

Guida all'uso di `Mc.xls`

Riccardo 'Jack' Lucchetti

23 aprile 2002

1 Cos'è `Mc.xls`?

`Mc.xls` è una cartella di lavoro Excel che uso nella didattica del corso di Econometria. Il suo contenuto è, in sostanza, una dimostrazione di come si calcola lo stimatore OLS. In più, si può effettuare un semplice esperimento di Monte Carlo in cui si mostrano le caratteristiche distribuzionali di $\hat{\beta}$ sotto le ipotesi classiche:

$$\left\{ \begin{array}{l} y = X\beta + u \\ u \sim N(0, \sigma^2 I) \\ X \text{ fissa in campioni ripetuti} \\ |X'X| \neq 0 \end{array} \right.$$

Poiché ottenere questi risultati è un lavoro che alla lavagna non richiede più di tanto, ci si potrebbe chiedere quale sia il senso dell'operazione. I motivi per cui questo attrezzino a me torna utile sono diversi. Molti dei miei studenti non sono avvezzi alle delizie dell'algebra matriciale. Un'espressione come

$$\hat{y} = P_X y = X\hat{\beta},$$

che per noi vecchie spugne è quasi noiosa, è in grado di gettare alcuni dei miei studenti in uno stato di superstizioso terrore, specialmente all'inizio del corso. Per questi studenti, alzare il cofano a macchina accesa e far vedere gli ingranaggi che girano è spesso una boccata di ossigeno. Di qui la scelta dello strumento: Excel non è certo lo strumento più adatto per fare dell'econometria, ma quasi tutti gli studenti hanno con Excel una certa familiarità; su un foglio elettronico, poi, si 'vedono' le matrici y , X , $X'X$ e così via, e molto si demistifica.

Un'altra virtù della visualizzazione esplicita delle matrici sta nel fatto che capire il concetto di 'variabili esplicative fisse in campioni ripetuti' diventa un gioco da bambini: ad ogni ricalcolo del foglio, si vede cosa cambia e cosa no. Quando si spiega agli studenti che, ogni volta che si preme **F9**¹, stiamo

¹In Excel, ogni volta che si preme **F9**, tutto il foglio viene ricalcolato. In particolare, se ci sono numeri pseudocasuali, questi vengono rigenerati.

estraendo un ω dal pentolone Ω , ossia stiamo osservando uno stato diverso del mondo, di solito le facce si illuminano.

Una conseguenza importante è che non si fa più molta fatica a far comprendere che differenza c'è tra uno stimatore e una stima: cosa banale una volta che uno l'ha capita, ma non proprio immediata per quegli studenti che, come alcuni dei miei, arrivano a frequentare Econometria con basi di statistica inferenziale, diciamo, traballanti (e uso un eufemismo).

Siccome i valori veri di β e σ^2 sono visibili dall'utente (e modificabili), lo studente capisce subito che i vari valori di $\hat{\beta}$ che si ottengono ad ogni ricalcolo non sono mai uguali a β , ma, guarda caso, gli oscillano intorno con una certa regolarità.

L'ingrediente finale è una macro che esegue un certo numero di esperimenti (ricalcoli del foglio), salvandosi i risultati in un foglio separato. Questo ha sostanzialmente due scopi: in primo luogo, mostrare le proprietà distribuzionali degli stimatori diventa banale, in quanto esse possono essere facilmente esemplificate con delle semplici statistiche descrittive. In secondo luogo si possono introdurre agli studenti le idee di base sugli esperimenti di Monte Carlo, con annesse problematiche di generazione di numeri pseudocasuali e tutto il resto, ma questo è opzionale, e suppongo rappresenti un'attrattiva solo per i patiti del genere (come me).

2 Organizzazione dei fogli

La cartella è organizzata su tre fogli diversi, che si chiamano “Livello visibile”, “Livello invisibile” e “Monte Carlo”. In più, c'è un foglio grafico (non indispensabile), che si chiama, appunto “Grafico visibile” e può essere visto come un'appendice al foglio “Livello visibile”. Li descrivo brevemente nel seguito. L'idea guida, però, è semplice: spesso è difficile far tenere a mente costantemente quali siano le grandezze osservabili da quelle di cui, semplicemente, postuliamo l'esistenza. Il caso più eclatante è dato dalla classica confusione, che gli studenti tendono a fare, tra disturbi e residui. Ancora più sottile è il caso della matrice varianze-covarianze di $\hat{\beta}$: mi capita spesso di trovare studenti che fanno fatica a capire che $s^2(X'X)^{-1}$ è la controparte osservabile di un'entità sfuggente, che può essere $\sigma^2(X'X)^{-1}$ nel caso di regressori deterministici, o addirittura $\frac{1}{T}\sigma^2 E(x_t x_t')^{-1}$ nel caso di regressori stocastici.

Di conseguenza, le grandezze presenti nella cartella `Mc.xls` vengono suddivise su fogli diversi secondo il criterio dell'osservabilità: sul “Livello visibile” ci sono tutte le grandezze che in una situazione reale (e non di simulazione) sono effettivamente visibili; nel foglio “Livello invisibile”, invece compare tutto ciò che noi non possiamo vedere, ma che postuliamo esista, come il vettore β e così via.

2.1 Il foglio Livello visibile

Come dicevo, questo foglio contiene tutte le quantità che sono visibili nel mondo reale, cioè i dati (y e X) e le statistiche da essi calcolate. In questo foglio non c'è nulla che attenga al “vero” processo di generazione dei dati, poiché in una situazione normale lo sperimentatore deve appunto inferire dai dati le cose che non sa.

I conti vengono fatti usando un campione di 100 osservazioni e una matrice X di tre colonne, inclusa l'intercetta. Attenzione! Le formule che calcolano $\hat{\beta}$ e tutto il resto sono tarate su questa scelta, cosicché se volete modificare l'ampiezza campionaria e/o il numero di regressori dovete correggere appropriatamente le formule relative, a partire da quelle contenute nelle colonne G,H e I.

I valori della matrice X non cambiano fra una replicazione ed un'altra. Questa non è una scelta politica, ma deriva semplicemente dal fatto che io introduco nel corso i regressori stocastici dopo l'esercitazione in cui uso `Mc.xls`. Se uno vuole introdurre in X una o più colonne stocastiche, non deve far altro che sostituire le espressioni che ci sono con altre basate sulla funzione Excel di generazione di numeri pseudocasuali, che è `=Casuale()`.

Le statistiche che vengono calcolate qui sono semplicemente:

Statistica	ossia	nelle celle
$\hat{\beta}$	$(X'X)^{-1} X'y$	G14:G16
i valori fittati \hat{y}	$X\hat{\beta}$	M2:M101
i residui e	$y - \hat{y}$	N2:N101
TSS	$y'y$	H22
ESS	$\hat{y}'\hat{y}$	H23
RSS	$e'e$	H24
s^2	$\frac{\text{RSS}}{T-k}$	H18
R^2	$1 - \frac{\text{RSS}}{\sum_{i=1}^T (y_i - \bar{y})^2}$	H19
$\widehat{V(\hat{\beta})}$	$s^2 (X'X)^{-1}$	G27:G29

e le statistiche t per ogni coefficiente. Ovviamente, se si vogliono aggiungere cose (o semplicemente organizzare l'output in modo più ordinato), si è padronissimi di farlo modificando il foglio. Si tenga presente, tuttavia, che il vettore $\hat{\beta}$ e lo scalare s^2 servono poi per il Monte Carlo, e possono essere spostati (le locazioni sono memorizzate con dei nomi), ma non eliminati e comunque non è consigliabile giocherellarci più di tanto.

La cosa interessante da fare in questo foglio è premere un po' di volte il tasto **F9** e notare come le quantità non stocastiche restano ferme, mentre quelle che variano ad ogni campionamento (la y e tutte le statistiche su essa basate) si muovono.

Assieme a questo foglio, c'è un grafico nel foglio “Grafico visibile”, che è spudoratamente ispirato ad Eviews, nel quale sono presentati graficamente

i tre vettori y , \hat{y} e \hat{e} . Ovviamente, anche qui i valori cambiano ogni volta che si preme F9.

2.2 Il foglio Livello invisibile

Qui sono ospitate le grandezze “vere”, cioè β , σ^2 ed il vettore dei disturbi u . In più, vengono calcolate grandezze che attengono alla dispersione di $\hat{\beta}$, come ad esempio la sua matrice di covarianze (questo mi serve per far vedere che, sotto le ipotesi classiche, la stima della varianza di $\hat{\beta}$ è corretta).

I disturbi u sono generati come numeri pseudocasuali normali utilizzando il metodo dell’inversione, visto che in Excel è disponibile la funzione `Inv.Norm()`. Chi volesse approfondire questo aspetto, può facilmente cambiare qui l’algoritmo di generazione, usando ad esempio il metodo di Box-Müller, o anche numeri casuali non normali. La seconda scelta può essere utile per introdurre le conseguenze della non normalità dei disturbi (un esempio che io trovo particolarmente gustoso è generare le u_i con la funzione `=SE(CASUALE()>0.99;99;-1); provateci`).

2.3 Il foglio Monte Carlo

Come dicevo, premendo un po’ di volte il tasto F9 mentre si guarda il livello visibile ci si rende conto “visivamente” del concetto di variabilità campionaria e di distribuzione di una statistica (stimatore o test che sia). Ad esempio, fissando a 0 uno degli elementi di β contenuti nel livello invisibile, ci sono buone probabilità (circa il 64%) che premendo 20 volte F9 nel livello visibile il test t associato a quel coefficiente risulti almeno una volta maggiore di 2 in valore assoluto.

Sarebbe interessante mettere da parte ogni volta i risultati per esaminarli con calma. Questo viene fatto nel foglio “Monte Carlo”. Questo foglio, infatti, accoglie i risultati di un semplice esperimento di Monte Carlo basato su un numero di replicazioni che lo studente può impostare. Il tutto viene fatto usando una semplice macro scritta in Visual Basic.

Premendo **Ctrl-M** si avvia l’esperimento: apparirà una finestra di dialogo in cui l’utente deve inserire il numero di replicazioni che vuole effettuare (200 al massimo²), dopodiché può andare a controllare sul foglio Monte Carlo cos’è successo. Vi consiglio di sperimentare per un po’ con una replicazione alla volta, poi lanciarvi.

Eseguendo la macro più volte, i risultati, che occupano le prime 4 colonne, vengono aggiunti ai precedenti nelle righe sottostanti, quindi il limite di 200 è più apparente che reale, ed ha il solo scopo di non far inchiodare una macchina nel malaugurato caso qualcuno scegliesse di fare 10000 replicazioni

²Volendo, questo limite può essere modificato cambiando il valore della variabile `MaxIter` nella subroutine `CommandButton1_Click()`, che fa parte della macro.

(non è che Excel brilli per prestazioni). Si può ripartire con un nuovo esperimento cancellando tutte le celle dalla riga 2 in giù nelle prime 4 colonne. **Avvertenza:** un po' per pigrizia, un po' perché non è che io sia proprio un mago del Visual Basic, non cancellate **per alcun motivo** la prima riga (quella con tutti zeri), altrimenti la macro dà di matto.

Per ognuno dei 4 parametri stimati vengono visualizzate semplici statistiche descrittive; anche in questo caso, se uno volesse aggiungerne, non fa molta fatica. Una cosa abbastanza sfiziosa (if I may say so myself) è un istogramma che dà la distribuzione empirica dello stimatore. Si controlla quale delle 4 distribuzioni viene mostrata nel grafico inserendo un numero da 1 a 4 nella cella H9.