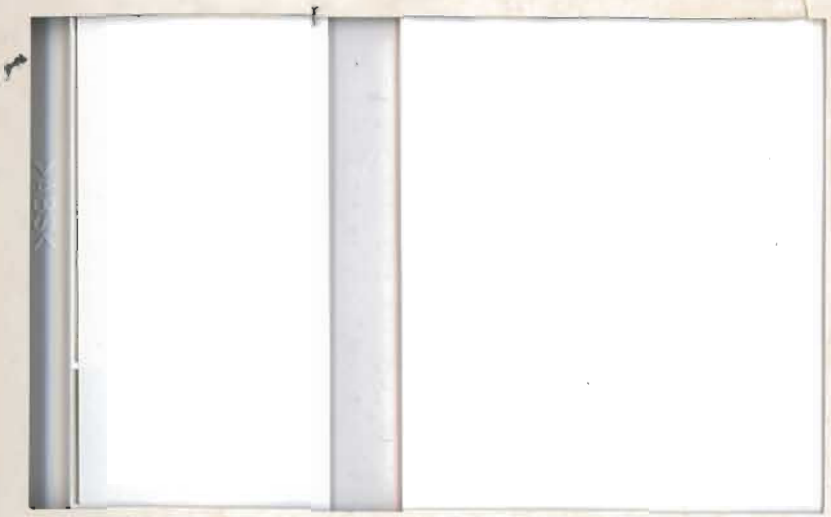

Michele Costa

Analisi fattoriale e criteri di informazione: una simulazione nell'ambito di applicazioni finanziarie

Serie Ricerche n. 3

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna
1992

Finito di stampare nel mese di Agosto 1992
presso le Officine Grafiche Tecnoprint
Via del Legatore 3, Bologna

1 - Introduzione

Il problema della scelta, tra un insieme di modelli alternativi, del modello "vero" non rappresenta solo uno fra i più classici argomenti del dibattito scientifico, ma individua anche uno dei temi che, recentemente, hanno suscitato maggiore interesse, stimolando numerose ricerche teoriche ed empiriche.

Il problema della selezione del "vero" modello fattoriale

$$X = f\Lambda + U$$

(dove $X_{N\times p}$ è la matrice osservabile dei dati, $U_{N\times p}$ è la matrice dei termini di errore normalmente distribuiti, $\Lambda_{k\times p}$ è la matrice dei coefficienti, $f_{N\times k}$ è la matrice, con $k < p$, dei fattori, N è il numero di osservazioni delle serie storiche utilizzate) si riduce essenzialmente alla determinazione del "vero" numero di fattori.

La ricerca del numero di fattori riveste, quindi, un ruolo di primo piano in tutti gli studi teorici relativi ai metodi dell'analisi fattoriale e, soprattutto, nelle verifiche empiriche che utilizzano questi metodi in una vastissima gamma di applicazioni.

In particolare, nell'ambito dei modelli dei mercati finanziari, il numero dei fattori rappresenta l'elemento discriminante tra i due modelli più noti in letteratura: il CAPM (Sharpe, 1964, e Lintner, 1965) e l' APT (Ross, 1976). Molto sinteticamente e rimandando ai riferimenti bibliografici per ulteriori approfondimenti, per il CAPM il rendimento del portafoglio di mercato rappresenta il solo fattore di rilievo nel processo di generazione dei rendimenti. Per il titolo i-esimo vale, allora, la nota relazione

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f)$$

dove R_i è il rendimento del titolo i-esimo, R_f è il rendimento di un titolo privo di rischio, R_m è il rendimento del portafoglio di mercato e $\beta_i = cov(R_i, R_m)/var(R_m)$.

L' APT, invece, assume che i rendimenti di p titoli siano conformi ad un modello lineare a k fattori, con $k < p$. Per il titolo i-esimo vale allora:

$$R_i = E(R_i) + \lambda_{i1} f_1 + \ldots + \lambda_{ik} f_k + e_i \qquad i = 1, \ldots, p$$

dove le f_k sono fattori comuni a media zero, le λ_{ki} rappresentano la sensibilità del rendimento del titolo i-esimo a movimenti nel fattore k-esimo, le e_i sono elementi di stocasticità a media zero, incorrelati fra loro e

indipendenti dai fattori f_k. L'assunzione fondamentale dell' APT, e cioè l'assenza di opportunità di arbitraggio prive di rischio, può essere tradotta nella seguente relazione tra il rendimento atteso e i parametri λ_{ki}

$$E(R_i) = y_0 + \lambda_{1i} y_1 + \ldots + \lambda_{ki} y_k \qquad i = 1, \ldots, p$$

dove y_0 viene interpretato come il rendimento di un titolo privo di rischio, mentre le y_j $(j = 1, \ldots, k)$ sono vettori $N \times 1$ di costanti.

Se la matrice dei dati $X_{N \times p}$ è costituita da p vettori N x 1 di rendimenti azionari R_i $(i = 1, \ldots, p)$, e quindi $X \equiv R_i$, i modelli CAPM e APT possono essere rappresentati come modelli fattoriali.

E' necessario, infine, sottolineare come la ricerca del numero di fattori latenti sottostanti i rendimenti azionari abbia dato luogo a risultati estremamente incerti e variabili.

L'identificazione di una struttura fattoriale stabile viene tradizionalmente effettuata utilizzando il rapporto di verosimiglianza, al quale, nel corso degli anni sono stati affiancati anche altri strumenti, come i criteri di informazione e la cross-validation.

Questo lavoro si propone, attraverso una simulazione, di chiarire il contributo del rapporto di verosimiglianza, di alcuni criteri di informazione e della cross-validation alla determinazione del "vero" numero di fattori del modello.

Uno dei maggiori risultati di questa ricerca è individuare in tutti i metodi considerati una tendenza a sottostimare il valore vero di k, al quale, peraltro, tutte le procedure convergono al crescere di N. Il criterio di informazione di Akaike, inoltre, sembra identificare la struttura fattoriale con maggiore precisione rispetto al rapporto di verosimiglianza e ai criteri di Hannan-Quinn e di Schwarz. Il rapporto di verosimiglianza, infine, risulta molto sensibile a variazioni del numero di titoli considerati mentre il criterio di Schwarz sembra essere piuttosto influenzato dalla lunghezza N delle serie utilizzate.

2 - Alcuni metodi di selezione di un modello fattoriale

2.1 - Rapporto di verosimiglianza

Il rapporto di verosimiglianza rappresenta il metodo tradizionalmente più utilizzato per sottoporre a verifica il numero di fattori del modello. La possibilità di effettuare una tale verifica costituisce, inoltre, una delle principali ragioni del successo, fra le procedure di analisi

fattoriale esplorativa, di quelle di massima verosimiglianza (Lawley e Maxwell 1971; Kim e Mueller, 1983; Anderson, 1984).

Se i fattori sono ortogonali fra loro la matrice di covarianza campionaria Σ dei dati X può essere espressa come $\Sigma = \Lambda'\Lambda + \Psi$ dove Ψ è la matrice diagonale $p \times p$ di varianza e covarianza degli errori U del modello fattoriale: $E(U'U) = \Psi$.

L'ipotesi nulla che viene sottoposta a verifica per determinare il numero di fattori è:

$$H_0: \quad \Sigma_p = \Lambda_k'\Lambda_k + \Psi_k$$

dove Σ_p è la matrice di varianza e covarianza campionaria costruita sulla base di p variabili, mentre Λ_k e Ψ_k sono, rispettivamente, la matrice dei coefficienti e la matrice dei varianza e covarianza degli errori del modello k-fattoriale.

In altri termini l'ipotesi nulla può essere espressa come

$$H_0: \textit{ sono sufficienti } k \textit{ fattori } (o \textit{ meno})$$

contro l'alternativa:

$$H_1: \textit{ sono necessari } p \textit{ fattori.}$$

Dato il logaritmo della verosimiglianza del modello

$$\log L = -\tfrac{1}{2} N p \log(2\pi) - \tfrac{1}{2} N \log |\, \Lambda'\Lambda + \Psi \,| - \tfrac{1}{2} N p$$

la statistica del rapporto di verosimiglianza assume la forma:

$$-2 \log \lambda = N^* (\log |\, \Lambda_k'\Lambda_k + \Psi_k \,| - \log |\, \Sigma_p \,|)$$

dove il numero di osservazioni viene corretto secondo la formula di Bartlett: $N^* = N - (2p + 4k + 11)/6$. Sulla base di condizioni sufficientemente generali, il rapporto di verosimiglianza si distribuisce come un χ^2 con $((p-k)^2 - p - k)/2$ gradi di libertà.

p è il numero di variabili considerate e, quindi, l'ipotesi più generale è data da un modello con un numero di fattori uguale al numero di titoli considerati. Se la differenza tra p e k fattori non è significativa, allora sono sufficienti k fattori per rappresentare le p variabili X. Fra tutti i valori possibili di k (e cioè fra quelli per i quali H_0 non è significativa) si sceglie il valore minimo. Si ottiene, così, una sequenza di test tra loro dipendenti, in quanto i rapporti di verosimiglianza

$$\lambda(k) = \frac{L(k)}{L(p)}$$

sono tutti raffrontati allo stesso termine L(p); ad esempio, per il caso più generale (k = 0), il rapporto

$$\lambda(0) = \frac{L(0)}{L(p)}$$

può essere pensato come il prodotto di

$$\frac{L(0)}{L(1)} \times \frac{L(1)}{L(2)} \times \ldots \times \frac{L(s)}{L(s+1)} \times \ldots \times \frac{L(p-1)}{L(p)}$$

L'utilizzo del rapporto di verosimiglianza per la determinazione del numero di fattori, tuttavia, non è affatto esente da critiche. In primo luogo vengono espresse forti perplessità per l'utilizzo di questa statistica nel caso di elevati ordini di grandezza di N: quando il modello non è quello esatto, è possibile, infatti, che alti valori di N portino all'inclusione di fattori in realtà trascurabili, tanto più che mentre il rapporto di verosimiglianza dipende da N, i gradi di libertà sono influenzati solo da p e da k. Un'ulteriore riserva sulla robustezza del rapporto di verosimiglianza si riscontra nella diversità del numero di fattori risultante dalle diverse ricerche empiriche.

2.2 - *Criteri di informazione*

Una delle ipotesi fondamentali dei criteri di informazione è assumere che il confronto tra il modello teorico e i dati possa essere ricondotto al confronto tra due distribuzioni di probabilità. Mentre un modello può essere rappresentato senza troppe difficoltà attraverso una distribuzione di probabilità, stimare, invece, la vera distribuzione a partire dai dati può causare problemi di non facile soluzione.

Il primo e più generale tentativo di misurare in modo obiettivo la distanza tra la distribuzione empirica e il modello è noto come quantità di informazione di Kullback-Leibler. Supponendo che $g(x)$ sia la funzione di densità di probabilità effettiva e che $f(x)$ sia la funzione di densità di probabilità ipotetica, il valore atteso di $\log(g(x)/f(x))$

$$I(g,f) = \int g(x) \log \frac{g(x)}{f(x)} dx$$

rappresenta la quantità di informazione di Kullback-Leibler della distribuzione effettiva rispetto al modello. In realtà la distribuzione "vera" è generalmente ignota e si dispone solo di dati appartenenti ad essa. Per poter confrontare la distribuzione del modello con la distribuzione effettiva si ricorre allora al logaritmo della verosimiglianza del

modello, la cui stima è sufficiente (Sakamoto *et al.*, 1986) per confrontare diversi modelli attraverso la quantità di informazione di Kullback-Leibler.

Sulla base del logaritmo della verosimiglianza viene costruito anche quello che è forse il più celebre criterio di informazione, l' Akaike Information Criterion (AIC), proposto da Akaike nel 1973 nella forma

$$AIC = -2 \log \max L + 2h$$

dove h è il numero di parametri del modello e, per la scelta del modello, si ricorre a una regola di minimizzazione, assumendo come valore vero di k quello in corrispondenza del quale il valore dell' AIC è minimo.

Nel modello fattoriale la matrice dei coefficienti Λ contiene kp parametri da stimare, mentre la matrice di varianza e covarianza Ψ, diagonale, ne ha p; pertanto, la matrice di varianza e covarianza campionaria Σ contiene $p(k+1)$ parametri da stimare. Tuttavia per assicurare l'identificabilità del modello è necessario imporre l'ulteriore restrizione $\Gamma = \Lambda'\Psi^{-1}\Lambda$, con Γ diagonale, che aggiunge $k(k-1)/2$ vincoli addizionali al modello. Pertanto il numero di parametri liberi di Σ, nell'ambito di un modello fattoriale ortogonale, risulta $p(k+1) - \frac{1}{2}k(k-1)$.

Per il modello fattoriale l'espressione dell' AIC è la seguente

$$AIC(k) = Np \log(2\pi) + N \log |\Lambda\Lambda' + \Psi| + Np + 2(p(k+1) - k(k-1)/2)$$

Il primo membro può essere visto come una misura della bontà dell'adattamento del modello a un insieme di dati, mentre il secondo membro può essere interpretato come una penalità per l'aumento del numero dei parametri, in linea con il principio di parsimonia. L'impiego dell' AIC viene fortemente limitato dalla tendenza di questo criterio ad indicare un elevato numero di fattori. Akaike (1987) attribuisce questa tendenza alle frequenti soluzioni improprie dell'analisi fattoriale di massima verosimiglianza e, introducendo una funzione di distribuzione a priori, propone una soluzione bayesiana che sottoponga a un controllo la funzione di verosimiglianza.

Una modifica all' AIC viene proposta nel 1980 da Smith e Spiegelhalter che, al posto di $2h$, suggeriscono di utilizzare come secondo termine dell' AIC un generico αh:

$$AIC = -2 \log \max L + \alpha h$$

Sempre nell'ambito delle procedure di massima verosimiglianza, ma con l'intento di presentare una procedura alternativa All' AIC e rifacendosi a motivazioni di natura bayesiana, nel 1978 Schwarz propone un nuovo criterio di informazione

$$SCH = -\log \max L + \tfrac{1}{2} h \log N$$

che, rispetto all' AIC, tiene conto dell'ampiezza N delle serie storiche utilizzate e risulta essere meno favorevole all'inclusione di più fattori.

Il criterio di Schwarz è stato utilizzato anche al di fuori del contesto bayesiano essendo indipendente da particolari specificazioni a priori e sembra essere meno incline all'inclusione di fattori di quanto non siano il rapporto di verosimiglianza e il criterio di Akaike.

Entrambi i criteri costituiscono una formalizzazione del celebre rasoio di Occam: "pluralitas non est ponenda sine necessitate", in quanto tendono a scegliere un modello con meno parametri, tuttavia è immediato notare la forte diversità delle indicazioni fornite dai due criteri e le similarità tra il criterio di Akaike e il rapporto di verosimiglianza, mentre il criterio di Schwarz, suggerendo un minor numero di fattori, si discosta dai metodi precedenti.

Nel 1979 Hannan e Quinn propongono un nuovo criterio di informazione, basato, come i precedenti, sulla minimizzazione di $-\log \max L + hC$

$$HQ = -2 \log \max L + 2 h\, c \log \log N$$

Nella proposta iniziale di Hannan e Quinn c è maggiore di uno, tuttavia, è interessante notare come, per $c = 1$, si ha una misura intermedia tra AIC e SCH.

2.3 - Cross-validation

Fra i contributi rivolti alla ricerca di un criterio di determinazione del numero di fattori privo degli inconvenienti riscontrati nel rapporto di verosimiglianza si segnala la proposta di Conway e Reinganum (1988), che indicano nella cross-validation una soluzione alternativa.

L'idea fondamentale della cross-validation può essere sintetizzata in una procedura a due stadi.

Nel primo stadio vengono calcolate, a partire da un dato campione di p variabili X, le stime di massima verosimiglianza dei parametri del modello, e cioè gli elementi di Λ e di Ψ.

Nel secondo stadio le stime ottenute non vengono confrontate con la rispettiva matrice di varianza e covarianza campionaria, Σ, ma con una Σ^* relativa a un diverso campione di p variabili X, assunto come campione di controllo.

In questo modo si cerca di isolare la parte stabile della struttura fattoriale dalle componenti accidentali.

Il rapporto di verosimiglianza

$$-2 \log \lambda = N^*(\log |\Lambda'\Lambda + \Psi| - \log |\Sigma|)$$

viene, quindi, modificato in

$$CV = N^*(\log |\Lambda'\Lambda + \Psi| - \log |\Sigma^*|) + N^*(tr((\Lambda'\Lambda + \Psi)^{-1}\Sigma^*) - p)$$

Il termine $N^*(tr((\Lambda'\Lambda + \Psi)^{-1}\Sigma) - p)$ non è presente nell'espressione del rapporto di verosimiglianza, in quanto, non sostituendo Σ con Σ^*, assume valore nullo.

3 - Una simulazione

Nell'ambito delle applicazioni dei metodi di determinazione del numero di fattori precedentemente illustrati, lo scopo di questo paragrafo è quello di presentare alcuni risultati relativi all'utilizzo di dati simulati, per i quali la struttura fattoriale sottostante è conosciuta esattamente.

Per ottenere le nuove variabili X^* simulate si fa riferimento al modello

$$X^* = f^* \Lambda^* + U^*$$

dove

- f^* è la matrice $N \times k$ dei nuovi fattori, ortogonali e incorrelati tra loro, ottenuti con estrazioni casuali da una distribuzione normale con media nulla e varianza unitaria effettuate con la funzione RANNOR del pacchetto SAS;
- U^* è la matrice $N \times p$ dei termini di errore, anch'essi estratti casualmente da una distribuzione normale con media nulla e varianza unitaria;
- Λ^* è la matrice $k \times p$ dei coefficienti, che vengono ricavati da una analisi fattoriale di massima verosimiglianza effettuata su un campione di 30 rendimenti azionari e ipotizzando la presenza di 5 fattori. Per ridurre al minimo il rischio di estrarre un campione di rendimenti "anomali" sono stati estratti casualmente (da un insieme di 100 titoli azionari quotati alla borsa di Milano dal 1986 al 1989) 10 campioni di 30 titoli ciascuno e, valutate le indicazioni fornite dai diversi criteri di informazione, per scegliere i coefficienti Λ^* è stato individuato un campione in linea con l'andamento generale.

I coefficienti Λ^* così ottenuti sono stati mantenuti fissi in tutto il corso della ricerca, mentre per quanto riguarda i fattori f^* sono stati utilizzati 3 campioni di 5 fattori ciascuno ma di lunghezza diversa: nel primo $N = 200$, nel secondo $N = 1000$ e nel terzo $N = 5000$.

Ricavato così il termine $f^*\Lambda^*$ le p variabili X^* simulate sono state ottenute attraverso successive p estrazioni casuali del vettore U^* dei

termini di errore.

In questo modo la struttura fattoriale risulta conosciuta in anticipo e la variabilità delle X^* è interamente ed esclusivamente attribuibile alle diverse determinazioni del vettore U^*.

I diversi metodi precedentemente illustrati sono stati quindi applicati a campioni di p variabili X^* simulate, analizzando, in particolare, la sensibilità delle diverse misure a variazioni delle numerosità p dei titoli considerati e N delle serie utilizzate.

Sono stati calcolati 3000 vettori X^*, ripartiti in 120 campioni, dei quali 20 con $p = 30$ e $N = 200$, 20 con $p = 30$ e $N = 1000$, 20 con $p = 30$ e $N = 5000$, 20 con $p = 20$ e $N = 200$, 20 con $p = 20$ e $N = 1000$ e 20 con $p = 20$ e $N = 5000$.

Nelle tavole 1 e 2 sono riportati i risultati relativi a $p = 30$ e a N=200, 1000, 5000.

Tav. 1 - Numero di campioni di 30 titoli suddivisi per numero di fattori e metodo di determinazione.

		LR			AIC			HQ1			SCH			CV		
numero fattori	N	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000
0																
1					1			17			20			2		
2		8			2			3				2		7		
3		9			10							16		7		
4		3	6		5				16			2		2	1	
5			14	20	2	17	15		4	20			20		19	20
6						3	5									
7																
8																
9																
10																

I metodi che sembrano riconoscere in modo più preciso la struttura a 5 fattori sottostante le variabili X^* sono il rapporto di verosimiglianza, il criterio di Akaike e la cross-validation. Si può notare come per N=5000 tutti i criteri considerati si concentrino attorno al valore vero k=5, mentre, passando a N=1000, si osserva come solo l'AIC e la cross-

validation mantengano l'indicazione k=5 e come, invece, il criterio di Schwarz subisca una modifica rilevante, indicando k=3. Per $N = 200$ tutti i metodi indicano un numero di fattori fortemente inferiori rispetto ai casi precedenti, tuttavia, i 3 fattori dell' AIC rappresentano l'indicazione più vicina ai 5 fattori. L' AIC risulta, pertanto, abbastanza robusto a variazioni della numerosità N delle serie utilizzate, mentre sia il criterio di Hannan-Quinn, sia quello di Schwarz vengono fortemente influenzati da variazioni di N.

Tav. 2 - Numero di campioni di 30 titoli suddivisi per numero di fattori e metodo di determinazione.

		AIC3			AIC4			HQ2			HQ3			HQ4		
numero fattori	N	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000
0																
1		10			20			20			20	20		20	20	
2		8							7							
3		2							13							
4			3			17							17			20
5			17	20		3	20			20			3			
6																
7																
8																
9																
10																

Modificando l' AIC, secondo i suggerimenti di Smith e Spiegelhalter e ricavando così AIC3, per $\alpha = 3$, e AIC4, per $\alpha = 4$, non si verificano particolari modifiche dei risultati, se non l'ovvio slittamento verso un minor numero di fattori. Analogamente per il criterio di Hannan e Quinn, porre, rispettivamente, c = 2, 3, 4, non porta a risultati di particolare interesse e già per N = 1000 il numero di fattori tende ad essere molto basso.

A titolo illustrativo in figura 1 viene riportato, per un campione di 30 titoli, l'andamento del criterio di Akaike, di Hannan - Quinn con c = 1 e di Schwarz.

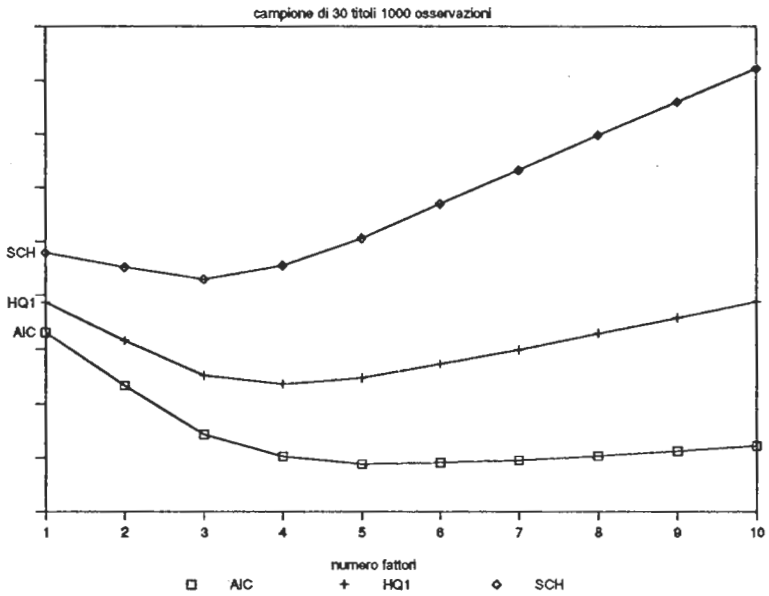

Fig. 1 - Criterio di Akaike, di Hannan-Quinn (c=1) e di Schwarz in un campione di 30 titoli e 1000 osservazioni.

Per valutare la sensibilità delle diverse misure a variazioni del numero p di titoli considerati, l'analisi è stata effettuata anche per p = 20 e i risultati sono riportati nelle tavole 3 e 4.

Rispetto alla situazione p = 30 il metodo che risulta essere più robusto a variazioni di p è il criterio di Akaike. Infatti, mentre per N = 5000, oltre all'AIC, anche la cross-validation e il criterio di Hannan e Quinn (c = 1) indicano 5 fattori, per N = 1000 solo l'AIC non si sposta verso un minor numero di fattori. Inoltre, per N = 5000, non si registra più la generale convergenza a k=5 osservata per il caso p=30 e, in particolare, il rapporto di verosimiglianza e il criterio di Schwarz indicano, rispettivamente, 3-4 e 4 fattori. Per N = 200, ancor più che nella tavola 1, tutti i metodi indicano un solo fattore, tranne l' AIC che ne presenta 2-3. Il metodo che risulta maggiormente sensibile a variazioni di p è il rapporto di verosimiglianza, le cui indicazioni risultano, in tal modo, maggiormente variabili e, di conseguenza, meno rilevanti.

Tav. 3 - Numero di campioni di 20 titoli suddivisi per numero di fattori e metodo di determinazione.

	LR			AIC			HQ1			SCH			CV		
numero fattori / N	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000
0															
1	20	10		2			19			20	5		14		
2		10		10			1				12		6		
3			15	8				18			3	2		3	
4			5		7			2	4			18		15	
5					**10**	**14**			**16**					**2**	**20**
6					3	6									
7															
8															
9															
10															

Tav. 4 - Numero di campioni di 20 titoli suddivisi per numero di fattori e metodo di determinazione.

	AIC3			AIC4			HQ2			HQ3			HQ4		
numero fattori / N	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000	200	1000	5000
0															
1	16			20			20	9		20	20		20	20	
2	4							11							
3		6			18				3			20			20
4		14			2	2			17						
5			**20**			**18**									
6															
7															
8															
9															
10															

Analogamente alla situazione rappresentata per $p = 30$ in tavola 2, le diverse varianti dei criteri di Akaike e di Hannan e Quinn non segnalano risultati particolarmente interessanti, se non un più accentuato spostamento verso un numero inferiore di fattori.

4 - Conclusioni

In questa ricerca il confronto tra diversi metodi per la determinazione del numero di fattori di un modello fattoriale viene effettuato facendo ricorso a una simulazione. Applicando il rapporto di verosimiglianza, i criteri di Akaike (con $\alpha = 2,3,4$), di Hannan e Quinn (con $c = 1, 2, 3, 4$), di Schwarz e la cross-validation a dati simulati, per i quali si conosce esattamente la struttura fattoriale sottostante, si osserva se e come le indicazioni fornite si allontanano dalla determinazione vera di k.
In particolare si è considerato un modello con k=5. Un primo interessante risultato è la convergenza, al crescere di N, delle indicazioni fornite dai diversi metodi al valore vero $k = 5$.

Il processo di convergenza, inoltre, risulta più veloce per la cross-validation e soprattutto per l' AIC, che sembra cogliere non solo con più precisione ma anche con maggiore rapidità la struttura fattoriale sottostante.

Uno dei rilievi che viene tradizionalmente mosso alle procedure di determinazione del numero di fattori del modello è l'inclusione di più fattori di quelli realmente presenti. Conoscendo il numero vero $k = 5$ di fattori sottostanti le X^*, si può, invece, osservare come i diversi criteri tendano a sottostimare, piuttosto che a sovrastimare, il numero vero di fattori: solo pochissime volte, infatti, viene indicato un numero di fattori superiore a 5, mentre un numero di fattori inferiore a 5 compare ben più frequentemente.

Questa tendenziale sottostima del numero vero di fattori, anche se legata a una particolare ampiezza del modello fattoriale (k=5), può costituire, se confermata in altri campi di ricerca, un importante elemento negli studi teorici ed empirici relativi all'analisi fattoriale.

Una ulteriore nota riguarda i coefficienti Λ^* utilizzati per la simulazione: essendo stati ricavati da una analisi fattoriale di massima verosimiglianza a 5 fattori su un campione casuale di 30 titoli, è probabile che i coefficienti relativi al primo fattore siano "alti" e quelli relativi al quinto fattore siano "bassi". In questo modo, moltiplicando nella simulazione Λ^* per f^*, potrebbero essere state generate delle variabili X^* dove i fattori rilevanti sono, in realtà, meno di 5.

In questa situazione l' AIC sarebbe certamente una misura precisa

della vera struttura fattoriale, ma coglierebbe anche la presenza di fattori eventualmente irrilevanti. D'altra parte il maggior peso attribuito al primo fattore rappresenta una caratteristica strutturale dell'analisi fattoriale e quindi sembra costituire un elemento dal quale non è opportuno prescindere. Una ulteriore verifica, nella quale i coefficienti Λ^* risultino ugualmente "pesanti", potrebbe chiarire anche questo punto.

Valori di riferimento, per k, diversi da 5, una maggiore numerosità, p, del campione di titoli considerati e, soprattutto, un maggior numero di simulazioni costituiscono, da un lato, problemi ancora aperti e rappresentano, d'altra parte, promettenti linee di ricerca.

Riferimenti bibliografici

H. Akaike (1979), A Bayesian Analysis of the Minimum AIC Procedure, *Annals of the Institute of Statistical Mathematics A*, n. 30, 9-14.

H. Akaike (1987), Factor Analysis and AIC, *Psychometrika*, vol. 52, n. 3, 317-332.

T.W. Anderson (1984), *An Introduction to Multivariate Statistical Analysis*, New York, Wiley.

M.S. Bartlett (1950), Tests of Significance in Factor Analysis, *British Journal of Mathematical and Statistical Psychology*, n. 3, 77-85.

D.H. Bower, R.S. Bower, D.E. Logue (1984), Arbitrage Pricing Theory and Utility Stock Returns, *Journal of Finance*, n. 4, 1041-1054.

H. Bozdogan, D.E. Ramirez (1987), *An Expert Model Selection Approach to Determine the Best Pattern Structure in Factor Analysis Models*, in Multivariate Statistical Modeling and Data Analysis, D. Reidel Publishing Company, 35-60.

D.E. Conway, M.R. Reinganum (1988), Stable Factors in Security Returns: Identification Using Cross-Validation, *Journal of Business & Economic Statistics*, n. 6, 1-15.

M. Costa, A. Gardini, P. Paruolo (1992), *Analisi econometrica di modelli finanziari a variabili latenti: un'applicazione al mercato italiano*, in corso di pubblicazione.

E.J. Hannan, B.G. Quinn (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society B*, vol. 41, n. 2, 190-195.

J. Kim, C.W. Mueller (1983), *Factor Analysis*, London, Sage.

D.N. Lawley, A.E. Maxwell (1971), *Factor Analysis as a Statistical Method*, London, Butterworths.

J. Lintner (1965), The Valuation of Risky Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics*, n. 47, 13-37.

P. Monari (1974), A proposito di analisi dei fattori, *Statistica*, n. 4, 755-766.

F. Panetta, E. Zautzik (1990), Evoluzione e performance dei fondi comuni mobiliari italiani, *Temi di discussione n. 142*, Banca d'Italia.

A. Rizzi (1991), *Inferenza statistica nell'analisi dei dati*, Atti del convegno Sviluppi metodologici nei diversi approcci all'inferenza statistica.

R. Roll, S. Ross (1980), An Empirical Examination of the Arbitrage Pricing Theory, *Journal of Finance*, n. 35, 1073-1103.

S. Ross (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, n. 13, 341-360.

Y. Sakamoto, M. Ishiguro, I. Kitagawa (1986), *Akaike Information Criterion Statistics*, D. Reidel Publishing Company.

G. Schwarz (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, n. 6, 461-464.

W.F. Sharpe (1964), Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance*, n. 19, 425-442.

A.F.M. Smith, D.J. Spiegelhalter (1980), Bayes Factors and Choice Criteria for Linear Models, *Journal of Royal Statistical Society B*, vol. 42, n. 2, 213-220.

K.C.J. Wei (1988), An Asset-Pricing Theory Unifying the CAPM and APT, *Journal of Finance*, n. 4, 881-892.