

Utilizarea tehnicilor de analiză cantitativă a datelor pentru estimarea riscului de faliment al corporațiilor

Ștefan Daniel ARMEANU

Academia de Studii Economice, București
darmeanu@yahoo.com

Georgeta VINTILĂ

Academia de Studii Economice, București
vintilageogeta@yahoo.fr

Maricica MOSCALU

Academia de Studii Economice, București
mari.moscalu@yahoo.com

Maria-Oana FILIPESCU

Academia de Studii Economice, București
oanadicea@yahoo.com

Paula LAZĂR

Academia de Studii Economice, București
lazar_paula@yahoo.com

Rezumat. *Alexandru D. Xenopol (1847-1920), personalitate marcantă a tradiției intelectuale românești la cumpăna dintre secole – academician, economist, filosof, istoric, pedagog, sociolog și scriitor – a rămas în memoria culturală universală ca un neobosit promotor al emancipării economice a teritoriilor locuite de români. Minte enciclopedică lucidă, deopotrivă vizionară, Alexandru D. Xenopol s-a dedicat căutării argumentelor compatibilității românilor cu modernitatea și globalizarea în istoria cultural-educațională și economică a românilor. Alexandru D. Xenopol nu a trecut neobservat; savanții acestei țări, precum Nicolae Iorga, i-au omagiat valoarea intelectuală deosebită: „crescut în cele mai bune tradiții ale școlii economice de la jumătatea veacului al XIX-lea, și mai presus de toate minte filosofică, deprins cu abstracțiile și cu o nesfârșită iubire față de subtilele legături dintre dânsle” (Iorga, 1975, p.190). Scopul acestei lucrări constă în evidențierea importanței răspândirii ideilor economice la români în conturarea stadiului de dezvoltare economică.*

Cuvinte-cheie: educație economică; industrie; instituții; protecționism; emancipare economică.

Coduri JEL: C81, D22, G30, G33.

Coduri REL: 9B, 11Z.

1. Introducere

Primul model de anvergură cu privire la analiza riscului de faliment, atât în literatura de specialitate, cât și în practica financiară i-a aparținut lui E. I. Altman, care l-a publicat în forma sa inițială în anul 1968, și este cunoscut sub denumirea de funcția-scor Z. Deși în aparență simplu, acest model a avut asupra predicției riscului de faliment de întreprindere un efect analog celui pe care celebrul model Black-Scholes l-a avut asupra evaluării instrumentelor financiare derivate.

Modelul propus de Altman are la bază analiza discriminantă, aceasta fiind utilizată pentru a elabora modele de clasificare/de predicție a apartenenței unor observații la anumite clase determinate a priori. În acest scop, în cadrul analizei discriminante este construit un clasificator pe baza unui set de observații și a unor indicatori caracteristici pentru aceste observații. În cazul modelului Altman setul de observații este reprezentat de un număr de firme clasificate de către autor în solvabile și insolabile, iar indicatorii considerați sunt anumite rate pe baza cărora este analizată starea financiară a companiilor.

Funcția-scor Z propusă de Altman este, de fapt, o aplicație a unui clasificator liniar (de tip Fisher), îmbrăcând următoarea formă:

$$Z(r_1, r_2, \dots, r_n) = \alpha_0 + \alpha_1 \times r_1 + \alpha_2 \times r_2 + \dots + \alpha_n \times r_n,$$

unde:

r_1, r_2, \dots, r_n = ratele considerate în vederea elaborării modelului de clasificare;

$\alpha_1, \alpha_2, \dots, \alpha_n$ = coeficienții ratelor financiare considerate;

α_0 = termenul liber (intercept-ul) funcției de clasificare.

Pe baza scorului obținut de fiecare companie analizată se realizează alocarea la una dintre cele două categorii, și anume companii falimentare sau companii nefalimentare. De asemenea, pe baza scorului Z se estimează și probabilitatea de faliment a companiei analizate.

Varianta originală a modelului Altman, propusă în anul 1968, este următoarea (Altman, 2002, p. 14):

$$Z = 1.2xr_1 + 1.4xr_2 + 3.3xr_3 + 0.6xr_4 + 1.0xr_5,$$

unde $r_i, i = \overline{1,5}$ sunt definite mai jos:

$$r_1 = \frac{\text{Active circulante}}{\text{Active totale}};$$

$$r_2 = \frac{\text{Profit reinvestit}}{\text{Active totale}};$$

$$r_3 = \frac{\text{Profit din exploatare}}{\text{Active totale}};$$

$$r_4 = \frac{\text{Capitalizare bursiera}}{\text{Valoarea contabila a datoriilor firmei}};$$

$$r_5 = \frac{\text{Cifra de afaceri}}{\text{Active totale}}.$$

Altman a definit trei regiuni în funcție de care întreprinderile se pot clasifica:

- $Z > 2,99$: zona sigură. Probabilitate de faliment foarte scăzută, practic neglijabilă.
- $1,8 < Z < 2,99$: zona nesigură. Risc de faliment mediu.
- $Z < 1,8$: zona riscantă. Probabilitate de faliment corporativ crescută.

Calibrarea modelului propus de Altman s-a realizat pornind de la un eșantion de 66 de companii comerciale cu capital de stat ce operau pe piața americană, aproximativ 50% dintre acestea fiind falimentare. Ulterior, Altman a elaborat un model ce putea fi aplicat și în cazul întreprinderilor private, prezentat în cele ce urmează:

$$Z' = 0.717xr_1 + 0.847xr_2 + 3.107xr_3 + 0.420xr_4 + 0.998xr_5,$$

unde:

$$r_1 = \frac{\text{Active circulante}}{\text{Active totale}};$$

$$r_2 = \frac{\text{Profit reinvestit}}{\text{Active totale}};$$

$$r_3 = \frac{\text{Profit din exploatare (EBIT)}}{\text{Active totale}};$$

$$r_4 = \frac{\text{Capitalizare bursieră}}{\text{Valoarea contabilă a datoriilor firmei}};$$

$$r_5 = \frac{\text{Cifra de afaceri}}{\text{Active totale}}.$$

Intervalele critice pentru această funcție Z' sunt:

- $Z' > 2,9$: zona sigură;
- $1,23 < Z' < 2,99$: zona nesigură;
- $Z' < 1,23$: zona falimentară.

Există și o a treia variantă a modelului Altman, actualizată și revizuită, ce prezintă avantajul aplicabilității și în cazul companiilor neindustriale și a celor ce operează în țări aflate în curs de dezvoltare:

$$Z'' = 6.56xr_1 + 3.26xr_2 + 3.72xr_3 + 1.05xr_4,$$

în care:

$$r_1 = \frac{\text{Active circulante}}{\text{Active totale}};$$

$$r_2 = \frac{\text{Profit reinvestit}}{\text{Active totale}};$$

$$r_3 = \frac{\text{Profit din exploatare}}{\text{Active totale}};$$

$$r_4 = \frac{\text{Activ net contabil}}{\text{Valoarea contabilă a datoriilor firmei}}.$$

În acest caz, regiunile critice sunt cele de mai jos:

- $Z'' > 2,6$: zona de siguranță;
- $1,1 < Z'' < 2,6$: zona incertă;
- $Z'' < 1,1$: zona cu risc ridicat de faliment.

Modelul Altman, foarte utilizat în practica financiară, are o acuratețe de clasificare de peste 70% (conform Stancu, 2007, p. 787).

Un model de clasificare asemănător cu modelul Altman a fost realizat de economiștii J. Conan și M. Holder în anul 1979 și îmbracă forma de mai jos:

$$CH = 0.24xr_1 + 0.22xr_2 + 0.16xr_3 - 0.87xr_4 - 0.1xr_5,$$

în care:

$$r_1 = \frac{\text{Excedent brut de exploatare}}{\text{Datorii totale}};$$

$$r_2 = \frac{\text{Capitaluri permanente}}{\text{Total bilanțier}};$$

$$r_3 = \frac{\text{Active circulante} - \text{Stocuri}}{\text{Total bilanțier}};$$

$$r_4 = \frac{\text{Cheltuieli financiare}}{\text{Cifra de afaceri netă}};$$

$$r_5 = \frac{\text{Cheltuieli cu personalul}}{\text{Valoare adăugată}}.$$

Conform modelului Conan-Holder, o valoare CH egală cu -0.21 înseamnă o probabilitate de faliment de 100%, un scor de 0,068 indică o probabilitate de 50%, iar scorul CH de 0,164 implică o probabilitate de faliment de 10%.

2. Analiza componentelor principale

Scopul acestui studiu îl reprezintă realizarea unei funcții scoring de tip Altman pe un eșantion de 60 de firme românești cotate pe piața românească de capital, pentru a pune în evidență atât puterea financiară a acestora, dar și capacitatea lor de a face față obligațiilor asumate. În acest sens am luat în considerare un număr de șapte indicatori economico-financiari reprezentativi pentru activitatea companiilor (*activ total* – AT, *cifra de afaceri netă* – CA, *rezultat din exploatare* – EBIT, *flux net de trezorerie (cash flow) din activitatea de exploatare* – CF, *profit net* – PN, *datorii totale* – Datorii totale și *capitalizare bursieră medie* – CB).

Pentru început, am standardizat indicatorii considerați. Tabelul 1 prezintă matricea de corelație pentru cele șapte variabile originale. Evident, elementele de pe diagonala principală a matricei sunt egale cu unitatea:

Tabelul 1

Matricea de corelație dintre variabilele originale

Variable	Correlations (baza de date + indicatori2010.sta)						
	Activ total	CA	Datorii totale	PN	EBIT	CB	CF
Activ total	1,000000	0,955671	0,904545	0,811757	0,883171	0,984581	0,864128
CA	0,955671	1,000000	0,967587	0,644614	0,741079	0,905970	0,803860
Datorii totale	0,904545	0,967587	1,000000	0,493520	0,608211	0,826175	0,686195
PN	0,811757	0,644614	0,493520	1,000000	0,990361	0,891671	0,839384
EBIT	0,883171	0,741079	0,608211	0,990361	1,000000	0,944330	0,876308
CB	0,984581	0,905970	0,826175	0,891671	0,944330	1,000000	0,899479
CF	0,864128	0,803860	0,686195	0,839384	0,876308	0,899479	1,000000

Sursa: calcule proprii.

Matricea de corelații arată legătura strânsă existentă între toate cele șapte variabile considerate, prognozând o mai bună reprezentare a acestora într-un număr substanțial redus de noi variabile, componente principale. Existența unor corelații puternice între variabilele analizate diminuează semnificația individuală a acestora din urmă, pe de o parte, și evidențiază existența redundanțelor informaționale, pe de altă parte: există o cantitate semnificativă de informație disipată în legăturile dintre variabile. În demersul nostru, ne propunem atât reducerea dimensionalității spațiului cauzal inițial, cât și eliminarea acestor redundanțe informaționale, și de aceea utilizăm metoda analizei componentelor principale.

Plecând de la standardizarea datelor inițiale, se prezintă mai jos valorile proprii rezultate în urma analizei componentelor principale. Este de menționat că doar valorile proprii mai mari decât unitatea sunt de reținut pentru că doar acele componente principale care au varianța mai mare decât cea a variabilelor originale standardizate (medie nulă și varianța egală cu 1) ar trebui extrase, potrivit criteriului lui Kaiser. Rezultatele sunt prezentate în tabelul 2:

Tabelul 2

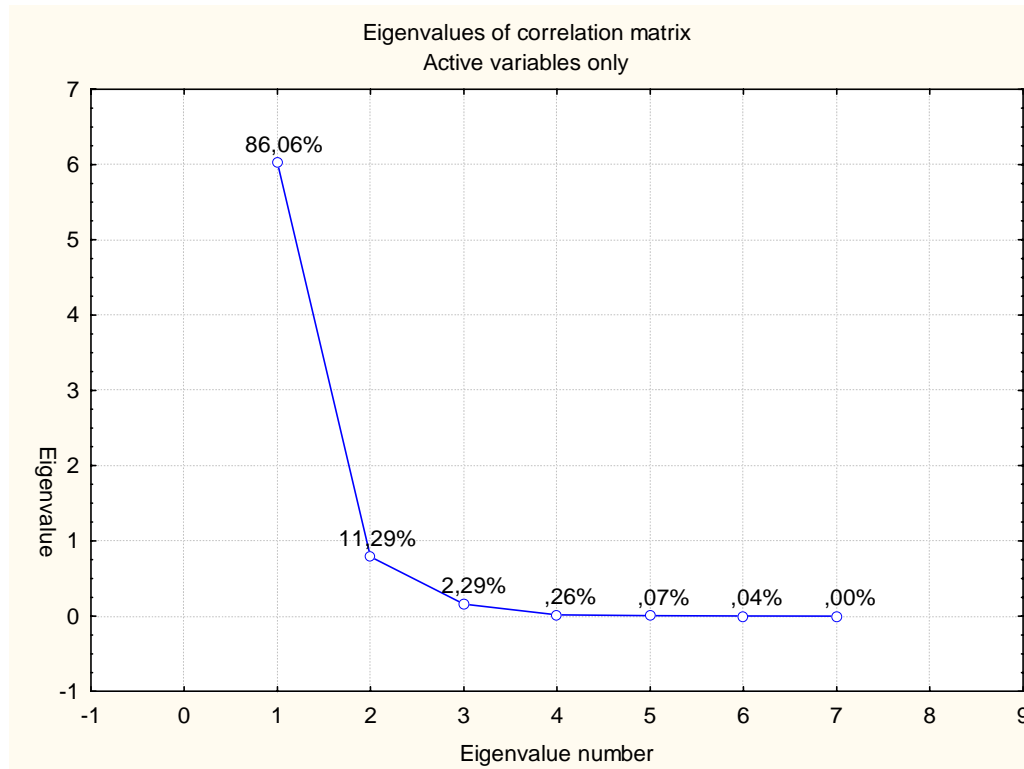
Valorile proprii supraunitare ale matricii de corelație

Value number	Eigenvalues of correlation matrix, and related statistics (baza de date + indicatori2010.sta) Active variables only			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	6,024163	86,05947	6,024163	86,0595
2	0,790360	11,29086	6,814522	97,3503
3	0,160099	2,28713	6,974622	99,6375
4	0,018130	0,25900	6,992751	99,8964
5	0,004561	0,06516	6,997312	99,9616
6	0,002558	0,03654	6,999870	99,9981
7	0,000130	0,00185	7,000000	100,0000

Sursa: calcule proprii.

Se observă astfel că doar prima nouă variabilă formată are valoarea proprie – care se interpretează în termeni de varianță sau de cantitate informațională – mai mare decât unitatea, așadar se reține în analiză doar prima componentă principală.

Aceeași decizie se poate formula și pe baza analizei graficului prezentat în figura 1, utilizând criteriul pantei:



Sursa: calcule proprii.

Figura 1. Graficul valorilor proprii ale matricei de corelație

Odată determinat numărul de componente principale reținute în analiză, mai departe se va proceda la încercarea de interpretare a componentelor principale. Astfel, se va determina matricea factor pentru singura componentă reprezentativă rezultată. Matricea factor este foarte importantă în analiza noastră, întrucât elementele sale (cunoscute și sub denumirea de *intensități ale factorilor*) sunt coeficienții de corelație între variabilele originale și componentele principale. Formula de calcul pentru un element oarecare al acestei matrici este:

$$f_{ij} = \frac{\sqrt{\lambda_j}}{\sqrt{\text{VAR}(x_i)}} \cdot \beta_{ij}, i=1, 2, \dots, n \quad j=1, 2, \dots, k$$

unde k este numărul de componente principale reținute.

Formula anterioară ne dă coeficientul de corelație dintre variabila originală i și componenta principală j. Demonstrația relației are la bază definiția coeficientului de corelație:

$$\rho_{x_i, z_j} = f_{ij} = \frac{\text{COV}(x_i, z_j)}{\sqrt{\text{VAR}(x_i)} \cdot \sqrt{\text{VAR}(z_j)}}$$

Trecând la nivel matriceal, relația anterioară se rescrie astfel:

$$F = VXW,$$

unde X este matricea de covarianță dintre vectorii x și w (vectorii variabilelor originale și componentelor principale), iar V și W sunt matrici diagonale, ale căror elemente de pe diagonala principală sunt egale cu inversele varianțelor variabilelor originale și, respectiv, componentelor principale:

$$V = \begin{pmatrix} \frac{1}{\sqrt{\text{VAR}(x_1)}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\text{VAR}(x_n)}} \end{pmatrix}$$

$$W = \begin{pmatrix} \frac{1}{\sqrt{\text{VAR}(z_1)}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\text{VAR}(z_n)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

Matricea X se determină pornind de la definiția matematică a covarianței:

$$X = \text{COV}(x, z) = E[(x - E(x)) \cdot (z - E(z))^t]$$

Vom face acum ipoteza simplificatoare a centrării variabilelor originale și a componentelor principale, știut fiind că operația de centrare nu influențează cu nimic varianța unei variabile aleatoare. Ținând cont și de relația anterioară, ecuația principală va deveni:

$$X = E(xz^t) = E(x(B^t x)^t) = E(xx^t B) = E(xx^t) B = \Sigma B$$

Înlocuind relațiile anterioare vom avea:

$$F = VXW = \begin{pmatrix} \frac{1}{\sqrt{\text{VAR}(x_1)}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\text{VAR}(x_n)}} \end{pmatrix} \Sigma B \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

Ținând cont de expresia matricei Σ dată și de configurația matricei A , rezultă că, într-adevăr, un element oarecare al matricei F este dat de coeficientul de corelație dintre variabila originală i și componenta principală j .

Matricea factor, ale cărei elemente (factor loadings) reprezintă coeficienții de corelație dintre variabilele originale și componentele principale, este prezentată în tabelul 3:

Tabelul 3

Matricea factor

Variable	Factor-variable correlations (factor loadings), based on corre (baza de date + indicatori2010.sta lations	
	Factor 1	
Activ total	-0,987380	
CA	-0,927273	
Datorii totale	-0,844738	
PN	-0,875847	
EBIT	-0,932918	
CB	-0,995254	
CF	-0,920787	

Sursa: calcule proprii.

Se observă că noua componentă principală prezintă corelații foarte puternice, dar în sens negativ cu toate cele șapte variabile inițiale, de peste 85%.

În tabelul 4 sunt prezentați coeficienții combinațiilor liniare ce definesc componentele principale (vectorii proprii ai matricei de corelație), pe baza cărora vom calcula scorurile observațiilor în spațiul componentelor principale:

Tabelul 4

Coeficienții componentelor principale

Variable	Eigenvector	
	Factor 1	
Activ total	-0,402287	
CA	-0,377798	
Datorii totale	-0,344171	
PN	-0,356845	
EBIT	-0,380098	
CB	-0,405495	
CF	-0,375155	

Sursa: calcule proprii.

Reprezentarea companiilor în noul spațiu dat de componenta principală este prezentat în figura 2:

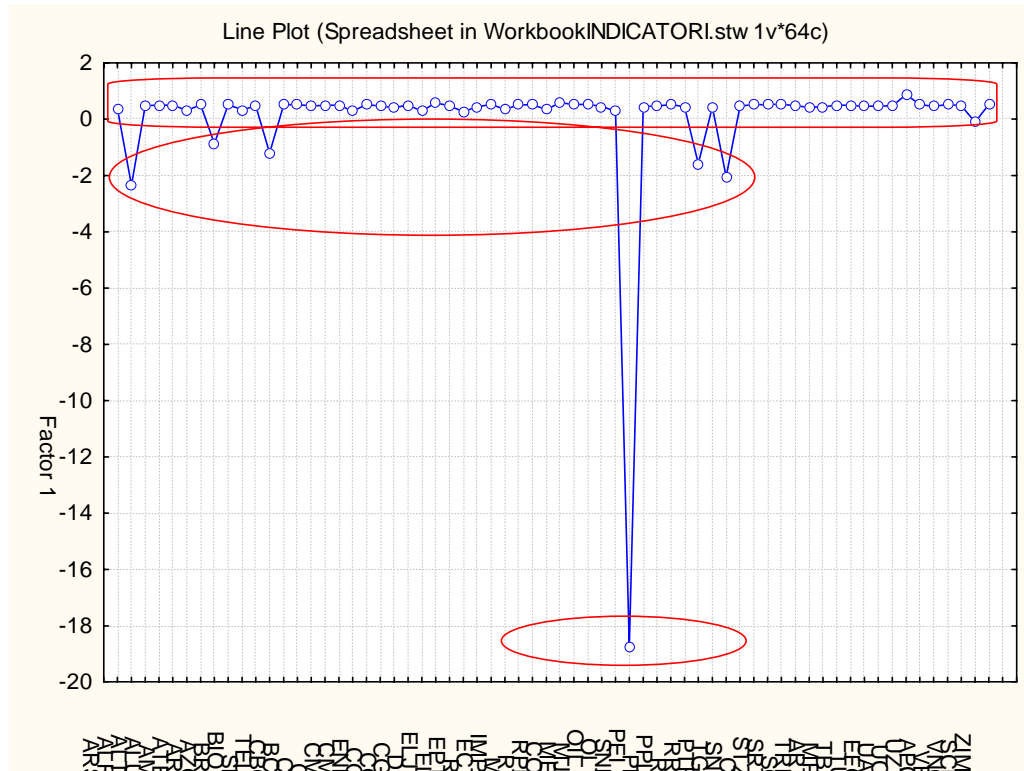


Figura 2. Reprezentarea companiilor în spațiul primei componente principale

Este de remarcă că SNP se detașează clar față de celelalte companii în funcție de această componentă principală, întrucât valorile înregistrate pentru toate variabilele sunt net superioare comparativ cu celelalte firme. O altă categorie de firme este reprezentată de ALR, AZO, TEL, RRC, TGN; acestea prezintă niveluri înalte ale tuturor indicatorilor, iar celelalte companii aflate în a treia categorie au valori mici și medii ale indicatorilor.

În urma aplicării analizei componentelor principale, ținând cont că am identificat o singură componentă principală ce sintetizează peste 86% din informația generată de indicatorii inițiali, identificăm un model de tip Altman care are următoarea formă:

$$Z = -0.402xr_1 - 0.377xr_2 - 0.344xr_3 - 0.356xr_4 - 0.38xr_5 - 0.405xr_6 - 0.375xr_7$$

unde $r_i, i = \overline{1,7}$ sunt definite mai jos:

$$r_1 = \text{Activ total}; \quad r_2 = CA; \quad r_3 = \text{Datorii totale}; \quad r_4 = PN; \quad r_5 = EBIT; \\ r_6 = CB; \quad r_7 = CF$$

Din analiza graficului reprezentării întreprinderilor în funcție de prima componentă principală și pe baza scorurilor obținute de acestea, identificăm trei regiuni în funcție de care firmele se pot clasifica:

- $Z < -2,34$: zona sigură. Probabilitate de faliment foarte scăzută, practic neglijabilă.
- $-2,34 < Z < -0,102$: zona nesigură. Risc de faliment mediu.
- $Z > -0,102$: zona riscantă. Probabilitate de faliment corporativ crescută.

3. Analiza cluster

Analiza cluster ocupă un loc important în cadrul metodelor de recunoaștere necontrolată a formelor (cunoscute și sub denumirea de metode de învățare nesupervizată). Scopul analizei cluster este reprezentat de clasificarea datelor (a observațiilor sau a formelor) în structuri informaționale semnificative, relevante, cunoscute sub denumirea de clase, grupe sau clustere.

O noțiune esențială utilizată în analiza cluster este, prin urmare, *cluster*-ul. Un cluster se definește ca o submulțime a mulțimii inițiale de obiecte (observații) care au proprietatea că gradul de disimilaritate dintre oricare două obiecte ce aparțin cluster-ului este mai mic decât gradul de disimilaritate dintre oricare obiect aparținând cluster-ului și orice obiect ce nu aparține cluster-ului respectiv.

Este necesar a fi menționate o serie de precizări de ordin tehnic. În primul rând, pentru evaluarea distanței (disimilarității) dintre obiecte (societățile listate la categoria I) sau dintre clustere va fi utilizată distanța Manhattan. Distanța Manhattan, numită și distanță rectangulară, distanță “City-Block” sau normă de tip L1, se calculează ca sumă a valorilor absolute ale diferențelor coordonatelor celor două obiecte sau celor două variabile analizate.

În al doilea rând, vom folosi ca metodă de clasificare ierarhică prin agregare metoda lui Ward. Această metodă este considerată a fi cea mai eficientă și mai performantă dintre toți “algoritmii” de clasificare ierarhică deoarece este singura care tratează explicit problema omogenizării claselor, adică a minimizării variabilității intracluster: la fiecare pas, se comasează acele două clustere pentru care variabilitatea cluster-ului rezultat este cea mai mică din toate posibilitățile de comasare de clustere. O premisă importantă a metodei lui Ward o reprezintă descompunerea varianței totale în varianță intra-cluster și varianță inter-cluster, astfel:

$$\sigma_T^2 = \sigma_w^2 + \sigma_b^2,$$

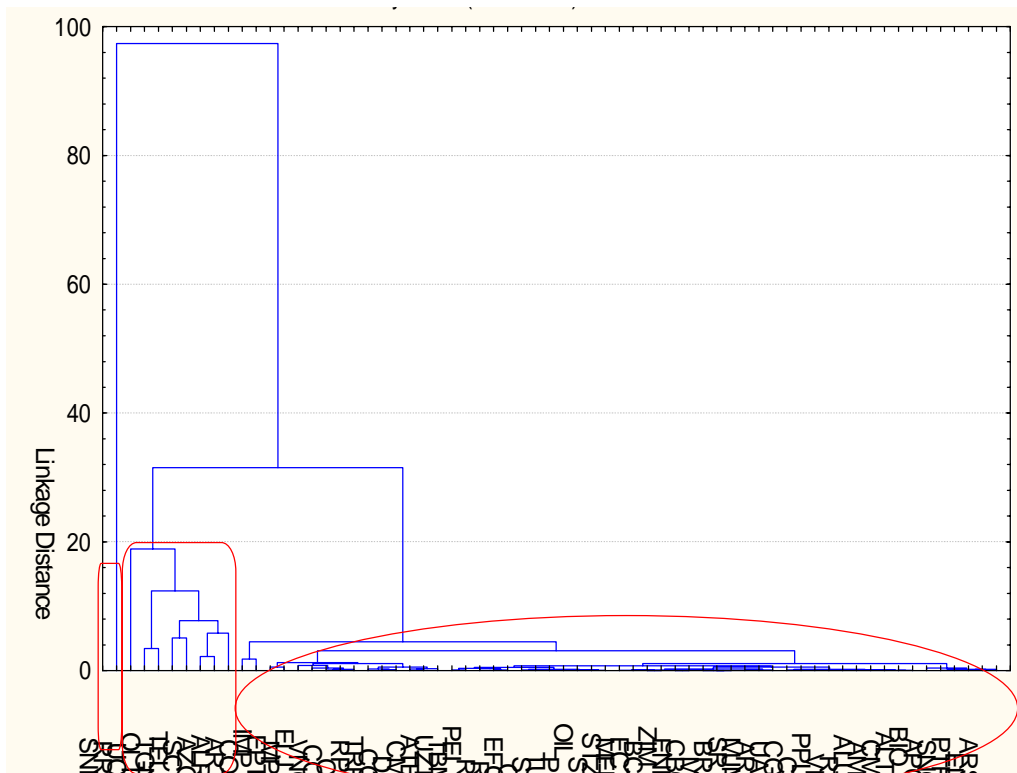
unde σ_w^2 și σ_b^2 sunt varianțele intra- și, respectiv, intercluster.

Metoda lui Ward se bazează pe următorul raționament: dacă la un moment oarecare al clasificării există p clase $\{\omega_1, \omega_2, \dots, \omega_p\}$, iar varianța totală intracluster este σ_w^2 , se vor comasa două clusteruri astfel încât variabilitatea intracluster rezultată (și care va fi obligatoriu mai mare decât cele două varianțe intracluster individuale, deoarece creșterea numărului de obiecte ce fac parte dintr-o clasă o face pe aceasta din urmă mai eterogenă, deci cu o variabilitate mai mare), pe care o vom nota $\tilde{\sigma}_w^2$, să fie cea mai mică posibilă, adică să reprezinte o soluție a următoarei probleme de optimizare:

$$\min(\tilde{\sigma}_w^2 - \sigma_w^2)$$

Argumentul funcției de optim este chiar distanța Ward.

Aplicarea acestei tehnici asupra setului nostru de date a generat (figura 3) rezultatele prezentate în dendrograma clasificării pentru cele 60 firme analizate (utilizând metoda de agregare a lui Ward).



Sursa: calcule proprii.

Figura 3. Dendrograma clasificării celor 60 de firme analizate

O caracteristică a tehnicilor de agregare ierarhică (inclusiv a metodei lui Ward) o reprezintă producerea mai multor soluții cluster, alegerea uneia dintre ele trebuind să se facă în funcție de obiectivele stabilite în procesul de analiză. Selectarea unei soluții cluster se realizează trasând o paralelă cu axa absciselor pentru diferite niveluri ale distanței de agregare. Astfel, considerând o distanță-limită de agregare relativ mică, vom obține trei cluster de societăți comerciale (marcate cu roșu pe graficul din figura 3), clusterelor fiind asemănătoare celor rezultate prin analiza componentelor principale. Astfel, primul cluster identificat include firma cu simbolul SNP. În cel de-al doilea cluster, pe lângă firmele incluse conform încadrării obținute prin analiza componentelor principale, mai apar firmele cu simbolurile UCM, OLT și SCD. În cel de-al treilea cluster sunt incluse toate celelalte firme din eșantion.

4. Analiză discriminantă

Analiza discriminantă presupune utilizarea unui ansamblu de metode, tehnici și algoritmi în scopul de a determina acele caracteristici ale unor obiecte care au cea mai mare relevanță din punct de vedere al recunoașterii apartenenței acestor obiecte la anumite clase prestabilite (avem de-a face, așadar, cu o tehnică de recunoaștere supervizată a formelor) și de a determina apartenența cea mai probabilă a obiectelor la diferite clase.

De reținut este că SNP a fost eliminat din analiză, întrucât discriminarea nu poate fi realizată cu clase conținând un singur element. Clasele au fost considerate cele obținute în analiza cluster.

Un prim rezultat al analizei discriminante este prezentat în tabelul 5:

Tabelul 5

Bonitatea modelului analizei discriminante și contribuția caracteristicilor individuale						
	Discriminant Function Analysis Summary (baza de date + indicatori2010.sta) No. of vars in model: 7; Grouping: Clasa* (2 grps) Wilks' Lambda: ,15952 approx. F (7,55)=41,399 p<0,0000					
N=63	Wilks' Lambda	Partial Lambda	F-remove (1,55)	p-level	Toler.	1-Toler. (R-Sqr.)
Activ total	0,174561	0,913815	5,18724	0,026663	0,043724	0,956276
CA	0,278340	0,573099	40,96945	0,000000	0,082668	0,917332
Datorii totale	0,175662	0,908089	5,56677	0,021882	0,008899	0,991101
PN	0,160368	0,994686	0,29381	0,589978	0,000951	0,999049
EBIT	0,159590	0,999539	0,02539	0,873991	0,001363	0,998637
CB	0,180197	0,885235	7,13040	0,009946	0,124843	0,875158
CF	0,313932	0,508124	53,24131	0,000000	0,352962	0,647038

Sursa: calcule proprii.

În primul rând, se observă că discriminarea de ansamblu este foarte puternică, după cum indică informațiile din antetul tabelului: statistica *Lambda* a lui Wilks are valoarea 0,15952 (cu cât valoarea statisticii este mai apropiată de zero, cu atât puterea de discriminare este mai mare; cu cât *lambda* este mai aproape de unu, cu atât puterea de discriminare se reduce), *p*-value este de sub 10^{-4} . Funcțiile de clasificare pentru fiecare dintre cele două grupuri sunt prezentate în tabelul 6:

Tabelul 6

Funcțiile de clasificare pe grupuri

Variable	Classification Functions; grouping: Clasa* (baza de date + indicatori2010.sta)	
	G_1:1 p=,12698	G_2:2 p=,87302
Activ total	26,8541	-36,2707
CA	-30,7169	12,2313
Datorii totale	49,8322	-0,7513
PN	65,1031	26,9284
EBIT	-47,9097	-35,8740
CB	-26,4144	16,5351
CF	14,3467	-4,9985
Constant	-13,8868	-2,4161

Sursa: calcule proprii.

Matricea clasificării arată corectitudinea discriminării în cele două clase rezultate în urma analizei cluster.

Tabelul 7

Matricea clasificării

Classification Matrix (baza de date + indicatori2010.sta)			
Rows: Observed classifications			
Columns: Predicted classifications			
Group	Percent Correct	G_1:1 p=,12698	G_2:2 p=,87302
G_1:1	87,5000	7	1
G_2:2	100,0000	0	55
Total	98,4127	7	56

Sursa: calcule proprii.

Matricea arată că toate companiile încadrate în clasa a doua în urma analizei cluster au fost alocate în aceeași clasă și în urma analizei discriminante; o singura companie, SCD, aparținând clasei 1, a fost alocată clasei 2, în urma discriminării. S-a obținut astfel un procent al corectitudinii clasificării de 98,41%.

5. Concluzii

Așa cum evidențiază Heffernan (2005), modelele de diagnosticare a riscului de faliment elaborate pe baza analizei discriminante (cum sunt modelul Altman și modelul Conan-Holder) pot induce ușor în eroare deoarece, în primul rând, ele se bazează pe date istorice. Chiar dacă la momentul elaborării lor aceste modele erau de o acuratețe rezonabilă, acuratețea lor se diminuează pe măsura trecerii timpului dacă nu se iau măsuri de actualizarea a variabilelor considerate și/sau de recalibrare a modelelor. Este plauzibil să considerăm că ratele financiare considerate se pot schimba în timp, chiar în funcție de piața pe care operează compania. Este necesar ca băncile să (re)testeze cu o frecvență suficient de ridicată modelele discriminante și să efectueze actualizări periodice ale modelelor de risc utilizate în practică (Heffernan, 2005, p. 161).

O problemă și mai dificilă ține de faptul că rezultatul impus de model este unul binar: fie debitorul este solvabil, fie nu. În practică, există mai multe situații posibile, ca de exemplu întârzieri în plata ratelor lunare, neplata acestora, neplata de comisioane sau dobânzi penalizatoare ș.a.. De cele mai multe ori debitorul înștiințează banca cu privire la dificultățile sale de ordin financiar și termenii creditului sunt renegociați, ori modelele de analiză discriminantă utilizate nu pot include stările de solvabilitate, insolvabilitate și restructurare simultan.

Tehnicile propuse de analiza multidimensională a datelor se dovedesc a fi extrem de utile atunci când cercetarea se realizează asupra unei mulțimi de obiecte caracterizate de un număr mare de variabile, ceea ce face ca studierea dependențelor cauzale și clasificarea obiectelor să fie greu de realizat. Acesta este și cazul nostru, obiectul analizei fiind constituit de companiile listate la Bursă de Valori București, pentru care am considerat un număr de șapte caracteristici reprezentative (activul bilanțier total, cifra de afaceri netă, rezultatul din exploatare – EBIT, profitul net, fluxul net de trezorerie din activitatea de exploatare, cuantumul total al datoriilor și capitalizarea bursieră medie).

După cum am văzut, cele șapte caracteristici individuale sunt caracterizate de niveluri ridicate ale volatilității, însă sunt și puternic corelate între ele, ceea ce înseamnă că, pe lângă conținutul informațional intrinsec al fiecărei variabile, există și o cantitate importantă de informație disipată în legăturile neobservabile direct dintre variabile. În acest context, analiza componentelor principale este un instrument util de studiu, deoarece reușește atât sintetizarea informațională, cât și eliminarea redundanțelor informaționale.

Aplicând metoda componentelor principale pe setul nostru de date, am obținut o componentă ce sintetizează aproximativ 86,10% din informația

conținută în spațiul cauzal inițial. Astfel, trecerea de la șapte variabile la numai una s-a realizat în condițiile unei pierderi informaționale minime, de circa 23%. Prima componentă principală recuperează 86,% din informația din spațiul original și este puternic corelată negativ cu toți indicatori considerați, furnizând astfel informații cu privire la volumul de afaceri, rentabilitatea companiilor (atât la nivelul exploatării, cât și la nivelul activității de ansamblu) despre valoarea de piață a acțiunilor emise de firme și se constituie într-o expresie a capacității companiilor de a genera excedente de trezorerie din activitatea de exploatare. În urma aplicării analizei componentelor principale, ținând cont că am identificat o singură componentă principală ce sintetizează peste 86% din informația generată de indicatorii inițiali, identificăm un model de tip scoring care are următoarea formă:

$$Z = -0.402xr_1 - 0.377xr_2 - 0.344xr_3 - 0.356xr_4 - 0.38xr_5 - 0.405xr_6 - 0.375xr_7$$

Din analiza graficului reprezentării întreprinderilor în funcție de prima componentă principală și pe baza scorurilor obținute de acestea, identificăm trei regiuni în funcție de care firmele se pot clasifica:

- $Z < -2,34$: zona sigură. Probabilitate de faliment foarte scăzută, practic neglijabilă.
- $-2,34 < Z < -0,102$: zona nesigură. Risc de faliment mediu.
- $Z > -0,102$: zona riscantă. Probabilitate de faliment corporativ crescută.

În consecință, constatăm că atât aplicarea analizei cluster, cât și a analizei discriminante relevă faptul că cele trei zone identificate ale funcției scoring de grupare a întreprinderilor sunt corecte.

Mulțumiri

În cadrul lucrării sunt diseminate o parte din rezultatele cercetării realizate prin Proiectul de cercetare exploratorie, PN II ID-PCE-2008-2, nr. 1764, CNCSIS, finanțat de la Bugetul de stat prin Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării (UEFISCDI).

Bibliografie

- Altman, E.I. (2002). *Managing the Commercial Lending Process*
- Armeanu, D. (2005). *Evaluarea riscului activității financiare cu aplicații pe economia românească*, teză de doctorat
- Damodaran, Aswath (2005). *Applied corporate finance*, 2nd edition, Wiley
- Everitt, B., Landau, S., Leese, M. (2006). *Cluster Analysis*, Arnold, Fourth edition
- Ghahramani, Z., *Unsupervised learning*,
<http://www.inf.ed.ac.uk/teaching/courses/pmr/docs/ul.pdf>
- Gujarati (2005). *Basic econometrics*, McGraw-Hill
- Heffernan, S. (2005). *Modern Banking*, Wiley
- Helfert, E. (2001). *Financial analysis tools and techniques – a guide for managers*, The McGraw-Hill Companies
- Moore, A. „K-Means and Hierarchical Clustering”,
<http://www.autonlab.org/tutorials/kmeans.html>
- Nicolae, M., „Utilizarea metodei componentelor principale la analiza „stării” de sănătate a societăților bancare”, *Revista Oeconomica*, 2004, București
- Ruxanda, Gh. (2005). *Econometrie II*, suport de curs masterat Management Financiar și Piețe de Capital”, București
- Ruxanda, Gh. (2007). *Analiza multidimensională a datelor*, Master Baze de Date – Suport pentru Afaceri
- Simar, L. (2004). *Applied Multivariate Statistical Analysis*, Springer
- Spircu, L. (2004). *Tehnici de analiza datelor*, Academia de Studii Economice București
- Spircu, L. (2006). *Analiza Datelor: Aplicații Economice*, Editura ASE, București
- Stancu, I. (2007). *Finanțe*, Editura Economică, București
- Vintilă, G., Toroapă, M.G., „Building a Scoring Model for Bankruptcy Risk Prediction on Multiple Discriminant Analysis”, The International Conference *Present issues of global economy*, 8th Edition, April 16th-17th, 2011, Annals of the “Ovidius” University, Economic Sciences Series Volume XI, Issue 1 /2011
- Vintilă, G. (2005). *Gestiunea financiară a întreprinderii*, Editura Didactică și Pedagogică, București
- “Cluster Analysis” -<http://www.statsoft.com/textbook/stathome.html?stcluan.html&l>
- “Cluster Analysis – Statnotes, from North Carolina State University”,
<http://www2.chass.ncsu.edu/garson/PA765/cluster.htm>
- “Notes on Cluster Analysis”, <http://www.uic.edu/classes/idsc/ids472/clustering.htm>
- “A Tutorial on Clustering Algorithms. K-Means Clustering”,
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html