ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

# Self- Organizing Map and Multidimensional Scaling in a Tandem Approach: a Visualization of Bankruptcy Trajectory

*by:*

Wieke Laura van der Lugt - 456034

**Msc Econometrics**

Business Analytics and Quantitative Marketing

*Supervisors:*

Andreas Alfons (ESE) & Jack Gannaway (PwC)

*Second Assessor:*

Patrick Groenen (ESE)

May 14, 2019

## Abstract

In this paper, two innovations to a trajectory approach using Self- Organizing Maps are introduced to analyze financial statements of U.S. listed companies between 2013-2018, approaching changes in these statements as movement patterns. The first innovation entails a projection of quarterly statements on multiple maps instead of a single Self- Organizing Map. The second entails the use of a tandem approach between the Self-Organizing Map and Multidimensional Scaling to allocate the neurons in a Self-Organizing Map a more optimal position on the grid. By applying Multidimensional Scaling to the neurons in a configured Self- Organizing Map the visual representation of this map portrays more intuitive insights into the neighboring relations between each neuron. Including both innovations in the trajectory approach resulted in a slightly more accurate prediction of bankruptcy compared to the traditional Self- Organizing Map approach. By its ability to capture both cross-sectional and longitudinal differences between companies and depicting these in a two-dimensional intuitive map, the SOM- MDS tandem is a highly relevant tool for anyone seeking support in their company viability analysis.

# Contents

# 1 Introduction

In January 2018, British multinational constructor Carillion filed for bankruptcy with an outstanding debt of almost a billion pounds. In the months leading up to the default, a number of large, important contracts were revealed to severely underperform. A £300 million cash injection would have saved the company, but neither the banks nor the government were willing to provide the money. Consequently, the government made a compulsory liquidation order against the firm on January 15th, 2018. Carillion was the second largest constructor in the UK and according to the Financial Times (2018) its bankruptcy cost 19,000 people their jobs. Not surprisingly, the liquidation caused a lot of commotion, reaching far beyond the construction industry alone.

A conceivably unexpected sector affected by Carillion's downfall was the audit sector. Commotion was caused by the sign-off of the 2016 accounts by Carillion's auditor, just four months prior to the bankruptcy (see a.o. Accountant (2018)). This was problematic, as the independent review of financial statements by an auditor is meant to serve as a safeguard measure and should provide additional assurance to stakeholders about the financial condition of a firm. Unfortunately, Carillion's 2016 audit report was not assuring as it did not raise any concerns and was released without additional comments, even though the valuation of a small number of contracts were decisive for the firm's survival.

On a more fundamental level the sign-off of the audit report was problematic as it was performed on a "going concern" basis. This is a hypothesis on the viability of a company: it assumes the firm to be a continuing business, active for at least twelve months after the sign-off of the audit report. The applicability of going concern should be reassessed by the auditor for each individual review and if there is material uncertainty about a firm's viability, an "emphasis of matter" paragraph has to be included at the end of the report, motivating why the firm can or should not be treated as a going concern (see section 570 of the ISAA (2016)). Clearly, no such paragraph was included in Carillion's audit report. In general, auditors are reluctant to issue a warning about going concern, as the mere act of raising these doubts could have a self-fulfilling prophecy effect. Nonetheless, auditors are liable for their review, meaning legal action can be undertaken against them should a firm go bankrupt while treated as a going concern. Exactly this is happening to Carillion's auditor at the moment (see The Guardian (2018)).

As the uproar around Carillion is not the first criticism the audit sector faced (i.e. during the financial crisis many banks went bankrupt or were bailed out without any previous warning on their deteriorating safety), public confidence in the sector has substantially decreased and questions have been raised about the function and quality of the audit review ACAA (2011). As a response, within the sector many initiatives have been set up aiming to restore public trust. Following Carrilion's collapse, United Kingdom's accounting watchdog launched a taskforce evaluating the amount of work an auditor should put in to reach a going concern conclusion FRC (2018). A similar taskforce has been set up by the Public Company Accounting Oversight Board in the United States, assessing whether there is need for regulatory action in the accounting profession to assure audit quality PCAOB (2018). In the Netherlands, the professional association for accountants addressed their recent scarred reputation and suggested including a mandatory 'continuity' clause in every audit report, forcing auditors to motivate the viability of every firm they audit, independent of any going concern issues being present or not NBA (2018). What these initiatives have in common is their focus on repairing the sector's credibility by improving the auditor's ability to detect early warning signs for going concern issues in the financial statements. Therefore, auditors have become large stakeholders in the search for accurate indicators of deteriorating financial performance and it (re)spiked interest in research on bankruptcy prediction.

In this paper, a trajectory approach using Self- Organizing Maps is used to analyze the financial statements of U.S. listed companies between 2013-2018, approaching changes in financial statements as movement patterns (i.e. forming a trajectory). A Self-Organizing Map (SOM) is a non-parametric, single-layer neural network which is widely used as a dimension reduction, visualization- and clustering tool. This research assesses whether the SOM trajectory approach can capture and analyze the time evolution of U.S. listed companies and whether it can discriminate between movements that result in bankruptcy and those who do not.

The methodology used in this paper builds on the trajectory SOM approach for bankruptcy prediction models previously explored by Schreck et al. (2007), Du Jardin and Séverin (2011) and Chen et al. (2013). The main idea of trajectory SOM is to project the financial statements of companies over time on a map, such that a firm is no longer represented by a single neuron on the map, but by a sequence of neurons on the map over time. The trajectory SOM analysis as presented in these papers relies on two steps. The first step comprises the construction of a feature-based SOM (where the features are the line items on the financial statements) characterizing the risk of failure for each company a year prior to potential bankruptcy. Additionally, the series of features from previous years are also projected on this map and the successive positions on the map for each company are stored in a so-called trajectory. The second step comprises the visualization and clustering of these trajectories using an additional SOM, aiming to detect differences in movements between healthy- and non-healthy firms.

Even though this bankruptcy model is innovative in its ability to dynamically assess bankruptcy as a path one can follow *throughout* a series of financial statements rather than it being a sudden event *following* a single statement, the ability of the Self- Organizing Map to capture and visualize relative differences between observations (companies) is limited. This limitation is caused by the fixed, discrete locations of the neurons in the map, a characteristic inherent to SOM. If one consequently clusters a series of these discrete, restricted locations as done in the bankruptcy SOM model as explored by Schreck et al. (2007), Du Jardin and Séverin (2011) and Chen et al. (2013) this could result in a substantial loss of available information affecting the cluster quality.

Therefore, this paper additionally investigates the potential of a tandem (sequential) approach between the Self-Organizing Map and Multidimensional Scaling (MDS) to aid visual analysis and improve the cluster quality of trajectories by capturing additional information on relative distances between observations. Whereas SOM and MDS are often presented as competing techniques for dimension reduction, a combination of the two techniques could enhance the trajectory SOM model by optimizing the location of each neuron in the map using MDS after configuration of the SOM. By releasing the restriction of the fixed position of these neurons a more intuitive visualization of relative distances between observations can be realized and furthermore additional information is stored in the same amount of dimensions as the regular SOM. It is therefore expected the tandem approach gives additional and more accurate insight in the differences between financial developments of bankrupt- and healthy companies.

The analysis of bankruptcy using a (tandem) SOM approach is unique in its ability to capture both cross-sectional and time-varying characteristics of companies without the need to specify an explicit functional form, yet still maintaining the ability to assess individual feature performance. This is different to older, more traditional models such as the famous Z-score proposed by Altman (1968) or the conditional probability models (i.e. Ohlson (1980), Chi and Tang (2006)) who treat bankruptcy as a static event and rely on cross-sectional variance only. It is also different to pure machine-learning based techniques such as Geng et al. (2015), Leshno and Spector (1996) who are not bounded by a restrictive functional form, but pay little attention to individual feature importance, providing mostly black-box predictions. The experimental results show promising evidence for the performance of the SOM-MDS tandem to capture and visualize upcoming bankruptcy for U.S. listed firms one year prior to the event and could therefore serve as a highly relevant tool for decision support in going concern analysis or any early warning system on financial distress.

The remaining sections will be structured as follows. First, a short recap on the related literature in the field of bankruptcy prediction is presented in Section 2. Following, in Section 3 the data is presented and some of its characteristics are highlighted. In Section 4.2 the Self- Organizing Map algorithm and the trajectory SOM approach are presented, along with their advantages and limitations. In Section 4.3 the SOM-MDS tandem approach is introduced, after which Section 4.4 explains how the tandem trajectory approach is applied to the data. Section 5 shows the experimental results and finally Section 6 and 7 present a summary of the research, its contribution to the literature and a proposition for future research steps. Additionally, for the interested reader Appendix IV shows how the (tandem) trajectory SOM approach relates to the industry standard approach: the Altman Z-score.

## 2 Literature Review

The analysis of corporate failure is an elaborately researched topic within the financial literature. This field of study dates back to the paper by FitzPatrick (1932), who first introduced the concept of identifying differences between successful and failed firms by comparing some of the line items on their financial statements. About thirty years later, Beaver (1966) formalized this notion, introducing a simple bankruptcy prediction model, classifying a firm as healthy or non-healthy using Univariate Discriminant Analysis. The field of corporate failure analysis truly put itself forward with the publication of the influential paper by Altman (1968), introducing the now well known Z-score. Altman extended Beaver's univariate model to a multivariate setting, mapping firms into eiter a "safe-", "grey-" or "distressed" zone based on linear combinations of five financial ratios (working capital/total assets, retained earnings/total assets, EBITDA/total assets, market value of equity/total liabilities and sales/total assets). Even though many more advanced techniques have been introduced since, Altman's Z-score (or its renewed version, the Zeta-score, see Altman et al. (1977)) is still frequently used in practice and serves as a benchmark model in many academic papers.

After Altman, a lot of papers were published extending the applicability of the Z-score (originally targeted at large U.S. manufacturing companies) to other industries or settings, such as small firms Edmister (1972), industrial companies Deakin (1972), Blum (1974) and stockbrokers Altman and Loris (1976). As more papers explored bankruptcy using Discriminant Analysis, the technique also became subject to increased criticism directed at the restrictive assumptions underlying the technique. For example, Multiple Discriminant Analysis requires predictors to be multivariate normally distributed and the covariance matrices of healthy and non-healthy firms to be homogeneous.

As a response to this critique a new set of models was put forward based on conditional probability. Ohlson (1980) was the first to apply a Logit model to distinguish between failed and non-failed firms. Similarly, Zmijewski (1984) constructed a Probit model to solve the task. Using a logistic distribution results in much simpler calculations as compared to using a cumulative normal one. Therefore, especially the Logit model was adapted by many researchers after Ohlson, for example Zavgren (1985), Peel and Peel (1987), Swanson and Tybout (1988), Johnsen and Melicher (1994).

Even though an unknown data distribution is less problematic for conditional probability models, they remain highly sensitive to multicollinearity, outliers and missing values. In general, financial statements are of high dimensionality (there are many line items in financial statements), have questionable data quality (especially for private firms whose statements do not have to be checked an independent auditor) and often contain missing values. It is therefore very difficult to make assumptions regarding any functional form when using such statements in a bankruptcy prediction model.

During the 1990s, research within the field of machine learning shifted from a knowledge-based approach to a data-driven one, which re-imagined the possibilities of the techniques and caused the models to become more popular ever since. As machine learning models usually require very little to no assumptions about the data generating process, many researchers investigating corporate failure readily adapted a variety of these techniques to gain insight into corporate failure. Odom and Sharda (1990) used the same five financial ratios as Altman to predict bankruptcy, but this time using a Neural Network to distinguish between failing and successful firms. More Neural Networks were tested by e.g. Altman et al. (1994), Leshno and Spector (1996) and Atiya (2001). Other examples of applied techniques were Decision Trees Frydman et al. (1985), Yeh et al. (2014), Support Vector Machines Min and Lee (2005), Wang et al. (2005) or cluster algorithms Telmoudi et al. (2011). With the introduction of this range of new approaches many comparative studies also emerged, comparing different machine learning techniques to each other (e.g. Geng et al. (2015), Martens et al. (2008)) or to the more traditional Z-score and conditional probability models (e.g. Yim and Mitchell (2003), Heo and Yang (2014)).

Although the papers mentioned above vary greatly in their econometric approach of modelling corporate failure, the set-up of their research is very comparable. All papers treat bankruptcy as as a static event, marking the end of a firm's life cycle. As bankruptcy is a legal event which can be directly observed, the model likelihood can be presented as a binary outcome with bankrupt and non-bankrupt firms separated in an exact way, which (hopefully) can be predicted by a predetermined set of predictors. In other words, these papers aim to predict bankruptcy at time $t$ using the statements from $t_{-1}$ and maybe $t_{-2}$ as predictors. The number of financial statements analyzed are often limited and especially the dynamics between them are disregarded. One exception to this are the so-called duration models for bankruptcy analysis. Shumway (2001), Kim and Partington (2015) and Cole et al. (2009) used a dynamic hazard model (with time-varying covariates) to predict the moment of bankruptcy for each firm. In contrast to the other models, the timing of bankruptcy is explicitly predicted by the dynamic hazard model rather than being implied by the model. However, it remains a static event.

A rather different perspective on bankruptcy is to approach it as a process, of which the legal filing of bankruptcy is the final stage. This corresponds to a view on company failure as a path which one can follow *throughout* multiple financial statements rather than it being a sudden event *following* a statement. This notion seems compatible with the reality of corporate failure, where most firms suffer through a period of financial difficulty before filing for legal bankruptcy. This period of financial suffering however takes many shapes and sizes and therefore it seems unnatural to assess bankruptcy following a uniform failure process. One should also keep in mind that many seemingly distressed firms actually recover and find themselves to be viable again at a later stage. Therefore, analyzing differences between healthy and non-healthy developments, but also differences within non-healthy developments can yield great additional insights into the processes behind company failure. Furthermore, by analyzing financial statements in such a way, one implicitly takes into account the previously disregarded dynamics *between* the statements. Consequently, even though assessing company failure as a dynamic process complicates the definition and categorization of bankruptcy, it seems reasonable and necessary to assume that the changes between- and development in financial statements are informative for the assessment of a firm being bound to file bankruptcy or not.

Despite their potential and the repeated call for their necessity, (see e.g. Agostini (2018), Balcaen and Ooghe (2006)) dynamic bankruptcy analysis has received relatively little attention in the literature. Laitinen (1991) was one of the first to introduce the rationale of different bankruptcy processes for different firms, emphasizing that an equal deterioration of a financial ratio has different implications for firms on different failure paths. Lukason (2012) compared simple changes in mean values of financial line items in the years preceding bankruptcy of companies in Estonia to see whether differences were detectable between firms of different size or in different industries. Ooghe and De Prijcker (2008) tried to define different failure processes, aiming to identify segments in the observed trajectories. Even though these papers touch upon the notion of bankruptcy as a process, they rely on qualitative analysis and are therefore not comparable to the quantitative yet static approaches discussed earlier in this section.

An unique exception to this is a quantitative approach using Self-Organizing-Maps as presented in Deboeck and Kohonen (1998), Du Jardin and Séverin (2011) and Chen et al. (2013). A Self-Organizing Map (also known as Kohonen map, named after Kohonen (1990) who first introduced the methodology) is an unsupervised, single-layer neural network which produces a low-dimensional, discrete representation of the input space by preserving the original topology of this input space as much as possible. After the introduction by Kohonen it has been used for applications in many different fields, ranging from medicine Bauer and Schöllhorn (1997), Lagerholm et al. (2000) to biology Törönen et al. (1999), Xiao et al. (2003) to text processing Honkela et al. (1997), Honkela et al. (1996) and economics Kourtit et al. (2012), Kuo et al. (2002). Its popularity can be attributed to SOM's ability to simultaneously reduce the dimensions of the input space and cluster observations, presenting observations in readily divided buckets.

To use the attractive properties of SOM to model bankruptcy as a process, Deboeck and Kohonen (1998), Du Jardin and Séverin (2011) and Chen et al. (2013) have combined Self-Organizing Maps with a rationale taken from trajectory analysis. Trajectory analysis is concerned with the analysis of object movements, describing the course of a measured variable over time. It is therefore popular in fields as visual surveillance Wang et al. (2013), traffic monitoring Wang et al. (2013) or robotic navigation Tellex et al. (2011). The main idea of trajectory SOM is to project financial statements of companies over time on a map, such that a firm is no longer represented by a single neuron on the map, but by a sequence of neurons on the map over time. Du Jardin and Séverin (2011), Chen et al. (2013), Deboeck and Kohonen (1998) find that analyzing or clustering these sequences yield valuable insights, not only into differences between healthy- and non-healthy companies, but also between a variety of different failure paths. Additionally, the trajectory SOM can assist in the identification of some of the dynamic drivers behind company failure.

# 3 Data

Data were taken from the Compustat Capital IQ database via the Wharton Research Data Service (WRDS). Specifically, all information from the period 2010 - 2018 was retrieved from the North-American daily updated database. This database comprises of aggregated, quarterly financial statement representations of all US listed companies obliged to report at the Securities and Exchange Commission (SEC), an independent agency with primary responsibility for the enforcement of federal security law on the stock- and option exchanges in the U.S.

In this section, the data preparation procedure is presented. First, in Section 3.1, the collection and construction of the bankrupt sample is presented. Following, Section 3.2 explains how the sequences of financial statements for each company were assembled and prepared for analysis. Third, in Section 3.3 missing values and data transformation are discussed and finally Section 3.4 presents some descriptive statistics.

## 3.1 Bankrupt sample selection

The construction of an adequate and representative sample relied on three important decisions regarding the focus of listed versus private companies, the retrieval process of the bankrupt sample cases and the timeframe from which to select the sample cases.

Analyzing listed companies in a bankruptcy prediction study, as opposed to analyzing private firms, substantially restricts the amount of available bankrupt sample observations. To illustrate, in the U.S. 89,686 private companies went bankrupt in 2018 (Economics (2018)) compared to only 58 public firms (BankruptcyData (2019)). From the auditor's perspective, it are however especially the large, listed companies which receive a lot of publicity. Therefore, so will the audit of their financial statements. Insights in failure drivers of listed companies are therefore of special importance to the auditor.

Additionally, from a more theoretical perspective, the analysis of listed statements should be more representative and reliable. In most countries, private companies without publicly traded debt are not obliged to report financial statements, nor do these statements need to be audited. For private companies one can therefore merely sample from firms who voluntarily report statements, whereas the study of listed firms allows the analysis of the entire "population" of listed companies. Additionally, listed statements are more reliable as they are verified by an independent auditor. For these reasons, listed company data was preferred in this study. The U.S. was selected specifically as the SEC allows public access to all fillings they receive through their Electronic Data Gathering, Analysis and Retrieval (EDGAR) platform.

In order to retrieve the bankrupt sample cases, a clear definition of bankruptcy had to adapted. In the United States a company in financial distress can file for two different types of bankruptcy, chapter 7 and chapter 11. Firms file for chapter 7 if there is no option for reorganization and is therefore also known as liquidation bankruptcy. If a company files for chapter 11 the firm still receives the option to try and reorganize its debt and possibly resurface as a healthy company. According to BankruptcyData, a large corporate provider of bankruptcy- and distressed securities publications, 307 US listed companies filed for either chapter 7 or 11 between 2015 and 2018 BankruptcyData (2019). Chapter 7 and 11 fillings were treated equally to minimize restrictions on the scarce selection of bankrupt sample cases.

Even though the number of listed bankruptcies is publicly available, after thorough research it was concluded that no exhaustive list of the representative company names is freely available for research. Therefore, as a best alternative bankruptcy information was taken from two sources: the UCLA-LoPucki Bankrupty Research Database (BRD) and the S&P CapIQ Portal. The BRD is a project of the UCLA School of Law to promote bankruptcy research and contains information on more than 1000 large public companies who filed for bankruptcy since 1979. CapIQ is a product of the research division of Standard&Poors and is created to serve as a leading provider of financial services research. Through sample access (BRD) and PwC's license (S&P) a list of total 206 bankrupt companies between 2015-2018 was retrieved. This list was matched with the company data from the Compustat Capital IQ database of WRDS. Even though WRDS claims to include all listed US companies reporting at the SEC, data on only 161 of extracted bankrupt companies was available in WRDS.

As mentioned in the previous paragraph, bankrupt companies were selected from the period 2015-2018 and treated equally. Adding additional years to the sample would have increased the number of bankrupt sample cases, but also created additional aggregation bias (for further explanation, see Section 3.2). Using 2015 as a cut-off yields two main advantages. First, this year coincides with the FED raising the federal fund rate from zero for the first time in almost seven year, stating the economy was on a path of sustainable improvement (FED 2015). It was also the year the U.S. unemployment rate returned to its natural level (Bureau of Labor Statistics (2015)). Second, in light of the 2008 financial crisis, it was considered unfavourable to compare statements from before 2010 (when the effects of the crisis were still felt strongly in the U.S. economy) to those after. As the research is concerned with analyzing the developments of companies over time, analyzing a company who filed for bankruptcy in 2015 also entails the study of its statements from before that time. With some generalization, adopting the 2015 cutoff guarantees the selection of a set of companies who went bankrupt in a relatively comparable, healthy financial climate (2015 and further) whose past trajectories can be analyzed up till five years back without including direct crisis effects (2010 and further).

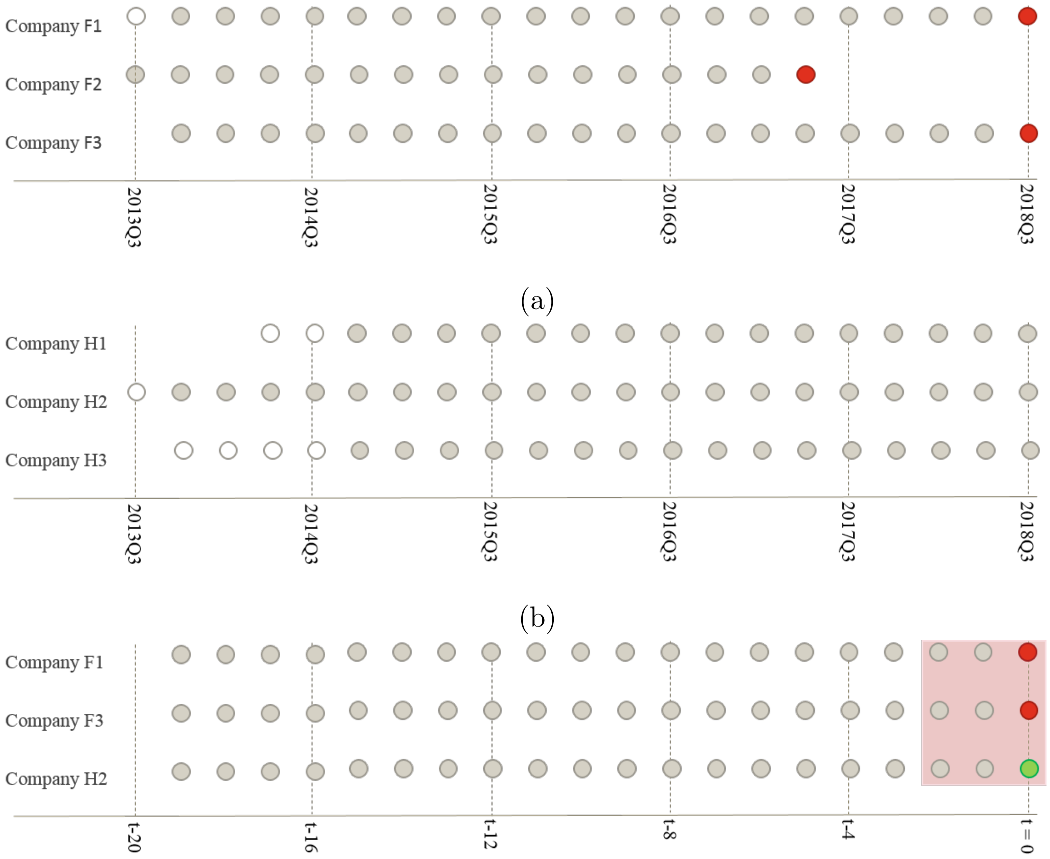
## 3.2 Constructing financial statement sequences

To analyze the financial statements of a company over time, a set of subsequent quarterly statements had to be selected for each company to form a sequence. Figure 1 illustrates this process.

In the upper panel (a) a selection of possible five-year bankrupt sequences are depicted. The red dot illustrates the last statement before bankruptcy and all grey dots the statements prior to this. As required for a five-year trajectory, all companies have at least 20 quarters of data available. In the middle panel (b), a selection of 'healthy' sequences are depicted: companies who are currently in business and have been for at least five year. In the lowest panel (c) an example of the sequence structure as used in the analysis is depicted. It shows three equal-length sequences, two of companies going bankrupt in the quarter following $t$ (F1 and F2, indicated by the red dot) and one who will not (H2, marked with a green dot). In this last panel the three most recent statements prior to bankruptcy are red shaded, as they are not included in the analysis. Going concern is about assessing the viability of a company in the upcoming twelve months: indicators of bankruptcy in quarters later than $t_{-3}$ are therefore 'too late' and not of interest.

To convert the upper two panels into the lower one, a few adaptions were made. First, the statements from calendar quarter 2018Q3 were selected for all healthy companies as the last available statement. The 19 preceding records were selected as well to form a five-year quarterly sequence. As can be seen, statements dating form before the start of the sequence (the lighter-grey dots in the upper two panels) are disregarded. Therefore, for the healthy companies, the most recent statement dates from 2018Q3, the second most recent from 2018Q2 and so on. For the bankrupt companies, the last available statement before bankruptcy was extracted, regardless of the calendar quarter or year, along with the 19 preceding records to also construct a five-year sequence.

It is important to understand that this approach makes some generalization: the bankrupt samples were moved down the timeline as if they had all gone bankrupt in 2018Q3. Even though this is a somewhat artificial adjustment, it was necessary because there are simply not enough bankrupt sample cases in a single quarter to perform a comparative analysis with healthy companies. By carefully considering the timeframe in which this time relocation took place (see Section 3.1) the effects of this generalization are deemed minimal.



(a)

(b)

(c)

Figure 1: Graphical representations of possible trajectories for (a) companies who went bankrupt between 2015 and 2018 and (b) companies who remain in business in 2018. Panel (c) depicts the trajectory structure as used in the analysis after transforming data from panel (a) and (b) following the procedure described in Section 3.2

Finally, the data was adjusted for seasonality. This was necessary, as though the healthy companies come from the same calendar quarter, the fiscal quarter corresponding to each statement can be different per company. Because companies have a tendency to boost their Q4 reports and are less concerned with their Q1 performance, the fiscal quarter potentially has an influence on the financial statement. Therefore, for each line item on each firm's statement, the value was regressed on the quarterly dummies and for those significant quarters, the data was corrected by the coefficient (note that this correction took place after the log-transformation of the data but prior to imputation - see Section 3.3).

Companies were disregarded from the analysis if one or more of the statements in the trajectory was not available. This was necessary as it is impossible to deduce from the data alone whether this statement was missing because the company failed to file a statement, WRDS failed to include it in the database or if there was some other reason for the statement to be missing. Following the same argumentation, bankrupt observations whose bankruptcy filing date and last filed statement prior to bankruptcy were more than 120 day apart (one quarter + 30 days of a granted 'delay' period) were also disregarded. Lastly, the few companies who report according to a half-year cycle were dismissed, as well as all financial companies. The latter is common practice in bankruptcy research and necessary as their balance sheet adheres to different standards and are therefore hard to compare them to other companies without substantial adjustments.

## 3.3 Data transformation and missing values

The WRDS database provides information on 319 different quarterly line items from either a firm's balance sheet, profit- and loss statement or cash flow statement. As many items are industry- or even firm specific, the number of items available per firm changes substantially. Additionally, the data contains quite some missing values.

First, to address the scattered nature of the data, the data were first transformed using a log-transformation. Let $\mathbf{x}_{ij}$ be a vector containing the financial line items of firm $i$ at time $j$. Each element $\mathrm{x}_{ij}$ (each line item value) in the financial statement was transformed according to

$$\mathrm{x}_{ij} = \begin{cases} \log(\mathrm{x}_{ij} + 1) & \mathrm{x}_{ij} > 0 \\ -\log(\mathrm{x}_{ij} - 1) & \mathrm{x}_{ij} <= 0 \end{cases} \tag{1}$$

Because all items are measured in (millions of) dollars, there is no direct need to normalize the data. In fact, doing so proved unproductive as the differences between the big group of smaller companies became too limited for the SOM algorithm to capture. Therefore, no normalization was applied.

To handle missing values, a trade-off was made between the required percentage of line items available and the amount of remaining variables left for analysis. In Figure 2, the number of variables remaining for analysis is depicted against different availability requirements. To illustrate, if one requires a variable to be available in at least 60% of the financial statements from all companies at all times, 145 variables remain available for analysis. From the Figure three important things can be noted. First, many variables are not even available for 10% of the observations due to the industry- or firm specific characteristic of some financial line items. Second, increasing the percentage availability requirement from 10 % to 50 % results in disregarding 46 variables, whilst increasing the requirement from 50 % to 90 % results in an omission of only 22 additional variables. Third, the same amount of variables remain available with a requirement of 80 % or 90%.

As comparing companies on variables who are not even available half of the time seems unprofitable and obtaining a very attractive availability rate of 90% only requires the omission of relatively few additional variables past this point, 90% was deemed the most favorable availability requirement and 110 variables were initially selected for analysis

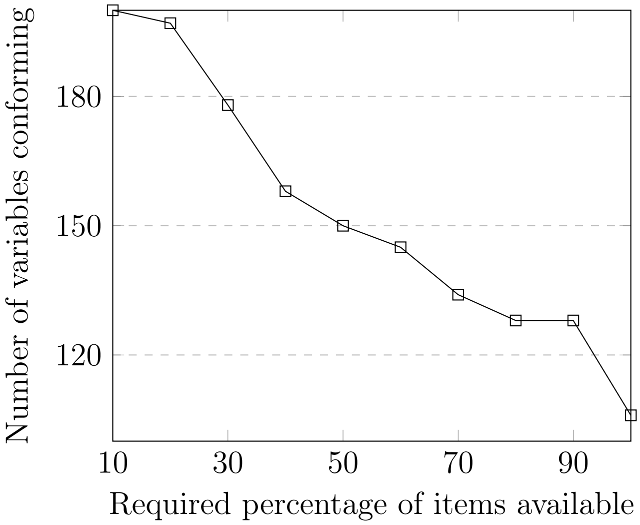


Figure 2: Graph representing the number of predictors available after deletion of variables failing to meet the percentage availability requirement

The remaining missing values were imputed using a using $k$-nearest-neighbor approach with $k = 6$. In this approach, first the six closest observations to the observation with the missing value are identified. Consequently, the missing value is imputed with the median value of this variable from the six nearest neighbors. Although $k$nn is a relatively basic, single-imputation approach, it is superior to mean- median- mode imputation as it leaves the multivariate structure of the data untouched. More importantly, the functionality of $k$nn is fairly similar to that of SOM. Both rely on (Euclidean) distances between observations to determine the level of similarity between them and use this sense of closeness to group observations. If a SOM has a hexagonal grid, each neuron has 6 adjacent neurons. Observations in neighboring neurons are considered similar and therefore, the number of neighbors was set to 6 to match this characteristic.

## 3.4 Descriptive statistics

In total, 3748 listed US companies were included in this study, of which 74 went bankrupt. In Table 1 the industry distribution according to the NAICS classification system is depicted (NAICS association (2017)). Due to the limited occurrences of bankruptcies and the restricted identification of them (see Section 3.1) not all industries are represented in the bankrupt sample. Nonetheless, the dominant industries in the healthy sample

("Manufacturing" and "Mining") were also dominant in the bankrupt sample. Additionally, most larger industries (with more than 100 firms) had one or more bankrupt sample counterparts, with the exception of "Real Estate and Leasing" and "Professional, Scientific and Technical Services". As the research is not focused on industry specific performance and the data were corrected for seasonality, the distribution was deemed satisfactory.

After following the variable selection procedure described in Section 4.4.1, 42 line items were found most important for distinguishing healthy- and non-healthy trajectories using SOM. In Table 4 in Appendix I the minimum, maximum, median and mean absolute deviation of each of these line item are depicted. The most important predictors were *Retained Earnings, Long-Term Debt, (Long-Term) Liabilities, Total Assets, Pretax Income* and *Invested Capital.*

Table 1: Overview of the industry distribution of both healthy and bankrupt companies

| NAICS | Description | Healthy | Bankrupt |
|---|---|---:|---:|
| 11 | Agriculture, Forestry, Fishing and Hunting | 13 | 0 |
| 21 | Mining, Quarrying, and Oil and Gas Extraction | 501 | 32 |
| 22 | Utilities | 191 | 1 |
| 23 | Construction | 63 | 0 |
| 31-33 | Manufacturing | 1500 | 20 |
| 42 | Wholesale Trade | 99 | 1 |
| 44-45 | Retail Trade | 134 | 6 |
| 48-49 | Transportating and Warehousing | 140 | 6 |
| 51 | Information | 385 | 2 |
| 53 | Real Estate and Leasing | 242 | 0 |
| 54 | Professional, Scientific and Technical Services | 121 | 0 |
| 56 | Administrative Support and Waste Mamagement | 73 | 0 |
| 61 | Educational Services | 14 | 1 |
| 62 | Health Care and Social Assistance | 58 | 3 |
| 71 | Arts, Entertainment and Recreation | 27 | 0 |
| 72 | Accomodation and Food Services | 71 | 0 |
| 81 | Other services (except Public Administation) | 8 | 0 |
| 99 | Other | 34 | 2 |
|  | **Total** | **3674** | **74** |

# 4 Methodology

This section will be structured as follows. First, the Self-Organizing Map is presented. Secondly, the specific application of Self-Organizing Maps for trajectory analysis is explained, along with its advantages and limitations. Third, the multi-map SOM-MDS tandem approach for trajectory analysis is introduced. Fourth, the way this approach has been applied in this research specifically is presented. Fifth and lastly, the performance measures for this research are presented.

## 4.1 Self- Organizing Maps

A Self-Organizing Map (SOM) is an unsupervised learning technique proposed by Kohonen (1990). The SOM is a neural network which generates a low-dimensional representation of the (higher-dimensional) input space, operating as a dimension reduction- and visualization tool. In it's essence, SOM aims to assign each observation to one of the neurons in a grid of pre-determined size and shape and organize them in such a way that observations mapped to neurons close to each other are more similar than those far removed. The intuition behind SOM is very similar to the construction of Voronoi regions Voronoi (1908), where a plane is divided in a predetermined number of regions and each region has one central site, located such that all points in the region are closer to that region's site than any other. SOM is a topology-preserving technique, meaning the neighborhood relations in the original space are retained, to the extent possible, in the projection space. The reduction of dimensions is therefore based on distance-preservation and thus similar to for example Multidimensional Scaling, but different to some other popular dimension reduction methods like Principal Component Analysis, which reduces dimensions whilst retaining as much of the original correlation structure as possible.

To formalize the technique in the context of this research, let $n$ be the number of companies, observed at $t$ points in time. Let $\mathbf{x}_{ij}$ be a $k$-dimensional vector of company $i$ at time $j$, representing $k$ items on the financial statement ($k$ predictors) with $i = 1, .., n$ and $j = 1, ..., t_{-2}, t_{-1}, t$. Furthermore, let $r$ be the number of neurons on a $p$-dimensional grid (with $p \leq k$) and let $\mathbf{l}_s$ be a $p$-dimensional location vector on the grid for each neuron $s = 1, .., r$. The position of each neuron on the grid is discrete, with each element of $\mathbf{l}_s$ belonging to the set of non-negative integers ($\mathrm{l}_s \in 0 \cup Z^+$). Next to a location index, each neuron also has a $k$-dimensional weight vector $\mathbf{w}_s$, for which initial values have to be specified by the researcher. For each iteration $u$, one must also specify a learning rate $\alpha_u$ with $0 \leq \alpha_u \leq 1$ and a neighborhood function, denoted by $\hbar(\mathbf{l}_s)$. The neighborhood function determines which neurons belong to the group of weight vectors to be updated, whilst the learning rate regulates by how much they are updated.

The SOM algorithm as proposed by Kohonen (1990) can be described as follows

1. Initialize $\mathbf{w}_s(0)$, the weight vectors at iteration $u = 0$ for each neuron $s$, set the learning rate $\alpha_0$ and pick a neighborhood function $\mathcal{h}(\mathbf{l}_s)$.

2. Randomly select one observation vector $\mathbf{x}_{ij}$

3. Calculate the distance between observation $\mathbf{x}_{ij}$ and each of the weight vectors $\mathbf{w}_s$ ($\delta_{ijs}$) and store it in the set of distances $\mathcal{D}_{ij}$ by

$$\delta_{ijs} = ||\mathbf{x}_{ij} - \mathbf{w}_s|| \tag{2}$$

4. Denote the neuron associated with ($\sim$) the smallest distance in $\mathcal{D}_{ij}$ as the best matching unit ($s^*$) to observation $\mathbf{x}_{ij}$

$$s^* = s \sim \min(\mathcal{D}_{ij}) \tag{3}$$

5. Update the weight vector of the best matching neuron $\mathbf{w}_{s^*}$ along with the weights of neurons in $s^*$' neighborhood, determined by the neighborhood function $\mathcal{h}_{s^*}(\mathbf{l}_{s^*}, \mathbf{l}_s)$, a function of the location vector of $s^*$ and the other neurons. The learning rate $\alpha_u$ determines by how much the weight vectors are updated, i.e. how much the neuron's weights change to be more like the observation. The update for each weight vectors $\mathbf{w_s}$ is computed according to

$$\mathbf{w}_s(u+1) = \begin{cases} \mathbf{w}_s(u) & r \notin \mathcal{h}_{s^*} \\ \mathbf{w}_s(u) + \alpha_u(\mathbf{x}_{ij} - \mathbf{w}_s(u)) & r \in \mathcal{h}_{s^*} \end{cases} \tag{4}$$

6. Empty set $\mathcal{D}_{ij}$ and update the learning rate. Repeat step 2 till 5 till all observations have been presented at least once and the weight vectors $\mathbf{w}_s$ converge, or the iteration limit is reached.

7. Store the SOM results in $\Phi = \{\mathcal{W}, \mathcal{L}, \mathcal{M}\}$, where $\mathcal{W} = \{\mathbf{w}_1, ..., \mathbf{w}_r\}$, $\mathcal{L} = \{\mathbf{l}_1, ..., \mathbf{l}_r\}$ and $\mathcal{M} = \{(1, \mathbf{l}_{s^*1}), ..., (n, \mathbf{l}_{s^*n})\}$ are the set of weight vectors, location vectors and mapping of observations to their best matching neurons respectively.

*The algorithm can also be run in a batch fashion. This means step 2-4 are repeated for each observation $i$, before updating all weight vectors simultaneously in step 5. Step 2-5 are repeated till the SOM converges.

As described in the algorithm, the positions of the neurons on the map $\mathbf{l}_s$ are fixed and only the weight vectors are updated in each iteration. The algorithm should result in a grid where the weight vectors of adjacent neurons are more similar than those on for example opposite sides of the grid. It is important to note that the grid itself does not move and $p$ does not prescribe the dimensionality of the weight vector (which is always $k$), but merely the amount of neighbors a neuron has. After all, the higher the dimensionality, the more sides a neuron has to attach neighbors to. The number of neurons, the shape of the grid and the initialization of the weight vectors are therefore important parameters to be set by the researcher. Running the algorithm in a batch fashion eliminates the importance of the order in which observations are presented to the map and reduces computational power. It does increase the influence of the initialization of the weight vectors on the final convergence of the SOM.

## 4.2 Trajectory analysis using Self-Organizing Maps

### 4.2.1 Analyzing bankruptcy trajectory using SOM

The Self-Organizing Map can be used to detect similarities and differences between companies by observing common- and opposing features between observations mapped close together (e.g. in the same neuron) or far apart (on opposite sides of the grid). Ideally the SOM will group together companies likely to go bankrupt, as they show similar detectable symptoms and companies who will not, as they lack these symptoms. Loosely speaking, the SOM would then result in a grid with 'dangerous areas' (neurons with relatively many bankrupt companies associated to them compared to healthy ones) as well as 'safe zones' (neurons associated with healthy companies only). The more bankrupt companies associated with a neuron, the higher the chance for a random observation associated with that neuron to be a bankrupt one and therefore, the more dangerous.

Schreck et al. (2007), Du Jardin and Séverin (2011) and Chen et al. (2013) used SOM to analyze changes in financial statements as a trajectory over time. In order to distinguish between firms going bankrupt at time $t$ and firms who do not, they constructed a SOM using a selection of financial ratios and line items from the financial statement a year prior to bankruptcy ($t_{-1}$). Following, the statements from the previous years $t_{-2}$ and $t_{-3}$ are also projected onto this map. This projection entails the process of locating the best matching neuron $\mathbf{l}^*$ for a firm at time $t_{-2}$ and $t_{-3}$ on the map of time $t_{-1}$ (note that the map itself not updated during this projection process). For each firm $i$, one ends up with a set of location vectors $\Lambda_i = \{\mathbf{l}_4, \mathbf{l}_6, \mathbf{l}_5\}$, where $\Lambda_i$ is now named the trajectory of company $i$. This particular trajectory would mean the best matching neurons for company $i$ at time $t_{-3}$, $t_{-2}$ and $t_{-1}$ were neuron 4, 6 and 5 respectively. Ideally, by analyzing these sequences of best matching neurons on the map, over time, it is possible to see whether a company is moving towards a dangerous area, whether it has just left a safe zone or if it is jumping all over the map, to name a few examples. The trajectories are used to cluster and categorize the trajectories into 'bad' and 'good' ones, i.e. trajectories associated with many firms going bankrupt versus trajectories linked to healthy companies only.

### 4.2.2 Advantages

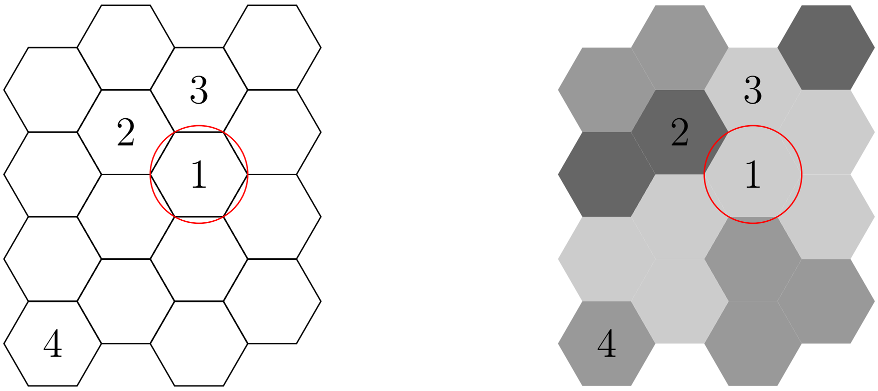
The rationale behind a trajectory approach is that by analyzing movements one can better distinguish between healthy and non-healthy firms as compared to clustering firms by a single point on the map. To clarify, imagine a company which has been moving around in a 'dangerous zone' on the map for two years but has now moved towards a 'grey' neuron in the third year. This grey neuron is not a particularly dangerous neuron, but it is certainly not completely save. Now compare this company to a firm which has always been in the same safe neuron, but suddenly makes a jump over the map to the grey neuron in question. If one would merely observe these two observations in the grey zone, both would be assessed as equally dangerous. However, knowing the companies' past movements, it seems that one firm is improving its position whilst the other company abruptly broke with its safe, stable past. The first is a movement to applaud, the second one a movement to further examine: what caused this disruption? Including the time-varying component additional to the cross-sectional element is therefore of utmost importance and an important benefit of the trajectory SOM approach.

Naturally, it would have also been possible to cluster the movements of a firm's financials as is. One can store the exact locations of each company in $k$-dimensional space, over time, and aggregate these movements in groups of 'good' and 'bad' ones. Note that this approach is the same as using each variable at each time as a separate predictor to classify a company as bankrupt versus non bankrupt. From the perspective of information available this would be more attractive than SOM, because SOM disregards quite some information by mapping each observation to a discrete summary point on a lower-dimensional grid. From a conceptual perspective, a SOM is however much more attractive for a number of reasons. First, if $p$ is set to two and one imagines the third dimension as time, one can visually analyze the development in the financials (actually 'see' the trajectory). Second, next to the mapping of observations to neurons, the $k$-dimensional weight vectors of these neurons can be interpreted as summary points of the original input space (they are analogous to the seeds in a Voronoi region). Therefore, they can give insight in the shape of the sample space and the distribution of observations throughout this space. Finally, by comparing the converged weight vectors of each neuron one can gain insights into why some neurons are more dangerous than others. The most distinguishing predictors between a 'dangerous' and 'safe' neuron are most likely important predictors of upcoming bankruptcy.

### 4.2.3 Limitations

As mentioned previously, SOM maps all observations to neurons in a fixed-size, fixed position grid. Throughout the iterations, the $k$-dimensional weight vector $\mathbf{w}_s$ of each neuron is updated, but the location $\mathbf{l}_s$ is set and is of dimensionality $p$. This location is furthermore discrete and every neuron is placed equally far from its neighbors, regardless of the differences in weight vectors. This particular characteristic of SOM has two main disadvantages, which are discussed below. Subsequently, an additional limitation of the trajectory approach as presented in Schreck et al. (2007), Du Jardin and Séverin (2011) and Chen et al. (2013) is discussed, which is not inherent to SOM but applies to the mapping process of the trajectories.

The first issue is illustrated in Figure 3. Depicted is a 4 × 4 hexagonal map, meaning each neuron has six neighbors and $p = 2$. The location vectors of the numbered neuron are given by $\mathbf{l}_1 = (3, 3)$ (third row from the bottom, third neuron from the left), $\mathbf{l}_2 = (3, 2), \mathbf{l}_3 = (4, 3)$ and $\mathbf{l}_4 = (1, 1)$. Let $k = 4$ and $\mathbf{w}_1 = (1, 2, 3, 4), \mathbf{w}_2 = (2, 2, 5, 6), \mathbf{w}_3 = (2, 2, 3, 4)$ and $\mathbf{w}_4 = (-3, -8, 0, 1)$. Imagine one merely observes Figure 3a. Following the logic behind SOM, one knows that companies mapped to neuron 1 are more similar to firms mapped to neuron 2 than to neuron 4, as 1 and 2 are neighbors whilst neuron 4 is much further removed. However, based on the location of the neurons alone, it is not possible to make a statement about the difference in closeness of neuron 1 to 2, and 1 to 3. All three neurons are adjacent and their location vectors are equally far removed from one another. From the weight vectors can be deduced that neuron 1 and 3 are closer, but this is not visually detectable on the map.



(a)

(b)

Figure 3: (a) Representation of neurons in an arbitrary 4 × 4 hexagonal grid (b) Visualization of relative distances between each neuron and its neighbors, also known as the U-matrix. The lighter the shade, the smaller the average distance to the neighbors.

To add information on relative distances between neurons, researchers often rely on the use of a U-matrix as depicted in Figure 3b (e.g. Deboeck and Kohonen (1998), Iturriaga and Sanz (2015)). The U-matrix uses color to indicate the closeness of each neuron, on average, to its neighbors. Neuron 2 has a darker shade than neuron 3, illustrating that its weight vector is further removed from that of neuron 1. In general, the lighter areas represent neurons that group well together (i.e clusters) and the darker neurons serve as separators. Even though the representation of the grid is now more insightful, note that these colors only give partial information as the shade is determined by an average. Furthermore, it seems cumbersome to visualize closeness with color when this can also be done by actually creating more or less distance between neurons.

The second issue with relative distances builds on the first limitation but is specific to applying SOM in a trajectory setting. In Chen et al. (2013) the trajectories consist of the sequence of best matching location vectors which are then categorized into 'good' and 'bad' sequences. By storing the movements over the SOM map in this way, the movement of a company from neuron 1 to neuron 3 would be equally far as a movement from neuron 1 to 2, whilst the distance travelled is much smaller. This information is disregarded by categorizing firms based on the movement in location vectors only. One may suggest to store the sequence of weight vectors for each company instead to include this information, but notice that this would result in $k \times t$ predictors for each firm. If one wants to use so many items, it is better to use (i.e. there is more information in) the 'raw' $k \times t$ variables taken directly from the financial statements rather than the summarized items from weight vectors (which are the same for every observation in the same neuron).

Finally the mapping process as proposed in Schreck et al. (2007), Du Jardin and Séverin (2011) and Chen et al. (2013) has a drawback. As explained in Section 4.2, the financial statements from time $t_{-2}, t_{-3}$ etc are all projected on the single SOM constructed with data from $t_{-1}$ only. With this single map projection they implicitly assume that the input space a year prior to potential bankruptcy is representative for the input space from which the previous bankrupt statements stem. This is a restrictive assumption: if simple things as for example the average size of companies change over the years, the map is no longer representative. Mapping observations to a grid that is too narrow or wide results in large discrepancies between the observations and their best matching neuron, thereby inaccurately reflecting the input space.

## 4.3 Trajectory analysis using a multi-map tandem approach

In Section 4.2.3 the limitations of trajectory analysis using SOM are discussed. In this paper a multi-map tandem approach of SOM and Multidimensional Scaling (MDS) is proposed to offset these limitations. For completeness, first MDS as a stand-alone technique is presented after which the SOM-MDS tandem is introduced.

### 4.3.1 Multidimensional Scaling

Multidimensional Scaling, like SOM, is a dimension reduction technique based on distance preservation first introduced by Torgerson (1952). MDS aims to visualize the level of (dis)similarity between observations by retaining the distance between observations in the input space, to the extent possible, in the lower dimensional projection. Imagine the set of weight vectors $\mathcal{W}_j$ of size $r$ from a certain converged SOM $\Phi_j$ of time $j$ to be the input observations. The MDS algorithm can then be described as follows

1. Let $\zeta_{syj}$ be the true dissimilarity between observations $s$ and $y$ at time $j$, for $s, y = 1, .., r$, computed by

$$\zeta_{syj} = ||\mathbf{w}_{sj} - \mathbf{w}_{yj}|| \tag{5}$$

   $\zeta_{syj} = \zeta_{ysj}$ and $\zeta_{ssj} = 0$ as the distance between observation $s$ and $y$ is equal to the distance between $y$ and $s$ and an observation is perfectly similar to itself.

2. Let $p$ be the desired dimensionality of the projection space (specified a-priori). Introduce $r$ $p$-dimensional vectors $\mathbf{q}_{sj}$, which are the randomly initialized coordinate points of the observations in the projection space at time $j$

3. Let the distance between observations in the projection space be given by

$$\mathrm{z}_{syj} = ||\mathbf{q}_{sj} - \mathbf{q}_{yj}|| \tag{6}$$

4. The optimal MDS projection, in its simplest form, is obtained by minimizing

$$\sigma_j = \sum_{s<y} (\zeta_{syj} - \mathrm{z}_{syj})^2 \tag{7}$$

   with $\sigma_j$ the so-called raw stress function, which measures the discrepancy between the distances of observations in the input- and the projection space. The MDS mapping is equal to those locations $\mathbf{q}_{sj}$ for which this discrepancy is the smallest. Minimization can be done is several different ways, but in the R smacof package this is done by majorizing Equation 7.

Instead of optimizing the raw stress function, the original distances can also first be transformed, such that $\zeta_{syj} \to \hat{\zeta}_{syj}$ according to some transformation. Several transformations are possible, of which the power- ($\hat{\zeta}_{syj} = \zeta^{c}_{syj}$), ratio- ($\hat{\zeta}_{syj} = b\zeta_{syj}$) and interval transformation ($\hat{\zeta}_{syj} = a + b\zeta_{syj}$) are only a few of the possibilities. Usually, the more free parameters in the transformation, the better the fit that can be obtained in the projection. If the original distances are transformed, the MDS solution can be found by an iterative approach, optimizing the transformation parameters and minimizing the raw stress function sequentially. It is also possible to add weights to the objective function, which allows the researcher to for example put more emphasis on accurately fitting highly reliable observations compared to less reliable ones.

### 4.3.2 SOM and MDS in a multi-map tandem approach

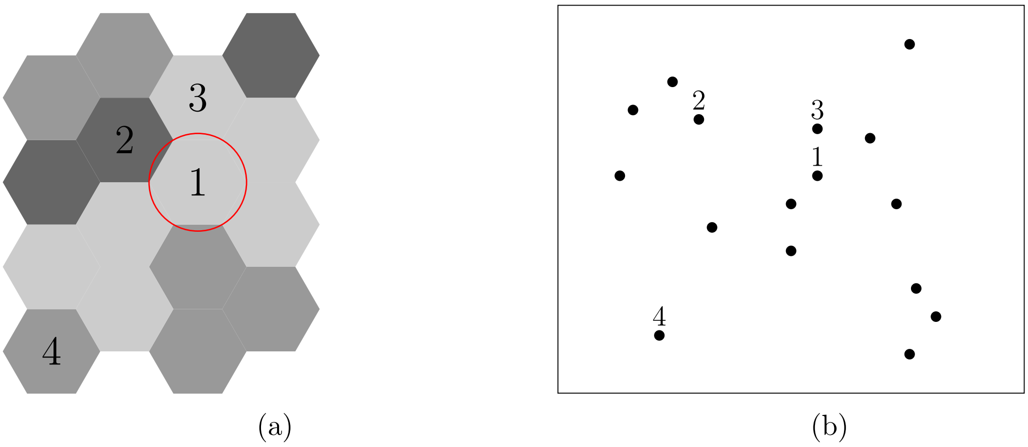
In the previous section attention was paid to the functionality of MDS as a stand-alone approach. Using its functionality, this section will explain how a tandem approach between SOM and MDS can offer a solution to the limitations of the trajectory SOM analysis (see Section 4.2.3). First, in the traditional SOM, the locations of the neuron in the map contain only limited information about the relative distances between neurons because they are fixed on the grid at discrete locations. The main idea of the SOM-MDS tandem is to use the weight vectors of each neuron, which do contain full information on relative distances, to re-position each neuron (i.e. change its location vector) after the SOM map has converged using Multidimensional Scaling. Second, the use of a single map to represent financial statements from different points in time, as done in previous research, could be too restrictive if the input space changes over time. Therefore this paper puts forward an approach using multiple maps. Both propositions will now be discussed in more detail.

The SOM and MDS tandem

In its essence, the SOM-MDS tandem consists of a relocation of the neurons on the SOM grid after convergence of the map. This relocation is equal to the MDS mapping of the weight vectors of each neuron. Because the weight vectors contain more information about the relative distances between neurons than the location vectors do, this should improve the visualization of relative distances between neurons and therefore observations.

After configuration of the SOM map as described in Section 4.1, the set of all weight vectors $\mathscr{W} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_s\}$ is extracted. Consequently, following the MDS methodology of Section 4.3.1 the distances between each element in $\mathscr{W}$ is computed. The optimal MDS mapping in $p$ dimensions is found by minimizing an interval transformed stress function, minimizing the discrepancy between the distances of neuron in the input- ($k$-dimensional weight vector) and projection space ($p$-dimensional location vector). Each neuron is assigned a new location vector $\tilde{\mathbf{l}}_s$, equal to the MDS projection of $\mathbf{w}_s$ in $p$ dimensions.

One should note that the new location vectors are now essentially a reduced version of the weight vectors, as opposed to a separate item. Using these optimized location vectors $\tilde{l}_s$ yields two main advantages. First, a visual inspection of the neuron's new position can give the researcher insights on the relative differences between neurons. This allows for better understanding of the input space and can furthermore serve as a quality check of the SOM. To illustrate this the U-matrix on the hexagonal grid from Figure 3b is reprinted in Figure 4a against the MDS representation of this same grid in Figure 4b. In the U-matrix, the darker shade of neuron 2 represented the fact that neuron 2 is more different from neuron 1 than neuron 3 is. In the MDS representation, this information is simply reflected by placing neuron 2 further out, a much more intuitive way to visualize differences than by the use of colors or the separate inspection of the weight vectors. Second, by storing the sequence of transformed location vectors $\tilde{\mathbf{l}}_s$ instead of the regular $\mathbf{l}_s$ , more information is captured in the same amount of dimensions (both are of dimensionality $p$). Because the location of the neurons are no longer restricted to be on a fixed and discrete location on the map, additional information on relative distances from the weight vectors can be transmitted to the location vectors. It can therefore be expected that clustering the optimized trajectories $\tilde{\Lambda}_i$ results in a more refined grouping of healthy- and non-healthy companies compared to using $\Lambda_i$, as there is more information available in the new trajectories.



(a)

(b)

Figure 4: (a) Representation of neurons in a 4 × 4 hexagonal grid in a U-matrix. The lighter the shade, the smaller the average distance to the neighbors.(b) Representation of neurons in a 4 × 4 hexagonal grid after optimization of their location using MDS

The SOM-MDS tandem approach has the potential to improve the quality of visual analysis and the cluster quality of the trajectories by capturing additional information on relative distances between observations compared to the traditional trajectory SOM approach. After thorough research, it was concluded the only papers who touched upon this notion were Deboeck and Kohonen (1998) and Törönen et al. (1999), who used a

Sammon mapping (an independent rediscovery of a particular variant of MDS) to assess the quality of the configured SOM map, but they did not emphasize the improvement such a mapping could yield for visual analysis. Additionally, the potential advancement in grouping quality of the trajectories has not previously been investigated.

The use of multiple maps

As mentioned in Section 4.2.3, the projection of all financial quarters on a single map is restrictive as it assumes the input space at time $t_{-1}$ to be representative for all other times $t_j$. In this research, it is proposed to train a SOM map for each quarter of data $j$ separately to accurately reflect the input space at all points.

One issue that comes with the use of multiple maps however is that of "label switching". To clarify, imagine a randomly initialized SOM is configured in such a way that two neurons in the upper left corner represent companies with high debt and a neuron in the bottom right represents companies with no debt. If one would run the algorithm again, it is not guaranteed that the high debt companies are in the upper left of the map again. For example, if the map is rotated by 180°, the SOM solution exactly the same, only the orientation is different. One can imagine that storing the location vectors of subsequent rotating maps is not that insightful as companies can jump all over the map without any change in their financial line items.

To guarantee a stable orientation of the maps, the weight vectors are initialized using the first two principal components of the data and the algorithm is run in a batch fashion. First, for each quarter of data $\mathbf{X}_j$ the first two eigenvectors- ($\mathbf{v}_{j1}$ and $\mathbf{v}_{j2}$) and values ( $\lambda_{j1}, \lambda_{j2}$) are computed. Second, the vector of column means $\mathbf{m}_j$ is computed. Then, the weight vector of each neuron is initialized as follows:

$$\mathbf{w}_{sj} = \mathbf{m}_j + \mathbf{v}_{j1}\lambda_{j1}\chi_s + \mathbf{v}_{j1}\lambda_{j1}\psi_s \tag{8}$$

with $\chi_s$ and $\psi_s$ depending on the location of the neuron on the x-axis and y-axis of the grid.

Naturally, the batch algorithm combined with PCA initialization will only solve the label switching issue if the first two principal components of each quarter of data are roughly in the same direction. Put differently: with this approach one assumes there is no structural break in the variables that explain most variance in the input space throughout the statements in the trajectory. The qualitative validation of this assumption is presented in Section 3.1, but naturally it is necessary to check the validity of this assumption in the data as well.

## 4.4 Research approach

Now the multi-map SOM-MDS tandem approach has been presented, this section will elaborate on how the approach has been applied in this research specifically. The procedure consist of two levels. The first level is referred to as the "base level" and entails the construction of a separate SOM map for each quarter of data. It additionally entails the optimization of the location vectors of each neuron on each map using MDS and the formation of the trajectories $\tilde{\Lambda}_i$ of best matching neurons on each map for each company. Consequently in the second step, referred to as the "meta level", the trajectories are grouped using an additional SOM. Furthermore, a Random Forest Breiman (2001) is trained to predict bankruptcy based on the trajectories only. The Random Forest is added as a predictive benchmark to assess whether MDS-optimized trajectories improve the predictive performance compared to the traditional SOM locations. For further details on the functionality of the Random Forest please see Appendix III. Below, first a broad overview of the two levels is presented after which subsequently the parameter details are presented.

### Base level

```
For all $j = 1, .., t_{-3}$

    1. Take the $n \times k$ input data matrix $\mathbf{X}_j$ containing $k$ financial line items of $n$ companies at time $j$

    2. Initialize $r$, $\alpha$, $\hbar$ and the weight vectors $\mathbf{w}_s$

    3. Perform the SOM algorithm as proposed in Section 4.1 in a batch fashion and store the output $\Phi_j$.

    4. Initialize $\mathbf{q}_{sj}$. Convert $\Phi_j \rightarrow \tilde{\Phi}_j$ by transforming $\mathcal{L}_j \rightarrow \tilde{\mathcal{L}}_j$. This is done by applying MDS on $\mathcal{W}_j$ as described in Section 4.3.1, with $\tilde{\mathcal{L}}_j$ equal to the MDS projection of the weight vectors. Update $\mathcal{M}_j \rightarrow \tilde{\mathcal{M}}_j$ by replacing all old location vectors of best matching neurons $\mathbf{l}_{i*}$ by their optimized MDS location vectors $\tilde{\mathbf{l}}_{i*}$.

    5. Store $\tilde{\Phi}_j = \{\mathcal{W}_j, \tilde{\mathcal{L}}_j, \tilde{\mathcal{M}}_j\}$

For all $i = 1, .., n$

    1. Construct the trajectory $\tilde{\Lambda}_i$ of the form $\tilde{\Lambda}_i = \{\mathbf{l}^*_{1s}, ..., \mathbf{l}^*_{t-3s}\}$, the sequence of best matching neurons in the maps $\tilde{\Phi}_j$ for $j = 1, ..., t_{-3}$
```

**Meta level**

The Trajectory SOM

*For all* $i = 1, .., n$

1. Take the $n \times t_{-3}$ input data matrix $\mathbf{\Lambda}$, constructed by pasting all transposed $\tilde{\Lambda}_i$ underneath each other.
2. Initialize $r$, $\alpha$, $\hbar$ and the weight vectors $\mathbf{w}_s$
3. Perform the SOM algorithm as proposed in Section 4.1 in a batch fashion and store the output $\Phi_{meta} = \{\mathcal{W}_j, \mathcal{L}_j, \mathcal{M}_j\}$.

Predictive benchmark: Trajectory RF

*For all* $i = 1, .., n$

1. Separate all observations in $n_{br}$ and $n_h$, companies who went bankrupt at time $t$ and those who do not. Construct a balanced sample of trajectories by randomly selecting a sample of size $|n_{br}|$ from the subset of all $\tilde{\Lambda}_i$ whose $i \in n_h$.
2. Train a Random Forest using 500 trees to classify trajectories belonging to either $n_{br}$ or $n_h$.

### 4.4.1 Base level parameters

The base level requires choices to be made on a series of parameters: the number of quarters considered ($j = 1, ..., t_{-3}$), the dimensionality of the grid ($p$), the number of neurons in the grid ($r$), the number of line items considered ($k$), the neighborhood function ($\hbar$), the learning rate ($\alpha(u)$) and the initialization of the MDS algorithm.

As discussed in Section 3.2, this research analyzes five-year quarterly trajectories. Therefore $t = 20$ represents the total number of quarters considered and $t$ is also the final statement before potential bankruptcy. The statements from time $t$, $t_{-1}$ and $t_{-2}$ are however disregarded from the analysis, as indicators of bankruptcy in quarters $j > t_{-3}$ are "too late" if one has to ensure going concern (meaning viable in the upcoming twelve months).

To decide on $p$, $r$ and $k$ several choices had to be made. For the dimensionality of the projection space, it makes sense to set $p = 2$, as projection of the financial statements on two dimensions allows for clear, visual comparison of the trajectories of different companies and is therefore highly attractive if one wishes to use the maps as a decision support tool for going concern. As the weight vectors of the neurons are initialized by the principal components of the data, one can depict the number of principal components against their proportion of total variance in the data explained to gain more support for the choice of $p$. If one can detect a cut-off value or 'elbow' at the $(p_{+1})^{th}$ dimension, it could make sense to only include the first $p$ dimensions, as the $(p_{+1})^{th}$ dimension (relatively) does not contribute as much towards increasing total fit.

Once one has decided on $p$, the number of neurons along each dimension in $p$ has to selected. Let us assume $p = 2$. If the first principal components has a relatively larger contribution to total fit than the second, it makes sense to allow for more freedom on the first axis compared to the second and therefore select a rectangular grid instead of a square one. Note that the number of neurons along each dimension also determines the total number of neurons on the grid $r$, as there are no empty spots on the grid. Additionally, optimal $r$ is codependent on the amount of variables, because additional variables bring additional variation to be captured in the neurons. Therefore, the number of neurons along each dimension is also a trade-off between the aggregation level and the similarity of each observation to its best matching neuron, meaning the optimal choice is research specific. For some studies a simplified, easily digestible version of the input space may be preferred, whilst other studies are interested in more subtle differences between observations, for which more neurons are required.

Due to these dependencies, an iterative approach was used to decide on $k$ and the number of neurons in each dimension $r_x$ and $r_y$ (assuming $p = 2$).

1. For all $j$, compute the first two principal components of $\mathbf{X}_j$ and store the absolute factor loadings of each variable, denoted by $\bar{f}_k$. Order the average factor loadings from low to high.

2. Define a set of potential $\{r_{xb}, r_{yb}\}$ and $\{r_{xm}, r_{ym}\}$ combinations, indicating the number of neurons among each dimension in the base-level and meta-level SOM.

3. Using all variables, perform a grid search on all possible $\{r_{xb}, r_{yb}\}$ and $\{r_{xm}, r_{ym}\}$ combinations and assess the performance on the multi-map tandem approach as described in Section 4.4 using the performance measures defined in Section 4.5. Choose the optimal base- and meta SOM size and denote this as $\{r^*_{xb}, r^*_{yb}\}$ and $\{r^*_{xm}, r^*_{ym}\}$

4. Backward elimination: Perform the multi-map tandem SOM with $\{r^*_{xb}, r^*_{yb}\}\{r^*_{xm}, r^*_{ym}\}$ $k_{-1}$ times, starting with all variables and deleting one variable at the time following the order determined in Step 1. Again assess performance for each variable set and decide on an accurate cut-off point $k^*$.

5. Repeat step 3 and 4 till $\{r^*_{xb}, r^*_{yb}\}$, $\{r^*_{xm}, r^*_{ym}\}$ and $k^*$ have stabilized.

For the neighborhood function and learning rate, the literature suggests two options each: a gaussian neighborhood versus a bubble (uniform) neighborhood and a constant versus an inverse time learning rate. The gaussian neighborhood and inverse time learning rate yielded the best results and were therefore selected. These are given by

$$\hslash = e^{-\frac{||\mathbf{l}_s - \mathbf{l}_y||}{2\gamma(u)^2}} \tag{9}$$

with $\gamma(u)$ the radius at iteration $u$. In this study the radius was constant and equal to $r_x$.

$$\alpha(u) = \frac{c\alpha(0)}{c+u} \tag{10}$$

with $c = \frac{\max(u)}{100}$ and max(u)$=$ 500.

Lastly, for the MDS algorithm, the initial coordinates of the weight vectors in the projection space have to be specified by the researcher. Naturally, it makes sense to set $\mathbf{q}_{sj} = \mathbf{l}_{sj}$, as the best guess available for the optimal placements of the weight vectors are the discrete location coordinates from the configured SOM. Additionally, initializing the projection space in this way ensures a stable orientation of the MDS solution.

### 4.4.2 Meta level

In the meta level, the trajectories from the base level are processed in two different ways. First, the trajectories $\tilde{\mathbf{\Lambda}}_i$ are grouped using an additional SOM. The rationale behind this approach is that the trajectories mapped to neurons in the same map are more similar than those far removed from each other. The SOM settings for the meta level are the same as the base level, but note that for the meta level SOM, only a single map is constructed.

Second, next to the meta-level SOM, the trajectories are also grouped using a Random Forest (for further details please see Appendix III and Breiman (2001)). The Random Forest was introduced as a predictive benchmark to assess whether the optimized neuron locations from the SOM-MDS tandem approach could result in an improved cluster quality. The SOM is less suitable for this, as despite its attractive visual characteristics it remains an unsupervised method and therefore less powerful to use for detecting differences in predictive performance between approaches.

The Random Forest was selected from the group of machine learning algorithms due to its high predictive power yet intuitive functionality. A machine learning algorithm was preferred over more traditional econometric methods as the trajectories are far from normally distributed and specifying any functional form on the data generating process of the trajectories would be overly restrictive. Other machine learning methods may be equally powerful in grouping the trajectories in "bankrupt" or "non-bankrupt" trajectories, but the Random Forest is attractive in its functionality as nothing more than a large set of decision rules. Additionally, the Random Forest provides a good indication of feature importance by storing the average contribution to the decrease of the gini-coefficient for each variable and could therefore give insights in the importance of each map (each quarter of data) in the trajectories.

## 4.5 Performance measures

Five performance measures were identified to assess the quality of the algorithm: four to quantify the performance of the Trajectory SOM and one to calculate the performance of the Random Forest classification of trajectories. For the Trajectory SOM first the quantization- and topology error were evaluated, as these are both inherent to the SOM algorithm. Two additional measures were introduced to assess the grouping quality of the SOM, the *bankrupt separation measure* and the *node purity.* For the RF, an $F_\beta$ score has been used.

1. The quantization error ($qe$) is given by

   $$qe = \sum_{i=1}^{n} ||\tilde{\Lambda}_i - \mathbf{w}_{s^*i}|| \tag{11}$$

   and measures the average Euclidean distance between the trajectories and their best matching neuron.

2. The topology error investigates to what extent the multidimensional topology of the input space is preserved in the projection by evaluating for each observation whether their second best matching neuron is a neighbor of the best matching one. The topology error ($te$) is given by:

   $$te = \sum_{i=1}^{n} \frac{u(\tilde{\Lambda}_i)}{n} \tag{12}$$

   $$u(\tilde{\Lambda}_i) = \begin{cases} 0 & \text{if } \mathbf{w}_{s^*i} \text{ and } \mathbf{w}_{s_2^*i} \text{ are adjacent} \\ 1 & \text{otherwise} \end{cases} \tag{13}$$

3. The bankrupt separation measure ($bsp$), identifies the smallest number of neurons in the map who contain at least 80% of all bankrupt trajectories. Ideally, all bankrupt trajectories are grouped together in a few neurons and hence a $bsp$ is desirable.

4. Merely assessing the $bsp$ is not enough. If all trajectories are mapped to one single neuron (i.e. all healthy and all bankrupt) the bankrupt separation measure is very

low, even though the quality of the SOM in general is poor. Therefore an additional measure was introduced, the node purity ($np$), specifying which share of the total number of trajectories in the neuron are bankrupt ones.

$$np = \frac{\sum i \in n_{br}}{n_s} \tag{14}$$

with $n_{br}$ representing the total set of bankrupt observations and $n_s$ the number of neurons in the respective neuron

5. To evaluate the performance of the RF, a weighted F-measure was used to weigh the importance of precision versus accuracy. The weighted F-measure is given by

$$F_\beta = (1 + \beta) * \frac{prec * rec}{(\beta * prec) + rec} \tag{15}$$

with $prec = \frac{tp}{tp+fp}$ and $rec = \frac{tp}{tp+fn}$. In consultation with PwC $\beta$ was set to 4 such that recall weighs four times heavier than precision (the costs and reputation damage of a false negative prediction are much higher compared to a false negative)

# 5 Results

The results section has been divided in three sections. First, the performance of the (multi-map) trajectory (tandem) SOM analysis is presented. Second, the output of the meta level SOM map is presented and its characteristics are explained. Finally an example of trajectory analytics based on the meta level SOM is shown. For the tuning results of the base level parameters for shape, size and dimensionality of the grid, as well as the selection of the optimal number of variables, please see Appendix II. The dimensionality of the SOM was set to $p = 2$, the base SOM maps were set to size $9 \times 6$, the meta SOM to size $10 \times 6$ and the first 42 variables with highest average factor loadings on the first two principal components were included in the analysis. For more details on the specific line items included in the analysis, see Table 4 in Appendix I.

## 5.1 Performance

In Section 4, two innovations were presented which could improve the analysis of bankruptcy using a SOM trajectory approach: the use of multiple SOM maps as foundation to the trajectory and the optimization of the location vectors using a tandem of SOM and MDS. Therefore, first the results of the Random Forest grouping of the trajectories are presented to assess whether these extensions actually improve predictive power. Following, an elaboration is made on the qualitative difference in performance of single- versus multiple maps. Lastly, the wider range of performance measures as introduced in 4.5 (as compared to merely predictive power) are evaluated for each of the extensions.

### 5.1.1 Predictive power: RF benchmark

To assess the contribution to predictive performance of the tandem approach and the use of multiple maps, Table 2 depicts the confusion tables of the RF classification for the variety of models. On the top left, the confusion table for the "Regular Single" model is depicted, i.e. the traditional SOM trajectory approach. On the top right, the performance of a model with regular location vectors is used, but *with* multiple maps, hence the "Regular Multi" model. Similarly, on the bottom left, the "Tandem Single" model uses transformed MDS location vectors but relies on single SOM map and finally, in the bottom right, the performance of the model with both innovations is depicted, referred to as the "Tandem Multi" model.

In table Table 2 the RF model performance on a test set is depicted. The "Tandem Multi" model predicted the most bankrupt trajectories correctly. It is however also imporant to assess how many false positive predictions were needed to do so. The "Tandem Multi" model wrongly marked 228 trajectories as bankrupt whilst they were healthy. This means that on average, for each correctly classified bankrupt trajectory 14.25 firms were wrongly classified bankrupt. This is definitely better than the 19.4 for the traditional "Regular Single" model, but also better than the 15.4 wrongly classified firms per true positive for the "Tandem Single" approach and 16.5 for the "Regular Multi" model. The predictive performance of the tandem approach therefore seems to be slightly better than for the regular approach.

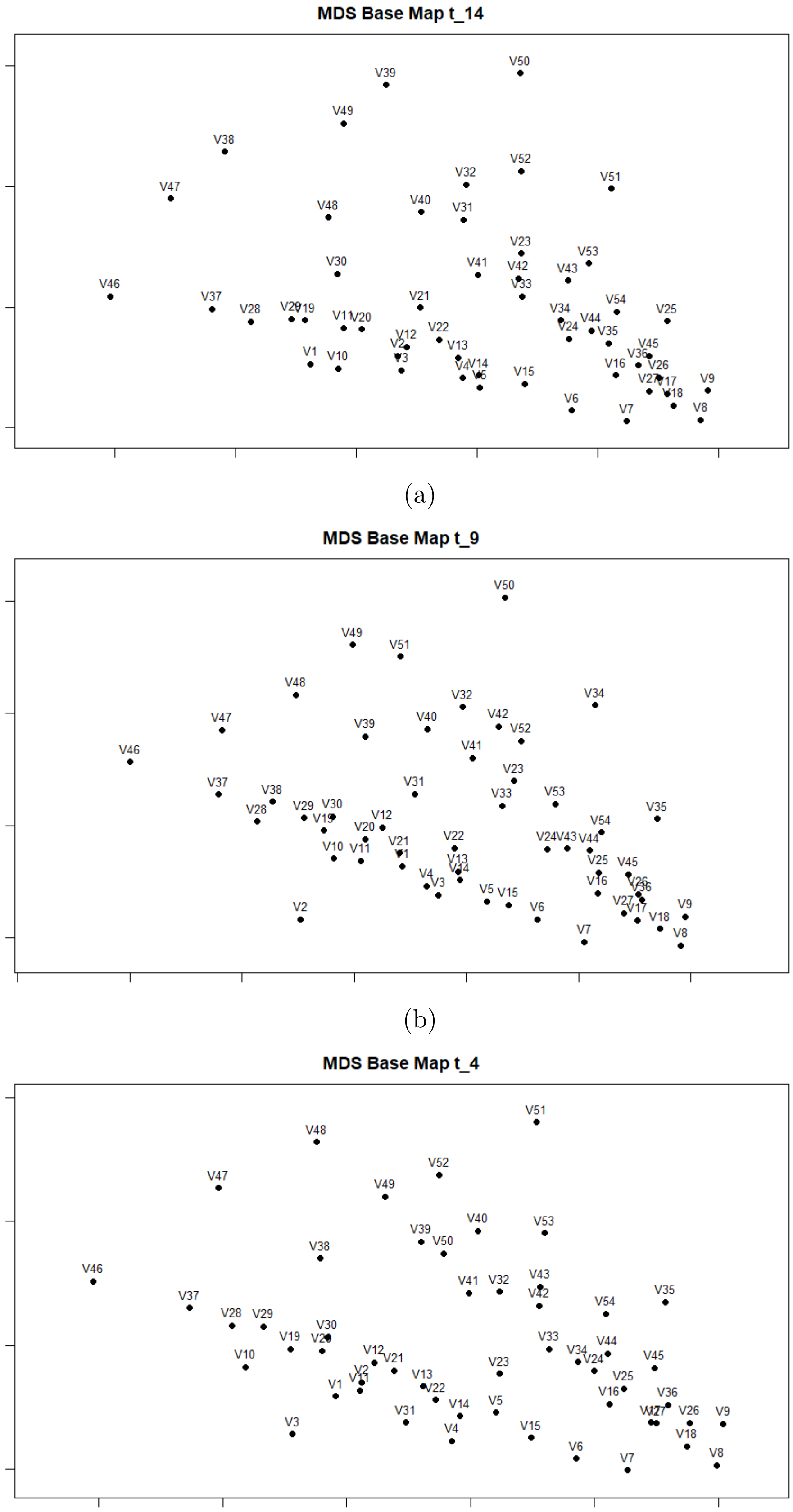
For the use of single or multiple maps the difference is less clear and depends on the focus of the researcher. If the true positive predictions are very important (as is the case in the going concern context) the multiple map is preferred, but if the false positive predictions gain more weight the use of a single map yields better performance.

Table 2: Out-of sample confusion tables for the trajectory Random Forest. $0_t, 1_t, 0_p$ and $1_p$ represent the true healthy, true bankrupt and predicted healthy and bankrupt classes respectively. Depicted is both the in- and out-of sample RF performance for regular or tandem approach trajectories based on single or multiple maps. The best results are underlined.

| | | Single map | | | Multiple map | |
|---|---|---|---|---|---|---|
| Regular | | $0_p$ | $1_p$ | | $0_p$ | $1_p$ |
| | $0_t$ | 685 | 233 | $0_t$ | 703 | 215 |
| | $1_t$ | 6 | 12 | $1_t$ | 5 | 13 |
| Tandem | | $0_p$ | $1_p$ | | $0_p$ | $1_p$ |
| | $0_t$ | 733 | 185 | $0_t$ | 690 | 228 |
| | $1_t$ | 6 | 12 | $1_t$ | 2 | 16 |

### 5.1.2 Single map versus multiple maps

The choice between the use of single and multiple maps does not merely depend on the predictive performance of the model, but also on the interpretation of specific neurons in the maps. To illustrate, in Figure 5 the MDS representations of the base level SOM maps of time $t_{-14}$ (a), $t_{-9}$ (b) and $t_{-4}$ (c) are depicted. Because the location of the neurons in the MDS representation are lower dimensional projections of the weight vectors of the converged SOM, they can be interpreted as summary points of the input space and be used to assess its shape and size. As can be observed from the Figure, the shape of the input space changes over time. Whilst all panels depict a somewhat fan-shaped distribution of the neurons, it seems that the closer to $t$, the more clearly separated the left neurons become from the right: a white separation area is more clear in panel (c) compared to (a). Additionally, in panel (c) there are more neurons who are placed further apart form the group, indicating that the input space was somewhat more scattered compared to (a) for example.



(a)

(b)

(c)

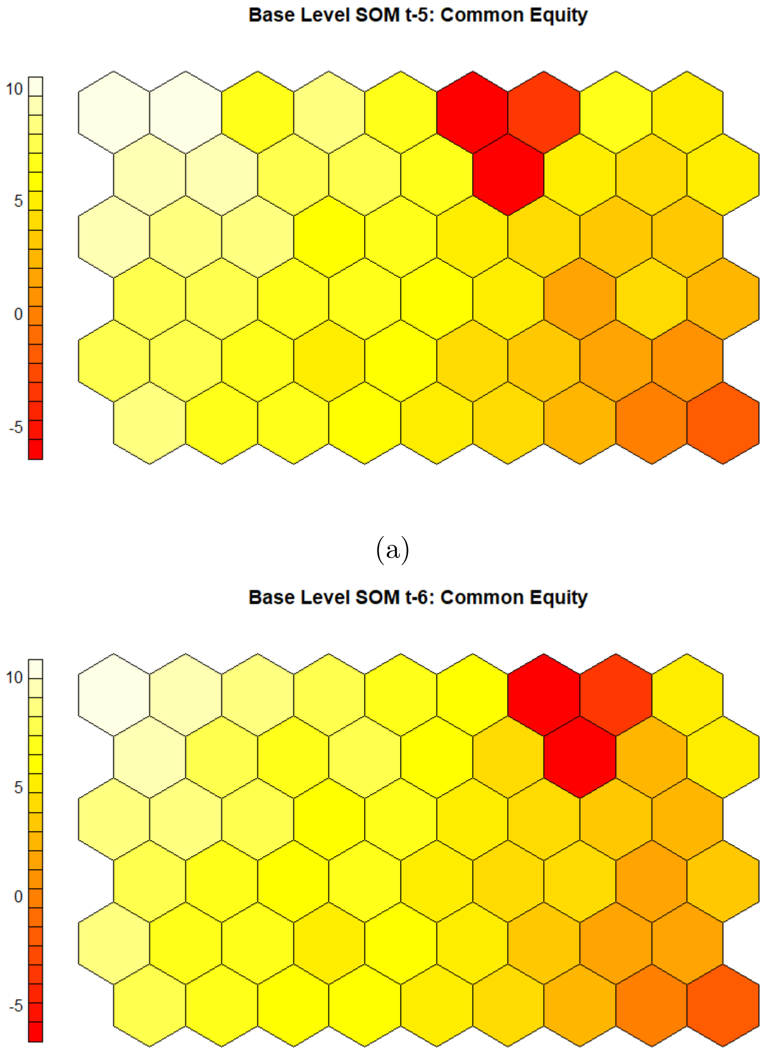
Figure 5: MDS representation of the base level SOM maps of quarter $t_{14}$ (a), $t_9$ (b) and $t_4$ (c), which can be interpreted as a 2-dimensional summary of the input space

The small changes in the range and shape of the input space become important when using a single map. If observations are projected to a SOM trained on a different quarter of data and the input space has changed, this results in a higher quantization error, as the SOM map simply fits the observations less well (for details on the quantization error, see Section 4.5). Table 3 confirms that the average quantization error for the singe map approach is higher than for the multiple map, supporting the hypothesis that a single SOM map can be too restrictive. If the SOM maps are used as a summary of the input space the use of multiple maps could therefore be a preferred choice.

Table 3: Average quantization error of the base maps for the single- and multiple- map approach, in- and out-of sample.

| | In Sample | | Out Sample | |
|---|---|---|---|---|
| | "Single" | "Multi" | "Single" | "Multi" |
| $\bar{qe}$ | 56.68 | 53.32 | 96.12 | 91.37 |

Nevertheless, even though the use of multiple maps results in a slightly more truthful representation of the input space, it does complicate the interpretation of the different maps. In Figure 12 this issue is demonstrated. Imagine you are using a regular trajectory SOM approach. The stability of the map (i.e. the orientation of the neurons) in this study was enforced by training the SOM base maps in a batch fashion and initializing the neuron's weight vector by the first two principal components. In Figure 12 the colour of each neuron indicates the value of Common Equity in the weight vector of the respective neuron (such a map is referred to as a 'component plane'). Due to this stability enforcement, the component planes at time $t_5$ (a) and $t_6$ (b) both show that firms with a high Common Equity were mapped to the top left of the map, whilst firms with a highly negative Common Equity were mapped to the top right. When looking more closely, one can however observe that the neuron on the second row from the top, fourth from the right in panel (a) represents companies with a negative Common Equity, whilst in panel (b) the neuron on this position represents firms with a Common Equity between 0 and 5. This means that even though the orientation of the maps is stable (roughly speaking, firms with low Common Equity are top right and firms with high Common Equity are top left), it does not guarantee that the meaning of each neuron at each position at all times is the same. Therefore, for interpretation purposes the use of a single map makes the model and its positions much easier to grasp. If the difference in average quantization error between the single and multiple map is not too large, a single map could be preferred.



(a)

(b)

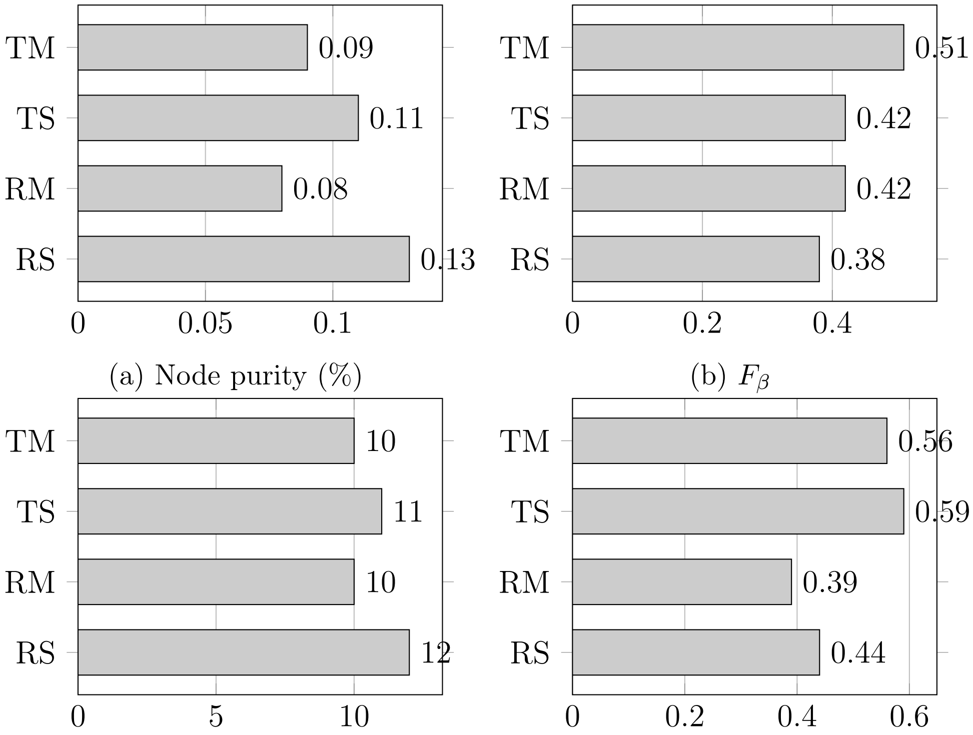
Figure 6: Common Equity component plane of the base SOM map at time $t_5$ and $t-6$. The colour of each neuron indicates the value of Common Equity in the weight vector of the respective neuron.

### 5.1.3 SOM-specific performance

In Figure 7 the performance measures as proposed in Section 4.5 are presented for each model. Panel (a) depicts the node purity, panel (b) the $F_\beta$ measure, panel (c) the bankrupt separation and (d) the topology error. From Table 2 it was already concluded that the "Tandem Multi" model had the highest predictive power, which also results in the highest $F_\beta$ value compared to the other models. For bankrupt separation, the minimum number of neurons containing 80% of the bankrupt observations, the "Tandem Multi" was also among the best performing models. From the perspective of node purity, the "Tandem Multi" however scores slightly worse. Looking more closely, one can observe that the models with a slightly lower bankrupt separation also have lower node purity, whilst for the models who have their 80% of bankrupt trajectories spread out over more neurons (higher bankrupt separation), these neurons are also more pure. The preferred model then again depends on the trade-off between the importance of true- and false positive bankrupt classifications.

Lastly in panel (d) the topology error is depicted. From the graph it seems that the tandem models perform much worse, as the average topology error in the tandem models was around 0.55 whilst the topology error for the regular models was around 0.42. Comparing the topology error across the regular and tandem approach however is not completely fair, as the MDS optimization of the SOM location vectors reorganizes the neurons, placing neurons with similar weight vectors close to each other, irrespective whether they were adjacent following their original SOM locations. The MDS optimization hence 'corrects' for wrong neighbor information, placing previously 'wrongly' neighboring neurons further apart because their weight vectors were not so similar after all. It is therefore surprising that the topology error based on the "old", "sub-optimal" neighbor information is higher than for the trajectories passing through the "optimal" neuron locations. It is possible to compare within the tandem / regular model classes, where both point at an improved performance for the multi map approach compared to the single map. Because the quantization error is similarly not comparable across the regular and tandem approach (the total range on which neurons can be located changed by applying MDS) and because it was close to identical for single and multi maps, it is not depicted.

The "Multi Tandem' approach has an superior out-of-sample $F_\beta$ score. On the other performance measures, the "Multi Tandem" approach does not score consistently better. On average it performs just as good as the other models. However, as this paper is an assessment of the functionality of the two innovations underlying the "Multi Tandem" model, for the remaining of this paper results from the "Multi Tandem" model will be interpreted.



(a) Node purity (%)

(b) $F_\beta$

(c) Bankrupt separation

(d) Topology error

Figure 7: Overview of the out-of-sample performance measures for each of the trajectory settings: TM ("Tandem Multi"), TS ("Tandem Single"), RM ("Regular Multi"), RS ("Regular Single")

## 5.2 Meta level: trajectory SOM

In order to interpret the output of meta- level SOM map, three important features of the map should be kept in mind. The first is the notion that all positions are relative. This means that the absolute distance in neither a regular SOM map nor the MDS representation can be interpreted. The explanatory power of each map lies in in the *relative* distances between the neurons and the trajectories they represent. Secondly, trajectory analysis is concerned with the recognition of patterns. When comparing different trajectories, it is possible to first observe similarities / differences by simply looking at differences in patterns (i.e. one trajectory is up whilst the other is down) before delving into what these differences represent. Lastly, note that for the meta level SOM, the weight vector of each neuron is a representation of the average location on each base map of all the firms projected to that neuron. For the base level SOM maps, the weight vector is a representation of the financial line items of the firms projected to the neuron at a specific time.

In Figure 8 a representation of the meta-level SOM is depicted. Inside each neuron, the weight vector of the respective neurons are depicted. As the location vector $\tilde{l}$ of each neuron in the base maps has an $x$ and a $y$ element, the trajectories are of length $t_{-3} \times 2 = 34$ and are of the format $\{\tilde{l}^x_{t_{17}}, \tilde{l}^y_{t_{17}}, ..., \tilde{l}^x_{t_3}, \tilde{l}^y_{t_3}\}$. Hence, the position on base map $t_{-17}$ is depicted by the most left two points and the position on base map $t_{-3}$ by the far right.

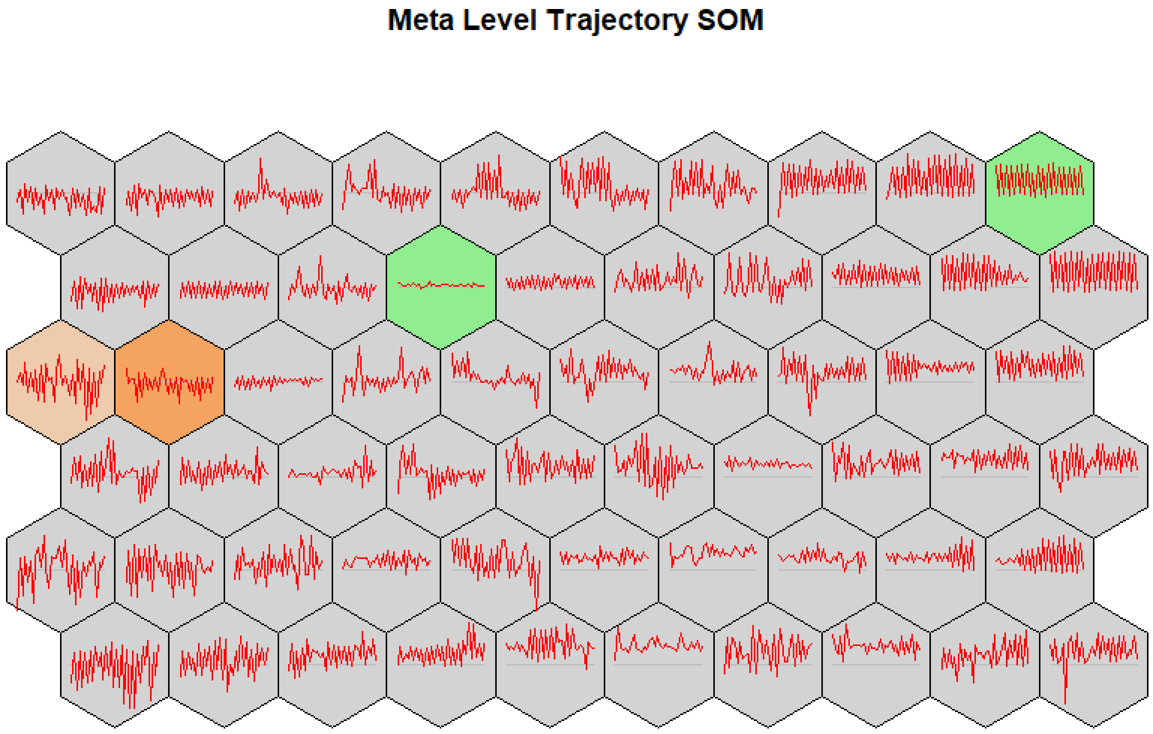


Figure 8: Representation of the 10 × 6 hexagonal metal level trajectory SOM. Within the neuron, the corresponding weight vector of the neuron is depicted, representing the average trajectory of the firms mapped to the respective neuron through the base level maps. The green neurons represent the most stable trajectories associated with solely healthy companies. The two red-shaded neurons contained the most trajectories ending in bankruptcy

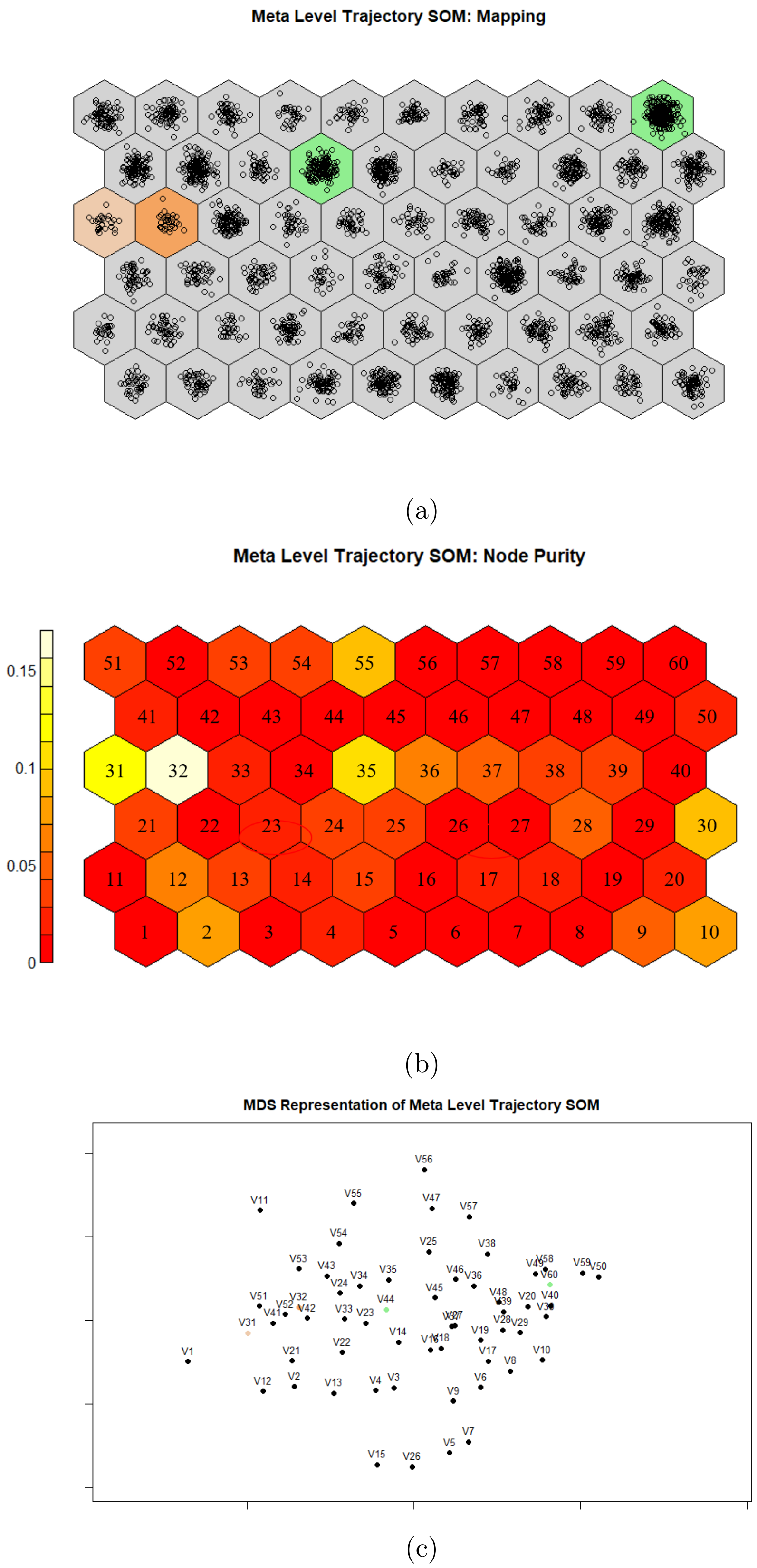
It is easy to see that the firms mapped to the most left green neuron are stable over time. The $x$ and $y$ coordinate in each base map are similar throughout all base maps. It is however important to understand that the upper right green neuron also depicts a very stable trajectory: the firm is at the left upper corner in each base map (low $x$-value in each map, high $y$-value). This is different to companies mapped to neurons such such as the bottom left. In this neuron one can see differences occur in the relative $x$ and $y$ positions over time: therefore it represents companies moving positions over the maps. Naturally, the representations of these weight vectors alone merely gives one a general idea about the level of movement in each trajectory. Before diving into the meta level SOM, the number of trajectories in each neuron, the purity of each neuron and the relative distances between the neurons are investigated to detect which neurons are most interesting for analysis.

In Figure 9a and 9b the mapping plot and node purity component plane of the meta SOM are depicted respectively. The mapping plot shows a jittered representation of the number of trajectories mapped to each neuron. The component plane shows the node purity per neuron: the lighter the shade, the higher the percentage of bankrupt trajectories, i.e. the more trajectories belonging to firms who went bankrupt at time $t$. As can be seen, neuron 32 has the highest node purity, with 16 percent. Even though the absolute percentage is not that high, one must remember the percentage of bankrupt companies in the sample is less than 2 percent and therefore the neuron is marked as relatively 'dangerous'. Neuron 66, 43 and 44 contain the most trajectories, of which none bankrupt. If one looks back at Figure 8, the most pure bankrupt neurons are therefore depicted in red, whilst the neurons with solely healthy trajectories projected to them are depicted in green.

In Figure 9c the MDS representation of the meta SOM map is depicted. As can be observed, the neighboring relation are a bit more complex than it appears on the SOM map. Especially the trajectories of neuron 11, 15 and 26 are very different to their adjacent neurons.

## 5.3 An example of trajectory analytics

In the previous sections some insights into the meta level SOM were already presented. The number of trajectories per neuron, the percentage bankrupt trajectories per neuron and the distances between the neurons give some initial idea about the different movements companies can make through their financial statements. The meta SOM however contains many more insights, which can be obtained by "delving deeper" into the map. In this section an example of such a deep dive is presented. Note that the information stored in the SOM map is by no means limited to this example. The specific example was selected because it clearly illustrates the power of the trajectory SOM method as well as the improvement in visualization the SOM-MDS tandem can achieve compared to the regular approach.



(a)

(b)

(c)

Figure 9: (a) Jittered representation of the number of trajectories in each neuron of the meta SOM. (b) Numbered component plane of the meta SOM depicting node purity. The lighter the shade, the higher the percentage of trajectories belonging to companies bankrupt at time $t$. (c) MDS representation of the meta SOM.

Before delving deeper into the meta SOM map, it is important to understand what to look for. If one revisits the "going concern" issue as presented in the introduction one is ultimately looking to distinguish firms who will not be viable in the upcoming year from those who are. Ideally, the trajectories of companies going bankrupt in the meta SOM should look different from those who remain in business, because in that case it is possible to separate the two. It would however be naive to think that all companies who go bankrupt develop themselves in the same way through their financial statements: some companies go bankrupt because of debt issues, some because of equity issues and so on. Therefore, when comparing trajectories to gain insights in going concern issues, one should be looking for two different things. First, focus on places in the trajectory where all bankrupt companies behave similar, but different to the healthy companies. These points in the trajectory can help detect features inherent to non-viable versus viable companies. Second, it is also interesting to focus on the 'bankrupt' trajectories in isolation (i.e. trajectories who relatively often result in bankruptcy with each other) and assess at what points in time these trajectories move differently. These points can help identify different possible "roads" to bankruptcy.

In Section 5.2 some of the specific neurons were highlighted. Neuron 31 and 32 had the highest node purity, meaning that the percentage of trajectories ending in bankruptcy compared to trajectories who did not was the highest for these two neurons. Therefore they were marked in red-like colors and referred to as "bankrupt" trajectories. Neuron 44 and 60 on the contrary contained a lot of trajectories of which none resulted in bankruptcy and therefore, these neurons were marked green and referred to as "healthy" trajectories. The color shade of the neurons will remain the same trough out the remaining figures in this section such that one can retrieve which information belongs to the trajectory in neuron 31 (pinkish-orange), neuron 32 (dark-orange) or neuron 44 (green).

To analyze the selected neurons from a more up-close perspective, in Figure 10 the trajectories of neuron 31, 32 and 44 are presented in more detail. Note that the patterns presented in this graph are exactly the same as the red-lined representation inside the neurons in Figure 8, only this time they are zoomed-in and colored differently to recognize which trajectory belongs to which neuron.

Turning to the healthy trajectories first, one can observe that it is relatively stable. At each point in time, the $x$ and $y$ coordinate are almost the same. This means that neuron 44 represents companies who have similar financial statements throughout the five years spanned by the trajectory: nothing substantial changes in the line items from quarter to quarter. Of course the green trajectory presented is an 'average' trajectory, i.e. the central point of all trajectories associated with neuron 44 of the meta SOM. This means some companies in the neuron may move a bit in position between maps throughout time, but on average they remain stable.

If one then turns to the bankrupt trajectories a different trend is visible. Contrary to the healthy trajectory, the two bankrupt trajectories move much more throughout the maps. For example, at time $t_{-12}$, the $x$-coordinate for trajectory 31 was a bit below zero. At time $t_{-11}$, it was roughly 0.5. This is quite a large rightward jump in position map-to-map. Companies associated with the trajectory of neuron 32 did not make such a jump: the $x$ and $y$ locations at time $t_{-12}$ and $t_{-11}$ were roughly the same. In general, neuron 31 seems to makes more and larger jumps than neuron 32, but both move much more over the maps than the healthy trajectory does.

It can now be concluded that the biggest difference between the healthy trajectory and the two bankrupt trajectories is that the financial statements of the bankrupt companies show much more fluctuation throughout the maps. The way in which the bankrupt trajectories deviate from the healthy trajectory is however different. To understand what these differences mean, one can again delve deeper into the meta level SOM to obtain more insights. The two bankrupt trajectories behave differently at multiple points in time, but two moments of specific interest are $t_{-6}$ and $t_{-5}$. The reason for this is that at time $t_{-5}$, both bankrupt trajectories are at the exact same height of the map (their $y$-coordinate is the same). At time $t_{-6}$ however, this was not the case yet. This means something must have changed in the financial statements of $t_{-6}$ and $t_{-5}$ which made the companies from two different trajectories to be more similar and therefore mapped close together. In other words: what caused the 'converge' of the two trajectories between $t_{-6}$ and $t_{-5}$?

To answer this question, the two base level SOM maps of time $t_{-6}$ and $t_{-5}$ are depicted in Figure 12. The red-like colors represent the neurons companies in trajectory 31 or 32 were mapped to. Note that the red-lined representation of the weight vectors in the neurons of this base-map are not trajectories, but representations of the financial line items. If one compares the weight vectors inside the two red-like neurons, at time $t_{-6}$ one sees a large dip somewhere half-way in the neuron belonging to trajectory 32 (dark orange), which is not present in the neuron of trajectory 31 (pink-orange). At time $t_{-5}$ this dip is no longer there and the weight vectors of the two red-like neurons are rather similar. Thinking back of the trajectories, this coheres with the 'convergence' of the two trajectories from $t_{-6}$ to $t_{-5}$. But what does this dip represent?

In Figure 11 a zoom of the base level weight vectors are depicted. In this Figure, it can be seen that the strong dip at time $t_{-6}$ for trajectory 32 corresponds to a dip in variables "teqq", "ceqq" and "seqq", which are all equity related variables (see Appendix I). It thus seems that equity related problems signify bankrupt trajectory 32 at time $t_{-6}$. At time $t_{-5}$, this problem is resolved and both the financial line items of trajectory 31 and 32 look rather similar.

It seems that now the reason for convergence between the two bankrupt trajectories between time $t_{-6}$ and $t_{-5}$ has been discovered. However, if one takes another look at positions of the associated base neurons in Figure 12, the red-like neurons are further apart at time $t_{-6}$ compared to $t_{-5}$. This is counter-intuitive, because the weight vectors of the two neurons are more similar at time $t_{-5}$ and therefore one would expect them to be closer together compared to $t_{-6}$. This discrepancy between the similarity in weight vectors but dissimilarity in neuron position stresses the limitation in interpretive value of the regular SOM locations of the neurons.

In Figure 13 the MDS representation of the base level SOM maps at time $t_{-6}$ and $t_{-5}$ are depicted. In these maps, the discrepancy between convergence in weight vectors and location is resolved. At time $t_{-6}$, one observes that the two red neurons are placed at a large vertical distance from each other, visually representing the differences in the weight vectors (one contained the dip, the other did not). At time $t_{-5}$, the dip in equity is restored and one indeed observes that the two neurons are now placed next to each other. In this way, the visual location and the similarity between the neurons is intuitive.

This example of trajectory analytics demonstrates that one can use the meta level SOM and dive deeper into its construction step by step to gain insights on the desired level of specificity. Naturally, it is also possible to start at the most specific level: the financial statement of a single firm at a single time and work the way back up. In this way one can see how the single firm relates to other firm in a bigger- and bigger picture. Secondly this examples demonstrates the importance of a SOM-MDS tandem to increase the quality of visual information that can be extracted from the locations of SOM neurons alone.

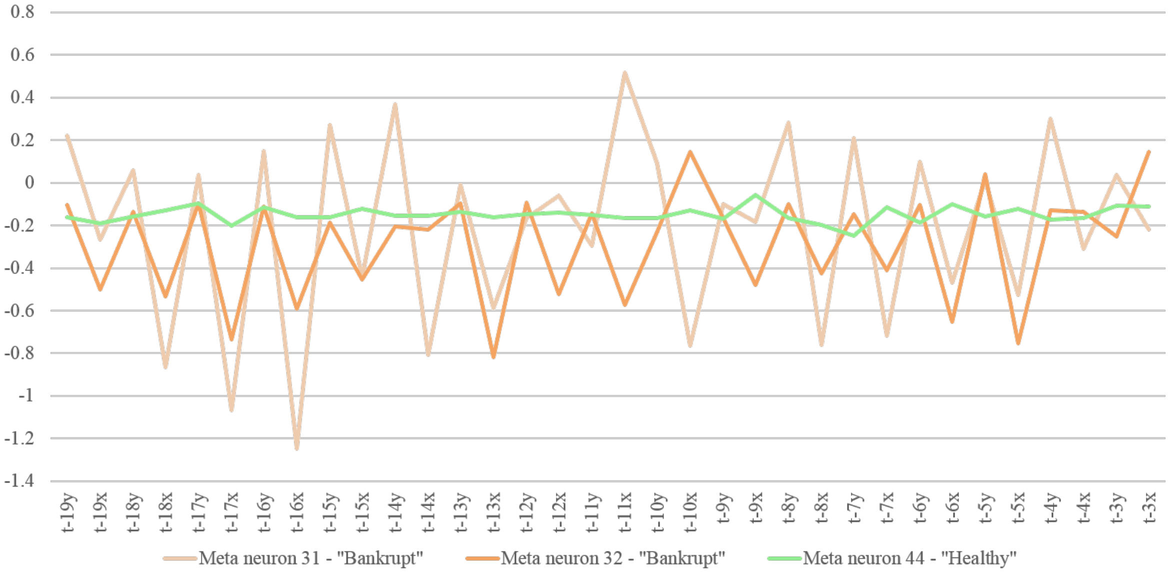


Figure 10: Zoom of the trajectories of Neuron 31, 32 and 44 in the meta level SOM (tandem multi-map). Depicted are the weight vectors for each neuron, presenting the average location on the base maps at each point in time for all companies mapped to the respective neuron.

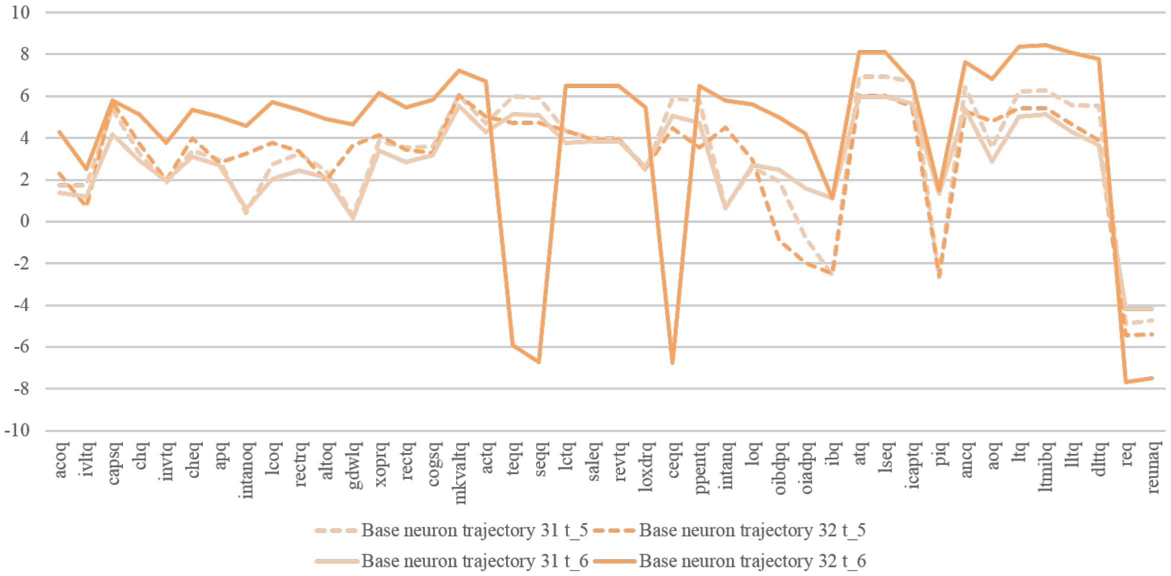
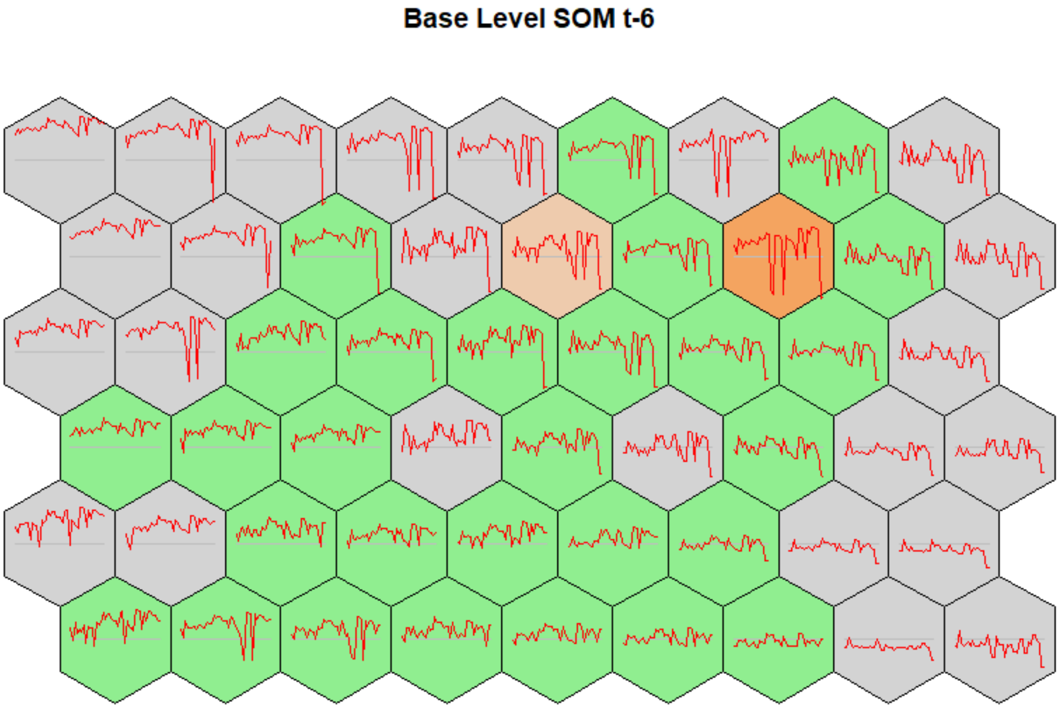
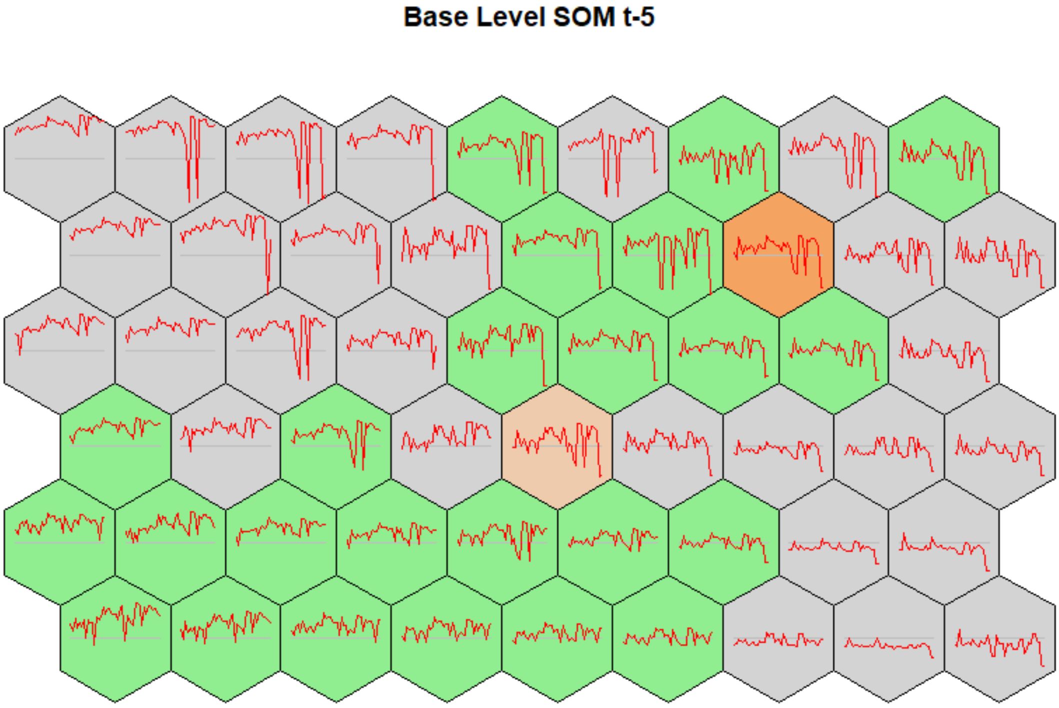


Figure 11: Zoom of the weight vectors of two neurons in the base level SOM maps of time $t_{-5}$ and $t_{-6}$. The x-axis represents the respective line items included in the analysis. The y-axis depicts the value of the weight vector for each of the line items. For full variable names with each abbreviation, please see Appendix I.

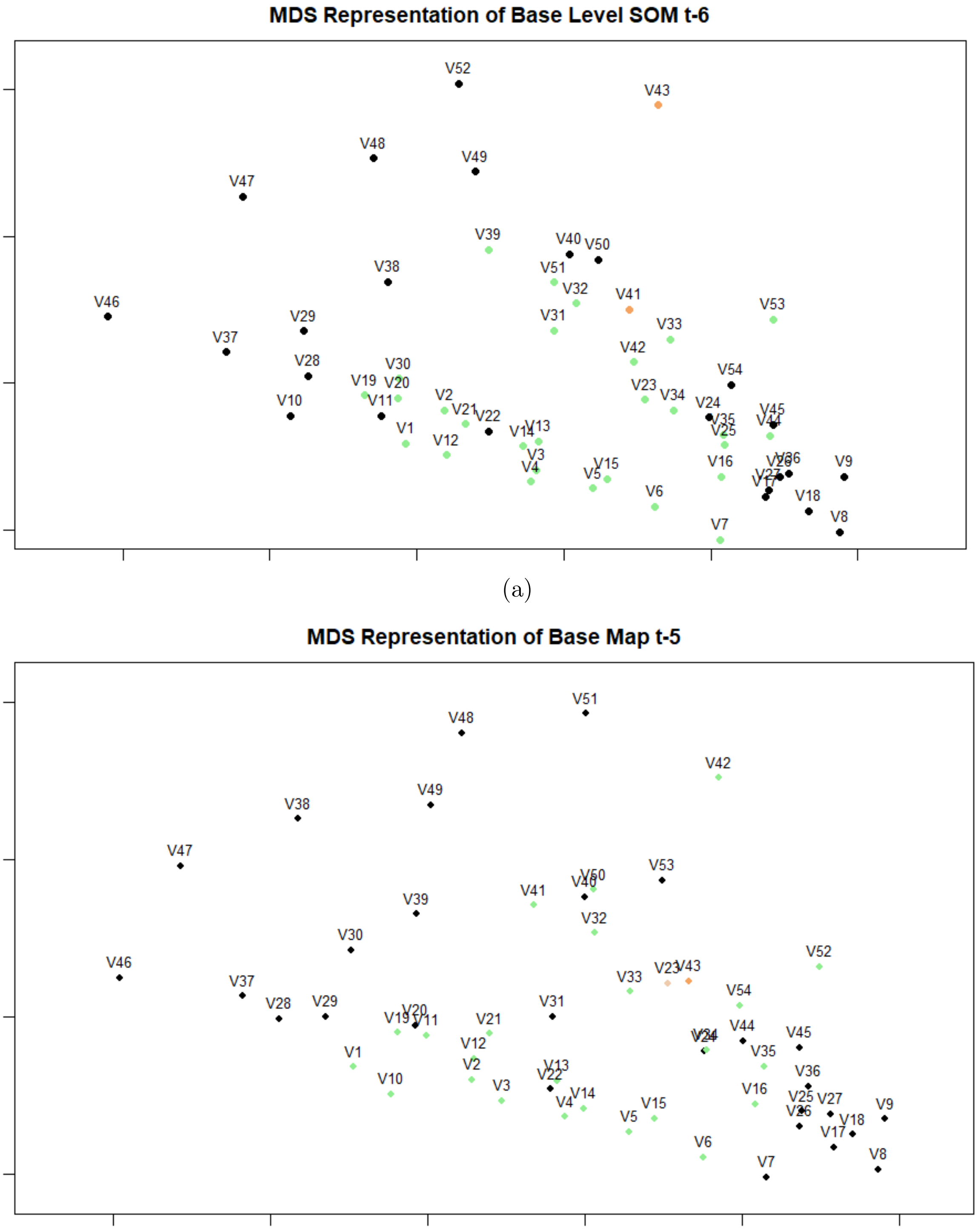


(a)



(b)

Figure 12: Regular representation of the base level SOM maps at time $t_{-6}$ and $t_{-5}$. The neurons depicted in green represent the location of the company trajectories mapped to neuron 44 in the Meta Level SOM map, whilst the two neurons depicted in red represent the location of the company trajectories mapped to neuron 31 and 32.



(a)

(b)

Figure 13: MDS representation of the base level SOM maps at time $t_{-6}$ and $t_{-5}$. The neurons depicted in green represent the location of companies mapped to neuron 44 in the Meta Level SOM map, whilst the two neurons depicted in red represent the location of companies trajectories mapped to neuron 31 and 32 in the meta level.

# 6 Conclusion

In this paper, the functionality of two innovations for financial statement trajectory analysis using Self-Organizing Map were investigated: the projection of quarterly statements on a single- versus multiple maps and the use of a tandem approach between Self-Organizing Map and Multidimensional Scaling to optimize the neuron's location vectors in each map. By using these innovations to analyze five year sequences of quarterly financial statements from U.S. listed companies between 2013 and 2018 across SOM maps, the Random Forest was able to predict bankruptcy with an average accuracy of 77.5 %.

The SOM-MDS tandem approach was compared against the performance of the regular SOM approach as used in Schreck et al. (2007), Du Jardin and Séverin (2011), Chen et al. (2013). For both the use of single and multiple maps, the RF performance of the tandem based trajectories was higher than for the other models. This could hint at an improved clustering performance due to the additional distance information captured in the MDS transformed location vectors, but as the improvement in performance was only marginal one should remain careful in generalizing these conclusions.

The true power of the SOM-MDS tandem lies in the improved visualization of distances between neurons, in the base maps as well as the meta level trajectory SOM. The MDS maps show that the neighboring relations in the converged SOM maps are not as straightforward as suggested by the fixed, discrete neuron representation in the traditional SOM. The MDS representation of the neurons give a more intuitive insight into the relative differences between observations mapped to each neuron, allowing to distinguish the closeness of two different adjacent neurons: something impossible to do with the traditional SOM map. The MDS representations revealed that the neighboring relations between adjacent neurons were not always accurate: some neurons were adjacent in the regular SOM map whilst their weight vectors were very far removed. MDS has the power to check the quality of the SOM map, the coherence between SOM maps over time and can be used to visualize the distribution of observations across the input space. It is therefore concluded that even though the SOM algorithm by itself is a highly attractive tool to *group* financial statement trajectories, the SOM-MDS tandem approach is additionally needed to obtain an accurate lower dimensional solution and to justfully interpret the differences and similarities of companies projected across neurons.

No clear preference for single- or multiple maps could be detected in the experimental results. The multiple map approach resulted in a smaller average quantization error through the base maps, indicating that on average, the map fitted the observations better and therefore was a more accurate summary of the input space. However, even though stability throughout the multiple map was enforced through training in a batch-fashion and an initialization of the neurons according to the first two principal components of the sample, this stability was not neuron specific. This means that even though the average value of a line item on the base maps was roughly the same in each area of the map (i.e. "top left" or "center"), the line item value in the weight vectors of specific neurons could be different from map to map. In a practical application of the methodology such as decision support for going concern, the single map could therefore be preferred. Using the single map, the trajectories will be easier to digest as they can simply be followed across the single map with the interpretation of each location fixed across time. However, if the SOM is also used as a tool to understand the shape, size and distribution of observations across the input space, the use of a single map is not sufficient and multiple maps are required.

Finally the example on trajectory analytics presented in this paper showed a small hint of all potential insights that can be retrieved from the meta level trajectory SOM. Depending on ones interest, one could start with the meta level SOM and consistently dive deeper into the base maps and the actual line items underlying the base maps to retrieve more detailed information. This transparency of the SOM, i.e. the ability to retrieve the position of each trajectory, at each point in time back to the raw line items is an unique and attractive feature of the trajectory SOM approach. The example additionally illustrated the increased quality of visual interpretation possible from the tandem SOM-MDS trajectories compared to the regular SOM locations. Not only are the regular SOM locations less meaningful, they can additionally cause confusion by misrepresenting the differences in weight vectors between neurons.

# 7 Discussion

The analysis of quarterly financial statements as a trajectory through a (series of) Self-Organizing Map(s) is an unique and innovative method to gain insight into the financial development of companies over time and their associated bankruptcy risk. The method is unique in its ability to capture both cross-sectional and time-varying characteristics of the data without the need to define a strict functional form as required in most panel models. The SOM model can be utilized without any prior knowledge about the nature of the relationship between each line item and a firm's performance: all insights are obtained by merely comparing the performance of one firm to the other companies in the sample.

This paper is the first to assess the functionality of a location-optimization of each neuron in the Self-Organizing Map using Multidimensional scaling. Where Multidimensional Scaling and Self-Organizing Map are often presented as competing techniques to reduce dimensions, this research shows the power in combining the attractive features of both algorithms by applying them in sequence. Even though the performance of a regular trajectory SOM is comparable to that of a tandem trajectory SOM, the trajectory SOM allows for a more accurate, representative visualization of neighboring relations and relative differences between neurons in the map. Where previous research relied on the coloring of neurons (U-matrix) to express distances between neurons, it is far more intuitive to express distance between neurons with *actual distance* in the projection as is done by the SOM-MDS tandem. This paper is also the first to investigate the use of a single map or multiple maps to construct the SOM trajectory. No convincing evidence for either of the approaches was however found and preference for either of the approaches will depend on the objective of the researcher.

This research knows two main limitations, the first data specific and the second concerning the generalizability of the constructed SOM trajectories across different time periods. Regarding the data, the number of bankrupt samples included was limited. To obtain a large enough sample, bankruptcies occurring at different time-points were artificially moved down the timeline to cohere with the healthy sample. Even though extensive effort was made to correct for this particular characteristic of the data (samples were taken from a period of homogeneous macro-economic circumstances and adjusted for seasonality) the effectiveness of these corrections should be evaluated against analysis with more complete bankrupt information. The second limitation concerns the stability of the maps and hence the possibility to generalize the maps to different samples and different time periods. The stability between maps in this research was ensured by applying the batch version of SOM and initializing the neurons by the first two principal components. Nevertheless, all insights in the model are based on the inter-relationships of the companies in the sample and therefore the inclusion of new firms as well as the removal of existing firms form the sample could result in a different segmentation of trajectories, especially if they influence the composition of the first two principal components. Therefore, if a substantial change in the input space occurs, the base- and meta maps should be re-estimated. Additionally, the setting of hyperparameters in the model influence the map. For example, a different distance measure than euclidean to measure the dissimilarity between financial line items can cause the SOM to converge in a different way.

Based on these limitations, two main suggestions for further research are proposed aimed at improving stability of the SOM maps. First, an ensemble algorithm could be constructed to train the base level maps. This would entail the training of multiple maps for each quarter based on bootstrapped samples of the firms. Because the SOM algorithm does not guarantee neurons in the same location of each map to have comparable weight vectors (the issue of label switching), neurons from the ensemble of maps will have to be merged according to their similarity of observations projected to that specific neuron. Neurons that contain a large share of similar observations across bootstrap samples can be said to represent roughly the same part of the input space and therefore it is justified to combine their weight vectors. An additional advantage of the ensemble algorithm would be the reduced influence of the a-priori specified number of neurons in the map, because redundant neurons could be aggregated into a single neuron during the merging process.

Secondly, the combined power of SOM and MDS could be further exploited. In this research, both methods were used in sequence and did not interact. It is however possible to apply the MDS transformation of the location vectors *within* the training process of SOM. This integration would entail an MDS transformation of the location vectors after each update of the weight vectors during the training phase. During the training process of the SOM, not only the best matching neuron for each observation is updated but also the neurons in the neighborhood. If initially neighboring neurons are updated to be less and less similar during the training phase, the integration of MDS within SOM would result in a re-positioning of the neurons outside of each other's neighboring space, ensuring that the neurons are no longer forced to be updated together. After the SOM has converged in this integrated approach, the locations of the neurons will already be optimized.

Finally, a suggestion for further research includes the search for the optimal grouping of trajectories. It is not necessary to group the trajectories with another (meta level) SOM as done in this research. The MDS-optimized trajectories are powerful and the arguments to prefer them over regular trajectories have been demonstrated in this paper. The trajectories can however be grouped according to any clustering technique which may yield far superior results than the meta SOM.

# Appendices

## Appendix I: Descriptive Statistics

Table 4: Descriptive statistics of the line items included in the analysis. Given are the minimum (min), maximum (max), median (med) and mean absolute deviation (mad) of each line item, for the bankrupt and healthy sample respectively. The variables are ordered by their average contribution to the total variance in the data (Imp).

| | | | Healthy | | | | Bankrupt | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Nr | Variable | Imp | min | max | med | max | min | max | med | mad |
| 1 | Unadjusted Retained Earnings | 7.68 | -69346 | 414540 | -18 | 293 | -13122 | 2961 | -244 | 506 |
| 2 | Retained Earnings | 7.66 | -69421 | 398278 | -15 | 321 | -13469 | 2974 | -270 | 494 |
| 3 | Long-Term Debt - Total | 3.72 | 0 | 125972 | 84 | 125 | 0 | 20582 | 478 | 709 |
| 4 | Long-Term Liabilities - Total | 3.59 | 0 | 220701 | 80 | 119 | 0 | 21995 | 547 | 809 |
| 5 | Liabilities - Total and Noncontrolling Interest | 3.42 | -96 | 353799 | 299 | 441 | 1 | 37007 | 801 | 1151 |
| 6 | Liabilities - Total | 3.41 | 0 | 350141 | 293 | 432 | 1 | 37007 | 801 | 1151 |
| 7 | Assets - Other - Total | 3.40 | 0 | 363234 | 101 | 150 | 0 | 11393 | 72 | 103 |
| 8 | Non-Current Assets - Total | 3.32 | 0 | 364951 | 210 | 310 | 0 | 19136 | 783 | 1134 |
| 9 | Pretax Income | 3.28 | -12603 | 7462 | 1 | 18 | -1873 | 343 | -25 | 37 |
| 10 | Invested Capital - Total | 3.20 | -1417 | 436480 | 428 | 633 | -57 | 26902 | 760 | 1112 |
| 11 | Assets - Total | 3.18 | 0 | 702095 | 633 | 930 | 1 | 43327 | 1016 | 1448 |
| 12 | Liabilities and Stockholders Equity - Total | 3.18 | 0 | 702095 | 633 | 930 | 1 | 43327 | 1016 | 1448 |
| 13 | Income Before Extraordinary Items | 3.05 | -10713 | 32551 | 0 | 16 | -1873 | 214 | -25 | 37 |
| 14 | Operating Income After Depreciation | 3.01 | -2030 | 9262 | 6 | 19 | -1808 | 585 | -9 | 24 |
| 15 | Operating Income Before Depreciation | 2.97 | -1082 | 13885 | 12 | 26 | -1748 | 917 | -2 | 19 |
| 16 | Liabilities - Other | 2.93 | 0 | 155598 | 19 | 28 | 0 | 7786 | 43 | 59 |
| 17 | Intangible Assets - Total | 2.87 | 0 | 219725 | 20 | 30 | 0 | 8166 | 1 | 1 |
| 18 | Property Plant and Equipment - Total (Net) | 2.85 | 0 | 261718 | 85 | 126 | 0 | 29179 | 627 | 927 |
| 19 | Common/Ordinary Equity - Total | 2.85 | -13475 | 348296 | 237 | 352 | -11399 | 9236 | 45 | 209 |
| 20 | Liabilities - Other - Excluding Deferred Revenue | 2.83 | 0 | 139558 | 11 | 16 | 0 | 2304 | 31 | 46 |
| 21 | Revenue - Total | 2.81 | -8 | 123179 | 100 | 148 | 0 | 2482 | 84 | 121 |
| 22 | Sales/Turnover (Net) | 2.81 | -8 | 123179 | 100 | 148 | 0 | 2482 | 84 | 121 |
| 23 | Current Liabilities - Total | 2.80 | 0 | 153786 | 92 | 135 | 1 | 4893 | 150 | 193 |
| 24 | Stockholders Equity > Parent > Index Fundamental | 2.76 | -13475 | 348296 | 242 | 360 | -11399 | 9236 | 63 | 215 |
| 25 | Stockholders Equity - Total | 2.75 | -13470 | 351954 | 245 | 364 | -11291 | 9769 | 66 | 218 |
| 26 | Current Assets - Total | 2.70 | 0 | 160917 | 185 | 270 | 0 | 4328 | 225 | 292 |
| 27 | Market Value - Total | 2.66 | 0 | 566023 | 503 | 730 | 1 | 12956 | 84 | 101 |
| 28 | Cost of Goods Sold | 2.64 | 0 | 88889 | 52 | 77 | 0 | 1734 | 62 | 90 |
| 29 | Receivables | 2.62 | 0 | 78643 | 46 | 68 | 0 | 1536 | 43 | 61 |
| 30 | Operating Expense | 2.61 | 0 | 115469 | 79 | 115 | 0 | 2292 | 93 | 129 |
| 31 | Goodwill | 2.59 | 0 | 140940 | 4 | 5 | 0 | 5618 | 0 | 0 |
| 32 | Other Long-term Assets | 2.58 | 0 | 107184 | 9 | 13 | 0 | 673 | 14 | 21 |
| 33 | Receivables - Trade | 2.55 | 0 | 74798 | 29 | 43 | 0 | 1245 | 24 | 35 |
| 34 | Current Liabilities - Other - Total | 2.52 | 0 | 46157 | 33 | 49 | 0 | 1525 | 46 | 61 |
| 35 | Other Intangibles | 2.50 | 0 | 114276 | 7 | 10 | 0 | 4096 | 0 | 0 |
| 36 | Account Payable/Creditors - Trade | 2.47 | 0 | 48274 | 22 | 32 | 0 | 2107 | 49 | 67 |
| 37 | Liabilities - Total | 2.32 | 0 | 122129 | 44 | 64 | 0 | 1460 | 25 | 35 |
| 38 | Inventories - Total | 2.28 | 0 | 50147 | 7 | 10 | 0 | 3472 | 8 | 12 |
| 39 | Cash | 2.26 | 0 | 50498 | 33 | 48 | 0 | 1250 | 15 | 22 |
| 40 | Capital Surplus/Share Premium Reserve | 2.21 | -2348 | 89563 | 192 | 277 | 0 | 10272 | 428 | 485 |
| 41 | Total Long-term Investments | 2.16 | 0 | 203353 | 0 | 0 | 0 | 2704 | 0 | 0 |
| 42 | Current Assets - Other - Total | 2.15 | 0 | 35528 | 8 | 11 | 0 | 817 | 17 | 24 |

# II: Tuning Results

## Dimensionality of the grid

The average eigenvalues of the first two principal components of each quarter as depicted in Figure 14 support the choice for $p = 2$. In this Figure, one can detect an 'elbow' at the third dimension, indicating that dimensions three, four and so on contribute relatively much less to the total fit than the first two. Looking at the absolute contributions, the first principal component contributes four times as much as the third and already twice as much as the second.

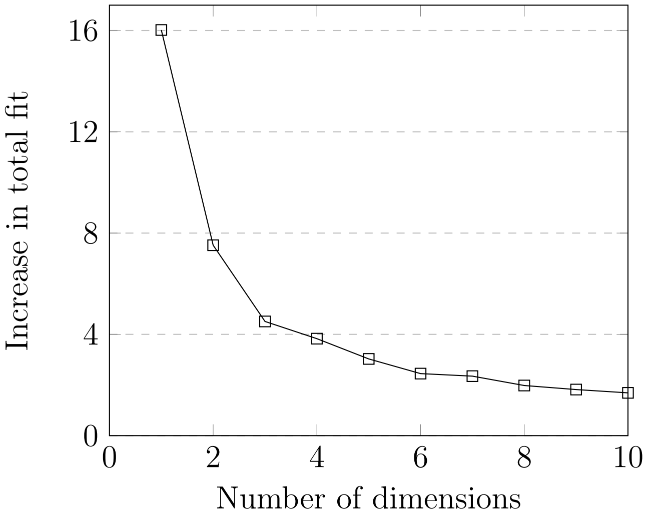


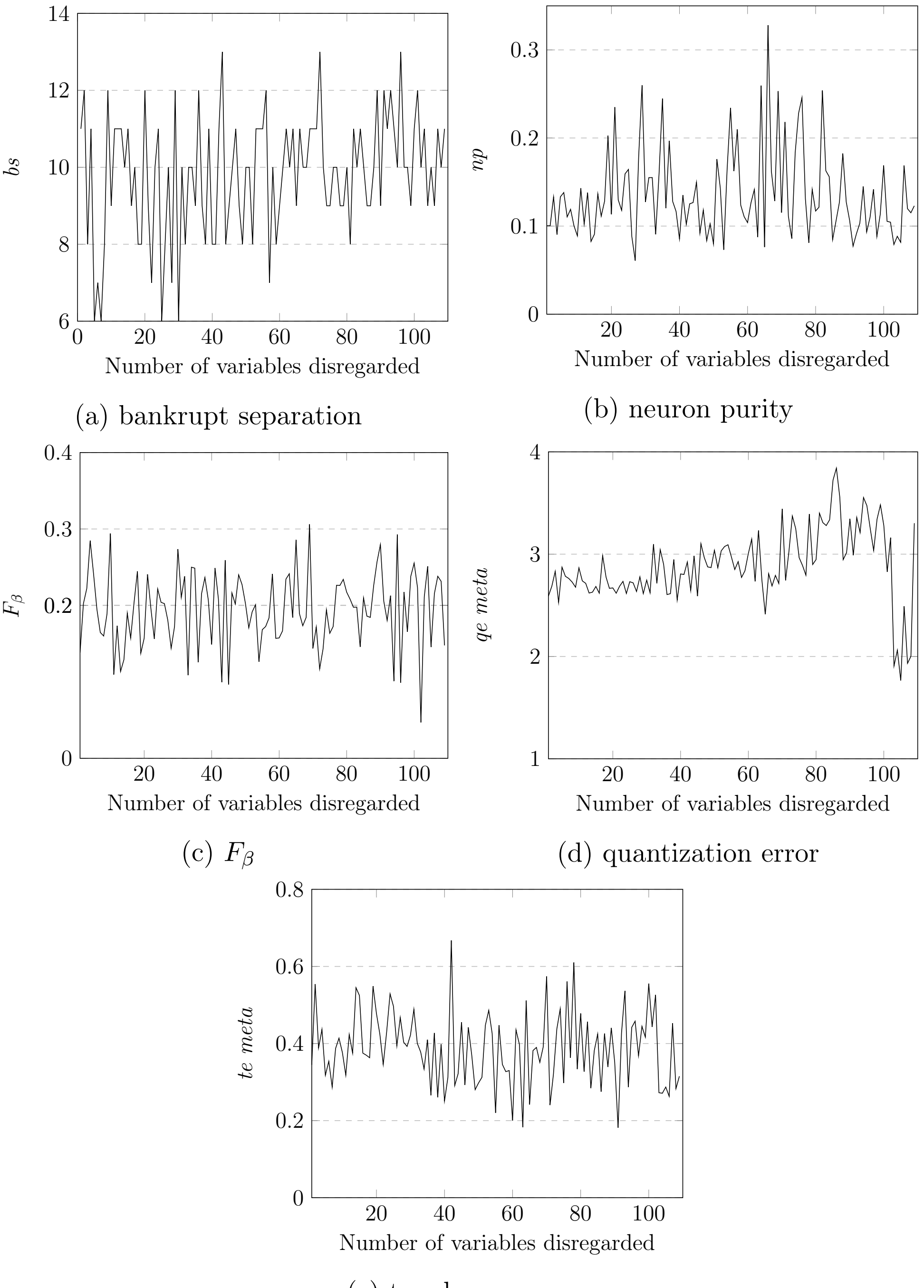
Figure 14: Scree plot depicting the average eigenvalues of the first 10 principal components of each quarterly statement

### Shape of the grid and number of variables

To decide on the optimal amount of neurons along each dimension of the SOM, the approach as explained in Section 4.4.1 was used. For both $\{r_{xb}, r_{yb}\}$ and $\{r_{xm}, r_{ym}\}$ the following combinations were evaluated: $\{6, 3\}$, $\{6, 4\}$, $\{7, 4\}$, $\{7, 5\}$, $\{8, 4\}$, $\{8, 5\}$, $\{9, 5\}$, $\{9, 6\}$, $\{10, 5\}$ and $\{10, 6\}$

In Figure 15 part of the variable selection results are presented (see Section 4.4.1 for full procedure). The figure represents the output of step 4 of the tuning process during the third and final iteration. Using base maps of size $9 \times 6$ and a meta map of size $10 \times 6$, the graphs show the performance measures on a validation set for each step in the backward elimination process. The x-axis of each subfigure depicts the number of variables eliminated, against one of the five performance measures on each y-axis. Based on these results, the optimal number of variables to disregard was set to 69. Looking at $F_\beta$ (panel c) only, this was the optimal cutoff, meaning the model with $110 - 68$ variables results in the best RF classification of the trajectories. Looking at the node purity however, a clear peak in node purity can be observed at a cutoff value of 66. The bankrupt separation (a) quantization error (d) and topology error (e) in general fluctuate around an average, and their optimal values do not coincide with each other, neither with optimal $F_\beta$ or node purity. At the cutoff of 66 or 69, all three performed conforming their average. Based on these observations, a cutoff of 69 was deemed more optimal than 66, because the decrease in $F_\beta$ performance for a cutoff of 66 compared to 69 was much higher than the decrease in node purity performance by switching from 66 to 69.

Alternatively optimizing the somsize and the optimal number of observations after the third iteration no longer changed the results. Therefore the base SOM maps were set to size $9 \times 6$, the meta SOM to size $10 \times 6$ and the first 42 variables with highest average factor loadings on the first two principal components were included in the analysis (for more details see Table 4 in I)



(a) bankrupt separation

(b) neuron purity

(c) $F_\beta$

(d) quantization error

(e) topology error

Figure 15: Output of the backward elimination process: validation set performance of the meta level depicted against a varying number of variables disregarded. The output presented is from the third iteration of backward elimination with base maps of size $9 \times 6$ and a meta SOM map of size $10 \times 6$

## Appendix III: Random Forest

A Random Forest (RF) is a supervised learning technique and consists of an ensemble of Decision Trees. Therefore, to understand the functionality of RF, first the functionality of the Decision Tree is presented.

Decision Trees

A Decision Tree (DT) is an supervised machine learning technique, proposed by Breiman et al. (1984). In its essence, a DT is a set of splitting rules, repeatedly dividing observations in groups of smaller, more similar sets of observations. This is achieved using two main steps: tree growing and tree pruning.

First, all observations are grouped together in the so-called root node. During the tree growing process, the root node is split into two child nodes, who are again divided in two child nodes on their own, and so on. For each split, all predictors are considered as possible splitting variables and the optimal splitting variable as well as the corresponding cutoff point are chosen by evaluating which split and cutoff result in the most homogeneous child nodes. This process is repeated until the tree is fully saturated, i.e. it is not possible to split the nodes in more homogeneous groups.

Put formally and in the context of this research, let all trajectories $\tilde{\mathbf{\Lambda}}_i$ for $i = 1, .., n$ be the set of observations. Furthermore, let $n_{br}$ be the group of firms going bankrupt following time $t$, label them as $c = 1$ and $n_h$ those who remain in business, labeled by $c = 0$. Imagine the observations in a random node $\rho$ need to be split and consider the element $\tilde{\mathrm{l}}_{ij}$ from $\tilde{\mathbf{\Lambda}}_i$ as a splitting variable $\tau$ at cutoff point $o$. Applying this split would result in two child nodes, $\rho_l$ and $\rho_r$ according to $\rho_L := \rho_l(o)$ and $\rho_r := \rho_r(o)$, where $\rho_l = \tilde{\mathrm{l}}_{ij} \leq o$ and $\rho_r = \tilde{\mathrm{l}}_{ij} > o$. Ideally, the split results in all bankrupt firms going to the left child node and all healthy firms to the right as this would mean the child nodes are completely pure (i.e. completely homogeneous). The gini coefficient $G$ measures the purity of the child nodes and for proposed split $\tau$ of node $\rho$ at point $o$ this given by

$$\mathrm{G}(\tau, \rho, o) = 1 - \sum_{c=0}^{1} \left[ \frac{1}{|\rho_l|} (\sum_{i \in \rho_l} \mathrm{I}_{i(c)})^2 + \frac{1}{|\rho_r|} (\sum_{i \in \rho_r} \mathrm{I}_{i(c)})^2 \right] \tag{16}$$

with $|\rho_l|$ and $|\rho_r|$ the number of observations in the left- and right child node respectively and $\mathrm{I}_{i(c)}$ and indicator function equal to 1 if observation $i = c$. $G$ is evaluated for each possible splitting variable $\tau$ at each cutoff point $o$. The split and cutoff that result in the largest decrease in $G$ is selected after which the process repeated to split the child nodes as well and so on, until the tree is fully saturated.

Naturally, the fully saturated tree is very much over-fit to the training sample, as the observations are split up to the very fine nuances in the sample which are not necessarily representative for the entire population. In the DT approach, this problem is resolved by pruning the tree: cutting off branches of the tree (splits) which add little predictive power. This is done by defining a cost function, which quantifies the trade-off between improved classification accuracy of adding a branch versus the over-fit and additional complexity these branches bring.

Random Forest

The Decision Tree has two main drawbacks. First, the Decision Tree is rather sensitive to small changes in the sample, meaning slightly different variable values could already result in a very different DT structure. Secondly, the DT is "greedy", meaning it always looks for split $\tau$ which yields the largest decrease in gini. It could however be the case that splitting on a variable which little decrease gini gives way to a much more productive split afterwards.

The Random Forest was introduced to counter the unrobust and greedy nature of the DT (see Breiman (2001)). The rationale behind RF is to train not a single tree, but a whole set of trees, i.e a forest. This is done in the following way. For each tree, a bootstrap sample (sample with replacement) of size $n$ is taken from the original set of trajectories $\mathbf{\Lambda}_i$. Each tree is therefore trained on a different sample and therefore prevents an overall overfit on the training data. It is therefore no longer necessary to prune the trees. Furthermore, for each split in each tree, a *random subset* of the predictor variables is considered for $\tau$ instead of *all* variables as done for a Decision Tree. This random element is meant to reduce the greediness of the algorithm as the splitting variables are no longer those variables that results in the largest decrease in gini overall, but merely a larger decrease than the other variables from the random subset.

After all trees are trained, the class prediction of each tree is stored and the final prediction for each observation is determined by majority vote. In this study, 500 trees were trained, as adding additional trees to the forest did not yield additional predictive power. The random number of variables considered in each split was set equal to 8 (as this was equal to the rounded square root of the number of variables $k = 54$)

## Appendix IV: The Altman Z-score

Even though the Altman Z-score has received a lot of criticism in the past, it is still widely used in practice and serves as the industry standard for quickly assessing the default risk of a company. To put the SOM-MDS methodology presented in this paper in some perspective, a comparison with the Altman Z-score performance is included in this section. The Altman Z-score is based on five financial ratios: working capital/total assets, retained earnings/total assets, EBITDA/total assets, market value of equity/total liabilities and sales/total assets. The Z-score is obtained by performing Linear Discriminant Analysis on these five predictor variable. Note that this is a stationary approach, meaning the financial ratios are considered from one financial statement only (a single time-period). Altman uses a matched sample to construct the Z-score, meaning the LDA is performed on a sample with an equal amount of bankrupt and healthy sample cases. This means for each bankrupt company a healthy company from the same industry was randomly selected. The Linear Discriminant is a value between 0 and 1, where 0 represents a firm with zero bankruptcy risk, whilst a value of 1 would represent a guaranteed failure.

Methodological recap

Linear Discriminant Analysis is a supervised clustering technique which aims to assign observations into groups by evaluating their score on the linear discriminant: a linear combination of the predictor variables. Essentially, to categorize a set of observations in $g$ groups, the method is concerned with finding $g-1$ separation lines, planes or hyperplanes (depending on the dimensionality) which separate the observations in different groups. The optimal separation lines/planes are those linear combinations of predictors who, after projection of the observations onto the lines/planes, minimize in-group variance and maximize between-group variance. This means observations in different groups should be projected as far apart as possible, whilst observations in the same group should be close. Consequently, a cut-off value can be defined at that point where observations from different groups are furthest removed. All observations on one side of the cut-off then belong to one group, and observations on the other side to a second group.

In the context of this research, let $\mathbf{X}_j$ be the $n \times k$ matrix containing information on $k = 5$ financial ratio's of $n$ firms at time $j$. Let $\mathbf{G}$ be an $n \times 2$ matrix with on each row a 1 in the first column if the firm belonged to class $c = 0$ (healthy) and a 1 in the second colun if the firm went bankrupt ($c = 1$) following time $t$. Lastly, let $\mathbf{M}$ be a $2 \times k$ matrix containing the respective group means on each variable for class $c = \{0, 1\}$.

To store the between-group variance, the $k \times k$ matrix $\mathbf{A} = \frac{1}{n-1}(\mathbf{GM} - \mathbf{1}\bar{\mathbf{x}}')'(\mathbf{GM} - \mathbf{1}\bar{\mathbf{x}}')$ is computed, with $n$-dimensional vector $\mathbf{1}$ containing ones only and $\bar{\mathbf{x}}$ an $k$-dimensional vector with the overall means (over all classes) for each line item. In the same way, $\mathbf{W} = \frac{1}{n-1}(\mathbf{X} - \mathbf{GM})'(\mathbf{X} - \mathbf{GM})$ represents the within-group variance.

The optimal LDA solution is the projection vector $\mathbf{p}$, which can be obtained by maximizing

$$L(\mathbf{p}) = \frac{\mathbf{p}'\mathbf{A}\mathbf{p}}{\mathbf{p}'\mathbf{W}\mathbf{p}} \tag{17}$$

constrained to $\mathbf{p}'\mathbf{W}\mathbf{p} = \mathbf{I}$. This can be achieved by defining $\mathbf{b} = \mathbf{W}^{1/2}\mathbf{p}$ and. $\mathbf{b}'\mathbf{b} = 1$. Equation 17 then equals

$$L(\mathbf{p}) = \mathbf{b}'\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\mathbf{b} = \mathbf{b}'\mathbf{E}\mathbf{b} = \mathrm{tr}\mathbf{B}'\mathbf{E}\mathbf{B}. \tag{18}$$

Subsequenty, equation 18 is maximized by performing an eigendecomposition on $\mathbf{B}$ and using this maximum, the optimal LDA projection $\mathbf{p}$ is found by $\hat{\mathbf{p}} = \mathbf{W}^{-1/2}\mathbf{B}$.

The Altman financial ratios

Altman defined five financial ratios to be important for assessing the bankruptcy rik of a firm. In Table 5 below, the five financial ratio's as well as some descriptive statistics are depicted for both the healthy and distressed group. As can be observed from the Table, the range of values is much larger for the healthy compared to the distressed group. The median and mean absolute deviation of each group however are rather similar.

Table 5: Descriptive statistics of the Altman financial ratio's. Presented are the minimum (min), maximum (max), median (med) and mean absolute deviation (mad) of each ratio.

| | Healthy | | | | Bankrupt | | | |
|---|---|---|---|---|---|---|---|---|
| Ratio's | min | max | med | mad | min | max | med | mad |
| Working Capital / Total Assets | -39.02 | 99.16 | 0.70 | 0.25 | -5.38 | 1.52 | 0.47 | 0.49 |
| Retained Earnings / Total Assets | -1539.26 | 45.80 | -0.51 | 1.63 | -10.80 | 1.02 | -0.85 | 0.51 |
| EBITDA / Total Assets | -697.01 | 37.89 | 0.33 | 0.43 | -6.70 | 1.83 | -0.19 | 0.56 |
| Market Value of Equity / Total Liabilities | -233.22 | 624.73 | 1.14 | 0.28 | -0.03 | 14.58 | 0.91 | 0.26 |
| Sales / Total Assets | -952.38 | 16.40 | 0.71 | 0.18 | -0.37 | 1.14 | 0.66 | 0.18 |

Results

The following coefficients were obtained for the five financial ratio's:
$\{-0.01, -0.23, -1.94, -0.43, 1.4\}$ for "Working Capital / Total Assets", "Retained Earnings / Total Assets", "EBITDA / Total Assets", "Market value of Equity / Total Liabilities" and "Sale / Total Assets" respectively. As healthy companies were marked as "0"

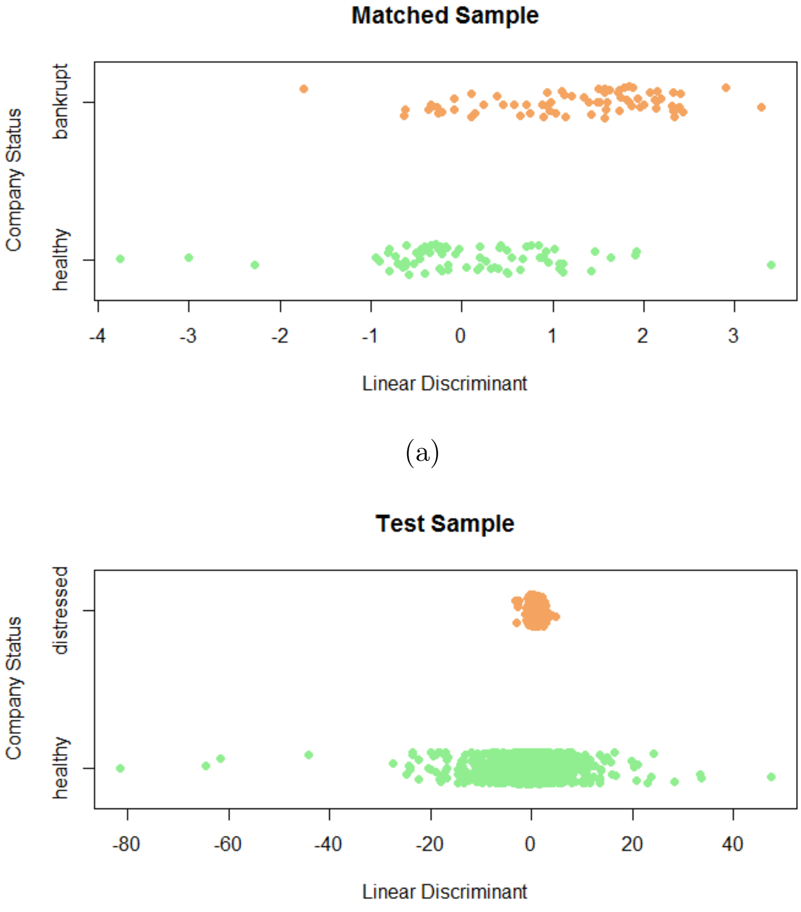
and bankrupt companies "1", this means a higher value for the first four ratios decreased the bankruptcy risk, while the results seem to indicate that higher "sales / total assets" increased bankruptcy risk. This is counter-intuitive, as high sale compared to total assets this normally indicates the firm can produce sales efficiently. Because LDA required the data to be multivariate normally distributed and the covariance matrices of healthy and non-healthy firms to be homogeneous, it is not surprising that data that does not meet these criteria produces change results. It is for that exact reason the SOM-MDS procedure was proposed in this study as it requires very few assumptions on the data distribution. Again, the Altman-Z score is however often used in practice and therefore added in this section merely as a comparison.

In Figure 16 the linear discriminant value of each observation is depicted. As can be seen from panel (a), for the matched in-sample observations, the average discriminant for healthy companies is lower than for bankrupt companies. This makes sense, as 0 represented the completely healthy and 1 the guaranteed bankrupt firm. However, there is a rather large overlap in the range of discriminant scores of both groups, and therefore it seems difficult to define a cut-off for classification. In panel (b) the discriminant scores are shown for the out-of-sample, unmatched set. The panel conforms to the in-sample findings: the range of the bankrupt linear discriminant scores is much smaller, but completely overlaps with the healthy range.

If one would allow the same amount of false positive bankrupt predictions as the best performing model in the trajectory SOM approach (the "Tandem Multi"), this would result an a classification as presented in Table 6. As expected, the performance of the Altman Z-score is worse than any of the trajectory SOM approaches.

Table 6: In- and out-of-sample confusion tables for the Altman-Z score

| In Sample | | | Out Sample | | |
|---|---|---|---|---|---|
| | $0_p$ | $1_p$ | | $0_p$ | $1_p$ |
| $0_t$ | 2865 | 806 | $0_t$ | 674 | 244 |
| $1_t$ | 19 | 55 | $1_t$ | 8 | 10 |



(a)

(b)

Figure 16: Multiple stripchart depicting a jittered projection of the Linear Discriminant score of each observation (a) depicts the matched sample projection (b) depicts an out of sample projection

# References

ACAA (2011). Audit under fire: a review of the post-financial crisis inquiries. Technical report.

Accountant (2018). Kpmg in vk onder vuur om carillion-debacle.

Agostini, M. (2018). *Corporate Financial Distress: Going Concern Evaluation in Both International and US Contexts*. Springer.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4):589–609.

Altman, E. I., Haldeman, R. G., and Narayanan, P. (1977). Zetatm analysis a new model to identify bankruptcy risk of corporations. *Journal of banking & finance*, 1(1):29–54.

Altman, E. I. and Loris, B. (1976). A financial early warning system for over-the-counter broker-dealers. *The Journal of Finance*, 31(4):1201–1217.

Altman, E. I., Marco, G., and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience). *Journal of banking & finance*, 18(3):505–529.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on neural networks*, 12(4):929–935.

Balcaen, S. and Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1):63–93.

BankruptcyData (2019). Bankruptcydata's 2018 corporate bankruptcy review: Public fillings fall for second consecutive year but looming debt issues and more trouble in retail point to uptick in 2019. Technical report.

Bauer, H.-U. and Schöllhorn, W. (1997). Self-organizing maps for the analysis of complex movement patterns. *Neural Processing Letters*, 5(3):193–199.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, pages 71–111.

Blum, M. (1974). Failing company discriminant analysis. *Journal of accounting research*, pages 1–25.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. wadsworth int. *Group*, 37(15):237–251.

Bureau of Labor Statistics (2015). Labor force statistics from the current population survey.

Chen, N., Ribeiro, B., Vieira, A., and Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications*, 40(1):385–393.

Chi, L.-C. and Tang, T.-C. (2006). Bankruptcy prediction: Application of logit analysis in export credit risks. *Australian Journal of Management*, 31(1):17–27.

Cole, R. A., Wu, Q., et al. (2009). Predicting bank failures using a simple dynamic hazard model. In *22nd Australasian Finance and Banking Conference*, pages 16–18. Citeseer.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting research*, pages 167–179.

Deboeck, G. and Kohonen, T. (1998). *Visual explorations in finance: with self-organizing maps*. Springer Science & Business Media.

Du Jardin, P. and Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems*, 51(3):701–711.

Economics, T. (2018). United states bankruptcies.

Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative analysis*, 7(2):1477–1493.

FED (2015). Federal reserve issues fomc statement.

Financial Times (2018). Why carillion was placed in liquidation.

FitzPatrick, P. J. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies*. Washington.

FRC (2018). Frc announces inquiry into going concern assessments.

Frydman, H., Altman, E. I., and Kao, D.-L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance*, 40(1):269–291.

Geng, R., Bose, I., and Chen, X. (2015). Prediction of financial distress: An empirical study of listed chinese companies using data mining. *European Journal of Operational Research*, 241(1):236–247.

Heo, J. and Yang, J. Y. (2014). Adaboost based bankruptcy forecasting of korean construction companies. *Applied soft computing*, 24:494–499.

Honkela, T., Kaski, S., Lagus, K., and Kohonen, T. (1996). Exploration of full-text databases with self-organizing maps. In *Neural Networks, 1996., IEEE International Conference on*, volume 1, pages 56–61. IEEE.

Honkela, T., Kaski, S., Lagus, K., and Kohonen, T. (1997). Websom—self-organizing maps of document collections. In *Proceedings of WSOM*, volume 97, pages 4–6.

ISAA (2016). G5 international standard on auditing (isa) 570 (revised) going concern. Technical report.

Iturriaga, F. J. L. and Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of us commercial banks. *Expert Systems with applications*, 42(6):2857–2869.

Johnsen, T. and Melicher, R. W. (1994). Predicting corporate bankruptcy and financial distress: Information value added by multinomial logit models. *Journal of Economics and Business*, 46(4):269–286.

Kim, M. H. and Partington, G. (2015). Dynamic forecasts of financial distress of australian firms. *Australian Journal of Management*, 40(1):135–160.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.

Kourtit, K., Nijkamp, P., and Arribas, D. (2012). Smart cities in perspective–a comparative european study by means of self-organizing maps. *Innovation: The European journal of social science research*, 25(2):229–246.

Kuo, R., Ho, L., and Hu, C. M. (2002). Integration of self-organizing feature map and k-means algorithm for market segmentation. *Computers & Operations Research*, 29(11):1475–1493.

Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., and Sornmo, L. (2000). Clustering ecg complexes using hermite functions and self-organizing maps. *IEEE Transactions on Biomedical Engineering*, 47(7):838–848.

Laitinen, E. K. (1991). Financial ratios and different failure processes. *Journal of Business Finance & Accounting*, 18(5):649–673.

Leshno, M. and Spector, Y. (1996). Neural network prediction analysis: The bankruptcy case. *Neurocomputing*, 10(2):125–147.

Lukason, O. (2012). Financial performance before failure: Do different firms go bankrupt differently? *International Journal of Trade, Economics and Finance*, 3(4):305.

Martens, D., Bruynseels, L., Baesens, B., Willekens, M., and Vanthienen, J. (2008). Predicting going concern opinion with data mining. *Decision Support Systems*, 45(4):765–777.

Min, J. H. and Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, 28(4):603–614.

NAICS association (2017). Naics identification tools.

NBA (2018). Green paper organisatiecontinuïteit. Technical report.

Odom, M. D. and Sharda, R. (1990). A neural network model for bankruptcy prediction. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pages 163–168. IEEE.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pages 109–131.

Ooghe, H. and De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: a typology. *Management Decision*, 46(2):223–242.

PCAOB (2018). Standard - setting update. Technical report.

Peel, M. J. and Peel, D. A. (1987). Some further empirical evidence on predicting private company failure. *Accounting and Business Research*, 18(69):57–66.

Schreck, T., Tekušová, T., Kohlhammer, J., and Fellner, D. (2007). Trajectory-based visual analysis of large financial time series data. *ACM SIGKDD Explorations Newsletter*, 9(2):30–37.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business*, 74(1):101–124.

Swanson, E. and Tybout, J. (1988). Industrial bankruptcy determinants in argentina. *Studies in Banking and Finance*, 7(1-25).

Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S. J., and Roy, N. (2011). Understanding natural language commands for robotic navigation and mobile manipulation. In *AAAI*, volume 1, page 2.

Telmoudi, F., El Ghourabi, M., and Limam, M. (2011). Rst–gcbr-clustering-based rga–svm model for corporate failure prediction. *Intelligent Systems in Accounting, Finance and Management*, 18(2-3):105–120.

The Guardian (2018). Kpmg to be investigated over carillion auditing.

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419.

Törönen, P., Kolehmainen, M., Wong, G., and Castrén, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS letters*, 451(2):142–146.

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik*, 133:97–178.

Wang, Y., Wang, S., and Lai, K. K. (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, 13(6):820–831.

Wang, Z., Lu, M., Yuan, X., Zhang, J., and Van De Wetering, H. (2013). Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168.

Xiao, X., Dow, E. R., Eberhart, R., Miled, Z. B., and Oppelt, R. J. (2003). Gene clustering using self-organizing maps and particle swarm optimization. In *Parallel and Distributed Processing Symposium, 2003. Proceedings. International*, pages 10–pp. IEEE.

Yeh, C.-C., Chi, D.-J., and Lin, Y.-R. (2014). Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, 254:98–110.

Yim, J. and Mitchell, H. (2003). A comparison of corporate failure models in australia: hybrid neural networks, logit models and discriminant analysis. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 348–358. Springer.

Zavgren, C. V. (1985). Assessing the vulnerability to failure of american industrial firms: a logistic analysis. *Journal of Business Finance & Accounting*, 12(1):19–45.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, pages 59–82.