A Review of Auditors’ GCOs, Statistical Prediction Models and Artificial Intelligence Technology

Bahaaeddin Alareeni
Middle East Technical University-Northern Cyprus Campus,
Güzelyurt, Mersin 10, Turkey
Bahaaedu@hotmail.com

Received October, 2018; December, 2018

Abstract. The main aim of this study is to give an overview of literature in accounting and finance regarding the performance of Auditors’ GCOs, Statistical Failure Prediction Models (SFPMs), and Artificial Intelligence Technology (AIT). The study reviews the accounting and finance literature regarding SFPMs and presents the most important types of SFPMs and AIT that have been developed to evaluate a company’s financial position from 1968 to date. The study focuses on studies that compare the relative performance of auditors’ GCOs with SFPMs and AIT. Our findings illustrated that SFPMs and AIT are better in predicting companies’ failure than auditors’ GCOs. We found that the prediction power of SFPMs is in many instances very high. Their accuracy differed from one model to another, depending on several factors such as industry, time period, and economic environment. The most commonly used and accurate models are the Altman models, logit models, and neural networks models, although overall the NNs models produce better results. We found that SFPMs and AIT can be very useful to users when assessing a company’s future position. Incorporating the use of SFPMs and AIT in the audit program can provide further evidence that the auditors exerted professional competence and due care. This study provides a comprehensive overview of research on Auditors’ GCOs, SFPMs, and AIT. The study provides a clear picture of the best tools used in failure/bankruptcy prediction in the last decades. Thus, it is an aid to future research in the area.

Keywords: Auditors’ GCOs; Bankruptcy Prediction Models; Decision Trees; Evolutionary Approaches; Rough Sets; Neural Networks; Artificial Intelligence Techniques

Type: Research paper

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: 10.51325/ijbeg.v2i1.30

1 Introduction

The bankrupt companies cause vast losses to countries and individuals. Some of these losses can be avoided or reduced when interested parties receive early warning signs about imminent danger. For example, auditors should issue a going-concern opinion (GCO) when they have doubts about the future prospects of a company they audit.

However, over the past few decades there have been a number of cases in which external auditors failed to inform about imminent business’ failure. This has led to criticism of auditors worldwide. Therefore, there is an obvious need for reliable instruments that can assist stakeholders in assessing a company’s future position status. In respect of the work
of auditors, auditing standards that are in use in most countries now include well-defined procedures and indicators that auditors should take into consideration to evaluate a company's ability to continue its activities. In spite of that, companies regularly go bankrupt after audit firms issue clean audit opinions. Although auditors' duties and responsibilities are not solely and even not primarily linked to bankruptcy prediction, it is clear that their failure to issue GCOs in cases that result in losses to investors and other stakeholders constitutes a threat to user confidence in auditors' opinions.

Over the past few decades, several statistical failure prediction models (SFPMs) and financial ratios have been developed to assist users in predicting a company’s future position status. Research shows that SFPMs can indeed anticipate a company’s failure within a few years before the failure, with high accuracy rates (e.g., Zmijewski, 1984; Zavgren, 1985; Altman, 1993; Pompe and Bildereek, 2005; Li and Sun, 2009; Alareeni and Branson, 2011; 2014). The accuracy rates differ substantially from one model to another (e.g. multiple discriminant analysis, logit, probit, decision trees, etc.). Additionally, accuracy rates fluctuate considerably when using samples from different time periods, from different industries, or in economic environments.

Multiple discriminant analysis (MDA) is one of the most common methods used in developing SFPMs. Among the most popular models based on MDA are the Altman models (Altman Z-score 1968; Altman Z''-score 1993). The Altman models have a high degree of accuracy in predicting business’ failure in the US and in some other countries (Altman et al., 1995). Other studies have used logit (logistic) regression analysis (LRA) to develop failure prediction models (e.g, Ohlson, 1980; Koh, 1991; Alareeni and Aljuaidi, 2014; Alareeni, 2017). LRA requires less restrictive assumptions than MDA. For example, LRA does not need the normality assumption on the predictors and it can calculate the failure’s probability directly.

A number of prior studies compare the performance of auditors’ GCOs with SFPMs and AIT. The research results are varied. Some studies (e.g. Sun, 2007) have shown that SFPMs perform better that auditors’ GCOs, while others (e.g. Hopwood et al., 1994) show that SFPMs and auditors’ GCOs are equivalent in their prediction. In some cases, the SFPMs succeed in classifying companies correctly with regard to their future failure or non-failure position while auditors’ GCOs are wrong. In other cases, it is the opposite. Also, sometimes these approaches fail/succeed to give investors and other stakeholders an accurate sign of a company’s future position. This obviously can be confusing for the users and thus cause losses. And if this problem remains unexplained, users could start losing confidence in both auditors’ GCOs and SFPMs.

These conflicts and convergences between auditors’ GCOs and SFPMs are worth further investigation. Therefore, this study reviews the accounting literature regarding auditors’ GCOs. We focus on the indicators used in assessing a company’s future position status and on the factors affecting the auditor’s GCO decision. The study also reviews the accounting and finance literature regarding SFPMs and presents the most important types of SFPMs that have been developed to evaluate a company’s financial position from 1968...
to date. The literature review also includes studies that compare auditors’ GCOs and SFPMs in predicting bankruptcy. In addition, the study shows how valid and accurate existing SFPMs compared to auditors’ GCOs.

The study includes four other sections. In the next section, we present the accounting literature regarding auditors’ GCOs. We consider the indicators auditors use in evaluating companies’ going-concern status and the factors affecting their GCO decision. In section three, we conduct a comprehensive literature review concerning SFPMs. In section four, we review the studies that compare auditors’ GCOs and SFPMs. In the final section, we present our conclusions.

2. Overview of SFPMs and AIT

In this section, we present the most important types of SFPMs and AIT that have been developed to evaluate a company’s financial position from 1968 to date. They are grouped into the following families:

2.1 Statistical failure prediction techniques (SFPMs),

2.2 Artificial Intelligence Technology (AIT) which includes the following techniques:

2.2.1 decision trees,
2.2.2 evolutionary approaches,
2.2.3 rough sets,
2.2.4 neural networks, and
2.2.5 case-based reasoning.

In our review, we highlight the relative performance of SFPMs and AIT in terms of prediction rates. We also provide information about the source of the sample size, the methodology and financial ratios (predictors) used and the countries and time period considered.

2.1 Statistical techniques

We review the most common statistical techniques, that is discriminant analysis and logit analysis as the following:

2.1.1. Discriminant analysis (DA)

DA is a statistical method for categorising observations into a set of classes. DA is used to state an observation’s class based on some input variables. DA is developed based on a set of observations for which the classes are known. Accordingly, DA develops a set of linear regression of the variables, known as discriminant analysis (Russell et al., 2000).

The development of SFPMs started in the US with the use of DA by Beaver (1966). Beaver used a univariate analysis of a number of financial ratios as bankruptcy indications, thus setting the ground for several attempts. He used 30 predictors of 79 failed firms and 79 non-failed firms in
the US during the period 1954-1964. He indicated that a set of predictors could differentiate between the failed and non-failed companies for five years prior to companies’ failure. Beaver was able to accurately classify 78% of failed and non-failed firms (i.e. overall classification accuracy rate is 78%) five years before failure. The function Beaver used to predict bankruptcy was:

\[ Z = 1.3A_1 + 2.4A_2 - 0.98A_3 - 6.787 \]

Where:
- \( A_1 \) = working capital funds flow/total liabilities;
- \( A_2 \) = net income/total assets; and
- \( A_3 \) = total liabilities/total assets.

2.1.1.1 Overview of the Altman models

In 1968, Altman used a Multiple Discriminant Analysis (MDA) with a list of twenty-two variables. He classified the variables into five financial ratio categories, including liquidity, profitability, leverage, solvency, and activity ratios. He used this method to a matched sample of 33 non-bankrupt and bankrupt firms in the US. He developed a model using as a linear combination of predictors that discriminate between bankrupt and non-bankrupt firms. Eventually, Altman (1968) reported the original model presented as the following:

\[ Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \]

Where:
- \( X_1 \) = the working capital/total assets;
- \( X_2 \) = the retained earnings/total assets;
- \( X_3 \) = the earnings before interest and taxes/total assets;
- \( X_4 \) = the market value equity/book value of total debt;
- \( X_5 \) = the sales/total assets, \( Z \) is the overall Z-score index.

Altman (1968) found that overall classification accuracy rate of the model was 95% and the Type I and Type II errors were 6% and 3% respectively, one year prior to bankruptcy.

2.1.1.2 Development and revision of Altman Z-score (1968) model

Considering some of the limitations and criticisms to his original Altman Z-score (1968) model, he conducted a set of studies to clarify the effectiveness and usefulness of his model in predicting bankruptcy.

In 1993, Altman revised his original model. He developed a \( Z' \)-score for private companies and a \( Z'' \)-score for non-manufacturing companies. Indeed, the original Altman Z-score (1968) model seemed to be inappropriate for use in all emerging markets and to be primarily suitable for manufacturers. He wanted to evaluate the model accuracy rate without the \( X_5 \) (sales/total assets) predictor.
The revised Altman Z’-score (1993) model for private firms is:

\[ Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5 \]

Where

- \( X_1 \) = the working capital/total assets;
- \( X_2 \) = the retained earnings/total assets;
- \( X_3 \) = the earnings before interest and taxes/total assets;
- \( X_4 \) = the book value of equity/book value of total debt;
- \( X_5 \) = the sales/total assets, and \( Z \) is the overall Z-score index.

The revised Altman Z”-score (1993) model for non-manufacturing firms is:

\[ Z'' = 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4 \]

Where

- \( X_1 \) = the working capital/total assets;
- \( X_2 \) = the retained earnings/total assets;
- \( X_3 \) = the earnings before interest and taxes/total assets;
- \( X_4 \) = the book value of equity/book value of total debt; and
- \( Z'' \) is the overall Z-score index.

### 2.1.2 (Logit) logistic regression analysis (LRA)

LRA is a multiple regression analysis but with a dummy variable and variables that are continuous or categorical (Field, 2009). This method does not demand normality assumption and equal covariance matrices as DA does. It makes a nonlinear transformation of the data that decreases the influence of outliers.

To avoid the limitations and problems of MDA and to produce a high accuracy rate in evaluating companies’ financial positions, several researchers have used LRA in constructing their SFPMs. In the US, Ohlson (1980) was the first to apply LRA by using 105 bankrupt and 2058 non-bankrupt firms during the 1970-1976 period. He found that the best predictors were working capital/total assets, total liabilities/total assets, net income/total assets, current liabilities/current assets, funds provided by operations/total liabilities and changes in net income. His model presented an overall classification accuracy rate of 96.12%, 92.84% for prediction within one year and two years before failure, respectively.

Zavgren (1985) developed logit models as a financial risk measure. His dataset consisted of 45 bankrupt and 45 nonbankrupt US firms during the 1972-1978 period. The models
developed achieved accuracy rate greater than 99% in classifying failed and non-failed firms.

Other studies have considered a non-US environment. Nam and Jinn (2000) investigated 92 non-financial listed Korean firms which included 46 companies that failed during the Asian financial crisis in 1997 and 1998. The authors used LRA for predicting bankruptcy and correctly classified 80.4% of failed companies and 73.9% of non-failed companies.

Grice and Dugan (2001) examined Ohlson’s (1980) model to samples of US failed and non-failed companies for time periods, industries, and financial conditions other than those used to develop Ohlson’s original model. They found that the model’s accuracy rates decreased when used to other environment. Ohlson (1980) showed 98.2% and 96.4% overall classification accuracies for his model using samples for the 1970–1978 period. The overall classification accuracies for the sample used in Grice and Dugan’s (2001) study ranged from 34.8% to 81.3%. The findings also suggested that the accuracy rate of the model lasts to drop moving from the periods of 1988–1991 to the 1992–1999. Thus, they came up that the relationship between predictors and financial failure changes over time.

2.1 Artificial Intelligence Technology (AIT):

As will be discussed in this section, AIT performed well in the domain of companies’ failure prediction. It includes the following methods:

2.1.1 Decision trees

Decision tree algorithms are employed for prediction problems (Gehrke et al., 1999). Decision trees form a part of ‘machine learning’ and as such are an important area of artificial intelligence. The main aim of decision trees technique is classification by separating observations into several subgroups, mutually exclusive subgroups and exhaustive subgroups. This separation is based on a set of independent variables that have the robust relationship with the dependent variable (Koh and Low, 2004). Decision trees use recursive partitioning algorithms (RPAs) to generate models on a given dataset[^1]. Frydman et al. (1985) used RPAs for financial analysis and compared them with DA within the context of firms in financial distress. They used a sample consisting of 58 bankrupt and 142 non-bankrupt US companies during the 1971-1981 period. They developed two alternatives of discriminant models (DAs) and matched them with two RPA models. They concluded that the RPA models outperformed DAs models in terms of overall classification accuracy rate and Type I and Type II error rates.

Edwards et al. (2005) explored alternative techniques of corporate failure prediction using decision tree analysis, MDA, LRA and neural networks[^9] to build SFPMs. They used 23
bankrupt and 23 non-bankrupt Australian companies from 1990-1998. They showed that the overall classification accuracy rates were 71.7%, 69.57%, and 78.26%, for the MDA model, decision tree model and logistic regression, respectively, one year prior to failure. The neural networks model achieved an overall classification accuracy rate of 93.75% one year prior to failure, and this was a better performance than with other statistical techniques.

2.1.2 Evolutionary approaches

Genetic algorithms (GAs) are stochastic search methods that can search large and complicated spaces on ideas from natural genetics and the evolutionary principle. GAs application has been successful in the field of bankruptcy expectation (Shin and Lee, 2002).

Varetto (1998) compared LDA for bankruptcy prediction with GAs. He analysed 1920 failed and 1920 non-failed Italian manufacturing companies for the 1982-1995 period. He found that the genetic linear function yielded a 97.88% Type I correct classification rate and LDA yielded 98.33% one year prior to bankruptcy. The Type II correct classification rate was 96.94% and 93.54% for the genetic linear function and LDA one year prior to bankruptcy, respectively.

Shin and Lee (2002) also developed a model using GAs for bankruptcy prediction. They illustrated how GAs could be used to predict companies’ future position. They reported that GAs can extract rules that are easy to comprehend for users such as expert systems. The data used contained 528 firms, 264 bankrupt and 264 non-bankrupt South Korean companies during the 1995-1997 period. The results showed that the five rules generated by GAs achieved a score of 80.8% for the Type I correct classification rate and 79.7% for the Type II correct classification rate.

2.1.3 Rough sets

Pawlak (1982) proposed the rough set theory assuming that with any object of a given universe there is some information associated. Objects categorised by similar data are indiscernible. This implies that we must take in our consideration equivalence classes of the indiscernibility relationship.

Dimitras et al. (1999) provided a model to distinguish between failed and non-failed companies by using the rough set method. The study sample consisted of 40 failed and 40 non-failed firms during the 1986-1990 period in Greece. They concluded that the rough set method performed better than DA and LRA in evaluating the firm’s failure risk. An overall classification accuracy rate was 65.8% for the rough set approach, 60.5% for DA and 60.53% for LRA.
McKee (2000) also used a set of predictors to develop a rough set based model. His sample consisted of 50 bankrupt and 50 non-bankrupt US companies during the 1986-1988 period. He concluded that the Type I correct classification rate of the model was 86% and the Type II correct classification rate 90% one year prior to bankruptcy. McKee’s study was in line with previous results of rough sets research that this method can produce a high quality bankruptcy prediction model.

### 2.1.4 Neural networks

Neural networks (NNs) have performed well in business classifications including bankruptcy prediction. In 1990, NNs was used for the first time in predicting bankrupt companies. Wilson and Sharda (1994) compared the accuracy rates of NNs with MDA. They used the Altman Z-score (1968) model with five variables. Their study sample consisted of 65 bankrupt and 64 non-bankrupt US companies during the 1975-1982 period. They concluded that NNs performed significantly better than MDA at predicting the positions of firms. The NNs produced an overall classification accuracy rate of 97.5% in both training and testing samples, while MDA’s overall classification accuracy rate was 88.25% only. Additionally, they reported that Type I and Type II correct classification rates of NNs were 97% and 98%, respectively.

The Type I and Type II correct classification rates of MDA were 79.75% and 96.75%, respectively.

Neophytou et al. (2004) developed SFPMs for industrial UK companies using the techniques of LRA and NNs. The study sample consisted of 51 matched-samples of failed and non-failed industrial companies during the 1988-1997 period. NNs presented the highest overall classification accuracy rate for all three years prior to bankruptcy (78%). The logit model achieved only 76% accuracy rate. They concluded that both these two models can be dependable models for bankruptcy expectation.

Pompe and Bilderbeek (2005) used data of small and medium-sized Belgian manufacturing firms in the 1986-1994 period. The study sample contained 1356 bankrupt and 3600 non-bankrupt firms. They found that economic decline coincided with the decline of a model’s performance. Additionally, they concluded that the NNs model outperformed MDA. The overall accuracy rate of the NNs model for one year prior to bankruptcy was 70.69%, for the MDA model it was only 63.78%.

### 2.1.5 Case based reasoning techniques

One of commonly used method in the field of AIT is case based reasoning (CBR), but is only just now beginning to be used in business contexts. Prior literature showed that CBR as a machine learning technique overcomes some of the deficiencies of SFPMs and NNs.
in evaluating companies’ future position failure or non-failure. Earliest studies done in this field indicates that the prediction accuracy of CBR is significantly high.

For instance, Jo and Han (1996) used a DA and two artificial intelligence models (NNs and CBR) in their study. Their sample consisted of 271 Korean bankrupt firms matched with a sample of comparable size and from the same industry, during the 1991-1993 period. The Type I correct classification rates were 84.15%, 88.03%, 87.68%, for DA, NNs, and CBR, respectively. They concluded that the NNs significantly outperformed the other two methods.

Yip (2004) used a CBR to predict business failure, with an analytic reasoning model called K-NNs, to predict the failure of 44 Australian firms during the 1991-2001 period. He compared the accuracy rate of CBR + K-NNs with that of DA. The results indicated that K-NN is an effective and competitive alternative for predicting companies’ statues.

Li and Sun (2009) aimed to improve the performance of CBR models for business failure prediction in terms of accuracy and reliability. Their sample consisted of 135 pairs of Chinese companies listed on during the 2000-2005 period. The results showed that the accuracy rate of CBR models had superior performance in predicting the business’ status in China.

3. Combining auditors’ GCOs, SFPMs and AIT

Some studies have attempted to compare the performance of auditors’ GCOs with SFPMs and AIT. For instance, Altman and McGough (1974) aimed to develop criteria to assist the auditor in identifying a company’s future position status. They compared auditors’ opinions prior to bankruptcy with the predictability accuracy of the Altman Z-score (1968) model for 34 US bankruptcies during the 1970-1973 period. Their results indicated that the Type I correct classification rate of the Altman Z-score (1968) model was 82%, whereas the Type I correct classification rate of the auditors’ opinions was only 44% one year prior to their entering bankruptcy proceedings. They concluded that the Altman Z-score (1968) model could be helpful to auditors in forming their judgment.

Lennox (1999) attempted to evaluate the accuracy and informativeness of auditors’ GCOs and a SFPM in identifying 160 failed UK companies. The data was obtained from stock exchange financial yearbooks during the 1987-1994 period. He compared the accuracy of a SFPM and auditors’ GCOs in both estimation and hold-out samples. He showed that a SFPM could be more accurate than auditors’ GCOs in a hold-out sample as well as in the estimation sample. He also showed that auditors’ GCOs did not signal useful incremental information about the probability of bankruptcy.

Sun (2007) re-examined performance of auditors’ GCOs and SFPMs in predicting bankruptcy by using a hazard model that incorporated some new features derived from Zmijewski (1984). The bankruptcy sample consisted of 243 firms matched within a
sample of 1165 non-bankrupt US firms during the 1991-2002 period. The results showed that the hazard model statistically outperformed the auditors’ GCOs. Auditors’ GCOs accurately predict 96.6% of non-bankruptcies (Type II correct classification rate) and 44% of bankruptcies (Type I correct classification rate). The estimated hazard model’s accuracy rate in predicting non-bankrupt firms was 82.7% and its accuracy rate in predicting firms status was 84%.

Padawy (2004) aimed to compare prediction capability of the Altman Z-score (1968) model with auditors’ GCOs in order to state the best approach in predicting financial failure. He compared auditors’ GCOs and the Altman Z-score (1968) model for 13 failed and 13 nonfailed Jordanian companies from the 1991-2001 period. He found that the Altman Z-score (1968) model is better at predicting the failure of companies than auditors’ GCOs. The Type I correct classification rate was 92.3%, 100%, 92% and 77.8% for one year, two years, three years and four years prior to failure, respectively. Meanwhile, the Type I correct classification rate of the auditors was low with 38.5%, 27%, 15.4%, and 11.1% for the same companies in the same years, respectively.

4. Discussion and conclusions

Our review of the literature clearly showed that auditors’ GCOs are less accurate than SFPMs and AIT in predicting companies’ going-concern status. In our opinion, that might be the case for the following reasons:

1. Some of the events or conditions covered by auditing standards may have been considered by the auditor, and they led the auditor to refrain from issuing a GCO. For example, management’s plans may suggest actions to solve the firm’s financial problems such as borrowing money, liquidating assets, or restructuring debt, delaying expenditures, or increasing capital, but in fact the company did not seek to implement any of these activities. In other words, the auditor might be misled by what management told him about their future plans.

2. The auditor did not appropriately consider the financial, operating and other indicators reported by auditing standards (i.e., ISA 570) regarding companies’ going-concern status.

3. The GCO decision of the auditor is affected by such factors as opinion shopping, auditclient tenure, audit fees, experience, fraud on the part of the company, NAS, legal liabilities and self-fulfilling prophecy, as discussed earlier.

4. SFPMs and AIT do not ‘care’ about the results of the prediction and potential self-fulfilling prophecy problems.

Our findings illustrated that SPFMs are better in predicting companies’ failure than auditors’ GCOs, although overall the NNs models produce better results. The auditors’ GCOs do not score well in predicting companies’ failure, although they do ‘predict’ non-failure well.
As a general conclusion, AIT can be very useful to investors and other stakeholders when evaluating a company’s future position status. Their predictive power is in many instances very high. For auditors specifically, AIT can serve as helpful analytical tools in forming their judgment. AIT are additional tools, not alternatives. Incorporating the use of AIT in the audit program can provide further evidence that the auditors exerted professional competence and due care. They can also protect the auditors from the risk of litigation, when used with due care, that is when considering the following:

1. SFPMs and AIT produce misclassification errors (Type I and Type II errors), which might damage users. Therefore auditors cannot depend only on these tools in issuing GCOs. There are other indicators reported by auditing standards that should be considered when evaluating the going-concern status of a company.

2. The SFPMs and AIT and their predictive power differ from one model to another, according to several factors such as sample sizes, industry, year of data, economic environment. Additionally, the performance of the predictors used by these techniques may change over time, and depending on industries and countries. Therefore, there is a real need for the development of specific models, in different economic environments using recent time periods and industries other than those used to develop the original models.

3. The most common and accurate models used are Altman models, logit models and NNs models, although overall the NNs models produce better results.

Our study has important implications for researchers, investors and other stakeholders when using SFPMs, AIT and auditors’ GCOs. It attempts to provide a comprehensive review of the performance of SFPMs, AIT, and auditors’ GCOs. It provides evidence that AIT are better than and auditors’ GCOs, and more specifically the NNs models produce better results and the most accurate method in predicting companies’ future position failure or non-failure. Beyond this, the study tried to explain why there are important differences between the methods used and auditors’ GCO in classifying companies with regard of their future failure or non-failure position. There is a set of other factors that might impact an auditor’s independence and competence (e.g., audit firm size, audit fees, non-audit services, value added), thus also audit quality and the GCO decision. Therefore, auditors should be careful when considering other non-financial information to evaluate companies’ position (e.g., management plan).

References


Alareeni, B. A., & Branson, J. (2013). Predicting listed companies' failure in Jordan using Altman models: A case study. *International Journal of Business and Management, 8*(1), 113-126. [https://doi.org/10.5539/ijbm.v8n1p113](https://doi.org/10.5539/ijbm.v8n1p113)


Intelligent Systems in Accounting, Finance & Management, 13(2), 95 -112. https://doi.org/10.1002/isaf.259