

Analysis of the Economic Sustainability of Companies in the Water Sector

Lyubomir Halachev¹, Aziz Guergachi², Yashodhan Athavale³, Sri Krishnan⁴

¹ *School of Business Management, Ted Rogers School of Management, Ryerson University, Toronto, Canada (lyubomir.halachev@ryerson.ca)*

² *Department of Information Technology Management, Ted Rogers School of Management, Ryerson University, Toronto, Canada (a2guerga@ryerson.ca)*

³ *Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada (yashodhan.athavale@ryerson.ca)*

⁴ *Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada (krishnan@ee.ryerson.ca)*

Abstract: Many recent studies look at the issue of sustainability from the environmental and the social points of view. This paper takes a different approach and proposes to look at the economic stability of organizations in the water sector. It analyzes a set of 140 companies from around the world involved in this sector, and attempts to develop a methodology to find out, based on stock market data only, which ones will likely remain active in the foreseeable future, and which ones will likely "die" out. This methodology uses machine learning techniques to handle the stock price time series of companies over a period of variable length. These techniques make use of tools such as support vector machines and kernel learning, and more specifically of the Fisher kernel. This methodology has been previously tested on other sectors and has shown good results. This paper expands the testing to the water sector on an international basis and thus links it to the important area of sustainability. The paper's results further confirm the potential of the methodology and provide prospective for future research.

Keywords: Fisher kernels; Machine learning; Support Vector Machines; Stock price; Sustainability

1. INTRODUCTION

The idea of sustainable development, developed by and still strongly supported by the United Nations [1987], states that social, environmental and economic goals should be pursued simultaneously. This idea implies that change in one of these areas would result in at least the possibility of an impact on the other two areas. Economic and financial stability of companies in certain industries will then have a strong impact on the environmental and social stability in the geographic areas where these companies operate.

While many studies look at the issue of sustainability from the environmental and the social responsibility point of views, this paper proposes to look at the economic sustainability of businesses in the water sector. It analyzes a set of 140 companies from around the world involved in this sector in about 20 different countries, and attempts to develop a methodology to find out which companies will likely remain active in this sector, and which ones will likely 'die' out (i.e., get delisted from the stock market). Only stock price data is considered for the development of this classification mechanism as it is the most readily and objectively observable financial metric of public companies. The ability to identify the companies that are likely to survive in the water sector would have financial value for investors as the knowledge about a potential corporate failure may prevent losses or even be profitable. Additionally, this information would have value from a social point of view, as unemployment rates and personal bankruptcies may arise as a result of companies' deaths. From an ecological point of view, "dying" companies, depending on

their impact (negative or positive) on the environment (for a discussion of one tool for measuring impact see Searcy [2009]), can be either publicly salvaged or left to themselves.

The methodology proposed in this paper uses machine learning techniques to handle the stock price time series of company over a period of variable length, and make decisions as to whether the company will survive as an independent entity over the long-term or not. These techniques make use of the tools of support vector machines, and kernel learning. More specifically, they implement the concept of Fisher kernel to be able to process inputs that are time series of variable length, and thus not of the vector-type. This methodology has been previously tested on some other economic sectors and has shown good results. Thus, a conference paper by Athavale et al. [2009] set the precedent in this research when it looked at the Technology, the Pharmaceutical and the Banking sectors. The authors found reasonable distinction between active and dead companies in the Banking sector. The current paper extends this direction of economic sustainability research by analyzing companies in the water sector on a global scale, as opposed to constraining it to the North American markets as was done so far.

It seems appropriate that the methodology is utilized in the water sector, as the "death" of a water company performing an essential service can result in significant financial, social and ecological consequences. Most of these may be neutralized if society has timely knowledge of future corporate deaths and uses it to prevent them. Therefore, it appears useful that this type of forecasting analysis is performed for the water sector in order to preserve water companies' welfare, as well as to promote social, environmental and economic sustainability.

2. DESCRIPTION OF THE APPROACH

The approach that was used in this research consisted of two stages – collecting the company stock data for a sufficiently long period and applying the Fisher kernel on the stock price time series data.

2.1 Data Collection

The data for this paper was collected through the use of the Mergent and the Thompson Reuters Datastream databases.

The Mergent database allows one to search for and find companies that have water as part of their business. Thus, this study does not limit its scope to water treatment, water distribution or water bottling companies, but rather to a much broader variety of companies that operate in or service the water industry. Companies that provide IT solutions, financing, piping and irrigation equipment, as well as engineering solutions to the water industry are examples of the included companies. These were selected on a pseudo-random principle, in which the authors looked at an existing exchange-traded fund that focuses on companies that are servicing the water sector. Such companies were then selected through Mergent in a way that replicated that fund's sector structure. Similar procedures were followed for both active and dead companies.

After a list of 140 companies was obtained, weekly stock price data for the past 25 years was downloaded through Datastream. Because of the global nature of the companies selected, all prices had to be converted to a single currency. In this paper the currency selected was US dollars. More importantly, in its definition of "dead" companies, Datastream does not distinguish between bankrupted and merged/acquired companies, which warrants a careful interpretation of any results of this methodology. This issue will be further discussed in the conclusions section of this paper. Finally, log returns were computed from the historical stock prices in order to use some theoretical properties of returns, as discussed in the Black-Scholes option pricing theory [Black and Scholes, 1973].

2.2 Support Vector Machines and the Fisher Kernel

To process the collected time series data, a two-step method was applied to extract relevant *features* from the stock price time series (step 1 – using Fisher kernel), and then, based on these features, design an algorithm to carry out the classification of a company as either ‘active’ or ‘dead’ (step 2 – using support vector machines (SVM)). The actual mathematical equations that describe in detail these two steps can be found in the paper by Hosseinizadeh et al. [2009]. In this article, we explain qualitatively the techniques that underlie the two steps.

Fisher kernel and its corresponding features: The Fisher kernel can be considered a feature extractor which transforms a stock price time series T into a vector of a few scalar quantities, called features. Features provide a characterisation of T and can be of many different types. For instance, the mean of T is a feature of the time series. The variance is another feature. The frequency spectrum of the time series signal also provides features for the time series. In this work, the features that have been considered are derived from the Fisher kernel [Jaakola and Haussler, 1998]. When we use the Fisher kernel, a probability density function (pdf) on the space of all time series T has to be selected. In this research, an assumption is made that the two groups of companies (dead and active) will be described by two different normal distributions. Thus, a Gaussian Mixture Model (GMM) composed of two components is used for the pdf for all time series, where each component is a normal distribution parameterized by a mean and variance. The two components are combined using two mixing coefficients, called priors. The parameters (means, variances, priors) are estimated using the Expectation Maximization Algorithm [Moon, 1996]. The features that are used in this research are the coordinates of the Fisher Score, which is obtained by computing the partial derivatives of the log-likelihood of the GMM with respect to the parameters (mean, variance, prior) of this GMM. The Fisher kernel itself is obtained by defining a dot product in the vector space generated by the Fisher Score.

Support vector machines (SVM): SVM is used in this paper as a classification algorithm just like decision trees or neural networks. The SVM algorithm uses hyperplanes to construct classifiers. The hyperplane that is selected by the SVM algorithm is the one that provides the maximum distance between the hyperplane itself and the closest points in the vector space of the feature vectors. The closest points are called the support vectors (SV), and the selected optimal hyperplane can be expressed in terms of the coordinates of these SV.

3. RESULTS

After the above approach was applied, the plot displayed in Figure 1 was obtained. In this figure, each point represents one company plotted using the summed features obtained from the Fisher Score; the circles represent the active companies, and the crosses stand for the dead companies. This tri-dimensional plot shows clearly that the features that have been used provide an effective method for clustering the active and dead companies. This seems to support the hypothesis that there is a distinction between these two types of companies. The Fisher scores of the two groups seem to differ noticeably, especially with respect to their variances. When SVM is applied to this data set, the classification error that is obtained is less than 10%.

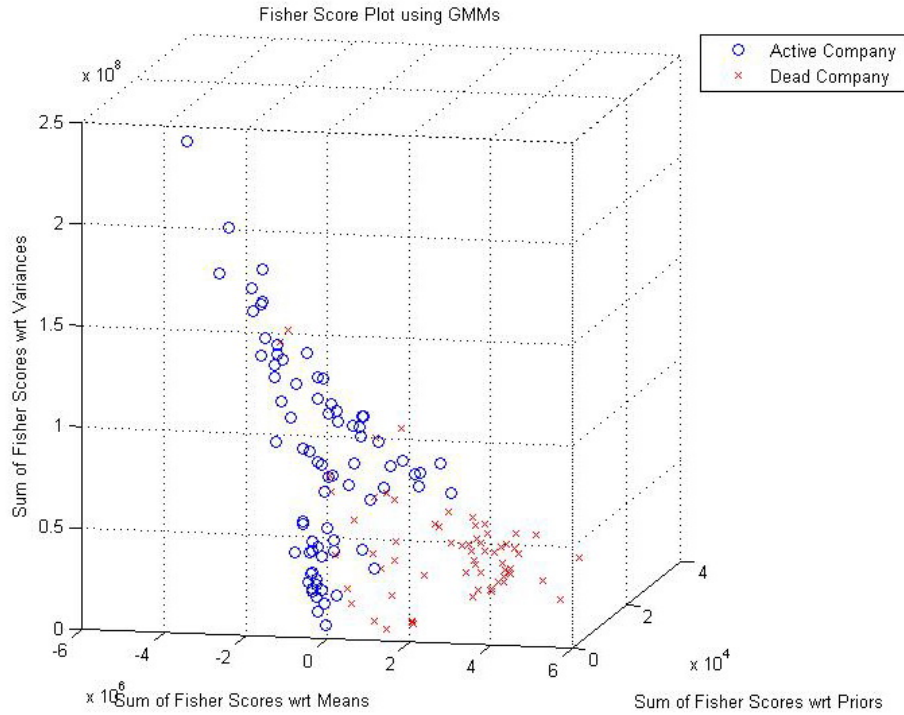


Figure 1. Both active and dead companies plotted using Fisher Scores

As a way to show that the labels ‘active’ and ‘dead’ do indeed bring a great deal of information to the task of separating a company that will likely survive from one that is likely to ‘collapse’ and to point out that such a distinction is very unlikely to happen randomly, the authors used two randomly selected groups created from the same total set of companies as the original. Both active and dead companies were combined into one list and assigned numbers from 1 to 140. Thereafter, the list was shuffled and two new groups were formed. There are about $5.6E+140$ unique combinations to choose these two equal groups. One such possibility was used and the test code was run on it. The two groups were now much more visually mixed and a reasonable distinction between the fictitious ‘active’ and ‘dead’ companies was no longer possible (Figure 2). It is important to note that this random simulation does not represent a form of calibration. Its purpose is to demonstrate that the discrimination we see in the first plot between the actual active and dead companies is not the result of pure chance, but is rather a consequence of the information contained in the historical labels ‘active’ and ‘dead’. The simulation uses the original set of companies in order to point out that in the vast majority of the cases one would not expect their Fisher scores to be as neatly classified as they are when grouped as active or dead. Thus, this classification seems to contain important ‘non-random’ information. It should also be noted here that, while the classification error when the actual sets of dead and active companies have been used is less than 10%, this error approaches, predictably, 50% in the case of the randomly selected sets.

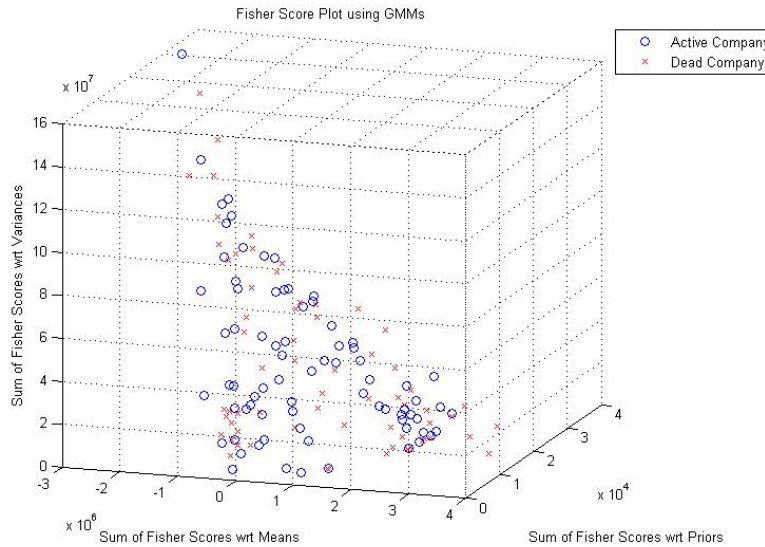


Figure 2. Plot obtained by randomly selecting two groups of companies

4. DISCUSSION, CONCLUSION AND FUTURE WORK

The results of this paper confirm that a methodology based on engineering methods and the simple measure of stock prices may have validity and usefulness for a variety of sectors. Extending previous research that has focused on several sectors in Canada and the US, this paper has demonstrated that a distinction is possible for publicly traded companies on a more global basis for companies that provide service to the water sector. Considering the importance of the water sector for sustainability worldwide, this result has important implications. Perhaps the most crucial one is that one may be able to distinguish early on between companies that have difficulties and are about to die out, and the ones that are likely to survive in a longer term. Then, similarly to the popular business practice of identifying and protecting corporations that are “too big to fail”, society may identify which of the ‘dying’ companies are “too important to fail” (because, for example, they contribute in a significant manner to the environmental bottom line), and attempt to protect them. Sustainability comes at a price, but preventive action and early knowledge about potential economic disruptions in sustainability may well be cheaper than reactive attempts to fix a situation.

Furthermore, financial theory suggests that investors, through their view of fair stock prices, account for all kinds of information they receive about companies in the water sector, and not merely about their technological or engineering potential. This idea fits well with the findings of Beck [2009a, 2009b] who argued that effective governance, inspired corporate leadership, and ability to learn and adapt in water-related companies is just as, if not more important than, its technological structure. Therefore, companies that find their Fisher scores closer to the ‘dead’ group may be interested in implementing both new engineering and management techniques to stay alive and serve their customers and society better. One important point should be made here. In this paper, the Fisher scores were computed using 25 years of weekly stock price data and so may be reflecting long term trends in companies’ financial health. This may mean that solutions or adjustments by companies may not have a quick effect on their Fisher score. Further research is currently in progress attempting to dynamically track companies’ scores through the years and to pinpoint the precise time intervals in which changes in the scores occur. This will allow one to look further into the events and the decisions that can cause these changes and to determine the speed of their effect. This would increase the usefulness of the current results and would allow for proper timing and planning of decisions regarding the fate of companies.

Furthermore, it should be noted that the notion of “death” is used in this paper to refer to the situation where the company is delisted from the stock market. The definition used by Datastream does not distinguish among the various ways in which a company gets delisted. Thus, an organization may be considered “dead” if it was acquired by another company, merged with a partner, or went bankrupt. Each of these events usually has different implications for investors and it would be important to discriminate among them. However, one can argue that from a social and environmental point of view all of these “deaths” can be considered negative to one extent or another.

Bankruptcies have obvious negative consequences, some of which include destroyed capital and increased unemployment, as well as potential long-term negative effects on sustainability if companies that provide important social service by improving the quality of water and/or making water distribution safer and more efficient disappear as a result of a bankruptcy. Furthermore, as Wein and Morris [1985] point out, in some cases a bankruptcy may discharge companies from cleanup and other environmental obligations, though in recent years courts have been stricter on those loopholes. However, acquisitions and mergers can also have negative implications. Possibly the major losers when a company gets delisted as a result of an acquisition or a merger are its employees. Often, new management will look for ways to improve synergies and reduce redundancy and will eliminate a number of jobs to make the new organization more efficient than the mere sum of its parts. Job losses can be considered a negative consequence on society. Mergers and acquisitions can also result in reduced competition, decreased quality standards and higher prices, which may further burden taxpayers. If society has the knowledge, a few years earlier, that there is a possibility that a “death” will occur, important companies may be salvaged.

Thus, while this paper provides interesting implications for the prediction of “death” as defined by Datastream, future research may focus on the implications of bankruptcies specifically, which potentially carry out the worst consequences on society and the environment. As previously mentioned, other interesting modifications of the current methodology would include a dimension of time. This would provide even more important information as to when a “death” is expected and how it can be prevented. Future research may also include other environmental sectors. For example, companies involved in the alternative energy or the waste management industries may be investigated. Nevertheless, while there is indeed a potential for improvement in the methodology, this paper may be an important first step towards a more sustainable management of socially important organizations.

REFERENCES

- Athavale, Y., Hosseinizadeh, P., Guergachi, A., and S. Krishnan, Identifying the potential for failure of businesses in the technology, pharmaceutical and banking sectors using kernel-based machine learning methods, paper presented at the IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA, Oct. 11-14, 2009.
- Beck, M.B., Grand challenges of the future for environmental modeling, white paper for the US National Science Foundation (NSF), 2009a (available at http://warnell.forestry.uga.edu/photos/beck/nsfwhitepaper_beck-05.pdf)
- Beck, M.B., Sustainability concepts paper, IWA SSG, 2009b.
- Black, F. and Scholes, M., The pricing of options and corporate liabilities, *Journal of Political Economy* 81 (3): 637–654, 1973.
- Hosseinizadeh, P., Guergachi, A. and V. Magness, Predicting system collapse: Two theoretical models, paper presented at the IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA, Oct. 11-14, 2009.
- Jaakkola, T.S. and Haussler, D., Exploiting generating models in discriminative classifiers, paper presented at the Conference on Advances in Neural Information Processing Systems, Denver, Colorado, USA, Nov.30- Dec.5, 1998.
- Moon, T.K., Expectation-maximization algorithm, *IEEE Signal Processing Magazine* 13 (6): 47-60, 1996.

- Searcy, C., Corporate sustainability performance measurement: Lessons from system of systems engineering, paper presented at the IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA, Oct. 11-14, 2009.
- United Nations, Report of the World Commission on Environment and Development, 11 December, 1987. Retrieved February 10, 2010 from <http://www.un.org/documents/ga/res/42/ares42-187.htm>
- Wein, H.J., and J.D. Morris, Emerging issues in bankruptcy and environmental law, *Environment*, 27(6), 4-5 & 43, 1985.