L'Analisi Fattoriale Dinamica in STATA

Alessandro Federici

Dipartimento di Scienze Economiche

Università di Roma *La Sapienza*

alessandro.federici@uniroma1.it

Abstract

Lo scopo del lavoro è quello di sviluppare una procedura in grado di implementare l'Analisi Fattoriale Dinamica (AFD) in STATA. L'AFD è una tecnica di analisi statistica multivariata[1] dove sono considerati dei vettori del tipo "unità x variabili x tempi":

$$\boldsymbol{X(I,J,T)} = \{x_{ijt}\},\ i=1\ldots I,\ j=1\ldots J,\ t=1\ldots T,$$

dove i è l'unità, j la variabile e t il tempo.

Tale metodologia consente di combinare, dal punto di vista descrittivo (non probabilistico), l'analisi cross-section attraverso l'Analisi in Componenti Principali (ACP) e la dimensione temporale dei dati attraverso il modello di regressione lineare.

Il lavoro è organizzato come segue: nella prima sezione sarà descritta la metodologia dal punto di vista teorico, mostrando gli elementi statistici ed econometrici che la procedura andrà successivamente ad implementare in STATA.

La metodologia proposta è stata introdotta da Coppi e Zannella (1978) e successivamente rivista da Coppi et al. (1986) e Corazziari (1997): nel presente lavoro sarà mantenuto l'approccio originario.

L'obiettivo della procedura è quello di scomporre la matrice di varianze e covarianze $\boldsymbol{S}$ associata a $\boldsymbol{X(IT,J)}$, matrice "unità x variabili", dove il ruolo delle unità è giocato dalle coppie "unità-tempo". La matrice $\boldsymbol{S}$ relativa alle $\boldsymbol{JxT}$ osservazioni sulle $\boldsymbol{I}$ unità, può essere scomposta nella somma di tre matrici di varianze e covarianze riferite alle variabili:

$$\boldsymbol{S} = {}^*\boldsymbol{S}_I + {}^*\boldsymbol{S}_T + \boldsymbol{S}_{IT}, \qquad (1)$$

dove:

[1] Tecnica in cui tre o più indici sono analizzati simultaneamente.

- $^*\boldsymbol{S}_I$ = matrice di struttura statica delle unità = matrice di varianze e covarianze dei centri medi delle unità rispetto ai tempi; riflette la variabilità data dalla struttura relazionale delle unità (misurate tramite le J variabili) a prescindere dal tempo.
- $^*\boldsymbol{S}_T$ = matrice di dinamica media del sistema = matrice di varianze e covarianze dei centri medi dei tempi; riflette la variabilità, dovuta al tempo, del centro medio collettivo delle unità nel suo complesso, a prescindere dalle variazioni connesse alla dinamica delle singole unità.
- $\boldsymbol{S}_{IT}$ = matrice della dinamica differenziale delle singole unità = matrice di varianze e covarianze delle interazioni tra unità e tempi; riflette la variabilità data dalla differenza tra dinamica del centro medio del collettivo (dinamica media), e la dinamica delle singole unità.

Sulla base della scomposizione fondamentale della variabilità totale (1), il dato elementare x_{ijt} può considerato visto come la somma di quattro componenti differenti:

$$x_{ijt} = \bar{x}_{\bullet j \bullet} + \left(\bar{x}_{ij\bullet} - \bar{x}_{\bullet j \bullet}\right) + \left(\bar{x}_{\bullet jt} - \bar{x}_{\bullet j \bullet}\right) + \left(x_{ijt} - \bar{x}_{ij\bullet} - \bar{x}_{\bullet jt} + \bar{x}_{\bullet j \bullet}\right), \qquad (2)$$

dove:

$\bar{x}_{\bullet j \bullet}$ = media generale della singola variabile;

$\left(\bar{x}_{ij\bullet} - \bar{x}_{\bullet j \bullet}\right)$ = effetto dovuto alla struttura statica delle unità;

$\left(\bar{x}_{\bullet jt} - \bar{x}_{\bullet j \bullet}\right)$ = effetto dovuto alla dinamica media;

$\left(x_{ijt} - \bar{x}_{ij\bullet} - \bar{x}_{\bullet jt} + \bar{x}_{\bullet j \bullet}\right)$ = effetto dovuto alla dinamica differenziale, cioè all'interazione tra unità e tempi.

La (2) rappresenta un modello di analisi della varianza a due fattori: il modello che sarà implementato nella sezione empirica del lavoro, il cosiddetto Modello 1, considera le differenti componenti della (2) ed i relativi elementi della variabilità totale (1) in termini di ACP e modello di regressione lineare.

Infatti, il Modello 1 si basa sulla seguente scomposizione della variabilità totale in due componenti:

$$\boldsymbol{S} = (^*\boldsymbol{S}_I + \boldsymbol{S}_{IT}) + {}^*\boldsymbol{S}_T = \boldsymbol{S}_T + {}^*\boldsymbol{S}_T, \qquad (3)$$

dove $\boldsymbol{S}_T$ è la matrice di dispersione media all'interno dei tempi (variabilità *within*), analizzata tramite l'ACP, mentre $^*\boldsymbol{S}_T$ rappresenta la variabilità tra tempi (variabilità *between*), analizzata attraverso il modello di regressione lineare:

$$\bar{x}_{\bullet jt} = a_j + b_j t + e_{jt}, \; j = 1\ldots J;\; t = 1\ldots T, \qquad (4)$$

dove i residui soddisfano la seguente condizione:

$$\operatorname{cov}\left(e_{jt}, e_{j't'}\right) = \begin{cases} w_j & j = j';\; t = t' \\ 0 & \text{altrimenti} \end{cases}. \qquad (5)$$

La condizione (5) deve essere tenuta in considerazione visto che la relazione esistente tra le j variabili deve essere spiegata soltanto dalla componente fattoriale, cioè dalla ACP relativa alla matrice $\boldsymbol{S}_T$: pertanto la dinamica media del sistema è distinta dalla dinamica media delle singole variabili.

La relazione (3) rappresenta una contrazione della scomposizione fondamentale (1), dovuta all'aggregazione di due fonti di variabilità: la struttura statica e la dinamica differenziale. Al fine di valutare la capacità esplicativa del modello, si possono andare a considerare degli indicatori in grado di misurare la variabilità spiegata da ognuna delle componenti fondamentali descritte in precedenza:

- $*I_t$: quota di variabilità del sistema di regressione rispetto alla variabilità complessiva;
- $I(t)$: qualità di rappresentazione della struttura fattoriale al tempo t; valuta quanto è ben rappresentata ogni singolo anno considerato;
- $*I_I$: qualità della rappresentazione fattoriale della struttura statica delle unità;
- I_{IT}: qualità della rappresentazione fattoriale della dinamica differenziale delle unità.

Nella seconda sezione del lavoro sarà sviluppata una procedura *ad hoc* al fine di eseguire su STATA l'AFD: sulla base dei comandi per l'ACP ed il modello di regressione lineare già presenti all'interno del software, la sezione svilupperà un nuovo comando per la stima del modello, con una serie di indicatori riguardo l'attendibilità dei risultati ricavati attraverso il modello stimato.

Nella terza sezione sarà presentata un'applicazione empirica di tipo macroeconomico, al fine di stimare un indice di innovazione generale per i paesi dell'OCSE.

Infine, saranno tratte alcune conclusioni, mettendo in evidenza l'ampio raggio di settori dove tale metodologia statistica (non ancora molto conosciuta nella letteratura anglo-sassone) può essere applicata.

Bibliografia essenziale

Coppi, R. (1986). *Analysis of three-way data matrices based on pairwise relation measures*. In *Compstat 1986 – Proceedings in computational statistics*, Physica-Verlag, Wien.

Coppi, R. (1988). *Simultaneous analysis of a set of multiway contingency tables*. In *Data Analysis and Informatics, V* (E. Diday ed.), North Holland, Amsterdam.

Coppi, R.; Di Ciaccio, A. (1994). *Multiway data analysis: software and application*. Special issue of *Computational statistics and data analysis*, vol. 18, North Holland, Amsterdam.

Coppi, R; Bolasco, S. (1989). *Multiway data analysis*. North Holland, Amsterdam.

Coppi, R; Zannella, F. (1978). *L'analisi fattoriale di una serie temporale multipla relative allo stesso insieme di unità statistiche*. In Atti della XXIX Riunione Scientifica della SIS, Bologna.

Corazziari, I. (1997). *Dynamic factor analysis*. In *Atti del convegno dell'IFCS, sezione italiana* (Pescara, 3-4 luglio 1997).

Gifi, F. (1990). *Nonlinear multivariate analysis*. J. Wiley, New York.

Law, H.G.; Snyder, C.W. Jr; Hattie, J.A.; McDonald, R.P. (1984). *Researh methods for multimode data analysis*. Preger, New York.