JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
GERMAN INTERNATIONAL JOURNALS COMPANY
ISSN: 2751-4390
IMPACT FACTOR (RESEARCH BIB) : 9,08. Academic research index

BUSINESS INSOLVENCY PREDICTION MODELS: A COMPARATIVE ANALYSIS OF MULTIVARIATE DISCRIMINANT ANALYSIS AND LOGISTIC REGRESSION

Olimjonova Munisa Nuriddin kizi
Tashkent state university of economics
Accounting faculty
olimjonovam50@gmail.com

Abstract: This study looks at predictive approaches that use logistic regression and multivariate discriminant analysis (MDA) to evaluate a company's insolvency. The Z-Score model developed by Edward Altman and William Beaver is the main focus of this historical review of bankruptcy prediction techniques. The use of a binary logistic regression model to assess the importance of individual financial ratios in predicting a company's bankruptcy risk is also covered in the paper. According to the results, MDA and logistic regression provide strong frameworks for comprehending the financial health of businesses, with applications in risk management and decision-making. To preserve accuracy and relevance in a changing economic environment, the study does stress the necessity of constant validation and adaptation.

Key word: Insolvency prediction models, multivariate discriminant analysis, logistic regression, risk mitigation, univariate analysis, bankruptcy, ratios, Altman.

Introduction.

Over the years, an interest arose among various finance professionals to analyze the risk of business (corporate) insolvency. Therefore, many experts developed diverse methods to predict bankruptcy of businesses.

The predictive capacity was carried out at two levels (Jiménez Cardoso et al., 2002, pp. 257-260):

a) Univariate analysis, based on the use of only one financial ratio for the elaboration of business bankruptcy prediction. The pioneering author resulting from this study was William Beaver (1966), who considered that through the use of *cash-flow/total debts* and *net profit/total assets ratios*, the bankruptcy situation of the company could be identified, noting that the former of these provides greater predictive value. A company would be in "bankruptcy" when it presents negative cash balances or the non-payment of obligations or dividends.

b) Multivariate analysis, corresponding to the integration of more than one independent variable (several financial ratios) with the objective of classifying companies according to their situation.

These comprehensive models aim to provide a more robust assessment of a company's financial stability by considering various financial dimensions simultaneously, thus offering a more holistic view of potential risks. They move beyond single indicators to capture the complex interplay of factors contributing to insolvency.

Literature review.

In this case, the precursor was Edward Altman (1968), who based his linear model on a multivariate approach, that is, he considered the simultaneous influence of two or more variables through a Multiple Discriminant Analysis (MDA) on a series of financial variables.

According to Jiménez Cardoso et al. (2002), "the classification of a company's situation and future is usually not done based on the value presented by a single financial ratio, which is why, in practice, it makes no sense to speak of a strictly univariate analysis" (p. 257). For this reason, this work will focus primarily on the multivariate analysis used in Edward Altman's model.

The choice has also been due to the fact that, as will be seen below, these ratios are calculated through any company's accounting data, which allows for simple calculation because cases will not be disregarded due to lack of financial information. It has also been one of the most applied prediction models both in finance and in accounting research, which is why it has been considered very useful for the study.

Likewise, the choice has also been made for the logistic regression or Logit model, which allows obtaining the significance of the ratios, used with the aim of identifying which of them, is the most relevant and requires more attention compared to the others. In addition, it will be observed if it has the same prediction capacity as the previous model.

In 1968, Edward I. Altman, a professor at New York University, wanted to develop a predictive method based on the Multiple Discriminant Analysis (MDA) statistical technique, given that it had previously been applied favorably in financial problems, including consumer credit evaluation and investment classification.

According to Marchese et al. (2006), "multivariate analysis is a set of statistical and mathematical methods designed to describe and interpret data that come from the observation of various statistical variables studied jointly" (p. 3).

Logistic regression analysis is a nonlinear econometric model used to predict the outcome of categorical variables (variables that can contain a limited number of classes) based on independent variables. Generally speaking, the probabilities obtained in the function should be between 0 and 1. There are two modes of logistic regression analysis (López-Roldán, P., and Fachelli, S., 2015):

Methodology.

This study employs a quantitative methodology to predict business insolvency, primarily utilizing two distinct statistical approaches: Multivariate Discriminant Analysis (MDA) and Logistic Regression. MDA is used to classify companies into solvent and insolvent groups by finding a linear combination of financial ratios that best discriminates between them. Complementing this, a binary Logistic Regression model is applied to estimate the probability of insolvency, allowing for the assessment of individual financial ratios' significance and their impact on a company's likelihood of bankruptcy, providing a comprehensive framework for risk assessment.

Results and discussion.

1) Altman's Z-Score Model

As mentioned, Marchese et al. define multivariate analysis as a collection of mathematical and statistical techniques intended to characterize and interpret data derived from the observation of many statistical variables under joint study. Therefore, this study allows examining the combination of one or two predefined groups. By virtue of this, "discriminant analysis is a multivariate analysis technique whose objective is to find the linear combination of independent variables that best differentiates the groups".

Therefore, Altman first selected two groups, which are made up of companies that are in bankruptcy (insolvent) and those that are not in bankruptcy (solvent). The prediction of each of them is established in the dependent variable of function (1). Once the grouping has been carried out, the financial data belonging to each of the companies that are part of the study are collected.

Subsequently, following Marchese et al. (2006), thanks to the MDA model, it allows determining a set of coefficients that best discriminate the set of groups, which are weighted and a linear equation is constructed.

Finally, he made use of a one-dimensional discriminatory function, which is given by the following equation:

$$Z = V_1 X_1 + V_2 X_2 + \ldots + V_n X_n \qquad \textbf{(1)}$$

where, as stated in the book by Edward I. Altman (1968), Z is the discriminant variable used to classify the company (global indicator); $V_1, \ldots, V_p$ are the weighting coefficients of the discriminatory variables; and $X_1, \ldots, X_p$ represent the independent variables corresponding to the company's financial ratios.

Once the MDA had been studied, the first questions that arose for the development of its model were the following (Altman, E. I., 1968):

1. What ratios are the most important for predicting such insolvency?

2. What weights should be assigned to each assigned proportion?

3. How should the obtained result be objectively established, once the previous questions have been resolved?

For his study, he based it on the classification of a total of 66 US manufacturing companies listed on the stock exchange, of which 33 belonged to insolvent companies, and the rest were healthy companies.

From the obtained sample, Altman selected a total of 22 financial ratios, but finally, he relied on the popularity of the literature and the potential for the study, choosing five of them, of which four variables correspond to the balance sheet and one of them to the income statement. These variables could be classified into five distinct categories: liquidity, profitability, leverage, solvency, and activity (Altman, E. I., 1968). And, for each of them, he established five different coefficients, which he combined using a linear discriminant function.

After the companies had been evaluated with their respective coefficients and ratios, the results obtained in said discriminant equations (denominated "Z") would be expressed through a series of areas, which are shown in Figure 6.

Table 1. Classification of Z-Score Areas.[1]

Danger zone	"Grey" zone	Secure zone
Insolvent	Uncertain	Solvent
$Z < A$	$A < Z < B$	$Z > B$

In a broader sense, each of these areas is defined in the following way:

- **Danger Zone:** The probability of presenting insolvency problems in the future is very high, so it can be considered that there will be an imminent bankruptcy.

[1] Author's elaboration, based on Altman, E. I. (1968)

- **"Grey" Zone:** There is a high probability that companies may enter into an insolvency situation in the next two years.

- **Secure Zone:** The company does not present insolvency problems in the future.

As represented in Table 2, Altman established five indicators in 1968 for those companies listed on the stock exchange, thus obtaining the equation called Z-Score. Subsequently, in 1983, Altman re-estimated the original model by changing both the parameters and the ratios, to adapt it to those companies not listed on the stock exchange, as the previously used financial ratio was not compatible with these societies. He called this new discriminant equation Z_1-Score.

It is worth mentioning that Altman also considered the so-called Z_2-Score based on the Z_1-Score, but excluding the ratio X_5 since it could distort the analysis, being the variable of sales a variable of easy manipulation by companies.

Table 2. Ratios employed by Altman in the Z-score model[2]

X_n	*Z-Score (1968)*	*Z_1-Score (1983)*
X_1	Working Capital / Total Assets	Working Capital / Total Assets
X_2	Retained Earnings / Total Assets	Retained Earnings / Total Assets
X_3	EBIT / Total Assets	EBIT / Total Assets
X_4	Market Value of Equity / Total Liabilities	Book Value of Equity / Total Liabilities
X_5	Sales / Total Assets	Sales / Total Assets

It is worth mentioning that Altman also considered the so-called Z_2-Score based on the Z_1-Score, but excluding the X_5 ratio since it could distort the analysis, with the sales item being a variable that is easily manipulated by companies.

Next, we explain each of these ratios and the reasons for their use by Edward I. Altman (1968):

X_1: This ratio measures a company's liquidity through the relationship between working capital and total assets. The numerator, working capital, is defined as the difference between current assets and current liabilities. The denominator contains total assets and allows for comparison with other elements of the balance sheet or income statement (Marchese et al., 2016). Therefore, it can be said that the greater the working capital relative to total assets, the higher the level of liquidity. Altman, of the three different ratios that measure liquidity, considered X_1 to be one of the most essential ratios for its calculation.

X_2: This ratio shows the company's cumulative profitability, since its definition uses the year-end profit, thereby determining whether the company is profitable. As Altman noted in his book, he considered young companies to be more discriminatory compared to others, because they have recently entered the market and have not had enough time to accumulate profits. Therefore, these companies have a high probability of facing bankruptcy. However, older companies may also incur losses for several consecutive years, which will result in a negative result in this ratio.

X_3: This is a measure of a company's profitability. Altman considered it an appropriate ratio for studying corporate failure. Therefore, companies with negative earnings before interest and taxes will have a lower ratio and a greater likelihood of bankruptcy than those with profits.

[2] Author's elaboration. Source: Altman, E. I., & Hotchkiss, E. (2006)

X_4: This is the variable that relates equity to total debt. This ratio refers to the company's financial structure; in other words, it measures solvency. In this case, the methodology applied to listed companies uses the market value of capital, which means it takes into account the number of shares the company holds and the market price. However, for unlisted companies, the book value of capital is chosen, which gives the total value of equity.

X_5: Represents the company's ability to generate sales from its assets; in other words, it measures the company's efficiency. In these cases, Altman estimated the following models with their corresponding classifications according to the area to which the results obtained in the model pertain, which are reflected in Table 3 below.

Table 3. Altman, EI, and Hotchkiss, E. (2006)[3]

Manufacturing Companies Listed on the Stock Exchange	*Area*
$Z=1.2X_1 \quad +1.4X_2 \quad +3.3X_3 \quad +0.6X_4 \quad +9.99X_5$ **(2)**	Danger Zone: Z<1.81 Gray Zone: 1.81<Z<2.99 Safe Zone: Z>2.99
Manufacturing Companies NOT Listed on the Stock Exchange	*Area*
$Z_1 \quad =0.717X_1 \quad +0.847X_2 \quad +3.107X_3 \quad +0.42X_4 \quad +0.998X_5$ **(3)**	Danger Zone: Z<1.23 Gray Zone: 1.23<Z<2.9 Safe Zone: Z>2.9
$Z_2 \quad =6.56X_1 \quad +3.26X_2 \quad +6.72X_3 \quad +1.05X_4$ **(4)**	Danger Zone: Z<1.1 Gray Zone: 1.1<Z<2.6 Safe Zone: Z>2.6

Finally, it is worth mentioning that, according to the studies conducted, reliability was demonstrated between 80 and 90 percent. Edward I. Altman (1968) added that "this method was not probabilistic, but rather descriptive-comparative." In other words, this model should be used as a warning tool for a company's potential insolvency, as it identifies its orientation toward one group or another.

Subsequently, thanks to the development of the Altman model, a similar version was created, but this time based primarily on the detection of insolvency in Spanish SMEs3, which was carried out by Amat et al. in 2016. To do so, they again applied linear discriminant analysis to obtain a combination of ratios that better discriminates, which is why it was necessary to readjust the coefficients and variables used by Altman.

In his analysis, he studied a group of 2,000 companies that received short-term loans during 2008, of which 144 presented repayment difficulties at the time in which the loans were granted (Amat et al., 2016). Finally, according to the results obtained for the development of the new methodology, through the following Table 4 we can observe equation (5) obtained by Amat et al., and the interpretation given to the results.

Table 4. Equation of the Amat et al. model and its meaning[4]

Equation	*Area*
$Z^* =-3.9+1.28X_1 \quad +6.1X_2 \quad +6.5X_3 \quad +4.8X_4$ **(5)**	Financial Problems: Z<0 Financial Health: Z>0

where the ratios used in the equation are reflected in the following Table 5.

[3] Author's elaboration. Source: Altman, EI, and Hotchkiss, E. (2006)
[4] Source: Amat et al. (2016)

Table 5. Ratios used by Amat et al.

X_n	*Amat et al. (2016)*
X_1	Current Assets / Current Liabilities
X_2	Own Funds / Total Liabilities
X_3	Net Profit / Total Assets
X_4	Net Profit / Own Funds

X_1: Similarly to the numerator of the Altman X1 ratio, it corresponds to the liquidity ratio. This refers to the working capital, so it analyzes the short-term financial stability of the company. According to Oriol Amat (2008), the lower this ratio, the more likely the company is to default. However, if it is too high, it would obtain idle current assets, which is the same as losing profitability.

X_2: This ratio is very similar to the debt ratio and divides equity by debt. It establishes the company's financial independence vis-à-vis third parties. That is, it analyzes whether the company is capable of meeting all its obligations and debts. The more autonomous the company, the more its capital structure will be comprised of its own resources, so the company will be less dependent on its creditors.

X_3: This is a company profitability indicator that examines net profits relative to total assets. In other words, it measures whether the company is using its assets efficiently.

X_4: Determines the financial profitability, or Return on Equity (ROE), which indicates the net return obtained from the company's equity after deducting interest and taxes. This ratio also indicates the return earned by shareholders, based on the value created by the company. The higher this ratio, the better the company's performance, as its equity will be more profitable.

2) Logistic regression model

To examine the relationship between the dependent binary outcome and the selected explanatory variables, a logistic regression model was employed. Logistic regression is a widely used statistical method for modeling binary response variables, where the outcome takes on values of 0 or 1. This model estimates the probability that a given event occurs as a function of one or more independent variables, using the logit transformation of the probability.

First, there is binary logistic regression, where the dependent variable is dummy or dichotomous, that is, it can only take two values (probability of an event occurring or not).

Second, multinomial logistic regression, which allows for qualitative explanations in a more general case. Depending on the categorical variable selected, a distinction can be made between nominal polytomous variables (e.g., the choice of a product brand) and ordinal polytomous variables (e.g., the level of satisfaction in a restaurant).

As will be seen later, this paper will focus solely on the study of a binary logistic regression, in which the dichotomous dependent variable will represent the value 1 if the company is filing for bankruptcy and 0 if it is currently in bankruptcy. Therefore, as previously discussed, the dichotomous dependent variable can be explained by one or more explanatory or independent variables, where the classification of each case can be predicted based on the group to which they correspond. This result is therefore possible thanks to the weights or coefficients associated with each of these variables in order to determine their significance in the study.

All necessary formulas of logistic regression model are given below in Table 5. The specification of the logit or binary logistic regression model is as follows **(formula 6)**, where Y is the dichotomous dependent variable, indicating the occurrence of the event (Y=1) or its non-occurrence (Y=0); β_1, β_2, …, β_k are the population coefficients or parameters; and finally X_1, X_2, …, Xk are the explanatory variables.

Table 5. Logistic Regression Model Formulas.

Formula Number	Mathematical Expression	Description
(6)	$P_i(Y=1 \mid X_1, X_2, ..., X_k) = \frac{1}{1+e^{-x_i'\beta}}$	Logistic regression probability function
(7)	$P_i(Y=1) = \frac{1}{1+e^{-x_i'\beta}} = P_i$	Probability of success (Y=1)
(8)	$P_i(Y=0) = 1 - \left(\frac{1}{1+e^{-x_i'\beta}}\right) = 1 - P_i$	Probability of failure (Y=0)
(9)	$\ln\left(\frac{P_i}{1-P_i}\right) = x_i'\beta$	Log-odds (logit) transformation
(10)	$\frac{\partial P_i}{\partial x_{ik}} = \frac{\exp(x_i'\beta)}{[1+\exp(x_i'\beta)]^2}\beta_k$	Marginal effect of x_{ik} on P_i
(11)	$\text{odds ratio} = \frac{P_i}{1-P_i} = e^{x_i'\beta}$	Odds ratio interpretation
(12)	$L = \prod_{y_i=0}(1-P_i) \cdot \prod_{y_i=1} P_i$	Likelihood function
(13)	$\ln L = \sum_{i=1}^{n} y_i \ln P_i + \sum_{i=1}^{n}(1-y_i)\ln[1-P_i]$	Log-likelihood function
(14)	$S(\beta) = \frac{\partial \ln L}{\partial \beta} = 0$	Score function (for MLE estimation)
(15)	$I(\beta) = \mathbb{E}\left(\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}\right)$	Information matrix
(16)	$\beta_1 = \beta_0 + [I(\beta_0)]^{-1}S(\beta_0)$	Newton-Raphson update formula

In the simple binary logistic regression model, as the dummy variable collects the probabilities between 0 and 1, the probability of the event occurring (Y=1) is called "P_i" and the probability of it not occurring (Y=0) is 1 minus the previous probability **(formula 7)**.

This logistic function is represented by a logistic or sigmoidal curvature, which has an S shape and is reflected as a growth curve since the Y axis increases monotonically with respect to the X axis **(formula 8)**.

As a consequence, as Pedro López-Roldán and Sandra Fachelli (2015) comment, "the interpretation associated with each of the coefficients of logistic regression differs from that of linear regression, because the coefficient is not the measure of how much "y" will vary in response to a variation in one unit of "x", but rather the change produced by a variation of one unit of "x" in the natural logarithm "ln" of the quotient of probabilities of the two events, the so-called logit transformation" (p. 18) **(formulas 9, 10)**

The population parameters (βi) are determined through the relationship between the probability of the event occurring with the probability of it not occurring, called odds ratio, and whose expression is given as (**formula 11).**

Therefore, the interpretation of the signs obtained in the coefficients is similar to that of linear models: if the beta is positive, the probability of the event occurring increases, and if it is negative, it decreases. If the coefficient is equal to zero, there will be no change in the odds ratio.

When evaluating the marginal effect of each of the explanatory variables on the probability of the event occurring, it will be necessary to do so for certain values of the explanatory variables in **(formula 11).**

Following Maddala, G. (1983) and assuming that the observations are independent, the likelihood function that captures the probability of obtaining the observed sample is the product of the marginal probabilities of the binary variables (yi). Therefore, this function is as **(formula 12)**

The population parameters (βi) are estimated using the maximum likelihood method, which consists of obtaining those parameter values that, given the sample, maximize ln L **(formula 13)**. The maximum likelihood estimator is obtained by solving the system of equations **(formula 14)**

This system of equations **(formula 14)** for deriving the estimators is a nonlinear system that must be solved by numerical optimization methods, i.e., an iterative procedure must be used to solve them. The information matrix is the second derivative of the function **(formula 15).**

The iterative procedure starts with an initial value of , called 0, and in each iteration the values ( ) and ( ) are calculated to obtain a new estimate of the parameters as follows **(formula 16).**

In this equation ( ) is positive definite at each iteration step s. Therefore, the iterative procedure will converge to a maximum of the function regardless of the starting value. If the final convergent estimates are denoted by ˆ, then the asymptotic covariance matrix (ˆ) is estimated by [(ˆ)]−1 .

The maximum likelihood estimator is consistent and normally distributed in large samples. Therefore, individual significance hypothesis tests are performed using the asymptotically distributed t statistic N (0,1) under the true null hypothesis.

Conclusion.

This study has carefully examined models for predicting business insolvency, highlighting the analytical strength of logistic regression and multivariate discriminant analysis (MDA). This work demonstrates their significant effectiveness in identifying corporate financial health and predicting distress by carefully examining their theoretical underpinnings and real-world applications, including influential works like Edward Altman's framework and its modifications. Stakeholders benefit greatly from the analytical insights obtained, which offer tools for risk reduction, well-informed lending choices, and proactive strategic reactions. Although these models offer strong analytical frameworks that have been shown to be reliable, their long-term usefulness requires constant improvement through validation against new data and safe adjustment to changing business complexity, regulatory changes, and economic landscapes. uch sustained scholarly engagement is crucial for enhancing their predictive accuracy, analytical relevance, and long-term efficacy in a dynamic global economy.

References:

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of

corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.
2. Altman, E. I., & Hotchkiss, E. (2010). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (Vol. 289). John Wiley & Sons.
3. Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of international financial management & accounting*, *28*(2), 131-171.
4. Amat, O., Manini, R., & Renart, M. A. (2017). Credit concession through credit scoring: Analysis and application proposal. *Intangible Capital*, *13*(1), 51-70.
5. Amat, O. (2008). *Análisis de estados financieros: 8a Edición*. Grupo Planeta (GBS).
6. Burkhanov, A. U., Kurbonbekova, M. T., Usmonov, B., & Nizomiddinov, J. Z. (2024). Assessment of the Financial Sustainability of Enterprises: The Case of Uzbekistan. In *Development of International Entrepreneurship Based on Corporate Accounting and Reporting According to IFRS* (Vol. 33, pp. 215-223). Emerald Publishing Limited.
7. Vivar-Arrieta, M. A., Haro-Altamirano, J. P., Carrillo Barahona, W. E., López Sampedro, S. E., Usmanovich, B. A., Usmonov, B., & Ulugbek Kizi, M. S. (2023). Multicriteria evaluation of ancestral family agricultural systems, Chimborazo Province, Ecuador. *Caspian Journal of Environmental Sciences*, *21*(5), 1123-1134.
8. Usmonov, B. (2022, December). The impact of the financial ratios on the financial performance. A case of Chevron Corporation (CVX). In *International Conference on Next Generation Wired/Wireless Networking* (pp. 333-344). Cham: Springer Nature Switzerland.
9. Código de Comercio. (1885). Libro IV. De las suspensiones de pagos, de las quiebras y de las prescripciones.
10. Fitó-Bertran, À., Plana-Erta, D., & Llobet Dalmases, J. (2018). Usefulness of Z scoring models in the early detection of financial problems in bankrupt Spanish companies.
11. López-Roldán, P., & Fachelli, S. (2015, February). *Metodología de la investigación social cuantitativa*.
12. Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics* (No. 3). Cambridge university press.
13. Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). Introducción a la Econometría.
14. Marchese, M. (2014). Entrepreneurial Activities in Europe-Finance for Inclusive Entrepreneurship.
15. Huo, Y., Chan, L. H., & Miller, D. (2024). Bankruptcy prediction for restaurant firms: A comparative analysis of multiple discriminant analysis and logistic regression. *Journal of Risk and Financial Management*, *17*(9), 399.
16. Michalkova, L., & Ponisciakova, O. (2025). Bankruptcy Prediction, Financial Distress and Corporate Life Cycle: Case Study of Central European Enterprises. *Administrative Sciences*, *15*(2), 63.
17. Navarro-Galera, A., Lara-Rubio, J., Novoa-Hernández, P., & Cruz Corona, C. A. (2025). Using Decision Trees to Predict Insolvency in Spanish SMEs: Is Early Warning Possible?. *Computational Economics*, *65*(1), 91-116.
18. Matenda, F. R., Sibanda, M., Chikodza, E., & Gumbo, V. (2021). Bankruptcy prediction for private firms in developing economies: a scoping review and guidance for future research. *Management Review Quarterly*, 1-40.