



Credit scoring system for small business loans

Ray Tsaih^{a,*}, Yu-Jane Liu^b, Wenching Liu^a, Yu-Ling Lien^a

^aDepartment of Management Information Systems, National Chengchi University, 64, Chih-nan Road, Sec. 2, Wenshan, Taipei 11623, Taiwan, ROC

^bDepartment of Finance, National Chengchi University, Taiwan, ROC

Received 1 January 2003; accepted 1 April 2003

Available online 28 June 2003

Abstract

A requirement of a credit scoring decision support system for small business loans is that the embedded scoring model can be easily altered in accord with the change of business environment. To satisfy such a requirement, this study proposes an N-tier architecture integrated with the idea of Model-View-Controller. With this design, the system engineers can avoid frequently investing considerable time and effort in communicating with the model managers for finalizing the scoring models, and model managers can easily alter the embedded scoring models later at any time.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Credit scoring system; Small business loans; N-tiers; XML

1. Introduction

Many banking institutions have stepped up their effort in developing relationships with small businesses. However, this lending involves high potential risks due to the information asymmetry and time-vary essentials of a small business. Unlike large firms that are easy to raise funds from public debt markets, small firms rely on lending from commercial banks. On the other hand, lending to small business is beneficial to commercial banks because the margins on small business lending are higher than on many other bank products [11]. But it is difficult for banks to obtain detailed information from small firms since the financial reports of small firms are mainly for tax purposes

[1,5]. For managing this asymmetric information problem, banks tend to gather more information through the credit records of small business from the credit information center. In addition, small firms are more subject to economic cycles, and they might be seriously affected by an economic downturn. The credit risk of small business alters because of the change of business environment. Hence, important features for small business lending are to reduce information asymmetry and to avoid the credit risk from lending to small firms.

Accordingly, it is better for a banking institute to accumulate timely information to create a small business credit-scoring model for identifying desirable small business borrowers. Furthermore, the credit-scoring model should allow for being altered easily and timely when it is necessary in accord with incorporating the most recent credit record and rapid change of business environment.

* Corresponding author.

E-mail address: tsaih@mis.nccu.edu.tw (R. Tsaih).

In literature, many studies focus on large business credit scoring. For example, Katz et al. [8] use discriminatory analysis. But discriminatory analysis fails to capture the sensitivity of small business environment. Ou and Penman [10] adopt Probit model to predict the bankruptcy risk, and Tam and Kiang [12] use a neural network to evaluate the operating performance of credit customers, respectively. Unfortunately, credit-scoring models developed for large business loans are also not suitable for small business loans because, as mentioned above, small business loans have associated risks involved with information asymmetry. Till now, only studies like Refs. [3,4] center on the role of discrimination in credit markets for small business.¹ To our best knowledge, there are barely few studies exploring the credit scoring for small business.

It is helpful for a financial institution to have a credit scoring system where small business credit scoring models are embedded. This system can be used as a decision support system for managing small business loan applications. Current small business loan applications are managed mainly via man, in a less efficient, less effective, and subjective manner. A small business scoring system will benefit banks and small business by supporting the examination and evaluation of small business loans, and rendering the process objectively and well time controlling. In accordance with the evaluation score, credit officers then decide whether it is worth giving loan to the small business applicants. By improving the efficiency and managing its essential risk through decision support system, small business loans can become a high potential business for commercial banks. The problem is, as far as we know, no academic study investigates the credit scoring decision support system for small business loans.

There are more advantages with a credit scoring system for small business loans. For instance, a small business scoring system can eventually make an on-line application of small business loan possible.

¹ Cavalluzzo [3] examines the relationship between market structure and discrimination. Cavalluzzo et al. [4] analyze credit applications, loan denials, and interest rates paid across gender, race, and ethnicity of small business owners and demographic groups to better understand the sources of observed differentials in the small business credit market.

Moreover, based upon the updated information of current business environment, such small business credit scoring system can automatically review in batch a large amount of client samples of bank. With the reviewing results, more proper scoring models can be re-built in accord with the change of business environment.

A key issue of such a small business credit scoring system is how to provide a mechanism to render model managers² to easily alter the embedded scoring models correspondingly with the change of business environment. This study focuses on such a credit scoring system.

Our proposal is as follows. At the stage of system design, the tasks of model-developing, model-installing, and loan-evaluating are processed in parallel and as independently as possible. Scoring models, the final product of the model-developing task, are the shared components of the model-installing and loan-evaluating subsystem. After reasoning out all scoring models, model managers define all scoring models via the graphic user interfaces (GUIs) provided from the model-installing subsystem. The model-installing subsystem then implements all scoring models via transforming them into XML files. Scoring models are transformed into a set of defined variables and put into the database back-end. Then, account officers or credit officers can use the GUIs provided from the loan-evaluating subsystem to create new loan applications and evaluate them. Within the evaluation process, the loan-evaluating subsystem parses associated XML files. At the end of evaluation, the corresponding evaluation score is stored and can be retrieved for credit officers whenever needed.

With such a design, the system engineers can avoid considerable time and effort in frequently communicating with the model managers for finalizing the scoring models, and model managers can

² There are two groups of end-users who manage two kinds of tasks (loan management and scoring model management) via small business credit scoring system: (1) account officers and credit officers who manage small business loans, including their creating, evaluating, modifying, querying, and crediting. (2) Model managers who format and define the scoring models, regularly monitor them, and modify them if necessary. In principle, account officers or credit officers will not be model managers. Financial and economical professionals are model managers who set up reasonable and suitable rules that consist of scoring models.

easily alter the embedded scoring models later at any time via the GUIs provided from the model-installing subsystem.

Here, we take the following advantage of XML technology in creating and maintaining the scoring models. First is the simplicity of XML programming. In order to develop an XML-based application, the primary things are to make a Data Type Description (DTD) or schema for the XML document and to have a parsing engine. DTD provides a set of metadata for properly recording the scoring models and correctly interpreting the XML representation. As the XML document is parsed, the data in the document become available to other application programs.

Second, although simple, XML is powerful enough to express complex semi-structured data. As the business environment changes, scoring models may need to be altered properly. The XML syntax is rather suitable for recording these time-varying scoring models. As mentioned in Ref. [7], XML technology can be used to encode business rules. Also, there is a standard of using XML as rule representative language: the Rule Markup Language (RuleML), which is designed to cover the entire rule spectrum from derivation rules to transformation rules to reaction rules.³ RuleML can thus specify queries and inferences in web ontologies, mappings between web ontologies, and dynamic web behaviors of workflows, services, and agents.

In the following section, we shall first describe the procedure of developing the scoring model. The conclusion is made from three financial professionals. Their experience and comments on the information asymmetry and time-vary essentials associated with a scoring model will be shown. Then, in Section 3, we introduce the design of our proposed credit scoring system. In the last section, we summarize our study and present some future work.

2. Developing scoring models

To reduce the asymmetric information problem of small business loans, we use information on bank-borrow relationship of small firms and personal trade

history of small firms' principal owners from the credit information center, in addition to the fundamental characteristics of small companies. Moreover, industry factors, including business cycles and macroeconomic factors, are considered in our model to mitigate the sensitivity of economic trend in small business lending.

We use the Probit regression⁴ to develop a credit-scoring model. The dependent variable y , which estimates the credit risk and the probability of default risk, equals one when the small firm is defaulted after the principal owner gets the loan; zero, when the small firm is not defaulted. y is modeled as a linear function of explanatory variables and an error term v . Those explanatory variables for the borrower's credit history are denoted x , k , m , z . In other words,

$$y = a + bx + ck + dm + ez + v. \quad (1)$$

where a , b , c , d , and e are corresponding constant vectors.

The vector x contains the firm identification information, including name, capitalization, age (number of years since the firm was founded), import revenues for the recent 3 years, export revenues for the recent 3 years, profitability of firms, and financial difficulty/merge (a dummy variable; the value equals one if the firm involves in financial difficulty/merge, zero otherwise).

The vector k holds the personal credit history of the principal owner, consisting of credit card accounts and personal loans (number of accounts, outstanding balance, payment history, and default history). k also indicates if the principal owner is on the board of other firm's board of directors (a dummy variable; the value equals one if the principal owner is on the board of other firm's board of directors, zero otherwise).

The vector m comprises variables related to the business credit history of the firm, including number of banks from which a firm borrows, number of relationship banks with high reputation, the frequencies of inquiries on the credit file of the small firm and the principal owner (recent 3 months, recent 6 months and recent 3 years), short-term and long-term loans, export loans, collateralized loans, guarantees, depos-

³ Details of RuleML is available at <http://www.dfki.uni-kl.de/ruleml/#Initiative>.

⁴ For details of Probit regression, please refer to Finney [6].

its, reliance on bank loan, and defaulted history of the payments from all of the relation banks.

The vector z contains industry effects. We use the sector classification as the dummies of industry, and changes of industry indices as the growth potential and volatility of industry indices as the sensitivity of economic cycle. z also involves macroeconomic leading indicators to evaluate the economic health at the timing of lending to small business. The econometric analysis assumes that v is an independently distributed normal random variable with zero mean.

Our samples cover 41,000 small firms⁵, including 6000 defaulted firms and 35,000 non-defaulted firms. We use the data from 1996 to 1999 as the estimation period, and data from the period of 2000 as the out of sample testing period. We rely on two different data sources to construct the samples. The primary data source is the Joint Credit information Center in Taiwan, available for the years 1996–2000. The data include a comprehensive inventory of the firm code, firm characteristics, financial statements, four-digit industry code, and data on bank relationship. Combining with the up-to-date information on the personal credit history of the principal owner and the business credit history of the firm, the dataset allows us to consider information asymmetry and time-vary characteristics of small business loans. We also use data from the Taiwan Economic Journal (TEJ) for the years 1996–1999, offering details on the industry classifications, industry indices, and macroeconomic leading indicators.

Based upon the data mentioned above, repetitively, Probit regressions have been done to choose the best fitness model with respect to the testing period. In the scoring model we selected, the Probit analysis correctly classifies and predicts 80% of the defaulted cases. The explanatory variables of the selected model include age, financial difficulty/merge, personal loans, whether the principal owner is on the board of other firm's board of directors, defaulted history of credit cards, the frequencies of inquiries on the credit file of a company and the principal owner, short-term and long-term loans, guarantees, industry sector classifications, changes in industry indices, and macroeconomic leading indicators.

⁵ Here, a firm is classified as a small firm if its capitalization is less than 200 million (NT\$).

3. Small business credit scoring system

A typical component-based N-tiers architecture integrated with the idea of Model-View-Controller [2,9] is adopted here. The thin client layer provides GUIs implemented using dynamically generating HTML. The user interface runs within popular web browsers embedding a Java virtual machine. The middle tiers include the web server, Management Application Server (MAS), Loan Processing Subsystem (LPS), and Model Installing Subsystem (MIS). A database back-end, containing a table of Defined Variables (TDV), an XML repository, and a Database, forms the final layer in this architecture. Fig. 1 depicts this architecture.

MIS is designed to manage the task of defining scoring models, whereas LPS to manage the task of evaluation. MIS contains Model Defining Application Server (MDAS) and Model Recording Module (MRM). LPS contains Case Processing Application Server (CPAS) and Evaluation Module (EM) in which the XML parser and model engine sub-modules are available. The MAS provides interfaces for system manager to manage TDV, Database and XML repository, and other management tasks.

Each scoring model needs to be decomposed into a set of defined variables. For instance, for the scoring model of Eq. (1), the defined variable ES corresponds to the evaluation score y for each loan application. The value of ES is calculated with Eq. (1), accompanying with all corresponding values of explanatory variables, EV_i . EV_i is also treated as a defined variable, and its value is also calculated with a simple rule, for example, like Eq. (1), accompanying with all corresponding values of EV_i explanatory variables, EV_{ij} . In sum, each (dependent or explanatory) variable is regarded as a defined variable whose value conforms to a raw data. Or, as mentioned above, each (dependent or explanatory) variable is treated as a defined variable that associates with a simple rule according to a set of its explanatory variables.

Each defined variable stands for a field of a table in Database, and its corresponding value of a loan application is saved back in Database. The reason of using database technology is the requirement of persistency property of all data in a banking institute.

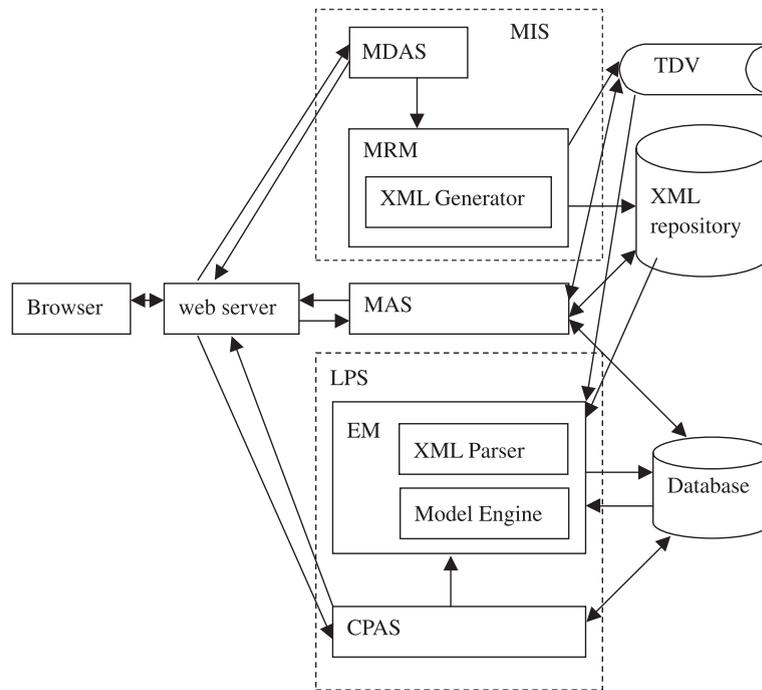


Fig. 1. High level architecture of Small Business Credit Scoring System.

A rule is simple if its associated computing action can be implemented with a specific program going together with a database management system.⁶ Common simple rules include the following:

- **EXPRESSION:** the associated computing action involves merely arithmetic operations of several explanatory variables, conditioned upon the values of some explanatory variables. For example, if $X_1 > 2$, then the value of defined variable X equals $X_2 + X_3 - X_4$, else X equals $2X_2 \times X_4$.
- **SQL:** the associated computing action is a regular database operation, which mainly accesses the explanatory variables' values from the database by means of a database query language.
- **DATE:** the associated computing action is relevant with date variables. For example, the value of defined variable X is the age of the firm.
- **SEQ_NUM:** the associated computing action generates a sequence number for the loan application.
- **RATIO:** the associated computing action calculates a ratio whose denominator or numerator may be zero.
- **OTHERS:** the associated computing actions do not match with those mentioned above.

Each defined variable has its own associated calculation priority, and the calculation priority of the defined variable X shall be defined to be lower than the calculation priorities of all explanation variables X_i . Within the evaluation process of a loan application, the calculation priority of each variable will determine when its corresponding value is calculated via parsing its associated XML file and implementing its associated computing action. The higher calculation priority a variable has, the higher priority of calculating its corresponding value is.

After reasoning out all defined variables required for a scoring model, the model manager is to define defined variables via the GUIs provided from MDAS.

⁶ We have implemented such a decision support system for a commercial bank via integrating the XML technology, the DB2 technology, and the Java technology.

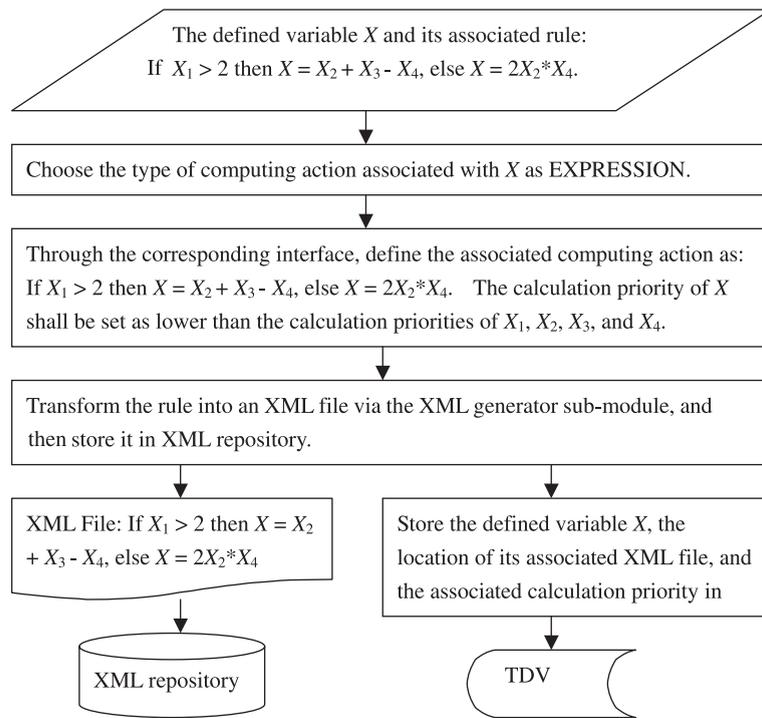


Fig. 2. An instance of defining a variable.

An instance of defining a defined variable is shown in Fig. 2.

Step 1

Choose the type of associated computing action suitable for the defined variable. Each chosen type of computing action will trigger a specialized application servlet used in Step 2 for inputting the detail definition of the rule associated with the defined variable.

Step 2

Define the detail of rules and the calculation priority associated with the defined variable through the corresponding servlet. The servlet processes the HTML request where all associated (explanatory) variables and the corresponding computing action shall be defined. The servlet passes the result on to the MRM to transform the definition of each rule and its associated computing action into an XML file via the XML generator sub-module.

Step 3

Store the defined variable. The MRM then stores the associated XML file in the XML repository as well as the defined variable, its associated calculation priority, and the location of associated XML file in TDV.

The MDAS listens for user http requests and delegates the specialized application servlet. The servlet processes the HTML request and then passes to MRM to transform the definition of each rule and its associated computing action into an XML file via the XML generator sub-module. MRM also manages the storage of defined variables, their calculation priorities, the associated XML files, and their locations. A scoring model is set up when all defined variables used in the model are defined via MDAS.

CPAS provides GUIs for account officers to input raw data of a loan application, save raw data into Database, and start EM. At the beginning of the evaluating process of each loan application, not only the raw data of the processed loan application shall be inputted and saved in Database, but also the corres-

ponding raw data of the private information on bank-borrow relationship of small firms and personal trade lines of small firms' chief executive officers shall be retrieved from the Joint Credit information Center and saved in Database.

Fig. 3 displays the later process of evaluation. With respect to each loan application, each defined variable will be accessed from TDV in turn. The access order of defined variables is in accordance with the decreasing order of their calculation priorities stated in TDV. The XML file associated with each accessed defined variable will be retrieved from the XML repository, and be parsed by the XML parser sub-module. The

XML parser will pass the result to the model engine sub-module that triggers a thread of the corresponding (Java) program to calculate the value of the accessed defined variable corresponding to the loan application.

Since the calculation priority of a defined variable X is defined to be lower than that of all explanation variables X_i , the corresponding values of these explanation variables X_i will be calculated first and saved in Database. Thus, in the computing action of the defined variable X , the corresponding values of these explanation variables X_i will be retrieved from Database for use; on the other hand, that of the accessed defined variable will be saved back into the Database.

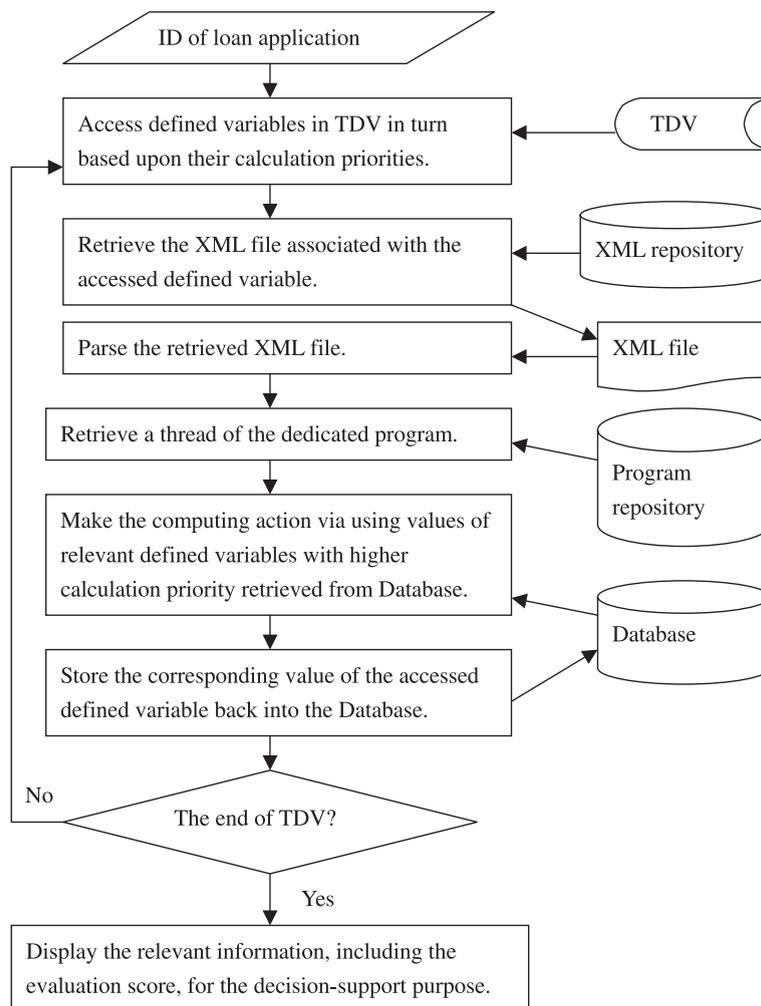


Fig. 3. Process of evaluating each loan application.

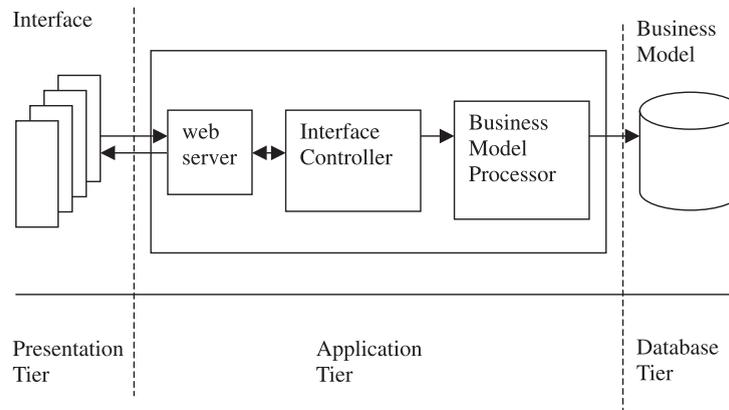


Fig. 4. The high-level architecture in the design of a subsystem.

At the end of evaluation, the corresponding evaluation score, which is also the value of a defined variable, is also stored back into Database.

The EM contains an XML parser sub-module that can parse XML files, and a model engine sub-module that provides threads of specific programs to make the associated computing actions and calculate the corresponding values of defined variables.

4. Summary and future work

In summary, as shown in Fig. 4, an N-tier architecture integrated with the idea of Model-View-Controller is adopted in the design of each subsystem. MAS, MDAS, and CPAS act as a kind of interface controller that gets the users' gestures, selects GUIs (view) shown in the client for response, and defines application actions. MRM and EM function as a kind of business model processor that maps application actions to updates of scoring models or implements scoring models. Scoring models, separated from and shared by both model processors (MRM and EM), are put into the database back-end via transforming scoring models into a set of defined variables. TDV and XML files constitute the scoring models.

Such an arrangement tries to facilitate the complexity involved in developing a credit scoring (decision support) system that manages the information asymmetry and time-vary essentials. This design renders the maintenance and upgrade of the scoring model easy via merely re-defining the rule or adding

new defined variables without managing any complex program. This design also allows the system engineers to avoid investing considerable time and effort in frequently communicating with the model managers for finalizing the scoring models. In summary, scoring models of this system can be easily altered to cope with the change of business environment.

One future work is to set up a subsystem including a data warehouse and a data mining mechanism that can regularly detect if the current scoring models properly manage the loan application.

Acknowledgements

The authors would like to thank Dr. Hsiou-Wei Lin and Dr. Robert Su for their help in developing the scoring models.

References

- [1] S. Bhattacharya, A. Thakor, Contemporary banking theory, *Journal of Financial Intermediation* 3 (1993) 2–50.
- [2] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, M. Stal, *Pattern-Oriented Software Architecture: A System of Patterns*, Wiley and Sons, New York, 1996.
- [3] K. Cavalluzzo, Market structure and discrimination: the case of small businesses, *Journal of Money, Credit, and Banking* 30 (4) (1998) 771–793.
- [4] K. Cavalluzzo, L. Cavalluzzo, J. Wolken, Competition, small business financing and discrimination: evidence from a new survey, business access to capital and credit, *Federal Reserve System Research Conference Proceeding*, 1999, pp. 180–266.

- [5] D. Diamond, Financial intermediation and delegated monitoring, *Review of Economic Studies* 51 (3) (1984) 393–414.
- [6] D. Finney, *Probit Analysis*, 3rd ed., Cambridge University Press, New York, 1971.
- [7] B. Groszof, Y. Labrou, H. Chan, A declarative approach to business rules in contracts: courteous logic programs in XML, *Proceedings of the 1st ACM Conference on Electronic Commerce*, 1999.
- [8] S. Katz, S. Lilien, B. Nelson, Stock market behavior around bankruptcy model distress and recovery predictions, *Financial Analysts Journal* 41 (1) (1985) 70–74 (Charlottesville).
- [9] G. Krasner, S. Pope, A description of the model-view-controller user interface paradigm in the smalltalk-80 systems, *Journal of Object Oriented Programming* (1988 August/September) 26–49.
- [10] J. Ou, S. Penman, Financial statement analysis and the prediction of stock returns, *Journal of Accounting and Economics* 11 (4) (1989) 295–324.
- [11] M. Petersen, R. Rajan, The effect of credit market competition on lending relationships, *Quarterly Journal of Economics* 110 (1995) 406–443.
- [12] K. Tam, M. Kiang, Managerial applications of neural networks: the case of bank, *Management Science* 38 (7) (1992) 926–947.



Ray Tsaih, also known as Rua-Huan Tsaih, is a professor at the Department of Management Information Systems, National Chengchi University, Taipei, Taiwan. He received his PhD in Operations Research in 1991 from the University of California, Berkeley. His research interests are Developing Neural Networks and Applying Neural Networks to Finance. His most recent work has been published in *Computer and Operation Research*, *Decision Support Systems*,

Mathematical and Computer Modelling, *Review of Securities and Futures Markets*, *MIS REVIEW*, *Taipei Economic Inquiry*, and *Sun Yat-sen Management Review*.



Yu-Jane Liu is a professor at the Department of Finance, National Chengchi University, Taipei, Taiwan. She received his PhD in Business Administration in 1991 from the National Sun Yat-sen University. Her research interests are Behavioral Finance and Market Microstructure. Her most recent work has been published in *Journal of Financial and Quantitative Analysis*, *Journal of Banking and Finance*, *Research in Experimental Economics*, *Review of*

Quantitative Finance and Accounting, *Journal of Business, Finance and Accounting*, *Global Finance Journal*, *Pacific Basin Finance Journal*, and *Journal of Financial Studies*.

Wenching Liu is an associate professor at the Department of Management Information Systems, National Chengchi University, Taipei, Taiwan.

Yu-Ling Lien is a master student majoring in Management Information System in National Chengchi University, Taipei, Taiwan.