

MONITORAGGIO DELLE IMPRESE IN BASE AI BILANCI: IL CASO DELLE AZIENDE TESSILI DEL DISTRETTO DI PRATO

The Monitoring of Textile Firms of the District of Prato

MARCO RIANI

Istituto di Statistica, Facoltà di Economia, Università di Parma

1. Introduzione

Gli studi sulla previsione delle insolvenze delle aziende sono ormai numerosi ma non si dispone ancora di una teoria consolidata. Tale carenza metodologica accresce notevolmente le difficoltà della ricerca, in quanto obbliga lo studioso a muoversi senza una rete interpretativa di riferimento attraverso cui filtrare i diversi risultati ottenuti con la strumentazione statistica. In letteratura si è cercato di supplire alla mancanza di una teoria dell'insolvenza con ricerche empiriche di carattere estensivo attraverso cui potenziare la capacità di prevedere quali sono le aziende a rischio. Nel nostro caso questo tipo di scelta si è concretizzato nel ricorso a più tecniche di analisi multivariata dei dati: componenti principali (PCA), *cluster analysis*, analisi discriminante (AD) lineare e quadratica Bayesiana.

Il caso su cui lavoriamo riguarda l'universo dei bilanci delle società di capitale (fonte Cerved), con sede nella provincia di Prato appartenenti ai comparti della Tessitura, Filatura, Finissaggio, Lanifici ed Altri (con riferimento a tali aziende nel seguito si userà la dicitura "aziende sane"). Gli anni considerati sono il 1993 ed il 1994. In questi 2 anni le "aziende sane" sono rispettivamente 843 e 971. Inoltre, abbiamo potuto disporre dei bilanci di 42 aziende che hanno manifestato sintomi di dissesto, sfociati in dichiarazioni di fallimento entro due anni dalla pubblicazione dell'ultimo bilancio ("aziende fallite").

2. Analisi preliminari dei dati e scelta degli indici

Un elemento importante in grado di condizionare le analisi successive è la qualità delle informazioni contabili disponibili. I dati dei bilanci sui quali abbiamo lavorato presentavano alcune volte valori mancanti. Dopo aver verificato se per questi *missing values* era soddisfatta l'ipotesi di casualità, abbiamo affrontato il problema della scelta degli indici. Partendo da una larga batteria di variabili, dopo aver effettuato su di esse una serie di analisi grafiche preliminari (Riani, 1995), siamo pervenuti alla scelta dei seguenti indici di bilancio:

$X1 = (\text{Attività totali correnti} - \text{Passività totali correnti}) / (\text{Totale attività})$

$X2 = (\text{Risultato operativo lordo}) / (\text{Totale attività})$

$X3 = (\text{Capitale Netto}) / (\text{Debito totale})$

$X4 = (\text{Ricavi}) / (\text{Attività totali})$

$X5 = (\text{Cash flow}) / (\text{Debito totale})$

$X6 = (\text{Oneri Finanziari}) / (\text{Valore aggiunto})$

Questi indici che rappresentano: la solvibilità a breve (X1), la profittabilità (X2), la struttura del capitale (X3), la produttività (X4), il *cash flow* (X5) e il servizio del debito (X6) sono

molto simili a quelli già utilizzati in altri lavori passati per la previsione delle insolvenze (ad es. Forestieri, 1986).

3. Analisi preliminari delle aree a rischio tramite il metodo delle componenti principali e le tecniche di *clustering* non gerarchico

Allo scopo di avere una prima indicazione sulle aree a rischio, abbiamo applicato la PCA (utilizzando le variabili in forma standardizzata) sulle "aziende sane", dapprima per ciascuno dei 5 comparti considerati. Dall'esame delle tabelle (qui non riportate per esigenze di spazio) che riportano i primi due *factor loadings* (correlazioni tra gli indici e le componenti) per i 5 comparti emerge che la prima componente è correlata in maniera diretta con le variabili X1, X2, X3, X4 e X5 e inversa con X6. Poiché le prime 5 variabili sono direttamente legate a quello che potremo definire "il grado di solidità dell'azienda" e che quest'ultimo aspetto a sua volta è correlato in maniera inversa con il grado di indebitamento (X6) appare chiaro che la prima componente principale può essere interpretata come un "indice sintetico di solidità aziendale". In altri termini: le aziende che si collocano ai primi (ultimi) posti nella graduatoria definita dalla prima componente principale sono quelle che possono essere considerate "più solide" ("meno solide"). Per quanto riguarda la seconda componente, essa risulta determinata per lo più dall'indice X4. Questo secondo indicatore sintetico deve essere interpretato, pertanto, come una variabile legata alla capacità dell'azienda di produrre ricavi. Valori elevati della seconda componente segnalano aziende con alta produttività e forte penetrazione delle vendite.

Osservazione 1: un aspetto interessante è la stabilità in termini di segno e di ordine di grandezza del primo *factor loading* nei vari settori considerati. Per quanto riguarda il secondo *factor loading* sia a livello settoriale sia a livello temporale (ad eccezione del comparto "Altri") risulta che il coefficiente di X4 è sempre quello più importante nel determinare la seconda componente.

Osservazione 2: l'applicazione del metodo delle componenti principali a ciascuno dei 5 comparti non è stato un puro esercizio, ma è servita per comprendere le affinità e le diversità tra i vari comparti. La stabilità in termini di interpretazione delle prime due componenti principali garantisce che i risultati dell'analisi aggregata non derivano da compensazioni tra diversi comparti.

La PCA è stata applicata successivamente al totale delle unità statistiche nei due anni considerati (per esigenze di spazio riportiamo solo i risultati riferiti al 1993). Dall'esame della Tab. 1 che riporta le percentuali relative e cumulate spiegate dalle prime tre componenti principali ed i primi 3 *factor loadings* emerge che l'interpretazione delle prime due componenti principali non cambia passando dai diversi settori all'intero insieme delle aziende.

Tab. 1. Risultati dell'analisi in componenti principali applicata ai 6 indici di bilancio (totale delle aziende sane): (anno 1993).

Indici di bilancio	Primo	Secondo	Terzo		quota di var.	quota di var.
	<i>factor loading</i>	<i>factor loading</i>	<i>factor loading</i>		spiegata	cumulata
X1	0.707	0.199	0.137	1° comp.	0.34	0.34
X2	0.495	0.401	-0.438	2° comp.	0.21	0.55
X3	0.790	-0.484	0.062	3° comp.	0.17	0.72
X4	0.260	0.827	-0.092			
X5	0.756	-0.295	-0.110			
X6	-0.203	-0.253	-0.877			

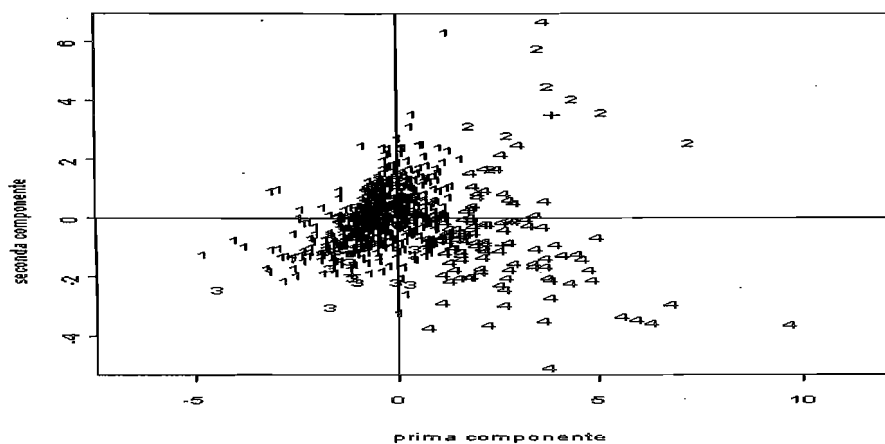
Dall'analisi condotta a livello globale emerge che anche la terza componente principale spiega una quota non trascurabile della varianza totale (17%). L'analisi del *factor loading* ad essa associato rivela che essa è determinata in maniera preponderante dall'indice X6 (servizio del debito). Tenendo presente che il coefficiente legato a X6 è negativo, è agevole constatare che le aziende che presentano i valori più elevati per X6 sono anche quelle meno indebitate.

Un passo ulteriore della ricerca è stato l'individuazione di gruppi di aziende omogenee con riferimento agli indici di bilancio considerati. A tale scopo abbiamo applicato il metodo di *clustering* non gerarchico delle *k*-medie. Per la scelta del numero ottimo dei gruppi ci si è avvalsi dei criteri suggeriti in letteratura (cfr. ad es. Milligan e Cooper (1985)) pervenendo ad una soluzione con 4 gruppi. L'algoritmo delle *k*-medie è stato applicato sia ai singoli comparti sia al totale delle aziende. Per meglio puntualizzare i risultati ottenuti e per poter individuare variabili discriminanti, ossia quelle che giocano un ruolo essenziale nella formazione dei 4 gruppi, si sono calcolati per ciascuna variabile i valori medi all'interno dei gruppi. Successivamente, per meglio individuare le caratteristiche delle aziende appartenenti ai 4 gruppi si è ritenuto interessante rappresentare nello spazio delle componenti principali i *clusters* emersi dall'analisi¹. Se si riporta per ciascuna impresa l'indicazione del gruppo di appartenenza e si rappresentano le aziende in base al valore che esse assumono riguardo alle prime due componenti principali per il 1993 si ottiene la Fig. 1. Dall'esame di tale grafico, ricordando l'interpretazione delle prime due componenti, emerge che:

- le aziende appartenenti al gruppo 1 sono caratterizzate da una solidità di bilancio inferiore a quella media globale;
- le aziende contrassegnate con il numero 2 presentano alta solidità aziendale ed alti ricavi;
- le aziende del gruppo 3 possono essere considerate come "a rischio", in quanto presentano valori inferiori alla media sia con riferimento alla prima che alla seconda componente principale.

le aziende del gruppo 4 presentano un grado di solidità aziendale superiore alla media, ma hanno una produttività non elevata.

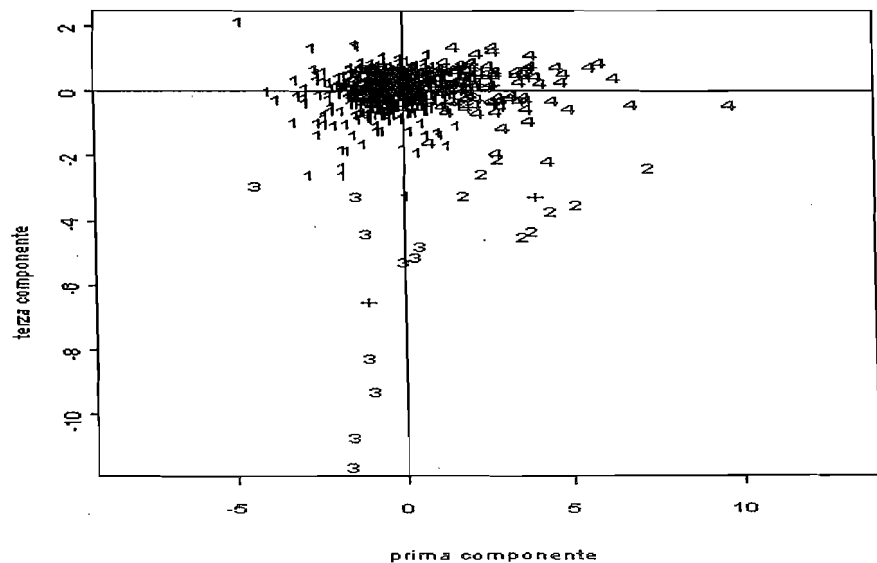
Fig. 1: Rappresentazione delle aziende nel piano delle prime due componenti principali e gruppo di appartenenza (anno 1993).



Osservazione: è interessante notare nella parte in alto del grafico un'unità statistica appartenente al gruppo 4 che nello spazio delle prime due componenti viene rappresentata molto vicina a quelle del gruppo 2. Ciò rappresenta un tipico caso di un'unità statistica per la quale la rappresentazione nello spazio delle prime componenti non risulta soddisfacente. E' bene ricordare, infatti, che la finalità principale del metodo delle componenti principali è l'individuazione di nuove variabili incorrelate (le componenti) che spieghino la maggior quota possibile della varianza delle variabili originarie e non quello di formare gruppi omogenei. L'applicazione successiva delle tecniche di *cluster analysis* combinata alle componenti principali, tuttavia, se da un lato permette di interpretare i *cluster* ottenuti alla luce degli indici sintetici trovati, dall'altra consente di individuare le unità la cui proiezione nel sottospazio definito dalle prime due componenti non risulta soddisfacente.

La lettura della Fig. 1 può risultare difficile a causa del fatto che le aziende facenti parte del gruppo 3 vengono rappresentate molto vicine a quelle dei gruppi 1 e 4. Allo scopo di cogliere meglio le differenze tra i vari gruppi si è ritenuto interessante rappresentare il totale delle unità statistiche nello spazio della prima e della terza componente (Fig. 2). Dall'esame di questa figura emerge che le aziende facenti parte del gruppo 3 sono chiaramente identificabili come quelle che presentano i valori più bassi per la terza componente (valori elevati per l'indicatore X6). Nella Fig. 2, inoltre, sono immediatamente riconoscibili le aziende facenti parte del gruppo 2. Tali aziende nello spazio delle prime due componenti erano state identificate come "molto solide" e con "forte penetrazione delle vendite". La rappresentazione nella Fig. 2 permette di affermare che queste aziende presentano un servizio sul debito non modesto.

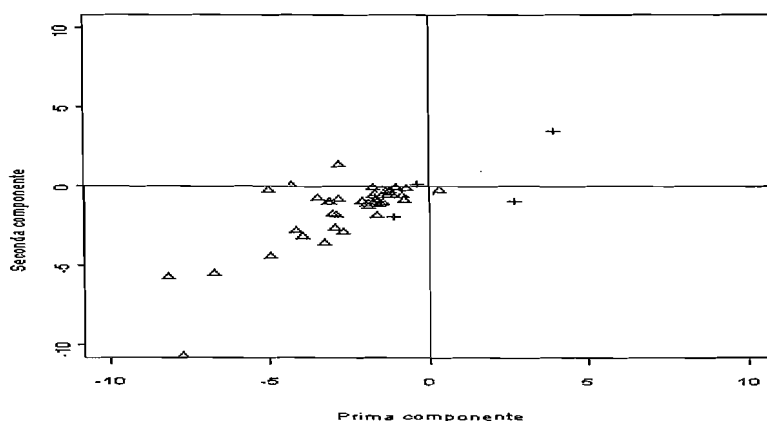
Fig. 2: Rappresentazione delle aziende nel piano definito dalla prima e dalla terza componente principale ed indicazione del gruppo di appartenenza (anno 1993)



Per avere un'ulteriore indicazione sul fatto che le aziende appartenenti al terzo quadrante della Fig. 1 possono essere considerate a rischio occorre analizzare la posizione che le aziende fallite avrebbero occupato in tale grafico. Dall'esame della Fig. 3 che riporta con triangoli la posizione delle aziende fallite nello spazio delle prime 2 componenti principali precedentemente identificate e con il simbolo "+" le posizioni dei centroidi dei gruppi, emerge che le aziende fallite, a parte qualche rara eccezione, tendono tutte a collocarsi nel terzo quadrante.

In conclusione, la PCA combinata alle tecniche di *clustering* si è rivelata un utile strumento per evidenziare aree critiche, cioè quelle aziende a rischio di dissesto che quindi dovrebbero essere monitorate attentamente. In ogni caso per avere informazioni più precise sulla previsione delle insolvenze è necessario ricorrere ad un'altra tecnica di analisi multivariata: l'analisi discriminante (AD).

Fig. 3: Rappresentazione delle aziende fallite nello spazio delle prime due componenti principali (anno 1993)



4. I risultati dell'analisi discriminante

Dati due universi, il primo costituito da aziende fallite ed il secondo composto dalle aziende "sane" l'obiettivo dell'AD (Forestieri, 1986) è quello di individuare una combinazione degli indici (la funzione discriminante) atta a distinguere le aziende appartenenti ai due universi. Nel caso in cui la funzione risulti soddisfacente con riferimento alle aziende per le quali si conosce l'universo di appartenenza si può proporre un'applicazione a scopo previsivo ad altre aziende con l'intento di valutare sulla base degli indici di bilancio considerati, se le aziende si trovano in uno "stato di salute" soddisfacente oppure precario.

In questo studio abbiamo applicato sia l'AD lineare che quella quadratica, in quanto (come afferma ad esempio Krzanowski, 1988: p. 358) la scarsa plausibilità dell'ipotesi di normalità dei dati in esame richiede l'applicazione congiunta di più tecniche. L'utilizzo del paradigma Bayesiano è motivato dalla necessità di disporre delle probabilità a posteriori per ogni assegnazione. Risulta molto importante, infatti, sapere se un'azienda è stata dichiarata

sana con una probabilità a posteriori di 0.51 oppure di 0.99. Un problema che abbiamo dovuto affrontare è stato quello del forte squilibrio in termini di numerosità dei due universi in esame. Da un lato, infatti, abbiamo "le aziende sane" (ossia quelle ancora in vita due anni dopo la pubblicazione del bilancio) che nel 1993 sono 843 e nel 1994 sono 971. Dall'altro lato "le aziende fallite" (ossia quelle che hanno presentato dichiarazione di fallimento entro due anni dalla pubblicazione dell'ultimo bilancio) che sono solamente 42. Lo squilibrio dei due universi ha reso necessaria una modifica nelle tecniche tradizionali, in quanto l'applicazione dell'AD Bayesiana con probabilità a priori non informative (pari cioè a 0.5 e 0.5) non consentirebbe di ottenere una corretta riclassificazione delle aziende. Si pone dunque il problema di pervenire ad una stima della probabilità a priori, da utilizzare nell'AD Bayesiana, che consenta al tempo stesso di tener conto della diversa probabilità a priori di appartenenza ai due universi imputabile alla differente numerosità dei due universi e di assicurare un ragionevole tasso di errate classificazioni. Per raggiungere tale obiettivo si potevano seguire diverse strade. Un criterio percorribile per la stima delle probabilità a priori era di fare riferimento alla importanza relativa in termini di numerosità delle aziende di ciascun gruppo sul totale delle aziende: ad esempio nel 1993 tali probabilità sarebbero pari rispettivamente a 0.95 (843/885) e 0.05 (42/885). Nel presente lavoro si è preferito determinare tali probabilità a priori in modo da garantire un certo tasso di errore di prima specie (percentuale di aziende fallite classificate come sane). A questo scopo abbiamo condotto in via preliminare l'AD su gruppi di aziende di uguale numerosità tenendo fisso il campione delle aziende fallite, ma estraendo diversi campioni stratificati proporzionali dall'universo delle aziende sane. La ripetizione nell'estrazione è resa necessaria dalla necessità di evitare che il "campione di confronto" contenga caratteristiche strutturali e funzionali tali da comportare differenziazioni indipendenti dalle anomalie di cui si vogliono studiare i sintomi. L'estrazione di campioni di uguale numerosità permette di ottenere una corretta valutazione degli errori di allocazione. Il passo finale della procedura consiste nel riproporre l'AD Bayesiana (lineare o quadratica) con riferimento al complesso delle aziende sane, scegliendo come probabilità a priori quella coppia che garantisce un tasso di errore di prima specie (aziende fallite riclassificate sane) il più vicino possibile, sia come ammontare che come composizione a quello ottenuto con l'analisi condotta su gruppi di uguale numerosità.

Una volta pervenuti per riclassificazione alla individuazione di un gruppo di aziende sane e successivamente di un gruppo di aziende fallite, il confronto fra la composizione di tali gruppi e quella dei gruppi di partenza consente di determinare gli errori che si commettono quando si procede ad una riclassificazione delle aziende utilizzando i risultati dell'AD. Dall'esame degli errori di riclassificazione per ciascuno dei 6 comparti considerati è emerso che il tasso di errore varia tra il 10% ed il 12%. Per approfondire ulteriormente il problema delle aziende fallite classificate come sane abbiamo riportato nella Tab. 2 in corrispondenza dei 6 campioni il numero d'ordine delle aziende classificate non correttamente. Dall'esame della Tab. 2 emerge che c'è una sostanziale stabilità per quanto riguarda le aziende fallite dichiarate sane. Ad esempio, è interessante notare che l'azienda 37 viene sempre classificata come sana da entrambe le funzioni discriminanti. E' utile osservare che tale azienda è l'unica tra le fallite che assume valori positivi per quanto riguarda la prima componente principale (Fig. 3). In questa azienda perciò, il profilo degli indici di bilancio non segnala una situazione prossima al dissesto. La stessa cosa si verifica in linea di massima per le altre aziende fallite dichiarate sane. Nella maggior parte dei casi, il fatto di allocare tali aziende al gruppo delle sane non costituisce una deficienza del metodo. E' verosimile ipotizzare che al fallimento di

tali aziende abbiano concorso cause esogene non quantificabili dal modello. Un criterio ragionevole di valutare le probabilità a priori di ogni assegnazione appare dunque quello di cercare la coppia di probabilità che garantisca un tasso di errore molto vicino a quello riscontrato con campioni di uguale numerosità. Dopo vari tentativi abbiamo constatato che la coppia di probabilità a priori 0.7 e 0.3 garantisce risultati molto vicini a quelli trovati sui 6 campioni. Dopo aver applicato l'AD sull'intero insieme di aziende a nostra disposizione utilizzando le precedenti probabilità a priori, abbiamo provveduto a valutare la bontà del modello tramite il metodo *jackknife* (Krzanowski, 1988; p. 362). Quest'ultima analisi ha segnalato che il tasso di errata classificazione si aggira intorno al 13%.

Tab. 2 Aziende fallite classificate sane applicando l'AD lineare e quadratica per ciascuno dei sei campioni (anno 1993).

Numero d'ordine	I		II		III		IV		V		VI	
	Lin.	Quad.	Lin.	Quad.	Lin.	Quad.	Lin.	Quad.	Lin.	Quad.	Lin.	Quad.
2		X		X		X		X		X		X
3	X		X		X		X		X		X	
6	X	X	X	X	X	X	X	X	X	X	X	X
11	X		X		X	X	X	X	X		X	
12		X				X						
24	X		X		X		X		X		X	
26	X											
37	X	X	X	X	X	X	X	X	X	X	X	X
41			X	X		X				X		
42										X		
Totale	6	4	6	4	5	6	5	4	5	5	4	3

Una notevole attenzione, a nostro avviso deve essere data alle "aziende sane" che sia l'AD lineare sia quella quadratica attribuiscono al gruppo delle fallite (in totale 29). Un risultato di rilievo è il fatto che di queste 29 società ben 17 non vengono bocciate l'anno successivo. I risultati ottenuti fanno quindi emergere il fatto importante che uno stato di forte debolezza economico-finanziaria non necessariamente assume per l'impresa un carattere di irreversibilità. Una congiuntura favorevole combinata a scelte aziendali azzeccate possono essere decisive per poter uscire da uno stato patologico che se protratto nel tempo avrebbe come conseguenza inevitabile l'insolvenza. Il confronto con i risultati ottenuti con il metodo delle componenti principali, infine, ha consentito di scoprire che molte delle aziende che erano state segnalate come a rischio in base a quest'ultimo metodo sono in buona parte le stesse che sono state "bocciate" dall'AD. Questo dimostra che l'utilizzo congiunto di diversi metodi statistici (anche se essi hanno presupposti e finalità diverse) può fornire un notevole contributo allo studio di un problema complesso quale la previsione delle insolvenze.

5. Conclusioni

In questo lavoro grazie all'impiego congiunto di svariate tecniche statistiche (componenti principali, *cluster analysis*, analisi discriminante Bayesiana) abbiamo individuato zone di debolezza dell'apparato tessile pratese. E' stato interessante osservare la concordanza sostanziale dei risultati derivanti dall'impiego delle diverse tecniche. Nella presente ricerca abbiamo utilizzato strumenti di analisi statistica multivariata. In futuro ci proponiamo di verificare anche in quale misura i metodi basati sulle reti neurali (Trippi e

Turban, 1992) siano in grado di integrare e/o gettare nuova luce sulla previsione delle insolvenze.

Un problema che abbiamo dovuto affrontare è stato quello del forte squilibrio in termini di numerosità dei due universi in esame. Per risolvere questo problema abbiamo proposto una strategia originale per la determinazione delle probabilità a priori da utilizzare nell'analisi discriminante, che si basa dapprima sull'estrazione di campioni di uguale numerosità e successivamente sull'analisi dei due universi.

In conclusione, l'essere riusciti con il conforto di metodologie diverse ad evidenziare un certo numero di situazioni di anomalia con largo anticipo (due anni nel nostro caso) fornisce agli operatori economici dei diversi comparti l'opportunità di intervenire tempestivamente per eliminare le cause di declino e mettere a punto i provvedimenti necessari per il riassetto delle imprese che vengono segnalate come "a rischio".

Note

⁽¹⁾ Un approccio alternativo, qui non trattato per esigenze di spazio, è quello che comporta l'impiego della tecnica di *fuzzy clustering*. In tal caso ogni unità viene assegnata ad un gruppo con un valore della funzione di appartenenza compreso tra 0 e 1 ed il valore di tale funzione per il *cluster* a rischio potrebbe essere interpretato come una stima della probabilità di fallimento.

Riferimenti bibliografici

FORESTIERI, G. (1986), a cura di, *La previsione delle insolvenze aziendali, profili teorici e analisi empiriche*, Giuffrè, Milano.

KRZANOWSKI, W.J. (1988), *Principle of Multivariate Analysis, A User's Perspective*, Clarendon Press, Oxford.

MILLIGAN, G.W., COOPER, M.C. (1985), An Examination of Procedures for Determining the Number of Clusters in a Data Set, *Psychometrika*, pp. 159-179.

RIANI, M. (1995), "Analisi grafiche multidimensionali: un'applicazione ad un insieme di aziende tessili", in P. Ganugi: *Produzione per comparti: analisi statistica dell'industria tessile*, Giuffrè, Milano.

TRIPPI, R.R., TURBAN, E. (1992), Ed., *Neural Networks in Finance and Investing*, Probus Publishing, New York.

Summary

MONITORAGGIO DELLE IMPRESE IN BASE AI DATI DI BILANCIO: IL CASO DELLE AZIENDE TESSILI DI PRATO *The Monitoring of Textile Firms of the District of Prato*

In this paper we use different techniques of multivariate analysis (principal components, cluster analysis, discriminant analysis) to evaluate the risk of bankruptcy of the textile firms of the district of Prato. In order to overcome the problem of different size of the universe of "good firms" with respect to that of failed firms, we suggest an original procedure in order to estimate prior probabilities to be used in Bayesian discriminant analysis.

Keywords

Bankruptcy forecast, Principal components, Cluster analysis, Bayesian discrimination.