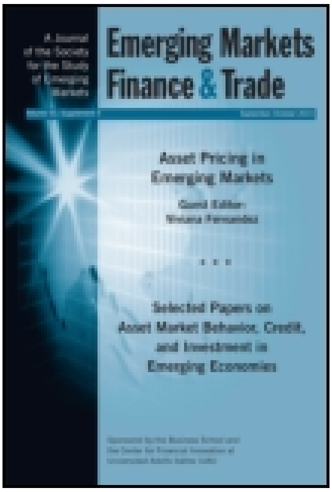

Emerging Markets Finance and Trade
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/mree20

The Application of Corporate Governance Indicators With XBRL Technology to Financial Crisis Prediction

Chien-Kuo Li[a], Deron Liang[b], Fengyi Lin[c] & Kwo-Liang Chen[d]

[a] Department of Information Technology and Management, Shih Chien University, Taipei, Taiwan

[b] Department of Computer Science and Information Engineering, and Software Research Center, National Central University, Taoyuan, Taiwan

[c] Department of Business Management, National Taipei University of Technology, Taipei, Taiwan

[d] Information Technology Division, Small & Medium Enterprise Administration, Ministry of Economic Affairs, Taipei, R.O.C.

Published online: 31 Mar 2015.

To cite this article: Chien-Kuo Li, Deron Liang, Fengyi Lin & Kwo-Liang Chen (2015) The Application of Corporate Governance Indicators With XBRL Technology to Financial Crisis Prediction, Emerging Markets Finance and Trade, 51:sup1, S58-S72, DOI: 10.1080/1540496X.2014.998888

To link to this article: http://dx.doi.org/10.1080/1540496X.2014.998888

Emerging Markets Finance & Trade, 51:S58–S72, 2015

ISSN: 1540-496X print/1558-0938 online
DOI: 10.1080/1540496X.2014.998888

The Application of Corporate Governance Indicators With XBRL Technology to Financial Crisis Prediction

Chien-Kuo Li[1], Deron Liang[2], Fengyi Lin[3], and Kwo-Liang Chen[4]

[1]Department of Information Technology and Management, Shih Chien University, Taipei, Taiwan; [2]Department of Computer Science and Information Engineering, and Software Research Center, National Central University, Taoyuan, Taiwan; [3]Department of Business Management, National Taipei University of Technology, Taipei, Taiwan; [4]Information Technology Division, Small & Medium Enterprise Administration, Ministry of Economic Affairs, Taipei, R.O.C.

ABSTRACT: The widespread adoption of eXtensible Business Reporting Language (XBRL) suggests that intelligent software agents can now use financial information disseminated on the Web with high accuracy. Financial data have been widely used by researchers to predict financial crises; however, few studies have considered corporate governance indicators in building prediction models. This article presents a financial crisis prediction model that involves using a genetic algorithm for determining the optimal feature set and support vector machines (SVMs) to be used with XBRL. The experimental results show that the proposed model outperforms models based on only one type of information, either financial or corporate governance. Compared with conventional statistical methods, the proposed SVM model forecasts financial crises more accurately.

KEY WORDS: corporate governance indicators, extensible business reporting language (XBRL), feature selection, financial crisis prediction, genetic algorithm, support vector machine (SVM)

Introduction

Beaver (1996) states that a financial crisis is indicated when a firm announces its bankruptcy, bond default, overdrawn bank account, or nonpayment of preferred stock dividends. Numerous financial scholars have defined a financial crisis as a situation in which publicly traded companies face a change in full value delivery or are legally engaged in transaction suspension, reconstruction, bankruptcy, or retreat from the public stock market (Altman 1968; Chen and Hsiao 2008; Min et al. 2006). Such distressed firms might exhibit signs of bankruptcy, which may reflect their financial statement features, such as financial leverage, long- and short-term capital intensiveness, return on investment, earnings per share (EPS), fixed asset turnover, profit growth rate, revenue per share, and debt coverage stability (Ahn et al. 2000).

Financial crisis prediction (FCP) is a vital research topic in accounting and finance. A wide variety of diagnosis techniques for predicting business crises have been developed, including traditional statistical and machine-learning methods. Multivariate discriminant analysis (MDA) and logistic regression (logit) are two widely applied statistical methods that have been adopted in a number of studies (Altman 1968; Lee et al. 1996; Ohlson 1980). MDA is often used to generate a set of models for classifying various groups and to obtain linear combinations with significant predicted variables. To reduce the quantity of necessary variables is greatly beneficial (Chen and Hsiao 2008). Logit regression can prevent self-correlation between the estimated residuals and collinearity that exist in independent variables. Statistical data requirements can also be normally distributed.

Address correspondence to Chien-Kuo Li, Department of Information Technology and Management, Shih Chien University, Taipei, Taiwan. E-mail: ckli@g2.usc.edu.tw

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/mree.

In addition to statistical methods, various machine-learning techniques have been developed for FCP, including decision trees (Tam and Kiang 1992), case-based reasoning (CBR) (Jeong et al. 2012; Li and Sun 2009), artificial neutral networks (ANNs) (Hu 2009; Ozkan-Gunay and Ozkan 2007; Tam and Kiang 1992), and support vector machines (SVMs) (Chandra et al. 2009; Ding et al. 2008; Hua et al. 2007). The processes of SVM development are performed by solving a linear and constrained quadratic programming problem to obtain the unique and global optimal solutions, leading to more favorable results than those of other methods (Min and Lee 2005; Ohlson 1980; Shin et al. 2005).

The design of prediction models in previous studies has been mainly based only on financial ratios. Scholars and practitioners have suggested that the diagnosis capability of a model is limited if the model considers only financial ratios (Lee 2007). Other categories of financial data that may contain critical information on various aspects of corporate operations should also be considered.

The Asian financial crisis in particular prompted studies on the effectiveness of corporate governance indicators, such as ownership structures, across Asian countries for FDP problems, in addition to classical financial ratios (Claessens et al. 2000; Hui 2005). Lee and Yeh (2004) consider a data set with four financial ratios and eleven corporate governance indicators (CGIs), whereas Wu (2007) considers a data set with six financial ratios and ten CGIs. Lee and Yeh (2004) examine the effect of each CGI by comparing the prediction accuracy of the model built using the set of financial ratios (FRs) only and the same model plus one CGI. They conclude that numerous CGIs were comparably helpful in conjunction with FRs. Wu (2007) reports that the model built using FRs (six ratios) and CGIs (ten variables) performed higher than the model only using FRs. Both studies suggest that CGIs are helpful for a fixed set of financial ratios; however, the generalizability of their conclusion is somewhat limited. Lin et al. (2010) propose an SVM-based prediction model that considers four financial ratios and six corporate governance indicators of fifty-four distressed Taiwanese firms from 2001 to 2005. Their study indicates that a model using corporate governance indicators performs more favorably than those that do not. However, their model could not handle a data set containing more than ten financial features because feature selection search in their model would be exhausted (Lin et al. 2011). Consequently, their conclusion can be considered preliminary only to a relatively small sample set. Huang and Liu (2006) propose an efficient feature selection method based on a genetic algorithm (GA) for a larger data set comprising forty-three financial ratios and three corporate governance indicators. Huang and Liu (2006) compare the performance of various feature selection algorithms when both financial ratios and corporate governance indicators are present, determining that the GA is a more favorable algorithm for feature selection than the gain ratio. However, Huang and Liu (2006) neither discuss the effectiveness of corporate governance indicators nor compare the prediction power among models with and without them.

In 2005, the U.S. Securities and Exchange Commission (SEC) issued Final Rule 33-8529, encouraging registrants to voluntarily file tagged financial statements on the EDGAR (Electronic Data Gathering and Retrieval) reporting system by using the eXtensible Business Reporting Language (XBRL) format. In XBRL, each piece of financial data is assigned a unique, predefined data tag. These data tags act as barcodes that identify the content and structure of the involved information (Hodge et al. 2004). The use of XBRL for regulatory bank reporting in the United States has achieved considerable success, according to the Federal Financial Institutions Examination Council (FFIEC). In a report on the benefits of XBRL, the FFIEC stated that its XBRL-based solution had achieved measurable benefits, including increases from 66 percent to 95 percent in data cleanliness, from 70 percent to 100 percent in accuracy, and from weeks to hours in timeliness and a 15 percent rise in the productivity of analysts. In 2004, the SEC (http://www.sec.gov/rules/final/2009/33-9002.pdf) began preliminary actions to phase in XBRL-created reports by public companies by using the EDGAR Web portal. The SEC has mandated a three-year phase-in report for all publicly traded companies to provide XBRL-tagged quarterly reports from 2011. The benefits and potential users of XBRL are still being explored and developed. The richness of the interactive data now allows users to investigate each account deeper and potentially retrieve information as far down as the transaction level (Cohen 2009; Gomaa et al. 2011).

The study of Premuroso and Bhattacharya (2008) indicates that early and voluntary filing of financial statements in the XBRL format usually signals superior corporate governance and operating performance vis-à-vis their nonvoluntarily filing peers. Several XBRL pilot projects are currently in process internationally, such as those spearheaded by the China Securities Regulatory Commission, Korea Stock Exchange, Tokyo Stock Exchange, and National Tax Agency of Japan. To date, members of the XBRL International include the following stock exchanges: Australia, China, Germany, Korea, London, New Zealand, and Tokyo. Their support for and involvement in XBRL adoption can be considered part of a global effort to reinforce corporate governance and information transparency by adopting XBRL for financial reporting. Although not yet an XBRL member, the government of Taiwan has endorsed several projects to promote XBRL financial reporting and convert the corporate data of listed companies into XBRL format (Hodge et al. 2004; Premuroso and Bhattacharya 2008). This study generated such financial features, including corporate governance indicators, by using XBRL technology and developing tools to search for predictors that might be effective in identifying distressed firms.

The objectives of this study are twofold: first, to study the feasibility of using financial statements in XBRL format for FCP; second, to investigate the role of corporate governance in FCP. To address the first research objective, we developed two tools, FeatureParser and XBRL FinancialAnalyzor (XFA), to code the financial ratios and corporate governance indicators into XBRL format to streamline the data analysis for subsequent prediction. For the second purpose, we designed a prediction model based on the GA and SVM. In our experiment, we examine whether the prediction accuracy can be significantly improved when corporate governance indicators are included in addition to traditional financial ratios. In comparison with previous studies (Huang and Liu 2006; Lin et al. 2010), the scope of our data set including the corporate governance information is substantially broader. The experimental design is more robust, causing the validity of the research results to be higher.

This study bears implications for both investors and government regulators. Investors will be able to obtain a clearer understanding of the roles of quantitative and qualitative features in predicting corporate financial crises. Government regulators might be able to detect and prevent potential financial crises at an early stage. In addition, our results reveal that XBRL can aid financial statement users by improving the transparency of the financial statement information of a firm and the information reporting choices of managers.

Research Background

Corporate Governance

Because financial factors are mostly retrospective, point-in-time measures and prediction models that examine only financial features have inherently been constrained; however, exactly how well these models perform out of sample (time, firm, or industry) is not clear. Financial scandals have increased in both frequency and size in recent years, requiring the specific interactions between different risk factors to be analyzed in detail.

According to relevant corporate governance literature, numerous listed companies in Taiwan still rely heavily on the support of their founding families to finance operations, in marked contrast to companies in industrialized countries. Current studies on firms with a concentrated ownership structure primarily use the divergence between control and ownership as a measure of agency conflict between majority and minority shareholders (Claessens et al. 2002). However, the divergence measure can be difficult for investors to calculate accurately, particularly when family-based controlling shareholders use pyramids and cross holdings to leverage control or to divert resources. Studies on companies with a concentrated ownership structure have indicated that intense agency conflicts and weak corporate governance are extremely likely to exist when the majority of directors and all of the supervisors belong to a controlling family. Therefore, the board structure of a firm can serve as a crucial indicator of whether a controlling family shareholder is committed to entrenching corporate governance.

Using a sample of 141 companies listed on the Taiwan Stock Exchange (TWSE; http://www.twse.com.tw/en/), Claessens et al. (2000) note that 34 percent were family controlled, with control defined as possessing a 20 percent shareholding. If the criterion for control is reduced to a 10 percent shareholding, then the percentage of family-controlled listed companies escalates to 47 percent. This percentage ultimately reached 67.5 percent because of the legal definition of insider shareholding. The extensive presence of family control in the listed companies of Taiwan renders corporate governance a particularly crucial concern in FCP.

Recently, corporate financial scandals in Taiwan have betrayed a common feature that is consistent with the conclusion of related studies: intense agency conflicts and weak corporate governance exist, particularly when board members are closely affiliated with the controlling family. In response to the extensive presence of concentrated ownership in corporate Taiwan, we apply shareholding ratios of different groups, ownership structures of board members or shareholders, and the number of times financial reports were restated within a year as designated corporate governance factors for our research. Moreover, we discover that several distressed Taiwanese firms, notably Rebar and Procomp Informatics, had changed their auditors frequently before filing bankruptcy. Thus, we also include the frequency of CPA change as one of our nonfinancial features.

Using XBRL for Data Reporting

XBRL is the product of XBRL International (www.xbrl.org), a nonprofit consortium of over 450 global financial service, technology, stock exchange, government, and accounting organizations. XBRL has become an international standard for addressing the challenging issues that businesses and governments face worldwide in current financial reporting and regulation.

A hard-copy financial report, such as a PDF or Excel file, cannot be fed automatically into a data analysis system. To transfer contextual information to other reports or systems, traditional methods for producing financial statements involve either rekeying data or producing a customized interface from one system to another that facilitates the proprietary interchange of electronic data. The introduction of XBRL tags enables reporting information content while capturing the relevant context, thereby facilitating the automation of business information processing by computer software (Pinsker and Li 2008). Computers can recognize the information in an XBRL document, select it, analyze it, store it, exchange it with other computers, and present it automatically to users in various ways.

Once data are received electronically in XBRL, information technology (IT) personnel can automate its handling and streamline the processes for collecting and reporting financial information (Pinsker and Li 2008). Companies can use XBRL technology to reduce costs by eliminating costly data collation and time-consuming reentry of information into their data processing systems. Furthermore, different types of reports that require using varying subsets of the data stored in the system can be produced with minimal effort when the data are gathered in XBRL format. Consumers of financial data, including investors, analysts, financial institutions, and regulators, can receive, find, compare, and analyze data rapidly and efficiently when the data are in XBRL format.

Timeliness is an additional benefit of using XBRL. XBRL improves the ability of firms to provide real-time data by eliminating the need to rekey data, thus improving the speed of data acquisition (Debreceny and Gray 2001). Furthermore, the software can immediately validate the data and highlight errors and gaps through the use of the Internet, financial information can be made available, not only at the end of a fiscal year or quarter but practically in real time (Debreceny and Gray 2001).

For regulation purposes, the adoption and use of XBRL are expected to help avoid the additional effort and complications associated with multiple reconciliations of local financial statements to the U.S. Generally Accepted Accounting Principles or the International Financial Reporting Standards (Premuroso and Bhattacharya 2008). XBRL allows regulators to promote the standardization and harmonization of international business reporting standards and assemble, validate, and review data more efficiently and usefully than they had previously (Premuroso and Bhattacharya 2008).

XBRL-coded tags allow individuals using software applications (e.g., search engines, parsers) to extract and simultaneously exhibit all identically coded information from financial statements and footnotes (Tay and Cao 2001). When XBRL is widely adopted, both humans and intelligent software agents can use highly accurate and reliable Web-based financial information (Debreceny and Gray 2001; Tay and Cao 2001). Hodge et al. (2004) suggest the feasibility of simultaneously extracting both financial and corporate governance indicators when the financial statements are presented in XBRL format. Therefore, this study focuses on processing financial data by using XBRL technology for FCP.

The Proposed Solution

This section presents the environmental framework and tools used to validate the feasibility of the proposed model. As shown in Figure 1, two tools, the FeatureParser and XBRL FinancialAnalyzer (XFA), were developed using VB.Net on Windows XP for processing the financial statements in XBRL format, called XBRL instances. We used FeatureParser to create both financial ratios and corporate governance indicators in XBRL format. The useful FR and CGI features were selected using XFA to build a prediction model based on GA-SVM.

We investigated whether incorporating corporate governance indicators with traditional financial ratios increases prediction accuracy. As illustrated in Figure 1, these data were used either as training data to construct the prediction model or as test data to validate the proposed model. The comparison was based on a training set with an equal proportion of distressed and nondistressed firms. The test data consisted of both distressed and nondistressed companies. The training and test sets were mutually exclusive, which is standard for a data mining experiment. The details of the data source and the functions of FeatureParser and XFA are presented further in this section.

Feature Extraction Using FeatureParser

Figure 2 presents the graphical user interface (GUI) of the FeatureParser. The XBRL-tagged reports enable decision makers to acquire necessary data and display them in their preferred format. The window on the left shows the financial ratios and CGIs that were defined by this parser. The middle window lists

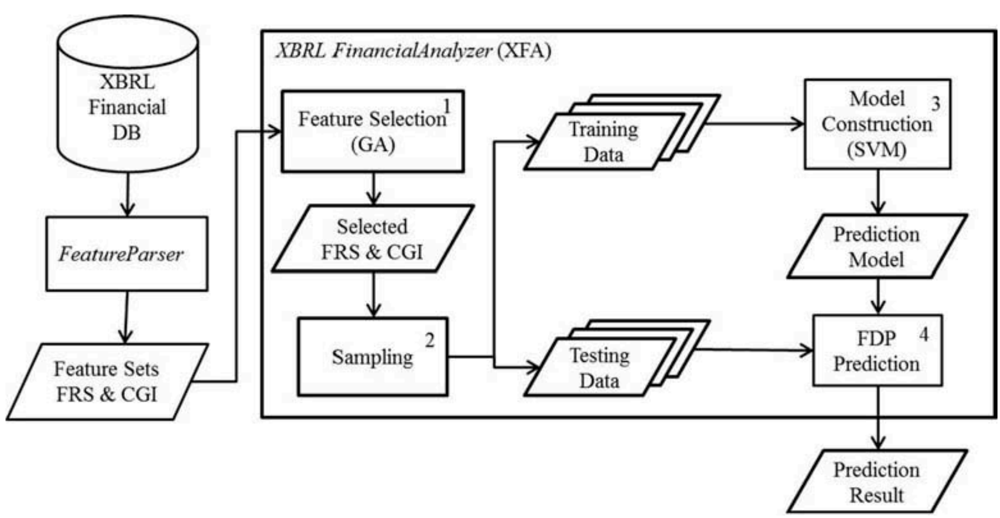

Figure 1. The framework of the prediction system based on the XBRL financial statements.

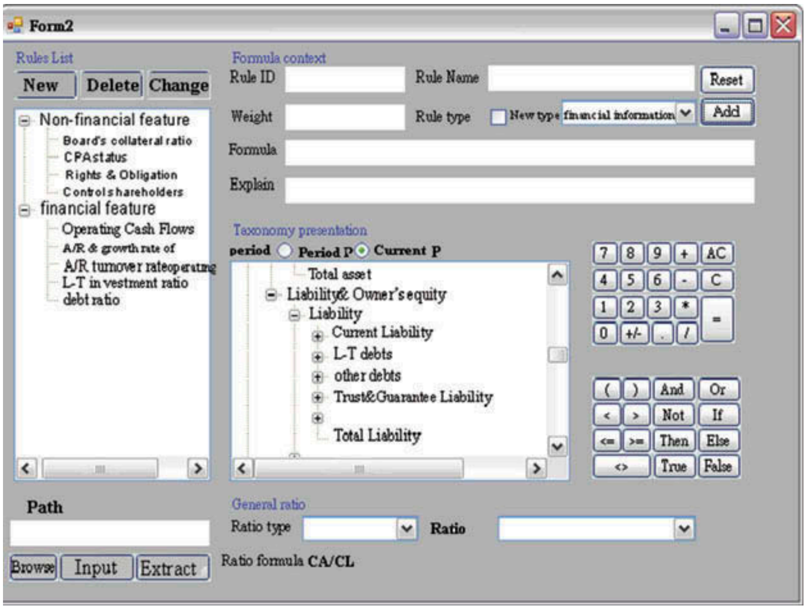

Figure 2. The GUI of the *FeatureParser.*

the account entries from a financial statement, such as "Total assets." A financial analyzer can use the tool boxes on the right to add, revise, and delete a ratio (or an indicator) shown in the left window. Therefore, a decision maker can identify the semantic information provided by the tag by using the taxonomy, regardless of whether the data are used in a spreadsheet or sophisticated statistical tools.

Feature Selection in XFA

In XFA, we implemented a GA-based feature selection algorithm to select the optimal feature subsets from the FRs and CGIs. We present the design details of this GA because feature selection profoundly affects the performance of XFA. GAs have been widely used to solve various nonlinear optimization problems since 1970 (Fogel 1994; Goldberg 1989; Srinivas and Patnaik 1994). A GA typically starts from a set of solutions, referred to as a population. A member belonging to this population is referred to as a chromosome. Mimicking the evolutionary processes in nature, each chromosome is evaluated by a fitness function within every generation to determine its chance of surviving to the next generation; a chromosome with a higher fitness value has a higher survival probability. A GA generally applies a set of genetic operations to reproduce a new population based on the current pool. The reproduction of the population continues from one generation to the next until a satisfactory solution is obtained. This approach is fundamentally different from traditional feature-selection methods, such as *t*-tests or stepwise regression methods, in which features are considered one at a time. Consequently, the multicollinearity issue is resolved once the GA is applied.

Designing an effective GA requires considering several issues including population size, genetic operators (selection, crossover, and mutation), and the stopping criteria. The size of a population influences both the performance and efficiency of the GA, which is typically set from 30 to 200 (Goldberg 1989; Grefenstette 1986; Srinivas and Patnaik 1994). The reproduction of the current population to the next generation starts with chromosome selection (Bäck 1996). The roulette wheel method and the tournament method are two standard approaches. The roulette wheel method selects a chromosome to survive with a probability proportional to the fitness value of this chromosome; the tournament method runs several "tournaments" to select winners based on the fitness values of the chromosomes entering the tournaments. It is suggested the roulette wheel method may sometimes fall into the premature convergence problem.

The chromosomes are randomly selected in pairs after the reproduction for crossover. The probability of being selected for crossover typically ranges from 0.5 to 1.0, and the commonly used crossover methods are single-point, two-point, and uniform crossover (Bäck 1996; Srinivas and Patnaik 1994). Following the crossover, the mutation operation determines whether a chromosome should be mutated in the next generation. The mutation operator produces small changes to the bit string by choosing a single bit at random then changing its value. The probability that a chromosome is mutated, or the mutation rate, ranges typically from 0.001 to 0.05. Commonly used mutation methods are point mutation, polynomial mutation, and uniform mutation. This reproduction process stops when a predetermined condition, called stopping criteria, is met. Stopping criteria shall be chosen carefully since they too have effects on the performance of the GA. The pseudo codes shown in Table 1 consist of the four main parts of the GA, which include the initialization (lines 1 to 2), the chromosome selection (lines 4 to 9), the crossover (lines 10 to 17), and finally the mutation (lines 18 to 22).

Table 1. The pseudo code of the genetic algorithm for the selection of the best feature set

Input:	**S Input feature set** **α size α of population** **β rate β of crossover**	**γ rate γ of mutation** **δ number δ of iterations**
Output:	**X* best solution**	

//Initialization
1. Randomly generate α feasible solutions (chromosomes), X_j, j =1~ α;
2. Save X_j's in the population *Pop*;

//Loop until the termination condition
3. *for* i =1 to δ do

//Chromosome selection
4. *for* j=1 to α do
5. Fitness value=average Accuracy based on the feature set of X_j;
6. *endfor*
7. Delete the bad solutions in *Pop*;
8. Copy the good solutions in *Pop* to replace the bad solutions;
9. Save them in the population *Pop*;

//Crossover
10. number of crossover $n=\alpha/2$;
11. *for* l =1 to n do
12. randomly select two solutions X_j and X_k from *Pop*;
13. *if* random <β
14. Generate $X_j{}'$ and $X_l{}'$ by one-point crossover to X_j *and* X_k;
15. Save $X_j{}'$ and $X_k{}'$ to *Pop*;
16. *endif*
17. *endfor*

//Mutation
18. *for* j =1 to n do
19. Select a solution X_j from *Pop*;
20. Mutate each bit of X_j under the rate γ to generate a new solution $X_j{}'$;
21. Update X_j with $X_j{}'$ in *Pop*;
22. *endfor*

//Updating
23. *endfor*

//Returning the best solution
24. *return* the best chromosome X^* in *Pop* as the recommended feature set;

The Model Construction in XFA

The support vector machine (SVM) is used to construct the prediction model in XFA. SVM was first developed by Boster et al. (1992) to provide better solutions than other classifiers such as artificial neural networks. SVM belongs to the type of maximal margin classifier in which the classification problem can be represented as an optimization process (Vapnik 1995). Support vectors are a subset of sample data used to define the boundary between two classes, or better known as the hyper plane. To find the optimal hyper plane, the input vector is mapped into a feature space, which is possible with a higher dimension. SVM applies the Lagrange multipliers in search of the optimal hyper plane.

Unlike most of the other classifiers that implement the empirical risk minimization principle, SVM seeks to minimize an upper bound of the generalization error rather than minimize the training error (Vapnik 1995). SVM can be generalized well even in high-dimensional spaces under small training sample conditions, indicating a learning ability independent of the feature space dimensionality. It is robust to outliers by reducing the effect of outliers using the margin parameter C to control the misclassification error. With the e-insensitive loss function, SVM can model nonlinear functional relationships difficult to be modeled by other techniques (Elish and Elish 2008). These characteristics make SVM a strong candidate in predicting financial crisis. We therefore implement XFA using SVM as the underlying classifier as FCP problem is perceived as a nonlinear problem.

The linear function and the radial basis function (RBF) are the two popular kernel functions suggested for SVM classifiers. As opposed to the linear kernel, the RBF kernel nonlinearly maps the samples into the high-dimensional space, which makes it feasible for nonlinear problems. One of the challenging problems using RBF kernel to build SVM model is the selection of parameter values for (C,σ) in order to ensure satisfactory prediction performance. The RBF model may yield poor performance if these parameters are not carefully chosen. Both kernels are supported in XFA using LIBSVM software (Chang and Lin 2011).

The benefits of taking advantage of the XBRL are obvious, and timeliness is most beneficial among all. Consumers of financial data, including investors, analysts, financial institutions, and regulators, can use the proposed tool, XFA, to streamline their analytical processes by collecting these XBRL financial statements from TWSE's website and then feeding them to XFA for further analysis without any rekeying. Other benefits include cost saving and data correctness by cutting out the time-consuming reentry of information to the data processing systems.

Experiment Design

A publicly listed firm is regarded to encounter financial crisis and turns into a distressed company when it is declared for full-value delivery, stock transaction suspension, reconstruction, and bankruptcy or goes out-of-market. According to the *Taiwan Economic Journal* (TEJ; http://www.tej.com.tw/twsite/) database, 240 distressed companies in Taiwan during the period 1999–2010 incorporated two additional criteria: (1) Sampled firms should have at least three years of complete public information before the financial crisis occurs; (2) There should be a sufficient number of comparable companies of similar size and in the same industry to serve as contrary samples. A total of 240 healthy companies were matched from the database by using randomly stratified sampling (Altman 1968). For a particular distressed firm, its financial statement from one year before the crisis occurred was used for analysis. The financial statement for the same year was used for the corresponding nondistressed firm.

Previous research on FCP has specified a considerable number of significant financial ratios of business crises (Altman 1968; Beaver 1966; Ohlson 1980). The XBRL database provided each firm with the following types of information: solvency, capital structures, profitability, turnover, and cash flow; ninety-five ratios in total were collected from the database. These ratios can be classified into one of two categories: (1) performance-related variables (current ratio, net profit margin, return on investment, and return on equity); or (2) structure-related variables (leverage ratio).

Studies on corporate boards of directors have generally been restricted to large firms in the United States, where investor protection is strong and ownership is dispersed and tends to treat board composition as being exogenous (Collier and Esteban 1999; Liang et al. 2013; Yeh and Woidtke 2005). Corporate governance is therefore seldom considered a contributing factor in corporate financial crises. However, studies focusing on emerging markets have indicated that corporate governance can be significant because ownership structures tend to be concentrated in countries outside the United States (Berkman et al. 2009; Cheung et al. 2013; Jian and Wong 2010; La Porta et al. 1999). This situation is particularly pronounced in Taiwan, as mentioned previously. Nonfinancial features have thus evolved with corporate governance issues. This study considers other corporate governance indicators including cash flow rights (Claessens et al. 2000; La Porta et al. 1999) and key persons retained (Albring et al. 2013; Dan 2010). Ninety-six CGI features were collected in the XBRL data set. The definition of each variable, regarding either FRs or CGIs, is listed in Oh (2012).

Feature selection and parameter settings play a crucial role in the performance of SVM models. The traditional binary coding scheme was used, in that the chromosome X is represented as $X = \{x_1, x_2, \ldots, x_n\}$, where $x_i = 1$ if the ratio r_i is selected, and $x_i = 0$ otherwise. The initial population was generated randomly, and the size was set at sixty, as suggested in Goldberg (1989), so that the convergence time and the population divergence are kept balanced. The average accuracy is used herein as the fitness value of each X under consideration. We apply the holdout method as the cross-validation method to obtain the accuracy of the training. The holdout number is set to be 100 according to our previous experience (Oh 2012), where the number is the minimum number with which we are able to obtain experimental results with satisfactory confidence level. For chromosome selection, the standard roulette wheel method is applied to retrieve half the size of the population; that is, thirty pairs of chromosomes. It follows by the one-point crossover to create new chromosomes with the possibility of 0.7. Then, uniform mutation is used with mutation rate of 0.005. Within the evolution process, every round is arranged with twenty generations. The algorithm halts when either it reaches convergence or all generations have completed. The parameters set in the GA are summarized in Table 2. The discussion of these settings is given in the previous section.

The ten-fold method partitions the data into two mutually exclusive subsets called a training data set and a test set. Training data are used for training and validation to select optimal parameters for the SVM and to prevent the over-fitting problem. When performing the cross-validation procedure for SVM, we have 10 percent of the training data used as a validation set. The SVM rests on the data generated from the year-end financial statements of the firms and is carried out to identify the most important predictors in bankruptcy classification. For each experiment, SVM is used to predict financial crisis for the sampled companies.

Table 2. Summary of the attributes set in the proposed GA

GA attribute	Value
Objective function	Average accuracy
Selection	Roulette wheel selection
Crossover method	Uniform crossover
Generation	20
Population size	60
Crossover rate (P_C)	0.7
Mutation rate (P_M)	0.005
Elite chromosome	2

The Analysis of Experiment Results

The purpose of this research is primarily to study the effectiveness of the incorporation of CGI with FRS in building a better prediction model. For the clarity of presentation, we denote M_i as the integrated model based on the proposed method that considers the feature sets: FRS and CGI. We construct three other models for benchmark purpose; M_c, M_f, and M_a. M_c and M_f are constructed in the same way as M_i except they include only CGI and FRS, respectively. The feature selection of the three models is based on the proposed GA as shown in Table 1.

Model M_a is based on MDA proposed by Altman (1968), which serves as an alternative to traditional ratio analysis for FCP. Model M_a inherits the five financial ratios (X_5, X_7, X_{17}, X_{28}, and X_{29}) as used in the original model (Altman 1968). Finally, the constants of the Z-score equation in M_a are recalculated according to our data set. The constants shown in the Z-score equation are obtained using the linear regression toolbox provided by *Matlab*. Table 3 and Table 4 summarize all the features used in building the models in this study, and Table 5 lists the FRS and CGI used in each of the four models. Z-score = $0.1591X_5+0.1291X_7+0.0764X_{17}+0.0581X_{28}+0.0475X_{29}$.

Figure 3 depicts the predictive accuracies of the four models. The results reveal that the proposed model M_i yields an average correct classification rate of 85.48 percent, which is significantly superior to models M_c (72.47 percent), M_f (83.31 percent) and M_a (76.82 percent). Empirical results show that the proposed model with both financial ratios and corporate governance indicators can serve as a promising alternative for existing financial crisis prediction models.

There are different penalty costs for each type of error. A Type I error occurs when an observation classified as “nondefault” actually does default. On the contrary, a Type II error happens when an observation classified as “default” actually does not default. In financial crisis prediction, Type I error is more important and serious since an erroneous prediction may lead to risky investment in a firm

Table 3. Summary of selected financial ratios by category

Variable	Definition	Variable	Definition
Solvency		**Profitability**	
X_1	Cost of interest-bearing debt	X_{20}	Operating income/net sale
X_2	Cash/total assets	X_{21}	Pretax income per share
X_3	Quick asset/current liabilities	X_{22}	Retained earnings/total assets
X_4	Long-term liability/current assets	X_{23}	Net income—exclude disposal gain or loss/net sale
X_5	Inventory/current liability	X_{24}	EPS-net income
X_6	Cash/current liability	X_{25}	Net income/total assets
X_7	Current ratio	X_{26}	(Liquidity-current assets)/total assets
X_8	Liabilities/total assets	X_{27}	Retained earnings/total assets
X_9	Interest-bearing debt/equity	X_{28}	Earnings before interest and tax/total assets
X_{10}	Pretax income/capital	**Cash flow ratios**	
X_{11}	Equity/liability	X_{29}	Operations cash flow/assets
X_{12}	Cash reinvestment ratio	X_{30}	Cash flow/total assets
X_{13}	Interest expense/total revenue	X_{31}	Cash flow/liability
X_{14}	Equity/liability	X_{32}	Cash flow/equity
Capital structure ratios		**Growth**	
X_{15}	Degree of financial leverage (DFL)	X_{33}	Realized gross profit growth rate
X_{16}	Interest coverage ratio (interest expense to EBIT)	X_{34}	Net income—exclude disposal gain or loss growth
X_{17}	Equity/total assets	**Others**	
Turnover ratios		X_{35}	Cash flow per share
X_{18}	Equity turnover	X_{36}	Operation income per employee
X_{19}	Sales/total assets		

Table 4. Summary of corporate governance indicators selected by all models

Variable	Definition	Variable	Definition
X_{37}	Shareholding ratio of outside unlisted company	X_{48}	Board seat rights/shareholding ratio control by ultimate controller
X_{38}	Shareholding ratio of foreign director and supervisor	X_{49}	Seats of directors serve as managers/number of directors
X_{39}	Seats of supervisor the ultimate controller control/Seats of ultimate controller served as individual director	X_{50}	Cash flow rights of ultimate controller/ shareholding ratio control by ultimate controller
X_{40}	Shareholding ratio control by ultimate controller/board seat rights	X_{51}	Shareholding ratio control by ultimate controller/ cash flow rights of ultimate controller
X_{41}	Number of times CPA was switched in the last three years	X_{52}	Seats of directors serve as managers/number of supervisors
X_{42}	Shareholding ratio of company manager and group manager	X_{53}	Amount of investments in other enterprises/ stockholder's equity
X_{43}	Number of times the financial report restate in a year	X_{54}	Number of times financial forecast published in a year
X_{44}	Shareholding ratio of ultimate controller through conglomerate	X_{55}	Deviation between voting rights and cash flow rights
X_{45}	Shareholding ratio control by ultimate controller	X_{56}	Turnover of spokesman within three years
X_{46}	Cash flow rights of ultimate controller	X_{57}	Turnover of internal audit within three years
X_{47}	Turnover of CEO within three years		

Table 5. Financial features selected in various prediction models

	FRS	CGI
M_i	$[X_1][X_2][X_3][X_4][X_5][X_6][X_7][X_8][X_9]$ $[X_{10}][X_{11}][X_{15}][X_{16}][X_{17}][X_{18}][X_{20}]$ $[X_{21}][X_{22}][X_{29}][X_{30}][X_{33}][X_{35}][X_{36}]$	$[X_{37}][X_{38}][X_{39}][X_{40}][X_{41}][X_{42}][X_{43}]$
M_f	$[X_7][X_8][X_9][X_{10}][X_{11}][X_{12}][X_{13}]$ $[X_{17}][X_{22}][X_{23}][X_{24}][X_{25}][X_{31}][X_{32}]$ $[X_{34}][X_{36}]$	
M_c		$[X_{41}][X_{42}][X_{43}][X_{44}][X_{45}][X_{46}][X_{47}]$ $[X_{48}][X_{49}][X_{50}][X_{51}][X_{52}][X_{53}][X_{54}]$ $[X_{55}][X_{56}]$ $[X_{57}]$
M_a	$[X_{14}][X_{19}][X_{26}][X_{27}][X_{28}]$	

later dragged into financial crisis. The worst scenario of a Type II error, however, is only to prevent the general public from investing in a healthy firm. The investors have plenty of other good choices in the market.

Figure 4 illustrates the Type I and Type II error rates in all four models. The experimental results reveal that the proposed model M_i (13.49 percent) is superior to model M_c (17.14 percent), model M_f (17.32 percent), and model M_a (18.73 percent) in terms of the Type I error rate. This implies that investors who rely on the proposed model M_i will incur considerably fewer losses caused by investing in distressed companies erroneously predicted to be healthy. Table 6 lists the results of Type II errors, where model M_i reports a significantly lower error rate of 15.56 percent compared with M_c (37.91 percent), M_f (16.06 percent), and M_a (27.64 percent).

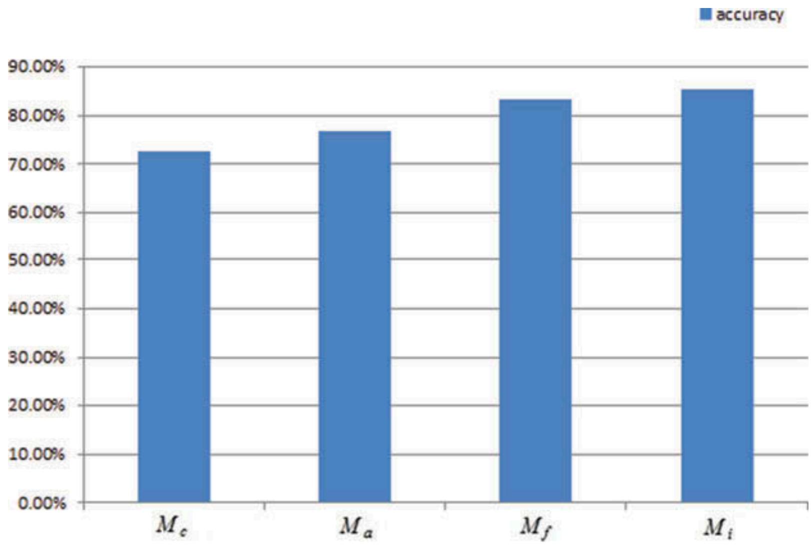

Figure 3. Accuracy rates of all models.

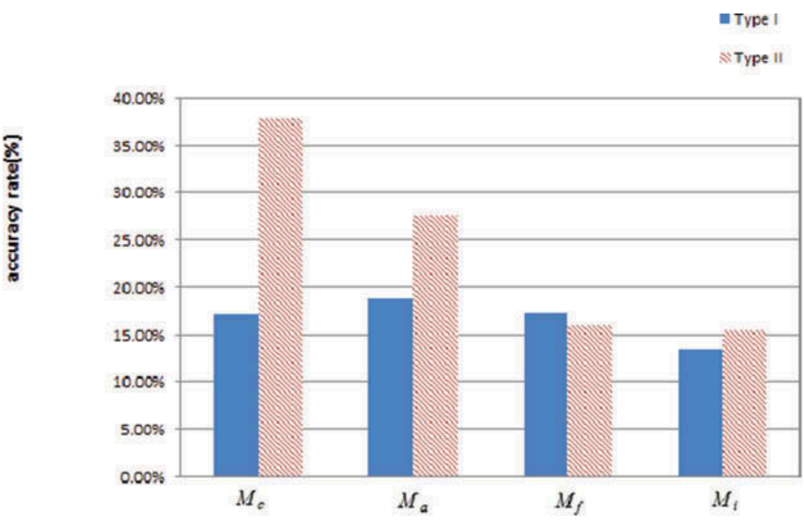

Figure 4. Type I and Type II error rates.

Table 6. Performance summary of various models

	M_i	M_c	M_f	M_a
Training Error				
Type I (percent)	12.72	15.49	16.83	17.81
Type II (percent)	14.38	35.35	14.66	27.63
Accuracy (percent)	86.45	74.58	84.25	77.28
Testing Error				
Type I (percent)	13.49	17.14	17.32	18.73
Type II (percent)	15.56	37.91	16.06	27.64
Accuracy (percent)	85.48	72.47	83.31	76.82

Table 7. Wilcoxon signed rank test for pairwise comparison of the average accuracy, Type I, and Type II errors

M_i	M_c	M_f	M_a
Type I	0.029044*	$8.02 \times 10^{-3**}$	$2.25 \times 10^{-3**}$
Type II	$1.07 \times 10^{-9**}$	0.556281	$1.52 \times 10^{-6**}$
Accuracy	$2.11 \times 10^{-9**}$	0.04544*	$7.36 \times 10^{-9**}$

*Significant at the level of 5 percent; **significant at the level of 1%.

We conducted a Wilcoxon signed-rank test to assess the significance of performance difference between the prediction results of distinct models (Demšar 2006). The Wilcoxon signed-rank test is a nonparametric alternative to the paired *t*-test that ranks the differences in performance of two classifiers for each data set, ignoring the signs and comparing the ranks of the positive and negative differences. As shown in Table 7, the proposed model M_i is significantly superior to other models at the levels of 1 percent or 5 percent in Type I, Type II, and average accuracy, except for the Type II error of M_f. As discussed, the significance of a Type II error is lower than that of a Type I error in accuracy.

In summary, the proposed model, encompassing both financial ratios and corporate governance indicators, enables predicting corporate financial crises more accurately than can a model based exclusively on either financial ratios or corporate governance indicators.

Conclusion

This study implements XBRL by using a GA-SVM model to generate financial statements, demonstrating that XBRL affords users the ability to retrieve data from financial statements without having to locate and retrieve data from the statements manually. Information in the notes of financial statements is more accessible when tagged according to a standard XBRL taxonomy. This study demonstrates that XBRL can facilitate using intelligent agents to gather and process financial information in Taiwan.

In addition to financial features, the features of corporate governance are examined to clarify whether they can be applied. The empirical results reveal that the proposed GA-SVM model is a promising hybrid SVM model for predicting financial crises, regarding both predictive accuracy and generalization ability. This study explores the role of corporate governance in FCP. The proposed SVM model for assessing corporate financial crises demonstrates a significantly improved accuracy compared with that of existing business failure classification models. The test results yield a high overall accuracy rate of 85.48 percent, superior to that of prediction models developed using only financial ratios or corporate government indicators. Because the extensive presence of concentrated ownership in the public listed companies in Taiwan has rendered corporate governance crucial for FCP, we incorporate corporate governance indicators such as the shareholding ratio of different groups, ownership structures of board members or shareholders, and the number of times the financial report restated input variables based on using the SVM method within a year. Including these nonfinancial features appears to enhance the performance of FCP. We believe that these nonfinancial features related to corporate governance may merit consideration in future studies, particularly those focusing on emerging markets populated with firms characterized by concentrated ownership.

This study has limitations that call for further research. Our models were inevitably affected by several factors. Moreover, the prediction accuracy might be improved further in the future by pairing sampled companies according to industry or by extending the survey period. In reaction to the recent outbreak of corporate financial scandals in Taiwan and overseas, we have paid special attention to the role of ownership structure and corporate governance in financial crisis prediction. Selection of nonfinancial features is therefore based on attributes related to corporate governance. This exclusive focus on corporate governance–related factors has prevented us from considering in our present study other potentially influential nonfinancial features, such as market share, management style, and industry prospect. Further research may be conducted to explore such potential nonfinancial indicators since the inclusion of nonfinancial features in financial crisis prediction has been proved to enhance prediction performance in our study. For XBRL financial reporting purposes, a complex set of nonfinancial data to the standard XBRL taxonomy will be needed.

References

Ahn, B.; S. Cho; and C. Kim. 2000. "The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction." *Expert Systems with Applications* 18, no. 2: 65–74.

Albring, S.; D. Robinson; and M. Robinson. 2013. "Audit Committee Financial Expertise, Corporate Governance, and the Voluntary Switch from Auditor-Provided to Non-Auditor-Provided Tax Services." *Advances in Accounting* 29, no. 1: 1–11.

Altman, E.I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance* 23, no. 4: 589–609.

Bäck, T. 1996. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. New York: Oxford University Press.

Beaver, W.H. 1966. "Financial Ratios as Predictors of Failure." *Journal of Accounting Research* 4, supp. 4: 71–111.

Berkman, H.; R.A. Cole; and L.J. Fu. 2009. "Expropriation Through Loan Guarantees to Related Parties: Evidence from China." *Journal of Banking and Finance* 33, no. 1: 141–156.

Boster, B.E.; I.M. Guyon; and V.N. Vapnik. 1992. "A Training Algorithm for Optimal Margin Classifiers." In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152. New York: ACM Press.

Chandra, D.K.; V. Ravi; and I. Bose. 2009. "Failure Prediction of Dotcom Companies Using Hybrid Intelligent Techniques." *Expert Systems with Applications* 36, no. 3: 4830–4837.

Chang, C.-C., and C.-J. Lin. 2011. "LIBSVM: A Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology (TIST)*, no. 3: 27.

Chen, L.-H., and H.-D. Hsiao. 2008. "Feature Selection to Diagnose a Business Crisis by Using a Real GA-Based Support Vector Machine: An Empirical Study." *Expert Systems with Applications* 35, no. 3: 1145–1155.

Cheung, Y.-L.; C.-W. Chung; W. Tan; and W. Wang. 2013. "Connected Board of Directors: A Blessing or a Curse?" *Journal of Banking and Finance* 37, no. 8: 3227–3242.

Claessens, S.; S. Djankov; and L.C. Xu. 2000. "Corporate Performance in the East Asian Financial Crisis." *World Bank Research Observer* 15, no. 1: 23–46.

Claessens, S.; S. Djankov; J.P. Fan; and L.H. Lang. 2002. "Disentangling the Incentive and Entrenchment Effects of Large Shareholdings." *Journal of Finance* 57, no. 6: 2741–2771.

Cohen, E.E. 2009. "XBRL's Global Ledger Framework: Exploring the Standardised Missing Link to ERP Integration." *International Journal of Disclosure and Governance* 6, no. 3: 188–206.

Collier, J., and R. Esteban. 1999. "Governance in the Participative Organisation: Freedom, Creativity and Ethics." *Journal of Business Ethics* 21, no. 2–3: 173–188.

Dan, L. 2010. "Does Auditor Tenure Affect Accounting Conservatism? Further Evidence." *Journal of Accounting and Public Policy* 29, no. 3: 226–241.

Debreceny, R., and G.L. Gray. 2001. "The Production and Use of Semantically Rich Accounting Reports on the Internet: XML and XBRL." *International Journal of Accounting Information Systems* 2, no. 1: 47–74.

Demšar, J. 2006. "Statistical Comparisons of Classifiers Over Multiple Data Sets." *Journal of Machine Learning Research* 7: 1–30.

Ding, Y.; X. Song; and Y. Zen. 2008. "Forecasting Financial Condition of Chinese Listed Companies Based on Support Vector Machine." *Expert Systems with Applications* 34, no. 4: 3081–3089.

Elish, K.O., and M.O. Elish. 2008. "Predicting Defect-Prone Software Modules Using Support Vector Machines." *Journal of Systems and Software* 81, no. 5: 649–660.

Fogel, D.B. 1994. "An Introduction to Simulated Evolutionary Optimization." *IEEE Transactions on Neural Networks* 5, no. 1: 3–14.

Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston: Addison-Wesley Professional.

Gomaa, M.I.; A. Markelevich; and L. Shaw. 2011. "Introducing XBRL Through a Financial Statement Analysis Project." *Journal of Accounting Education* 29, no. 2: 153–173.

Grefenstette, J.J. 1986. "Optimization of Control Parameters for Genetic Algorithms." *IEEE Transactions on Systems, Man and Cybernetics* 16, no. 1: 122–128.

Hodge, F.D.; J.J. Kennedy; and L.A. Maines. 2004. "Does Search-Facilitating Technology Improve the Transparency of Financial Reporting?" *Accounting Review* 79, no. 3: 687–703.

Hu, Y.-C. 2009. "Bankruptcy Prediction Using ELECTRE-Based Single-Layer Perceptron." *Neurocomputing* 72, no. 13: 3150–3157.

Hua, Z.; Y. Wang; X. Xu; and B. Zhang; L. Liang. 2007. "Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and Logistic Regression." *Expert Systems with Applications* 33, no. 2: 434–440.

Huang, P.-W., and C.-L. Liu. 2006. "Exploiting Corporate Governance and Common Size Analysis for Financial Distress Detecting Models." In *Proceedings of the Joint Conference on Information Science 2006*, pp. 6–11. Taipei: Atlantis Press.

Hui, T.-K. 2005. "Day-of-the-Week Effects in U.S. and Asia–Pacific Stock Markets During the Asian Financial Crisis: A Non-Parametric Approach." *Omega* 33, no. 3: 277–282.

Jeong, P.W.; J.H. Min; and M.L.S. Kim. 2012. "A Tuning Method for the Architecture of Neural Network Models Incorporating GAM and GA as Applied to Bankruptcy Prediction." *Expert Systems with Applications* 39: 3650–3658.

Jian, M., and T.J. Wong. 2010. "Propping Through Related Party Transactions." *Review of Accounting Studies* 15, no. 1: 70–105.

La Porta, R.; F. Lopez-de-Silanes; and A. Shleifer. 1999. "Corporate Ownership Around the World." *Journal of Finance* 54, no. 2: 471–517.

Lee, K.C.; I. Han; and Y. Kwon. 1996. "Hybrid Neural Network Models for Bankruptcy Predictions." *Decision Support Systems* 18, no. 1: 63–72.

Lee, T.S., and Y.H. Yeh. 2004. "Corporate Governance and Financial Distress: Evidence from Taiwan." *Corporate Governance: An International Review* 12, no. 3: 378–388.

Lee, Y.-C. 2007. "Application of Support Vector Machines to Corporate Credit Rating Prediction." *Expert Systems with Applications* 33, no. 1: 67–74.

Li, H., and J. Sun. 2009. "Predicting Business Failure Using Multiple Case-Based Reasoning Combined with Support Vector Machine." *Expert Systems with Applications* 36, no. 6: 10085–10096.

Liang, Q.; P. Xu; and P. Jiraporn. 2013. "Board Characteristics and Chinese Bank Performance." *Journal of Banking and Finance* 37, no. 8: 2953–2968.

Lin, F.; D. Liang; and W.-S. Chu. 2010. "The Role of Non-financial Features Related to Corporate Governance in Business Crisis Prediction." *Journal of Marine Science and Technology* 18, no. 4: 504–513.

Lin, F.; D. Liang; and E. Chen. 2011. "Financial Ratio Selection for Business Crisis Prediction." *Expert Systems with Applications* 38, no. 12: 15094–15102.

Min, J.H., and Y.-C. Lee. 2005. "Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters." *Expert Systems with Applications* 28, no. 4: 603–614.

Min, S.-H.; J. Lee; and I. Han. 2006. "Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction." *Expert Systems with Applications* 31, no. 3: 652–660.

Oh, G.W. 2012. Using Genetic Algorithm for Feature Selection in Financial Distress Problem. Master Thesis, National Central University, Taiwan.

Ohlson, J.A. 1980. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research* 18, no. 1: 109–131.

Ozkan-Gunay, E.N., and M. Ozkan. 2007. "Prediction of Bank Failures in Emerging Financial Markets: An ANN Approach." *Journal of Risk Finance* 8, no. 5: 465–80.

Pinsker, R., and S. Li. 2008. "Costs and Benefits of XBRL Adoption: Early Evidence." *Communications of the ACM* 51, no. 3: 47–50.

Premuroso, R.F., and S. Bhattacharya. 2008. "Do Early and Voluntary Filers of Financial Information in XBRL Format Signal Superior Corporate Governance and Operating Performance?" *International Journal of Accounting Information Systems* 9, no. 1: 1–20.

Shin, K.-S.; T.S. Lee; and H.-j. Kim. 2005. "An Application of Support Vector Machines in Bankruptcy Prediction Model." *Expert Systems with Applications* 28, no. 1: 127–135.

Srinivas, M., and L.M. Patnaik. 1994. "Genetic Algorithms: A Survey." *Computer* 27, no. 6: 17–26.

Tam, K.Y. and M.Y. Kiang. 1992. "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions." *Management Science* 38, no. 7: 926–947.

Tay, F.E.H., and L.J. Cao. 2001. "Improved Financial Time Series Forecasting by Combining Support Vector Machines with Self-Organizing Feature Map." *Intelligent Data Analysis* 5, no. 4: 339–354.

Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.

Wu, J.-L. 2007. "Do Corporate Governance Factors Matter for Financial Distress Prediction of Firms? Evidence from Taiwan." Master Thesis, University of Nottingham, England.

Yeh, Y.-H., and T. Woidtke. 2005. "Commitment or Entrenchment?: Controlling Shareholders and Board Composition." *Journal of Banking and Finance* 29, no. 7: 1857–1885.