**ECON 473-Applied Data Science**

**Research Project**

**The determinants of bankruptcy for companies in Taiwan**

Prof. Sedefka Beck Zongo Ivan

**Abstract**

A lot of research have been undertaken to predict bankruptcy in companies and financial institutions through the use of the decision tree method based on ratios. These studies used to focus more on specific ratios, which are well known around the world nowadays. However, the methods used to determine the causes of bankruptcy are not unique. For instance, apart from the decision tree method, the progress in fields such as data mining also brings some other tools able to manage and predict the likelihood of bankruptcy for these institutions.

One of the most critical research issues in finance is building effective corporate bankruptcy prediction models because they are essential for the risk management of financial institutions. Researchers have applied various data-driven approaches to enhance prediction performance, including statistical and artificial intelligence techniques, and many have proved useful.[[1]](#footnote-1)

**Introduction**

The purpose of our research is to determine what are the determinant of bankruptcy of Taiwanese companies using previous models and, by a discrimination method, select the one which includes the variables that can estimate with accuracy the bankruptcy.

Most of the prediction models used in this perspective are logistic regressions, linear discriminant analysis, and naïve Bayes. But the main difference between these studies is the use of feature selection which is the selection of a subset of representative features from the training dataset for use in the model construction.[[2]](#footnote-2) In addition, the definition of bankruptcy is quite different according to the countries.

In our framework, we will use two different logistic models to assess which one is more suitable for predicting the bankruptcy of companies in Taiwan. According to the results, we will be able to find the underlying determinants of bankruptcy in Taiwan. The notion of bankruptcy can be different according to the country.

It is important then to consider the main variables which can determine bankruptcy in this particular situation.

**Data**

The dataset[[3]](#footnote-3) we will be studying is from Kaggle, a crowd-sourced platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning, and predictive analytics problems. It has over 536,000 active members from 194 countries, and it receives close to 150,000 submissions per month. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.[[4]](#footnote-4)

We chose Kaggle as our source of the dataset as these specific types of datasets are difficult to find in classic data sites in the sense that they are very well structured and their usability is often high. For example, most of the datasets are reorganized to make it easier for the user to exploit.

Notwithstanding all the advantages cited above, there are a couple of issues that need to be mentioned in our dataset. First of all, the variables in the dataset are, for some, more than 32 characters and use special characters. These issues need to be addressed when using the platform S.A.S.

Secondly, the dataset[[5]](#footnote-5) is specific to the country's companies and does not represent the situation of other companies in the world. This is one of the points we will try to show through the research.

It is also important to note that this leads to a sort of selection bias in the sense that only the companies of one country are represented in this dataset. In addition, we may potentially have some measurement bias in the sense that some of the ratios are not exactly the same as the ones which are commonly used. For instance, some of these variables are the inverse of the common one. Then, their interpretations may differ from the others.

**Model/Methodology**

The model which will be used in our study is the logistic model, and our dependent variable is a binary variable with the following outcomes: 1 for bankruptcy and 0 for no bankruptcy.

The logit model is an estimation technique for equations with dummy dependent variables that avoids the unboundedness problem of the linear probability model. For instance, we will use in our specific case two logit models with different explanatory variables to find which of these two logit models is more efficient in explaining the bankruptcy of these companies.

For instance, in countries such as the U.S.A., where the insolvency laws are debtor-oriented, corporate bankruptcy procedures encourage companies in financial difficulty to continue as going concerns(Franks et al., 1996). Therefore, it is possible for companies that file for bankruptcy to reorganize and emerge from bankruptcy or to merge with another entity as a going concern (Shultz, 1995). This is in contrast to the insolvency procedures in creditor-oriented countries such as the UK, Germany, Australia, and New Zealand. Prediction model might not be the best proxy for assessing the going concern, especially in creditor-oriented countries.[[6]](#footnote-6)

Thus, bankruptcy is not always a synonym for liquidation, and then this may change the parameters, which can lead to bankruptcy.

To be more explicit, the first model we will use is the model called the liquidation model,[[7]](#footnote-7) in which the company has no other options than being liquidated, in contrast to the bankruptcy model,[[8]](#footnote-8) in which the company still has some solutions to emerge from bankruptcy. It is important to mention that the dependent variable of these two models is the variable "bankruptcy"; the only difference between the two models is the notion of bankruptcy which has underlying causes and the explanatory variables that are used to determine whether or not there is bankruptcy. Finally, we will also try to find through these two models whether or not this country is credit or debit oriented based on the significance of the two models.

We chose this model as the dependent variable is a binary dependent variable. In addition, whatever the value of these variables will be, the logit model will be able to restrict the changes of the dependent variable between 1 and 0, which are the important values of the dependent variable for us.

The only caveats of this model are related to the fact that the O.L.S. estimation cannot be used. The estimations produced by the O.L.S. model are very easy to interpret. Unfortunately, the logit model is different. For instance, the logit model allows us to create a link between the dependent variable and the explanatory variables in a linear functional form. In addition, logit models, which use the maximum likelihood method for estimations, require us, at first, to make a distributional assumption. Finally, most of the maximum likelihood estimators cannot be solved "by hand."

**Results of the study**

**Liquidation model**



According to the Liquidity model table, the variables which we renamed respectively X1, X2, X3, X4, X5, X6, X7, and X8 are the independent variables, and bankrupt is the binary dependent variable.

Our results suggest that X1, X4, X6, and X7 have a positive correlation effect on bankruptcy. In other terms, these variables increase the likelihood of Taiwanese companies being bankrupt. In contrast, previous research showed that the variable X6 is negatively associated with

On the other hand, our results show that X3, X5, and X8 are negatively correlated with the dependent variable bankruptcy.

For the variable X2, we get different results according to whether we use the Linear Probability Model, the Logit model, or the Probit model. For instance, using the L.P.M. model leads us to a positive correlation Based on the literature, X2 or the current ratio variable is negatively correlated with the dependent variable bankrupt.

**Bankruptcy model**



From the Liquidity model table, the main independent variables are respectively SIZE, TLTA, WCTA, CLCA, OENEG, NITA, FUTL, CHIN, and bankrupt as the binary dependent variable.

The findings suggest that TLTA, OENEG, and CHIN have a positive correlation effect on bankruptcy. In other terms, these variables increase the likelihood of Taiwanese companies being bankrupt. Our findings are in accordance with the literature.

On the other hand, our results show that WCTA, CLCA, and NITA decrease the likelihood of observing bankruptcy. The higher these coefficients are, the less likely these companies will be bankrupt.

For the variables SIZE and FUTL, we get different results according to whether we use the L.P.M. or the Logit/Probit model. The L.P.M. model gives us a positive relationship between the two variables (SIZE and FUTL) and the dependent variable bankruptcy. However, using the Logit or Probit models leads us to a negative relationship between the two variables and bankruptcy. Based on the literature, SIZE and FUTL are both negatively correlated with the dependent variable bankrupt. Thus, we can say that the bigger a company is in size, the less likely it is the probability for this company to experience bankruptcy, holding other variables constant. The same goes for the FUTL variable. The higher this ratio is, the less likely a company will be bankrupt, holding other variables constant.

**Comparison of our two models with the literature**

**First Model**



**Second Model**



By comparing the two models with the literature, we concluded that the bankruptcy model is the most suitable in predicting the bankruptcy of Taiwanese companies as we can see a better fitting of the independent variables of the bankruptcy model with the literature. This leads us to also conclude that Taiwan is a debtor-oriented country.

After finding that the bankruptcy model is the model to be used to assess the bankruptcy of Taiwanese companies, the next question is what can be the main and significant variables that determine whether or not these companies are likely to be bankrupt.

In addition, the most recurrent model that was used to assess bankruptcy was the Logit model. Then, we decide to use this model over the Probit model to have more comparable and close results even if the Probit model provides quite the same results.

**Examination of the marginal effects of the bankruptcy model's variables**



**Interpretation:**

Based on the board above, we can say the following for each variable

**TLTA**: An increase in the TLTA by one unit increases the probability of bankruptcy (bankrupt=1) by 2.49%, holding other variables constant

**WCTA**: An increase in the WCTA by one unit decreases the probability of bankruptcy (bankrupt=1) by 5.96%, holding other variables constant

**CLCA**: An increase in the SIZE by one unit decreases the probability of bankruptcy (bankrupt=1) by 0.59%, holding other variables constant

**NITA**: An increase in the SIZE by one unit decreases the probability of bankruptcy (bankrupt=1) by 15.44%, holding other variables constant

Through a discriminatory process, we end up with three main variables which are crucial in determining the bankruptcy of Taiwanese companies. These variables are TLTA, WCTA, and finally, NITA.

**Model Fit Statistics**

**Table

Description automatically generated**

Smaller values of these criteria indicate a better model. Adding predictor variables to a model can never increase the value of the -2Log(L), and this can be observed in our case.

Thus, penalized versions of these statistics (intercept only) are used and compared to adjusted versions which include the number of parameters used k (and, in the case of S.C., also for sample size n)

**Global Null Hypothesis test**

**Table

Description automatically generated**

The significant value of the likelihood ratio provides evidence that at least one of the regression coefficients for the explanatory variables is non-zero.

The Score and the Wald tests also confirm the previous test.

**Maximum likelihood estimatesTable

Description automatically generated**

**Interpretation:**

For our variables of interest, we have:

**TLTA**: A one-unit increase in the TLTA corresponds to a 21.98 increase in the log odds of bankruptcy, holding other variables constant.

**WCTA**: A one-unit increase in the WCTA corresponds to a 7.32 decrease in the log odds of bankruptcy, holding other variables constant.

**CLCA**: A one-unit increase in the CLCA corresponds to a 6.51 decrease in the log odds of bankruptcy, holding other variables constant.

**NITA**: A one-unit increase in the NITA corresponds to a 19.12 decrease in the log odds of bankruptcy, holding other variables constant.

**Odds Ratio Estimates**

**Table

Description automatically generated**

The performance of logistic regression in the presence of sparse data is questionable. In such a situation, a common problem is the presence of high odds ratios (ORs) with very wide 95% confidence interval (CI) (OR: >999.999, 95% CI: <0.001, >999.999).

The problem of separation is by no means negligible and may occur even if the variable values are low in numbers. To address this issue, penalized logistic regression (P.L.R.) method is often used

Although this method is shown to be another option as compared to the classic logistic model, it is not used in regular practice due to the computational difficulty, provision of degenerate estimates, and not allowing the inclusion of continuous type covariates.[[9]](#footnote-9)

**Association of predicted probabilities and observed responses**

Table

Description automatically generated

Generally, a higher percentage of concordant pairs and a lower percentage of discordant pairs indicate a more desirable model. Thus, our model fits the dataset well.

In addition, using probabilities from this model allow us to correctly differentiate between a company that is bankrupt and a company that is not about 92.3% of the time.

Synthesis

Previous research has been undertaken with the purpose of determining the determinants of bankruptcy. The environment has often changed the expected results of the basic models, which were somehow considered master keys that can be used everywhere. Our research paper, in that sense, posits a specific model which can be used in a specific environment.

"Despite some universal directions of change, the bankruptcy laws of individual countries still show many different features. The main criterion differentiating them is the friendliness of regulations towards debtors and creditors. Therefore, in the subject literature, a distinction is made between legal systems that are more debtor or creditor-friendly and the so-called hybrid systems. One of the first information and research aimed at separating bankruptcy systems friendly to debtors from those creditor-friendly was presented by P.R. Wood (1995), Q. Hussain, and C. Wihlborg (1999), E. Berglöf, H. Rosenthal, and E.L. von Thadden (2001), R.R. Bliss (2003), M. Falke (2003), G. Recasens (2004), S. Franken (2004), C. López-Gutiérrez, M. Olalla García and B. Torre Olmo (2005). In these publications, the authors both presented criteria that differentiated the two systems and attempted to qualify some countries as bankruptcy systems with regard to their friendliness to debtors and creditors. However, the classification of countries into individual bankruptcy regimes concerned a relatively small number of countries, apart from the research conducted by Z.R. Azar (2007), who proposed P.D.I. (pro debtor index) and PCI (pro creditor index) and based on several criteria and data from 2003 assigned 50 countries to more or less debtor or creditor-friendly systems.( Kuruppu, Nirosh, et al, 2003)"[[10]](#footnote-10)

Conclusion

To sum up, a plethora of methods exists when it comes to assessing the potential bankruptcy of a company. However, each method has some limits and can only be applied to certain circumstances. Furthermore, it is important to study the environment in which a company is before applying any methods. In our case, in terms of bankruptcy prediction potentiality, the bankruptcy model is more suitable as Taiwanese companies evolve in a debt-oriented environment. Moreover, this model can be used as a proxy or extended with additional ratios to get more accuracy. Finally, this study shows how the environment can change the effectiveness of a model and the need for models which are more dynamic and adaptive to a changing environment.

"I have neither given or received, nor have I tolerated others' use of unauthorized aid."

Zongo Ivan

BIBLIOGRAPHY

Ahn, Hyunchul, and Kyoung-jae Kim. "Bankruptcy Prediction Modeling with Hybrid Case-Based Reasoning and Genetic Algorithms Approach." Applied Soft Computing, vol. 9, no. 2, 2009, pp. 599–607., <https://doi.org/10.1016/j.asoc.2008.08.002>.

Lin, Wei-Chao, et al. "Feature Selection in Single and Ensemble Learning-Based Bankruptcy Prediction Models." Expert Systems, vol. 36, no. 1, 2018, <https://doi.org/10.1111/exsy.12335>.

Kuruppu, Nirosh, et al. "The Efficacy of Liquidation and Bankruptcy Prediction Models for Assessing Going Concern." Managerial Auditing Journal, vol. 18, no. 6/7, 2003, pp. 577–590., <https://doi.org/10.1108/02686900310482713>.

<https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction?select=data.csv>

<https://www.kaggle.com/getting-started/44916>

Ohlson, James A. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." Journal of Accounting Research, vol. 18, no. 1, 1980, p. 109., <https://doi.org/10.2307/2490395>.

Liang, Deron, et al. "Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study." European Journal of Operational Research, vol. 252, no. 2, 2016, pp. 561–572., <https://doi.org/10.1016/j.ejor.2016.01.012>.

Olson, David L., et al. "Comparative Analysis of Data Mining Methods for Bankruptcy Prediction." Decision Support Systems, vol. 52, no. 2, 2012, pp. 464–473., <https://doi.org/10.1016/j.dss.2011.10.007>.

Devika, S, et al. "Analysis of Sparse Data in Logistic Regression in Medical Research: A Newer Approach." Journal of Postgraduate Medicine, Medknow Publications &amp; Media Pvt Ltd, 2016, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4944325/>.

1. Ahn, Hyunchul, and Kyoung-jae Kim. “Bankruptcy Prediction Modeling with Hybrid Case-Based Reasoning and Genetic Algorithms Approach.” Applied Soft Computing, vol. 9, no. 2, 2009, pp. 599–607., <https://doi.org/10.1016/j.asoc.2008.08.002>. [↑](#footnote-ref-1)
2. Lin, Wei-Chao, et al. “Feature Selection in Single and Ensemble Learning-Based Bankruptcy Prediction Models.” Expert Systems, vol. 36, no. 1, 2018, <https://doi.org/10.1111/exsy.12335>. [↑](#footnote-ref-2)
3. <https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction?select=data.csv> [↑](#footnote-ref-3)
4. <https://www.kaggle.com/getting-started/44916> [↑](#footnote-ref-4)
5. Liang, Deron, et al. “Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study.” European Journal of Operational Research, vol. 252, no. 2, 2016, pp. 561–572., <https://doi.org/10.1016/j.ejor.2016.01.012>. [↑](#footnote-ref-5)
6. Kuruppu, Nirosh, et al. “The Efficacy of Liquidation and Bankruptcy Prediction Models for Assessing Going Concern.” Managerial Auditing Journal, vol. 18, no. 6/7, 2003, pp. 577–590., <https://doi.org/10.1108/02686900310482713>. [↑](#footnote-ref-6)
7. Kuruppu, Nirosh, et al. “The Efficacy of Liquidation and Bankruptcy Prediction Models for Assessing Going Concern.” Managerial Auditing Journal, vol. 18, no. 6/7, 2003, pp. 577–590., <https://doi.org/10.1108/02686900310482713> [↑](#footnote-ref-7)
8. Ohlson, James A. “Financial Ratios and the Probabilistic Prediction of Bankruptcy.” Journal of Accounting Research, vol. 18, no. 1, 1980, p. 109., <https://doi.org/10.2307/2490395>. [↑](#footnote-ref-8)
9. Devika, S, et al. “Analysis of Sparse Data in Logistic Regression in Medical Research: A Newer Approach.” Journal of Postgraduate Medicine, Medknow Publications &amp; Media Pvt Ltd, 2016, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4944325/>. [↑](#footnote-ref-9)
10. Kuruppu, Nirosh, et al. “The Efficacy of Liquidation and Bankruptcy Prediction Models for Assessing Going Concern.” Managerial Auditing Journal, vol. 18, no. 6/7, 2003, pp. 577–590., <https://doi.org/10.1108/02686900310482713>. [↑](#footnote-ref-10)