seconda soluzione per affrontare il problema di eteroschedasticità coincide con gli stimatori dei minimi quadrati pesati (WLS, *Weighted Least Squares*). Si parta di nuovo dal modello multivariato:

$$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i.$$

Si assuma di essere in presenza di un problema di eteroschedasticità con questa struttura: $\sigma_i^2 = \sigma^2 h(\mathbf{x_i}) = \sigma^2 h_i$ (la varianza non è costante e, in particolare, dipende dalle x_j). Di nuovo, valgono tutte le assunzioni del modello di regressione lineare (tranne A5). Di conseguenza, è facile vedere che:

$$E\left(\frac{\epsilon_i}{\sqrt{h_i}}\middle|\mathbf{x_i}\right) = 0$$

$$Var\left(\frac{\epsilon_i}{\sqrt{h_i}}\middle|\mathbf{x_i}\right) = \frac{\sigma^2 h_i}{h_i} = \sigma^2.$$

Stimiamo allora il modello trasformato:

$$y_i^* = \beta_1(1/\sqrt{h_i}) + \beta_2 x_{2i}^* + \cdots + \beta_k x_{ki}^* + \epsilon_i^*,$$

dove $y_i^* = y_i/\sqrt{h_i}$, $x_{ji}^* = x_{ji}/\sqrt{h_i}$ e $\epsilon_i^* = \epsilon_i/\sqrt{h_i}$. Il modello trasformato soddisfa tutte le ipotesi del modello di regressione lineare da A1 ad A5, e gli stimatori dei minimi quadrati sono sia corretti sia efficienti. Questi stimatori WLS sono un esempio di stimatori GLS (*Generalized Least Squares*), che vedremo meglio nel prossimo paragrafo e che sono una classe generale di stimatori validi quando non vale la condizione $E(\epsilon\epsilon') = \sigma^2 I$. Il problema principale nell'applicazione della procedura WLS appena descritta è che in pratica non è facile conoscere *ex ante* il valore dei pesi h_i, tranne rari casi in cui i pesi sono dati da uno specifico problema di eteroschedasticità (si pensi all'esempio dei **dati raggruppati** analizzato in classe).

In pratica, se uno vuole applicare la procedura WLS, deve trovare un modo per stimare i pesi h_i. In tal caso, stiamo usando stimatori FGLS (*Feasible Generalized Least Squares*, si veda il prossimo paragrafo). Si stimi

$$\log e_i^2 = \delta_1 + \delta_2 x_{2i} + \cdots + \delta_k x_{ki} + \epsilon_i$$

e si calcolino i valori fittati $\hat{g}_i$. La stima dei pesi WLS sarà quindi data da: $\hat{h}_i = e^{\hat{g}_i}$. Dopodiché, si applichi la procedura come sopra, cioè stimando il modello trasformato. In presenza di eteroschedasticità, gli stimatori FGLS sono consistenti e più efficienti degli OLS asintoticamente (in grandi campioni).

5 Stimatori GLS

La procedura dei WLS, che abbiamo appena discusso per risolvere il problema di eteroschedasticità, come detto, è un caso particolare di stimatore GLS. Questa classe di stimatori ha proprietà migliori degli OLS in presenza di **errori non-sferici**. Se per qualche ragione $E(\epsilon\epsilon') = \sigma^2\Omega \neq \sigma^2 I$, allora gli OLS smetteranno di essere BLUE. In particolare, gli stimatori dei minimi quadrati rimarranno corretti e consistenti, visto che per dimostrare queste proprietà sono sufficienti le assunzioni da A1 ad A4 fatte introducendo il modello di regressione lineare. Ma il teorema di Gauss-Markov (che richiede anche A5) non sarà in genere valido. E la varianza degli stimatori OLS diventerà:

$$Var[\hat{\beta}] = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1},$$

rendendo errate le procedure inferenziali presentate nell'argomento 1 di questi appunti.

Che fare, dunque? Si assuma di avere una matrice di trasformazione T, non singolare e di dimensioni $[N \times N]$. E di applicarla al nostro modello di regressione lineare:

$$T\mathbf{y} = TX\beta + T\epsilon.$$

In questo caso: $E(T\epsilon\epsilon'T') = \sigma^2 T\Omega T'$. Se potessimo scegliere T in modo da avere $T\Omega T' = I$, gli errori del modello trasformato sarebbero sferici e i minimi quadrati di nuovo BLUE. In effetti, se si conosce Ω, è possibile individuare una matrice T con queste caratteristiche. Dato che la matrice Ω è simmetrica e definitva positiva, esiste una matrice P non singolare tale che: $\Omega = PP'$. Da cui: $P^{-1}\Omega P'^{-1} = I$. Quindi, ponendo $T = P^{-1}$ possiamo applicare i minimi quadrati al modello trasformato e ottenere uno stimatore che è BLUE anche in presenza di errori non-sferici. Dato che $\Omega^{-1} = T'T$, lo **stimatore GLS** può essere scritto come segue:

$$\hat{\beta}_{\mathbf{GLS}} = (X'T'TX)^{-1}X'T'T\mathbf{y} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathbf{y}. \tag{7}$$

Dato che $\hat{\beta}_{\mathbf{GLS}} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon$, abbiamo che $E[\hat{\beta}_{\mathbf{GLS}}] = \beta$ e che

$$\begin{aligned} Var(\hat{\beta}_{\mathbf{GLS}}) &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] = \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}. \end{aligned}$$

Assumendo che $\epsilon \sim N(\mathbf{0}, \sigma^2\Omega)$ possiamo applicare le consuete procedure di test delle ipotesi in campioni finiti. Se non vogliamo introdurre l'assunzione di normalità, possiamo applicare i risultati asintotici già visti per gli OLS e applicare quelle stesse procedure di test per ottenere risultati efficienti in grandi campioni.

In particolare, possiamo usare il test F per testare J restrizioni lineari allo stesso tempo, $H_0 : R\beta = \mathbf{q}$. La statistica F nel caso GLS diventa:

$$\frac{(R\hat{\beta} - \mathbf{q})'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\hat{\beta} - \mathbf{q})/J}{s^2} \sim F_{J,N-K}, \tag{8}$$

dove $s^2 = (\mathbf{e}'\mathbf{e})/(N-K) = (\mathbf{y}'\Omega^{-1}\mathbf{y} - \mathbf{y}'\Omega^{-1}X\hat{\beta})/(N-K)$, dato che il vettore di residui $\mathbf{e}$ deriva dal modello GLS trasformato.

Si noti, infine, che la matrice Ω non è in genere osservata. Si pone quindi il problema di stimarla a seconda del problema di non-sfericità degli errori che si assume rilevante[1]. Se nell'equazione (7) mettiamo la stima $\hat{\Omega}$ al posto di Ω, otteniamo lo **stimatore FGLS** (*Feasible Generalized Least Squares*), così chiamato per enfatizzare il fatto che l'implementazione della procedura GLS richiede la stima di Ω. Lo stimatore FGLS è asintoticamente equivalente allo stimatore GLS, ma non è BLUE in campioni finiti.

6 Autocorrelazione

L'eteroschedasticità, come visto, è uno dei casi in cui viene violata l'assunzione di sfericità degli errori ($E(\epsilon\epsilon') = \sigma^2 I$). In quel caso, la violazione di tale assunzione dipende dal fatto che gli elementi sulla diagonale principale della matrice di varianza e covarianza degli errori non sono uguali tra loro. Un'altra possibile violazione nasce quando gli elementi della diagonale principale sono sì uguali (omoschedasticità), ma gli elementi al di fuori della diagonale principale sono diversi da zero, di modo che: $Cov(\epsilon_i, \epsilon_j) \neq 0$ per $i \neq j$. In questo caso, si parla di **autocorrelazione** tra gli errori del modello.

[1] Il modello con dati raggruppati discusso per il problema di eteroschedasticità è un raro caso in cui la matrice Ω è conosciuta.

Il problema di autocorrelazione è più facile da comprendere in un contesto di **serie storiche**, dove osservazioni diverse si riferiscono a periodi temporali diversi, con $t = 1, \ldots, T$. Infatti, anche se è possibile pensare a esempi di dati *cross-section* in cui osservazioni contigue sono spazialmente correlate tra loro, è nelle applicazioni con serie storiche che il problema di autocorrelazione ha una probabilità di emergere vicina ad uno. Visto che le realizzazioni passate possono influenzare quelle future (specialmente se vicine), è altamente probabile che gli shock che influenzano la variabile dipendente siano persistenti nel tempo, ovvero che $Cov(\epsilon_t, \epsilon_{t-1}) > 0$. In presenza di serie storiche, non tutte le assunzioni che abbiamo fatto per il modello di regressione lineare con dati *cross-section* ($i = 1, \ldots, N$) hanno senso. In particolare, l'assunzione di campionamento casuale (A3) è difficile da sostenere, almeno che non si creda che la natura agisca come una sorta di statistico benevolo (fornendoci in ogni periodo un'osservazione estratta casualmente da una popolazione di riferimento)! Senza A3, l'assunzione A4 non è più sufficiente per garantire la condizione di esogeneità forte e l'assunzione A5 non è più sufficiente per garantire la sfericità degli errori (si riveda l'analisi condotta nell'argomento 1 di questi appunti). Diventa quindi necessario riadattare le assunzioni del modello di regressione lineare per il caso con serie storiche.

- A1. La relazione tra la $\mathbf{y}$ e le X è **lineare**.
- A2. Possiamo escludere l'esistenza di **multicollinearità perfetta**, cioè: $rango(X) = K < T$.
- A3'. C'è assenza di **autocorrelazione** negli errori: $Cov(\epsilon_t, \epsilon_s|X) = 0$, per ogni $t \neq s$.
- A4'. Possiamo assumere l'esistenza di **esogeneità forte** delle X: $E(\epsilon|X) = \mathbf{0}$.
- A5. C'è assenza di **eteroschedasticità** negli errori: $Var(\epsilon_t|X) = \sigma^2$, per ogni t.
- A6. La distribuzione delle ϵ è una **normale**.

Appare chiaro che queste assunzioni sono più restrittive rispetto a quelle per dati *cross-section*. Mimando l'analisi condotta allora, è facile dimostrare che: i) sotto le assunzioni A1, A2 e A4', gli stimatori OLS sono corretti e consistenti[2]; ii) sotto le assunzioni A1, A2, A3', A4' e A5, gli stimatori OLS hanno varianza minima nella classe degli stimatori lineari e corretti (Gauss-Markov); iii) aggiungendo l'assunzione di normalità A6, possiamo applicare le procedure inferenziali standard anche in campioni finiti.

6.1 Processi stocastici autocorrelati

La presenza di autocorrelazione (al pari di quella di eteroschedasticità), quindi, non pone un problema di correttezza o consistenza, ma solo di efficienza. Per affrontare questo problema, possiamo applicare la procedura GLS, a patto di conoscere la struttura dell'autocorrelazione (cioè a patto di conoscere la matrice Ω o essere in grado di stimarla). Come nel caso di eteroschedasticità, infatti, non possiamo identificare tutti i termini della matrice di varianza e covarianza degli errori con sole N osservazioni (T nel caso di serie storiche). E aumentare il campione in questo caso non serve: all'aumentare di T aumentano anche i parametri da identificare, dato che la matrice di varianza e covarianza degli errori ha dimensione $[T \times T]$.

[2] In verità, assumere la presenza di esogeneità forte è altamente implausibile in molte applicazioni, dato che equivale a escludere la presenza di effetti ritardati delle variabili esplicative (a meno che non siano inclusi nel modello) o di futuri *feedback* della variabile dipendente su quelle esplicative. Inoltre, l'assunzione di esogeneità forte è violata per definizione in presenza di autocorrelazione con una variabile dipendente ritardata tra i regressori. Tuttavia, è possibile dimostrare che gli stimatori OLS sono ancora consistenti in presenza di un'assunzione più debole, quella di **esogeneità contemporanea** ($E(\epsilon_t|\mathbf{x_t}) = 0$), a patto che la y e le X siano stazionarie. Tali aspetti non verranno approfonditi in questo corso.

Dobbiamo per forza imporre una qualche struttura al processo di autocorrelazione. A tale scopo, in quanto segue, definiremo alcuni processi stocastici autocorrelati, la cui struttura potrebbe catturare (più o meno bene) quella degli errori del nostro modello di regressione.

Il processo AR(1)

Si assuma che:

$$\epsilon_t = \phi\epsilon_{t-1} + \eta_t,$$

dove $\eta_t \sim WN(0, \sigma^2)$ per ogni t. Questo processo è chiamato autoregressivo di ordine uno, o AR(1). Assumendo che il processo sia debolmente stazionario (cioè, $E(\epsilon_t) = E(\epsilon_{t-k})$, $Var(\epsilon_t) = Var(\epsilon_{t-k})$ e $Cov(\epsilon_t, \epsilon_{t-k}) = f(k)$ per ogni $k > 1$), è facile far vedere che:

$$E(\epsilon_t) = 0,$$

$$Var(\epsilon_t) = \frac{\sigma^2}{1-\phi^2},$$

che ha senso se e solo se $|\phi| < 1$. Quest'ultima è la condizione di stazionarietà del processo AR(1). Sempre sotto la condizione di stazionarietà, è facile far vedere che:

$$\gamma_k = Cov(\epsilon_t, \epsilon_{t-k}) = \phi^k \frac{\sigma^2}{1-\phi^2}$$

Di conseguenza, il correlogramma del processo AR(1) si configura come:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k.$$

Sotto la condizione di stazionarietà, è possibile riscrivere il processo AR(1) come un processo a media mobile di ordine infinito, MA(∞):

$$\epsilon_t = \sum_{j=0}^{\infty} \phi^j \eta_{t-j}.$$

Il processo MA(1)

Si assuma che:

$$\epsilon_t = \eta_t + \theta\eta_{t-1},$$

dove $\eta_t \sim WN(0, \sigma^2)$ per ogni t. Questo processo è chiamato a media mobile di ordine uno, o MA(1). In questo caso, è facile far vedere che:

$$E(\epsilon_t) = 0,$$

$$Var(\epsilon_t) = (1+\theta^2)\sigma^2,$$

$$\gamma_1 = \theta\sigma^2,$$

$$\gamma_k = 0 \quad \forall k \geq 2.$$

Si noti che i processi AR(1) e MA(1) hanno strutture di autocorrelazione molto diverse tra loro (come si può vedere dai rispettivi correlogrammi). Nel caso AR(1), $\rho_k \to 0$ se $k \to \infty$, ma la correlazione è sempre diversa da zero per ogni k. Invece, nel caso MA(1), solo ρ_1 è diverso da zero ($= \theta/(1+\theta^2)$), mentre i coefficienti di correlazione per $k \geq 2$ sono tutti uguali a zero.

Sotto la condizione di invertibilità ($|\theta| < 1$), il processo MA(1) può essere riscritto come AR(∞):

$$\epsilon_t = \eta_t + \sum_{j=1}^{\infty} (-1)^{1+j} \theta^j \epsilon_{t-j}.$$

I processi ARMA(p,q)

I processi visti sopra sono casi particolari di processi stocastici più generali. Un processo autoregressivo di ordine p, AR(p), è espresso come:

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_p \epsilon_{t-p} + \eta_t.$$

Un processo a media mobile di ordine q, MA(q), è espresso come:

$$\epsilon_t = \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2} + ... + \theta_q \eta_{t-q}.$$

In maniera simile, un processo ARMA(p,q) è espresso come:

$$\epsilon_t = \phi_1 \epsilon_{t-1} + ... + \phi_p \epsilon_{t-p} + \eta_t + \theta_1 \eta_{t-1} + ... + \theta_q \eta_{t-q}.$$

In questo corso, non approfondiremo la struttura di autocorrelazione di questi processi.

6.2 Stimatori GLS in presenza di autocorrelazione

Assumiamo, adesso, che nel modello di regressione multivariato gli errori seguano un processo AR(1). Per i risultati visti sopra, è facile calcolare la matrice di varianza e covarianza degli errori autocorrelati:

$$Var[\hat{\beta}] = E[\epsilon\epsilon'] = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \dots & \phi^{T-1} \\ \phi & 1 & \dots & \phi^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \dots & 1 \end{pmatrix} = \sigma_\epsilon^2 \Omega.$$

Se conosciamo ϕ, possiamo calcolare la matrice Ω e applicare gli stimatori GLS: $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathbf{y}$. Se usiamo una stima del coefficiente $\hat{\phi}$, applichiamo gli stimatori FGLS: $\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\mathbf{y}$.

Nel caso del processo AR(1), in ogni modo, esiste un modo alternativo per applicare la procedura GLS o FGLS. Si consideri il modello trasformato:

$$y_t - \phi y_{t-1} = (\mathbf{x_t} - \phi \mathbf{x_{t-1}})'\beta + \epsilon_t - \phi \epsilon_{t-1} = (\mathbf{x_t} - \phi \mathbf{x_{t-1}})'\beta + \eta_t,$$

con $t = 2, \dots, T$. Dato che gli errori del modello trasformato soddisfano le assunzioni di Gauss-Markov, gli stimatori dei minimi quadrati sono BLUE. Non del tutto in verità: vista la perdita d'informazione dovuta al fatto che non possiamo usare $t = 1$, questa trasformazione, detta di **Cochrane-Orcutt**, è solo asintoticamente equivalente allo stimatore GLS. Per recuperare le proprietà GLS anche in campioni finiti, si deve aggiungere al modello di cui sopra per $t = 2, \dots, T$ questa trasformazione per la prima osservazione:

$$\sqrt{1-\phi^2} y_1 = \sqrt{1-\phi^2} \mathbf{x_1}'\beta + \sqrt{1-\phi^2} \eta_1.$$

In questo caso, si parla di trasformazione di **Prais-Winsten**. Entrambe le trasformazioni, tuttavia, si basano sul presupposto di conoscere ϕ. Se così non è, dobbiamo stimarlo. Visto che l'equazione $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$ soddisfa le condizioni di Gauss-Markov, anche se non osserviamo gli errori ϵ_t, un approccio intuitivamente sensato è quello di stimare $\hat{\phi}$ dal modello: $e_t = \phi e_{t-1} + \tilde{\eta}_t$, dove e_t sono i residui OLS dal modello di regressione originario. Da cui:

$$\hat{\phi} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e_{t-1}^2}.$$

Ovviamente, con $\hat{\phi}$ al posto di ϕ, lo stimatore FGLS non è BLUE, ma asintoticamente equivalente allo stimatore GLS. Come regola pratica, se le stime OLS e FGLS sono simili, meglio affidarsi alle stime FGLS per ragioni di efficienza. Altrimenti, meglio rifugiarsi nelle stime OLS per paura di distorsioni.

La trasformazione di Cochrane-Orcutt suggerisce anche un modo semplice per applicare lo stimatore GLS nel caso di errori autocorrelati con un processo AR(p) di ordine $p > 1$. Per esempio, con errori AR(2), basta applicare la trasformazione

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = (\mathbf{x_t} - \phi_1 \mathbf{x_{t-1}} - \phi_2 \mathbf{x_{t-2}})'\beta + \eta_t$$

per ottenere un modello con errori sferici e applicare gli stimatori dei minimi quadrati. In aggiunta, i risultati ricavati sopra sul processo MA(1) mostrano come ricavare la struttura di Ω in presenza di errrori autocorrelati a media mobile. Un ultimo commento riguarda il fatto che a volte la presenza di autocorrelazione negli errori può essere provocata da una cattiva specificazione funzionale (come il mancato uso di variabili logaritmiche) o dinamica (come la mancata inclusione di variabili ritardate tra i regressori): in tal caso, piuttosto che usare lo stimatore GLS, si deve riflettere meglio sulla natura del problema e cambiare specificazione al modello.

6.3 Test di autocorrelazione

Adesso che abbiamo visto i problemi creati dalla presenza di autocorrelazione agli stimatori OLS e come usare la trasformazione GLS per risolverli, non ci resta che chiederci come individuare l'autocorrelazione degli errori del modello. Infatti, è pericoloso usare stimatori FGLS se non ce n'è bisogno (cioè, in assenza di autocorrelazione), rinunciando alle proprietà OLS in campioni finiti. Vediamo, allora, alcuni test per l'ipotesi nulla di assenza di autocorrelazione: nel caso AR(1), $H_0 : \phi = 0$.

Test di Durbin-Watson. La statistica test è data da

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=2}^{T} e_t^2}.$$

È facile far vedere che: $DW \approx 2(1 - \hat{\phi})$. Quindi,

- se $DW \approx 2$, $\hat{\phi} \approx 0$ (assenza di autocorrelazione),
- se $DW \approx 0$, $\hat{\phi} \approx 1$ (autocorrelazione positiva),
- se $DW \approx 4$, $\hat{\phi} \approx -1$ (autocorrelazione negativa).

Sotto tutte le assunzione del modello di regressione da A1 ad A6, esiste la distribuzione esatta di questo test (che però dipende dalle X). Per via di questa dipendenza dalle X, esistono soltanto dei limiti inferiori o superiori ai valori critici del test. In altre parole, per ogni livello di significatività, non si conosce il valore critico d_C, ma solo i limiti d_L e d_U: $Pr(DW < d_L) \leq Pr(DW < d_C) \leq Pr(DW < d_U)$. Di conseguenza, assumendo di volere testare $H_0 : \phi = 0$ contro l'ipotesi alternativa $H_1 : \phi > 0$ (autocorrelazione positiva),

- se $DW_{oss} \leq d_L$, rifiutiamo H_0;
- se $DW_{oss} \geq d_U$, non rifiutiamo H_0;
- se $d_L < DW_{oss} < d_U$, non possiamo concludere niente.

Uno dei limiti di questo test, quindi, è che esiste una regione di indecisione.

Test t per errori AR(1). Un altro modo per individuare la presenza di errori AR(1) è quello di usare i residui del modello di regressione originario per stimare $e_t = \phi e_{t-1} + \eta_t$ e testare $H_0 : \phi = 0$. È possibile includere una costante o anche i regressori X (se non vogliamo fare l'assunzione di esogeneità forte) nella specificazione di questa regressione ausilaria sui residui. Il test si basa su proprietà asintotiche (visto che stiamo usando i residui al posto dei veri errori ϵ_t, che ovviamente non osserviamo).

Test F per errori AR(p). Il test precedente suggerisce un modo per individuare la presenza di errori AR(p) di ordine $p > 1$. Semplicemente, si può stimare il modello $e_t = \phi_1 e_{t-1} + \cdots + \phi_p e_{t-p} + \eta_t$ e implementare un test F per testare l'ipotesi nulla $H_0 : \phi_1 = \cdots = \phi_p = 0$. Se rifiutiamo H_0, abbiamo problemi di autocorrelazione provenienti da errori autoregressivi.