Proposal

Factors influencing time to bankruptcy of companies: a survival analysis perspective

Promotor: Prof. Dr. Ingrid Van Keilegom
Co-Promotor: Mr. Wei Miao

Khadar Mohammed (r0927157)
Margaux Sculsky (r0636357)

KU Leuven
Faculty of Science
Master of Statistics and Data Science
2024-2025

1 Introduction

An accurate prediction of bankruptcy for a given company has many (socio)economic applications: a bank can decide whether or not to give out a loan, governments can assess risks, investors can make more informed decisions, etc. (Liang et al., 2016). For this thesis the aim is to build a model using survival analysis methods to predict the time until bankruptcy.

2 History of predicting bankruptcy

The subject of predicting bankruptcy has been studied many times using different techniques: e.g. discriminant analysis (Altman, 1968), principal component analysis (Achim et al., 2012), logistic regression (Ohlson, 1980), support vector machine methods (Horak et al. 2020).

The dependent variable used in these techniques is usually seen as dichotoumous, meaning these techniques have been used to predict whether a company will go bankrupt or not, regardless of the time frame in which they go bankrupt. For example, in their comparison of different input variables for modelling bankruptcy, Liang et al. (2016) used a data set of 478 companies that were active between 1999-2009, of which half have gone bankrupt. There was however no distinction made between companies that went bankrupt within 4 years or companies that survived 9 years. Considering the difference in implications of an in short or long term upcoming bankruptcy, we believe it significant to not only predict whether a company will go bankrupt but also in what time frame. For this reason we propose using survival analysis techniques which take the time until an event (in this case, bankruptcy) as a dependent variable.

3 Data

3.1 Extraction and variables

The data used was extracted from the Compustat North America and CRSP databases, providing fiscal data for North American companies. The dataset contains 10,982 different companies, active some time between 01/01/1985 and 31/12/2023. The following variables of interest are included, most of which are time dependent i.e. they change every fiscal year:

- *ipodate*: Initial public offering date, this is used as the start date to calculate the dependent variable survival time.
- $ENDDAT$: End of stock date, this is used as the end date to calculate the dependent variable survival time.
- $DLSTCD$: delisting code, the reason for which a stock is delisted.
- Ratio 1: *liquidity* (WCAP/AT): The working capital over total assets ratio, measures the net liquid assets of a business as a percentage of it's total assets. Given by $\frac{working\ capital}{total\ assets}$
- Ratio 2: Retained Earnings/Total Assets (RE/AT): Measures profitability that reflects the company's age and earning power
- Ratio 3: *profitability* (EBIT/AT): The Basic Earning Power ratio (BEP) is given by $\frac{Earnings\ Before\ Interest\ and\ Taxes}{total\ assets}$ This is a version of ROA ratio, which measures the company's efficiency to make profits from assets.
- Ratio 4: Market Value of Equity / Book Value of Total Liabilities ($MKVVALT/LT$): Adds market dimension that can show up security price fluctuation as a possible red flag.
- Ratio 5: *Asset turnover ratio*: Measures the efficiency of a company's assets in generating sales $\frac{net\ sales}{total\ assets}$
- *survtime*: The time in years for which the company remained active, $ipodata - ENDDAT$
- *censoring*: Indicates whether a company has gone bankrupt or is censored, set to be '1' if $ENDDAT =$ 29/12/2023 and '0' in other cases.
- Values of the given ratios above are calculated per company per year, and ratio-year combination is considered as a co-variate, adding (5*number of years) number of co-variates

3.2 Exploring the data set

The data set contains 10,982 companies with a start date between 01/01/1985 and 31/12/2023 of which 8,009 have gone bankrupt before 31/12/2023. Figure 1 shows a histogram of the end of stock data in the data set. It shows a peak in bankruptcies around the year 2000. Note that the high count at the year 2023 is because of censoring, these companies were still active in 2023, but the data set cuts off after the year 2023.

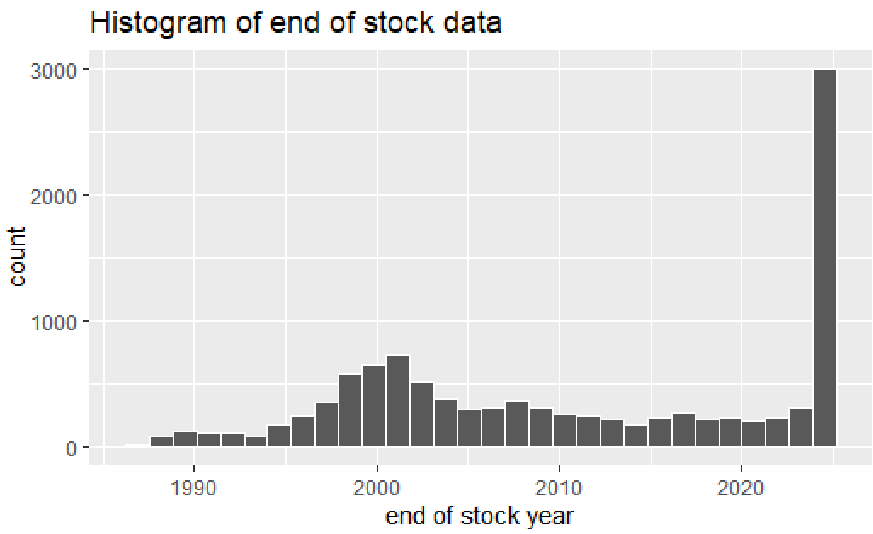

Figure 1: The histogram shows the last stock date. A last date of 2023 indicates censoring.

Figure 2 shows a histogram of survival time for bankrupt and non-bankrupt companies. It shows that most companies went bankrupt within the first 10 years. It also shows a peak of young companies in the data set that have existed for less than 5 years, but have not yet gone bankrupt.

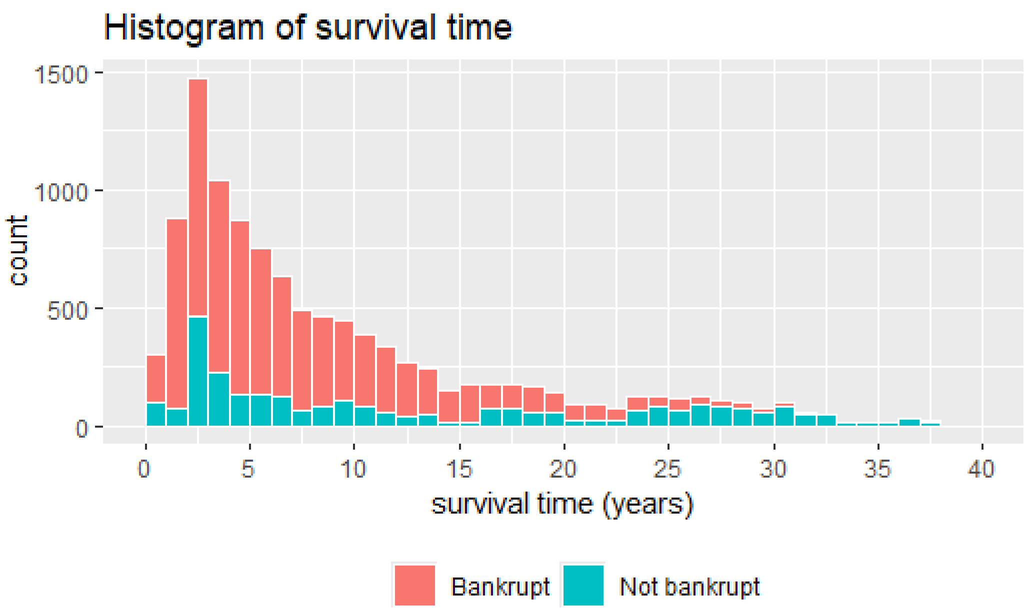

Figure 2: The histogram shows the survival time by bankruptcy status.

As mentioned above, the explanatory variables in the data set are time dependent and change every fiscal year. There are is a considerable amount missing values throughout the data set, as shown in Figure 3 and 4. Figure 3 shows the count of missing data. The figure shows that variables *liquidity* and *mkvalt/lt* are missing most often, and *mkvalt/lt* is missing for all years before 1998. Figure 4 shows the amount of missing data as a percentage of the full data set. It shows that the variables *liqudity* and *mkvalt/it* are missing for 20% to 25% of the data, while the other variables have less missing values with a percentage of around 5%.

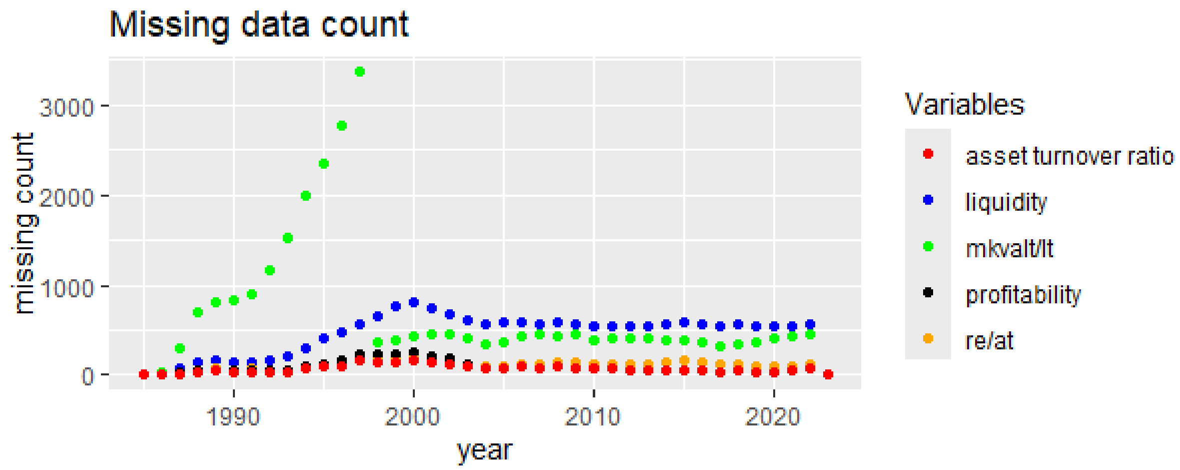

Figure 3: The scatterplot shows the count of missing data for each variable. Note that *mkvalt/lt* is not available before 1998.

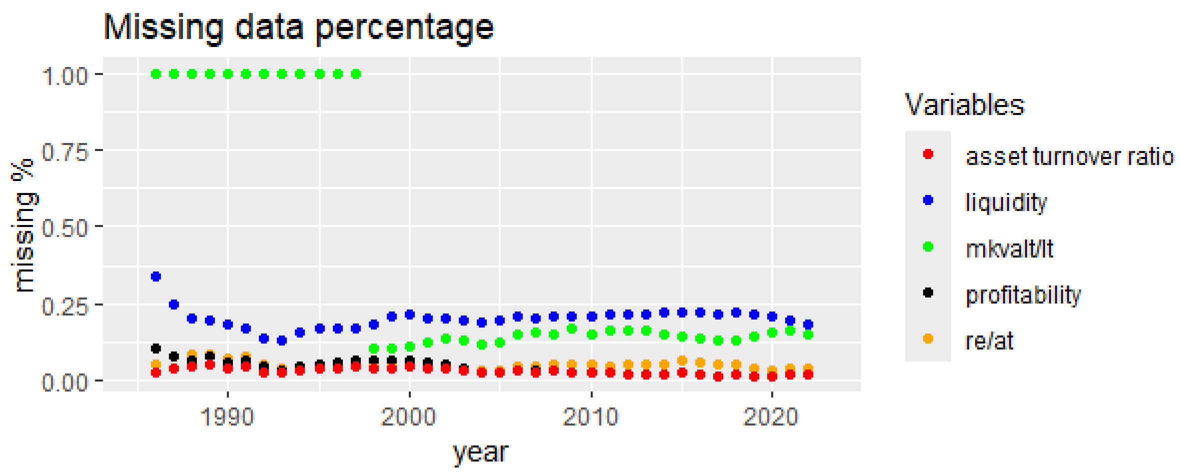

Figure 4: The scatterplot shows the percentage of missing data for each variable. Note that *mkvalt/lt* is not available before 1998.

4 Survival analysis techniques

In survival analysis the main goal is to estimate the time until an event through estimating the survival or hazard function, where the survival function *S(t)* describes the probability to survive until a given time t, and the hazard function *h(t)* describes the instantaneous risk at a given time t. One of the most common used models is the semi-parametric Cox Proportional Hazards Model (Kleinbaum & Klein, 2012), where the hazard function is given by:

$$h(t, \boldsymbol{X}) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i}$$

with explanatory variables

$$\boldsymbol{X} = (X_1, X_2, ..., X_p)$$

Because the explanatory variables in our data set change every year, we have to use an extended version of the typical Cox Proportional Hazards Model where we take into account the changing of the variables:

$$h(t, \boldsymbol{X(t)}) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i(t) \ + \ \sum_{j=p+1}^{p+q} \delta_j X_j(t)}$$

with p the number of time-independent predictors with coefficients given by β_i,
q the number of time-dependent predictors with coefficients given by δ_j.

and with explanatory variables given by:

$$\boldsymbol{X(t)} = (X_1, X_2, ..., X_p, X_{p+1}(t), X_{p+2}(t), ...X_{p+q}(t))$$

where the values for X_1 to X_p do no depend on time t,
and the values X_{p+1} to X_{p+q} do depend on time t.

5 References

Achim, M. V., Mare, C., & Borlea, S. N. A (2012). Statitical Model of Financial Risk Bankruptcy Applied for Romanian Manufacturing Industry. *Procedia Economics and Finance, 3*, 132-137. https://doi.org/10.1016/S2212-5671(12)00131-1

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of finance, 23(4)*, 589-609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x

Horak, J., Vrbka, J., & Suler, P. (2020). Support Vector Machine Methods and Artificial Neural Networks Used for the Development of Bankruptcy Prediction Models and Their Comparison. *Journal of Risk and Financial Management; 13(3)*, 1-15. http://dx.doi.org/10.3390/jrfm13030060

Kleinbaum, D. G., & Klein, M. (2012). *Survival Analysis: A Self-Learning Text* (Third edition). Springer.

Liang, D., Lu, C., Tsai, C., & Shih, G. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Optional Research, 252(2)*, 561-572. http://dx.doi.org/10.1016/j.ejor.2016.01.012

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research, 18(1)*, 109-131. https://doi.org/10.2307/2490395