A Decision Support Approach for Accounts Receivable Risk Management

Desheng D. Wu, David L. Olson, and Cuicui Luo

***Abstract*—Financial disasters in private firms led to increased emphasis on various forms of risk management, to include market risk management, operational risk management, and credit risk management. Financial institutions are motivated by the need to meet increased regulatory requirements for risk measurement and capital reserves. This paper describes and demonstrates a model to support risk management of accounts receivable. We present a decision support model for a large bank enabling assessment of risk of default on the part of loan recipients. A credit scoring model is presented to assess account creditworthiness. Alternative methods of risk measurement for fault detection are compared, and a logistic regression model selected to analyze accounts receivable risk. Accuracy results of this model are presented, enabling accounts receivable managers to confidently apply statistical analysis through data mining to manage their risk.**

***Index Terms*—Accounts receivable, data mining, decision support system, enterprise risk management, risk measures.**

I. Introduction

ENTERPRISE risk management (ERM) [1] is a systematic, integrated approach to managing all risks facing an organization. Its development in banking and accounting was motivated in part by traumatic events such as 9/11/2001 and business scandals including those of Enron [2] and WorldCom [3]. Consideration of risk has always been with business, leading to insurance. As AIG learned in 2008 [4]–[6], even financial field has involved high risks [7].

Risk itself is not always bad, as businesses exist to deal with those risks in their area of specialization, and insurance a means to protect against risks they are not as competent to deal with. Businesses exist to cope with specific risks efficiently: uncertainty creates opportunities for businesses to make profits. ERM seeks to provide means to recognize and mitigate risks throughout business activity. The field of insurance developed to cover a wide variety of risks, related to external and internal risks covering natural catastrophes, accidents, human error, and even fraud. Financial risk has been controlled through hedges and derivatives and other tools over the years, often by investment banks. With time, it was realized that many forms of risk could be prevented, or their impact reduced, through loss-prevention and control systems, leading to broader views of risk management. Contingency management has been widely systematized in the military and systematic organizational planning has recently begun to include scenario analysis, giving executives a means of understanding what might go wrong, and some opportunity to prepare reaction plans. A complicating factor to comprehensive risk management is that organization leadership is rarely a unified whole, but rather consists of a variety of stakeholders with potentially differing objectives; ERM offers a way through this, by opening risk management to strategy decisions through integration of various operations, and vice versa. Data mining models have been widely applied in banking for applications such as bankruptcy prediction and credit scoring [8].

This paper demonstrates a decision support model of a common banking risk management problem: Accounts Receivable risk management. This section introduced the problem of risk in organizations and the role of operations and finance in ERM. Section I discusses ERM in the context of finance operations. Section III discusses various data mining tools available to support financial ERM. Section IV presents variables and data, Section V discusses results and discussions. Conclusions are presented in Section VI.

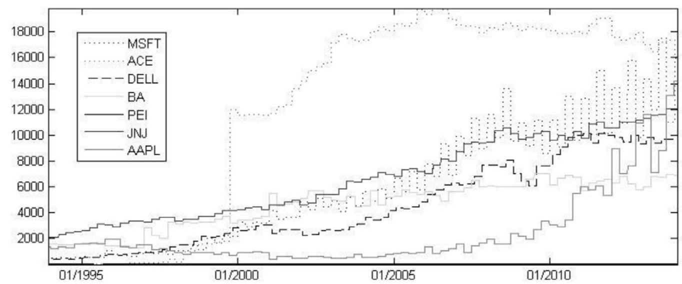


Fig. 1. Accounts receivable for representative companies; data from Bloomberg.

II. Accounts Receivable Risk Management

For most firms, accounts receivable is the largest asset on their balance sheet. Therefore, any inappropriate risk management for accounts receivable could have a serious impact on the firm's financial statements.

Fig. 1 demonstrates the amount of accounts receivable for seven giant companies from 1993-12-1 to 2014-1-31. The

Manuscript received December 18, 2013; revised March 25, 2014; accepted April 12, 2014. Date of publication July 21, 2014; date of current version November 13, 2014. This paper was recommended by Associate Editor T.-M. Choi.

D. D. Wu is with Stockholm Business School, Stockholm University, 114 18 Stockholm, Sweden, and also with RiskLab, University of Toronto, Toronto, ON M5S 2E4, Canada (e-mail: dash@risklab.ca).

D. L. Olson is with the University of Nebraska-Lincoln, Lincoln, NE 68588 USA (e-mail: david.olson@unl.edu).

C. Luo is with the Department of Mathematics, University of Toronto, Toronto, ON M5S 2E4, Canada (e-mail: cui.luo@mmf.utoronto.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2014.2318020

TABLE I
HISTORICAL EVOLUTION OF ACCOUNTS RECEIVABLE CREDIT RISK MITIGATION

Period	Phase	Events
I 1991-1996	Bank domination	Banks seek credit risk reduction by individual Insurance companies the primary investors **Total Return Swap** **Credit Default Swap** on single corporation
II 1996-2000	Banks continue dominance	Banks reduce credit risk on portfolios Focus on reduction of regulatory capital Money managers become investors **First-to-Default Swap** **Credit Default Swap** on basket of corporations
III 2000-2002	Shift from Banks to Investors	Banks focus on economic capital Insurance firms and banks joined by reinsurers Hedge funds emerge Money manager activity grows **Credit Default Swap on Asset-based Swaps**
IV 2002-2008	Investors dominate	Many banks diversity credit portfolios by going long on credit risk other than their lending portfolios Full range of institutional investors **Index of CDS** **Options on CDS** **Standard Index Tranches**
V 2008-current	Response to crash	Banks (those that survived) become ultraconservative

trend shown by curves of the figure suggests that the amount of accounts receivable for most companies jumps to a very high level since last financial crisis in 2007. Moreover, this high level of accounts receivable amount keeps a long time. For example, this very high amount has not decreased for Microsoft Corporation (MSFT) much since the financial crisis in late 2007. It has been constantly increasing for Apple Inc. (AAPL).

Rating of accounts receivables provides foundation for trade receivables to be securitized. Katz [9] reported that there were about \$10 trillion in accounts receivable in the U.S. alone, with a comparable amount in Europe, but that only about \$60 billion were securitized. Trade credit insurance is available to protect firms from bad debt losses on accounts receivable [10]. Smithson [11] reviewed the emergence of credit risk as shown in Table I.

The traumatic events of 2008 have made all financial risk management critical. Firms have responded in part by holding on to cash much more than in the past [12]. Thus, tools to aid accounts receivable risk management decisions are needed.

III. Statistical Modeling Approaches

The decision support system we propose relies on the development and application of standard data mining models for classification using well-known data mining methods [13], [14]. Systems to manage risk involved with accounts receivable have been published for applications in Brazil [15], which faced relatively more variable environments, as well as study of the supply chain related cash flow risks seeking to improve a cash conversion cycle [16].

Data mining studies typically involve applying logistic regression, neural network, and decision tree models to the data, selecting the specific model that the decision maker is most comfortable with unless one model seems to fit specific datasets much better [17]. Other data mining models have often been applied to financial fraud detection as well [18]. A few studies, see [36], employed a supervised learning approach for management of accounts receivable collections. We will briefly discuss all three of these standard models.

A. Logistic Regression

One key popular technique in recent years applied to financial risk management is logistic regression [19], [20]. It has been applied to fraud detection [21], and in management of accounts receivable in other areas such as electronic health care records [22], [23]. Logistic regression belongs to the larger class of generalized linear models that link the expected value of the target variable to a linear predictor through a function $\log(\cdot)$. In logistic regression the response variable is a binary variable that indicates if an event has taken place. For example, suppose the binary variable Y takes the value 1 to represent that an account or customer has defaulted and assumes the value 0 otherwise. Denote by $p = P(Y = 1 \mid x)$. the conditional probability of the case default given a set of predictors or explanatory variables $x = (x_1, \ldots, x_n)$, the logistic regression model is

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j \tag{1}$$

where n denotes the number of predictors. The term on the left of the above equality is called the logit [24] link function (or log of odds) and transforms the domain [0, 1] into the real line $(-\infty, +\infty)$ through the linear predictor. We assume that we have data consisting of N observations

$(x_i, y_i) = (x_{i1}, \ldots, x_{iN}, y_i)$ for $i = 1, \ldots, N$. The regression parameters α_i are estimated from this modeling dataset. The logistic regression procedure in SAS fits a given model through maximum likelihood estimation, either using Fisher-scoring or the Newton-Raphson optimization algorithm, resulting in the fitted model

$$\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \alpha_0 + \sum_{j=1}^{N} \alpha_j x_{ij}, \quad i = 1, \ldots n \tag{2}$$

where Λ denotes that parameters Λ are estimated ones.

B. Decision Trees

Decision trees are another type of model commonly applied in financial risk management [25], [26]. A decision tree represents a segmentation of the data that is created by applying a set of rules. Each rule assigns an observation to a segment based on the value of one input. One rule is applied after another, resulting in a hierarchy of segments within segments. The hierarchy is called a tree, and each segment is called a node. A decision is made and applied to all observations in the final nodes (leaves) of the tree. The type of decision depends on the context. In predictive modeling, the decision is simply the predicted value.

Decision tree has two main advantages over other models available in SAS Miner. First, it produces a model that may represent interpretable logic statements; second it can classify observations with missing data. Three measures are used for evaluating a splitting rule: a statistical significance test, i.e., an F test or a Chi-square test; the reduction in variance; and entropy or Gini impurity measure. To build a decision tree, the space of values for the predictors $x = (x_1, \ldots, x_n)$ is first partitioned into a collection of M regions $\omega_1, \ldots, \omega_M$ that correspond to a tree's terminal nodes or leaves. The data consists of N observations $(x_i, y_i) = (x_{i1}, \ldots, x_{iN}, y_i)$ for $i = 1, \ldots, N$. In a node m suppose there are N_m observations and let

$$\hat{p}_{ml} = \frac{1}{N_m} \sum_{\mathbf{x}_i \in \omega_m} I(y_i = l) \tag{3}$$

denote the proportion of class l $(l = 0,\ 1)$ observations in that node for the response y_i. The observations within node m are then classified to the majority class $l_m = \arg\max_l \hat{p}_{ml}$. Different measures of node error or impurity include [26] the misclassification error, the Gini index, and the cross-entropy or deviance measure. The purpose is to minimize these measures of error.

C. Neural Networks

The third basic data mining model is neural networks. These models also have been commonly used in financial risk management and ERM [27]–[30]. Feed-forward neural networks are a class of flexible nonlinear regression, discriminant analysis, and data reduction models. By detecting complex nonlinear relationships in data, neural networks can help to make predictions about real-world problems.

A neural network consists of neurons or nodes and connections between those neurons. There are three types of nodes: input nodes, hidden nodes for internal computations, and output nodes that compute the predicted values and compare them with the target variable values. Most connections in a network have an associated numeric value called a weight or parameter estimate. The training methods attempt to minimize the error function by iteratively adjusting the values of the weights. Most nodes also have one or two associated numeric values called bias and altitude, which are also estimated parameters adjusted by the training methods. Hidden and output nodes use two functions to produce their computed values. A single value is yielded by feeding all the computed values from previous nodes into a given node using a combination function. An activation function then transform the value produced by the combination function into outputs which involves no weights or other estimated parameters. The perceptron, a type of linear discriminant model, belongs to the earliest neural network architectures. This simple neural network has a single input node or independent variable, a single target or dependent variable, and a single output node or predicted value. Extensions include using logistic functions as the activation function, which makes a perceptron equivalent in functional form to a logistic regression model.

We now present a neural network utilized for binary classification as follows. Assume a target binary variable Y located at the top of a neural network diagram. Features Z_m are obtained from linear combinations of the input variables $x = (x_1, \ldots, x_n)$ and then the target Y is modeled by linear functions of Z_m $(m = 1, \ldots, M)$

$$\begin{aligned} Z_m &= \Phi(\alpha_{0m} + \alpha_m \cdot x) \\ f(x) &= g(\beta_0 + \beta \cdot Z) \end{aligned} \tag{4}$$

where $Z = (Z_1, \ldots Z_M)$. The activation function $\Phi(\cdot)$ is frequently chosen as a sigmoid $\Phi(v) = 1/(1 + e^{-v})$. In the case that $\Phi(\cdot)$ is the identity function, the neural network reduces to a linear model. Here, neural networks can be thought of as nonlinear generalizations of linear classification models. The parameters or weights in the model above consist of $\{\alpha_{0m}, \alpha_m; m = 1, 2, \ldots, M\}$ and $\{\beta_0, \beta\}$, where the above sets are collections of $M(N+1)$ and $M+1$ parameters, respectively. The output activation function $g(\cdot)$ allows a final transformation of the output vector $\beta_0 + \beta \cdot Z$ into output values. The middle nodes in the network $Z = (Z_1, \ldots Z_M)$ are not directly observable and therefore known as the hidden nodes, which are nonlinear functions of linear combinations of the observable inputs. Denote by $(x_i, y_i) = (x_{i1}, \ldots, x_{iN}, y_i)$ the training/modeling dataset, the widely employed measure of fit is either squared error $\sum_{i=1}^{N}(y_i - f(x_i))^2$ or cross-entropy $\sum_{i=1}^{N} y_i \log f(x_i)$ for classification problems.

The multilayer perceptron (MLP), known as universal approximators, is one of the most popular neural networks. MLPs can be used when you have little prior knowledge of the relationship between inputs and targets. A MLP has any number of inputs, one or more hidden layers with any number of nodes, any number of outputs with any activation function, connections between the input layer and the first hidden layer, between the hidden layers, and uses linear combination functions in the hidden and output layers or sigmoid activation functions in the hidden layers.

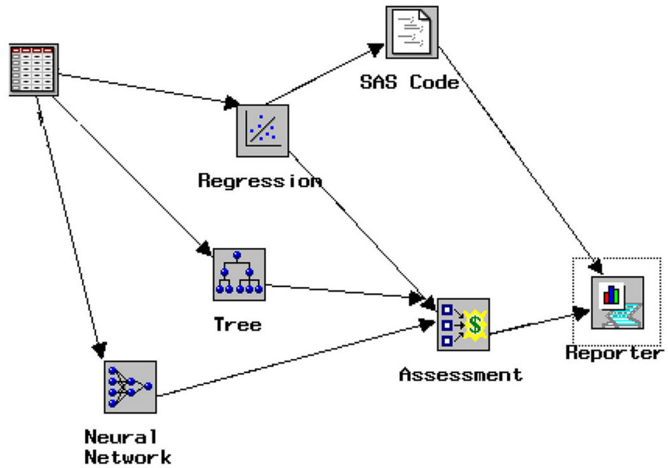


Fig. 2. SAS miner screen for objects used in the study.

D. Integrated system

SAS Miner is a standard industry data mining software. It is an excellent tool enabling use of objects to represent data, models, and output. It thus serves as an excellent decision support system. Fig. 2 demonstrates SAS Miner Screen for objects used in the study. Multiple data mining models, i.e., Logistic Regression, Regression Trees, and Neural Networks are built based on data sample provided. These data mining models are then assessed and results are reported in the object of reporter.

IV. Decision Support of Risk Management for Accounts receivables

In general, good indicators of portfolio credit quality are historic delinquency and write-off performance of associated receivables. A bank will usually carry delinquent trade receivables far shorter than most companies would before writing them off. However, it might be hard to determine the value of delinquent receivables in case of a discretionary charge-off policies, which may also be subject to manipulation. Therefore, the trade receivable criteria in a bank focuses on analyzing early-stage delinquencies as credit quality indicators.

An account is defined as "bad" if over the next five months as of observation date falls into any of the categories: 1) commercial risk rate is downgraded to 6 or above; 2) sent to the Special Lending Unit (SLU); 3) sent to National Credit Centre; and 4) sent to external collection agencies. The response variable is binary (1 if "bad," 0 otherwise). Nonbad accounts are classified as "Good." By computing the performance of how it predicts the bad accounts within five months, the effectiveness of the decision support scorecard will be measured. Statistic values of Lorenz curve, K–S, and divergence are calculated.

The approach used to develop the key decision support model is logistic regression, which is then compared with two benchmarking data mining models: decision tree and neural network. For model development, the validation has both an Observation Window and an Outcome Window. A score on the Score Date is computed using the historical data in an account from the Observation Window, which then predicts the account's performance during the Outcome Window. When an account is observed, the Observation Window defines the period during which data is used to compute an account's behavior score. The Outcome Window defines the period when the performance of the account is observed.

Fig. 3 depicts shifting observation and performance windows used in the model.

We exclude the accounts that fall into any of the following four categories: 1) less than six months of history; 2) less than five months of history after being scored; 3) accounts without a credit limit at score date; and 4) for each bad definition, the corresponding bad accounts at the score date.

A. Data and Variables

Table II illustrates the Bad Rates by Line Size for two samples from two sources of the sample data: credit bureau and the bank database, where the amount and percentage for both samples are reported. The first sample consists of the four types of Business lines of credit products, overdraft privilege, and mandatory-review accounts. The second sample only contains business lines of credit accounts. The overall bad rate of Overlimit is between 8% and 9.5%, while the bad rate of "35+ Days Since Last Deposit and Overlimit" accounts is between 1% and 2.8%. Table II shows that the bad rate for the second sample was higher than that of the first one: 9.44% versus 8.42% for "Overlimit" and 2.64% versus 1.76% for "of 35+."

The predictors in the model are variables that have to be correlated to the response and collected over the observation window. Account transaction activity information was analyzed to determine the most predictive characteristics. Table III gives variables and sources selected during the model building process.

Because no response variable is available at the time of subsequent validations, we have to evaluate alternative possible response variables: bad1 = overlimit; bad2 = overlimit and 90 days since last payment, and bad3 = overlimit and 35 days since last payment. We present this result for variable selection in Table IV. Only the variables selected in the final model were available in all data sets. We then analyze the distributions and correlations of explanatory variables across the two samples. From Table IV, we see that in sample 2 the frequency of the characteristic A1 has changed in the less than $50 M category, rendering this characteristic not usable in the model scoring. Based on a further multivariate analysis, the variables are reduced to two clusters, explaining 61% of the variation in development versus 59% in sample 2. We then observe slight decrease in the amount of variance explained by the chosen variables. Note that a multivariate analysis is preferred to a univariate analysis, since the latter does not take into account partial associations. Since the raw data for model development was not available, variable redundancy analysis was not possible.

B. Benchmarking

In order to provide independent benchmarks for the logistic regression models, each of these models were run on

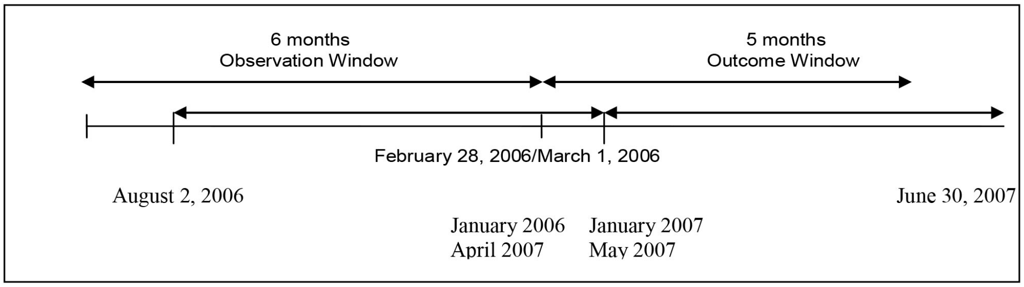


Fig. 3. Observation and performance windows.

TABLE II
BAD RATES BY LINE SIZE FOR TWO SAMPLES

Sample	Line Size	# of Overlimit	Bads # of Overlimit	Bad Rate of Overlimit	# of 35+	Bads # of 35+	Bad Rate of 35+
Sample 1	≤ \$50M	59,332	5,022	8.46%	61067	1,127	1.85%
	\$50-\$100M	6,777	545	8.04%	7,000	69	0.99%
	Total	66,109	5,567	8.42%	68,067	1,196	1.76%
Sample 2	≤ \$50M	61,183	5,790	9.46%	63,981	1,791	2.80%
	\$50-\$100M	6,915	637	9.21%	7,210	88	1.22%
	Total	68,098	6,427	9.44%	71,191	1,879	2.64%

Note: 35+ denotes "35+ Days Since Last Deposit and Overlimit"

TABLE III
VARIABLES AND SOURCES

Label	DESCRIPTION	TYPE
L1	No deposit when utilization > 100% in current month	Binary
L2	No credit balance in current month	Binary
L3	Utilization ≥80% in current month	Binary
L4	Age of account (in months)	Counts
L5	# of no deposit when outstanding>0% in past 5 months	Counts
L6	# of utilization > 100% in past 6 months	Counts
L7	# of utilization ≥ 90% in past 6 months	Counts
L8	# of utilization ≥ 50% in past 3 months	Counts

TABLE IV
VARIABLE SELECTION

	Type	Training	Validation
Sample 2 (≤\$50M)	Bad1	8.55%	8.25%
	Bad2	0.68%	0.62%
	Bad3	1.59%	1.41%
Sample 2	size	A1 = 0	A1 = 1
Development (≤\$50M-Training)	40,853	98.86%	1.14%
≤\$50M (training only)	41,510	100%	0%
≤\$100M (training + validation)	128,733	98.97%	1.03%

TABLE V
COMPARISON OF FITS FOR STANDARD DATA MINING MODELS

Charge-Off Model	Modeling RMSE	Validation RMSE
Logistic Regression	0.4519	0.4500
Classification Tree	0.4520	0.4505
Neural Network	0.4500	0.4482

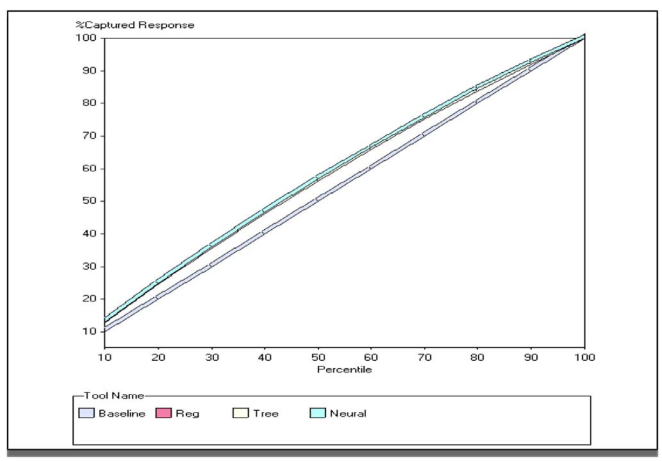


Fig. 4. Lorenz curve—validation dataset.

the modeling dataset and scored using the validation dataset. Table V presents a comparison of the root mean square error (RMSE) on the models and shows for that both datasets, all three standard data mining models performed very much the same for this data set.

Fig. 4 shows the Lorenz (cumulative accuracy profile—CAP) curve comparison [31] for the Validation dataset by running the Enterprise Miner software. It shows that logistic regression and neural network provide some lift over random and their risk discrimination capabilities are almost identical.

V. RESULTS AND DISCUSSION

The Akaike information criterion (AIC) [32], Schwartz criterion (SC) [33], and the -2 Log Likelihood [34] are used when competitive models are evaluated. Table VI also shows

TABLE VI
STATISTICAL TESTS

Data set	Partition	Bad Definition	% Concordant	% Discordant	% Tied	Hosmer-Lemeshow Goodness- of-Fit Test Prob> ChiSq
Development	Training	BADIND	88.3	8.8	2.9	<0.0024
	Hold-out	BADIND	88.8	8.1	3.1	0.04
Validation – Sample 2 (≤\$50M)	Training	BAD1	82.1	15.6	2.3	<0.0001
	Validation	BAD1	80.8	16.7	2.5	<0.0001
	Training	BAD2	86.3	6.4	7.3	0.12
	Validation	BAD2	83.7	6.3	9.9	0.07
	Training	BAD3	85.8	10.3	4.0	<0.0001
	Validation	BAD3	83.1	11.5	5.3	0.004
Beacon	Training	BAD1	69.3	29.5	1.2	<0.0001
	Training	BAD2	62.5	23.8	13.7	0.15
	Training	BAD3	67.9	26.5	5.6	0.0019

the Hosmer–Lemeshow [35] goodness-of-fit test results, which might be interpreted with caution because of its dependence on the way the observations are grouped. In addition, logistic regression has been fitted by using the above three bad definitions as response variables and maximum Beacon score as an explanatory variable. Beacon score is the credit rating used by Equifax Credit Bureau to rank an individual's creditworthiness. The performance of Beacon was compared only in predicting the response variables. As a result, the values of the AIC, SC, and −2 Log L tests for model fit suggest that Beacon has a lower power to explain variation. This was expected since Beacon data never include internal banking information.

A final Logistic regression model was generated based on a comparison of the statistic. Denote by p the probability of becoming bad. The final behavioral scoring model used in the scorecards is

$$\begin{aligned}\text{Log}\left(\frac{p}{1-p}\right) &= -6.0295 + 0.5996 * \text{L1} + 0.7967 * \text{L2}\\ &\quad + 0.5141 * \text{L3} - 0.0038 * \text{L4} + 0.2022 * \text{L5}\\ &\quad + 0.2398 * \text{L6} + 0.0793 * \text{L7} + 0.4862 * \text{L8}.\end{aligned}$$

The decision support model was then assessed using two criteria: 1) the score distribution and characteristic analysis and 2) the scorecard's ability to discriminate between good and bad accounts, that is, how well the scorecards perform.

A. Score Distribution and Characteristics

The population stability indices in Fig. 5 for the \$50 000 and \$100 000 segments show a steady increase from early 2006 to the middle of 2007. It seems to suggest that for each passing month, the score distribution of the development data becomes more unlike that of the most current population. However, the values of stability indices almost all remain below the benchmark value of 0.25, suggesting a significant shift in the score population.

Table VII presents percentage of Accounts for different types. As time spans and the portfolio ages, more accounts become less risky because they are assigned lower values, thus contributing to a shift in the overall score. For the January 2006 sample with a limit of less than \$50 000, 68%

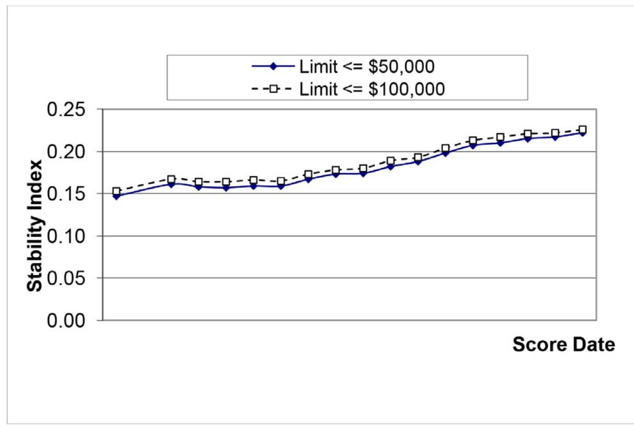


Fig. 5. Stability indices form Jan. 2006–Jun. 2007.

of accounts were on books for more than three years and 77% of accounts in June 2007. Of all the characteristics, the January 2006 sample has shown the greatest disparity between the Development sample and the most recent one, i.e., June 2007. From the table, it seems that this disparity continues to increase.

Percentage of accounts without Credit Balance is presented in Table VII, showing another contributing factor in the scoring period of time. A credit balance does not belong to 27.60% of accounts as of October 2007, and to 33.33% of accounts in June 2007. This suggests that more accounts receive higher scores to indicate riskier behavior. It is noted that this gap decreases over time.

It can be seen from Table VIII that the score distribution is very skewed. In the Development sample, 19% of accounts yield the score of either 21 or 22. This seems to imply that the stability index can be influenced by even a migration of one point down or up. As time spans we expect a shift in the score distribution, which is only inflated in the case of a large accumulation of accounts within a narrow range of scores. As of June 2007, 9.7% the accounts receivable had a score of either 21 or 22. Although the score shift is in general attributed to the distribution change, the degree of the increasing population shift is exaggerated by the nonscaled score in our case.

TABLE VII
PERCENTAGE OF ACCOUNTS FOR DIFFERENT TYPES

type	Population	Development: October 2007	January 2006	June 2007
Time on Books Greater than 3 Years	Limit ≤$50,000	54.20%	67.53%	77.46%
	Limit ≤ $100,000	54.20%	68.09%	77.85%
no Credit Balance	Limit ≤ $50,000	27.60%	38.62%	33.33%
	Limit ≤ $100,000	27.60%	39.73%	34.60%

TABLE VIII
SCORE DISTRIBUTION

Score	Development		June 2007		Index
	Count	Interval Percentage	Count	Interval Percentage	
394 – 10,000	5,467	10.00%	9,358	8.84%	0.001
153 – 393	5,471	10.00%	12,787	12.08%	0.004
62 – 152	5,461	10.00%	11,683	11.04%	0.001
30 – 61	5,840	10.70%	10,409	9.83%	0.001
23 – 29	7,621	13.90%	6,865	6.48%	0.056
22	6,109	11.10%	4,565	4.31%	0.064
21	4,340	7.90%	5,723	5.40%	0.009
19 – 20	5,343	9.70%	12,937	12.22%	0.006
16 – 18	4,074	7.40%	15,797	14.93%	0.053
0 – 15	5,086	9.30%	15,680	14.81%	0.026
Total	54,812	/	105,804	/	0.222

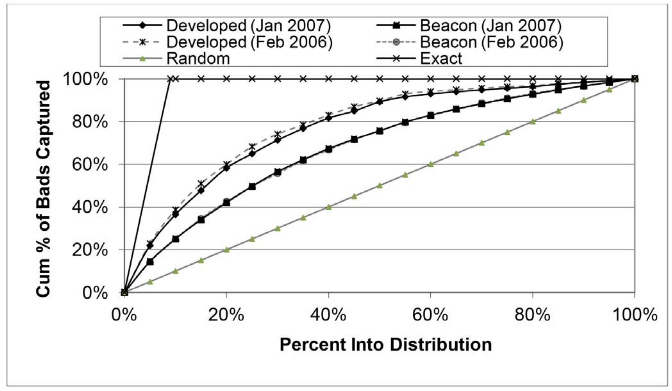


Fig. 6. Lift curve of overlimit sample performance.

Table VIII suggests that the account management strategy can be affected by the population shift. By comparing the interval percentage for the Development and June 2007 sample in Table VIII, we see that as scores migrate, the existing cut-offs used in the strategy may become less appropriate. Big gaps are observed for the score band "23–29," "22," "23–29," and "0–15." The shifted score distribution expressed in the column of Interval Percentage of Table VIII also shows that the population used to develop the model is different from the recent population in June 2007. This seems to imply that the weights assigned to each characteristic value might not be the ones most suitable for the current population used for modeling in June 2007.

B. Model Performance

Figs. 6 and 7 depict the Lift Curve of Sample Performance for "Overlimit" and "35+ Days Since Last Deposit & Overlimit," respectively. The Lorenz curves in Figs. 6 and 7 show that the developed scorecard captures bad accounts much more quickly than the Beacon score. Table IX gives performance statistics including mean score, standard deviation score, Divergence, K-S Statistics, and Lorenz ratio. As can be seen from the table, for either sample, the decision support scorecard was better able to predict a bad account than the bureau score. For the Overlimit definition, the development sample captured 58% of bad accounts, while Beacon captured 42%. The maximum personal Beacon score associated with each account was used as a benchmark of the business risk assessment. All the performance statistics based on the February 2006 score sample are better than those based on the January 2007. For example, in the case of the 35+ Days sample, the scorecard produced K-S statistic values of 58.53% and 64.58% for January and February 2007, respectively. For both bad definitions, there seems to be possible evidence of performance deterioration of the developed scorecard.

A comparison of Figs. 6 and 7 suggests that both scorecards, especially the developed model, perform better based on the sample with 35+ Days. There are two possible reasons for this observation. First, it can be better to use this bad definition for the sample of 35+ Days to approximate the original bad specification used to develop the scorecard. Second, it seems that this definition specifies a more severe degree of delinquency, resulting in a bigger gap between the goods and bads. Both figures seem to indicate that accounts that have less frequent deposits are more identifiable by the scorecard since they are more likely to exhibit lower aggregate credit balances. Consequently, the scorecard is more likely to assign high and risky scores to accounts that exhibit this type of behavior. Table IX presents performance statistics including divergence, K-S Statistics, and Lorenz values for Revision model, which defines firm's percentile ranking score based on analyst sentiment, and Beacon score. Table IX shows that for the January 2007 score date, 22.87% of Overlimit bad accounts had an aggregate credit balance, while only 9.95% had a balance for the 35+ Days definition. This implies that the latter accounts are much riskier: 700 versus 345. In contrast, the 35+ Days definition demonstrates a greater degree of delinquency for each of the characteristics.

VI. CONCLUSION

Accounts receivable involve financial risk, which was recognized as critically important after scandals and market disruptions centering around 2000, and has become even more critical since 2008. Accurate management of risk enables firms to profit, even in turbulent markets.

TABLE IX
PERFORMANCE STATISTICS

Statistic		Overlimit				35+ Days Since Last Deposit and Overlimit			
		Revision (Jan 2007)	Beacon (Jan 2007)	Revision (Feb 2006)	Beacon (Feb 2006)	Revision (Jan 2007)	Beacon (Jan 2007)	Revision (Feb 2006)	Beacon (Feb 2006)
Stat.	Good # Accounts	60,542	60,542	61,671	61,671	66,871	66,871	69,312	69,312
	Good Mean Score	108.89	738.71	127.30	734.67	137.4	734.28	171.81	729.23
	Good Std.	172.74	60.18	203.26	63.53	221.22	62.78	284.21	66.66
	Bad # Accounts	5,567	5,567	6,427	6,427	1,196	1,196	1,879	1,879
	Bad Mean Score	344.90	693.13	439.63	685.79	699.82	678.03	995.65	663.2
	Bad Std.	321.53	69.45	387.24	73.27	570.77	75.42	756.34	76.08
Bad Rate		8.42%	8.42%	9.44%	9.44%	1.76%	1.76%	2.64%	2.64%
Divergence		0.836	0.492	1.020	0.508	1.688	0.657	2.079	0.852
K-S	Score	78	726	136	716	233	726	394	707
	Ratio	45.76%	29.84%	48.82%	29.73%	58.53%	32.88%	64.58%	38.44%
Lorenz	Score Area	0.7713	0.6791	0.7842	0.6789	0.8659	0.7178	0.8857	0.7416
	Exact Curve Area	0.9579	0.9579	0.9528	0.9528	0.9912	0.9912	0.9868	0.9868
	Ratio	80.52%	70.89%	82.31%	71.26%	87.36%	72.42%	89.75%	75.15%

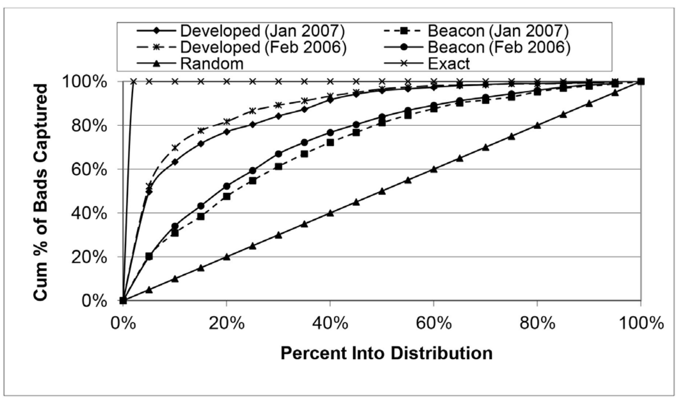


Fig. 7. Lift curve of performance sample of 35+ days.

A number of alternative statistical modeling approaches are available. We selected logistic regression as it provides a well-studied format providing sound models for financial data. While alternative conventional data mining models such as decision trees or neural networks provided very similar fits for training and validation data, logistic regression was used for the decision support system we propose.

Our scorecard model was developed to predict the likelihood of a delinquent accounts recovering within 5 months. We have presented detailed data and model development. Our data was divided into three subpopulations, the first consisting of cases over lending limits, the second over limits and having at least 90 days since last payment, and the third over limits and within 35 days since last payment. Benchmarking data was used for validation. SAS Miner logistic regression models were developed. We identified the eight variables associated with these changes. Research results indicate that the scorecard based on our logistic regression model identifies bad accounts in a superior manner relative to the Beacon score.

The decision support system based on our logistic regression offers a means for lending institutions to monitor accounts receivable more accurately, enabling them to maintain profitability in turbulent financial market times.

REFERENCES

[1] D. L. Olson and D. Wu, *Enterprise Risk Management Models*. Berlin, Germany: Springer, 2010.
[2] B. McLean and P. Elkind, *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. New York, NY, USA: Portfolio Trade, 2004.
[3] W. Pavlo, Jr. and N. Weinberg, *Stolen Without a Gun: Confessions from Inside History's Biggest Accounting Fraud—The Collapse of MCI Worldcom*. Tampa, FL, USA: Etika Books, 2007.
[4] R. Shelp and A. Ehrbar, *Fallen Giant: The Amazing Story of Hand Greenberg and the History of AIG*. Hoboken, NJ, USA: Wiley, 2009.
[5] R. Boyd, *Fatal Risk: A Cautionary Tale of AIG's Corporate Suicide*. Hoboken, NJ, USA: Wiley, 2011.
[6] A. S. Blinder, *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*. New York, NY, USA: Penguin Press, 2013.
[7] D. L. Olson and D. Wu, *Enterprise Risk Management*. Singapore: World Scientific, 2008.
[8] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 421–436, Jul. 2012.
[9] A. Katz, "Accounts receivable securitization," *J. Struct. Finan.*, vol. 17, no. 2, pp. 23–27, 2011.
[10] D. Edwards, "Trade credit insurance and captives," *Risk Manage.*, vol. 51, no. 8, p. 45, 2004.
[11] C. Smithson, "The promise of credit derivatives in nonfinancial corporations (and why it's failed to materialize)," *J. Appl. Corp. Finan.*, vol. 18, no. 4, pp. 54–60, 2006.
[12] T. W. Bates, K. M. Kahle, and R. M. Stulz, "Why do U.S. Firms hold so much more cash than they used to?" *J. Finan.*, vol. 64, no. 5, pp. 1985–2021, 2009.
[13] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Amsterdam, Netherland: Morgan Kaufmann, 2005.
[14] D. L. Olson and Y. Shi, *Introduction to Business Data Mining*. New York, NY, USA: McGraw-Hill/Irwin, 2006.
[15] G. Montibeller, V. Belton, and M. V. A. Lima, "Supporting factoring transactions in Brazil using reasoning maps: A language-based DSS for evaluating accounts receivable," *Decis. Support Syst.*, vol. 42, no. 4, pp. 2085–2092, 2007.

[16] C.-Y. Tsa, "On supply chain cash flow risks," *Decis. Support Syst.*, vol. 44, no. 4, pp. 1031–1042, 2008.

[17] W. Zhou and G. Kapoor, "Detecting evolutionary financial statement fraud," *Decis. Support Syst.*, vol. 50, no. 3, pp. 570–575, 2011.

[18] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–589, 2011.

[19] S. B. Caudill, M. Ayuso, and M. Guillén, "Fraud detection using a multinomial logit model with missing information," *J. Risk Insuran.*, vol. 72, no. 4, pp. 341–371, 2005.

[20] S. Viaene, M. Ayuso, M. Guillén, D. Van Gheel, and G. Dedene, "Strategies for detecting fraudulent claims in the automobile insurance industry," *Eur. J. Oper. Res.*, vol. 176, no. 1, pp. 565–583, 2007.

[21] J. Perols, "Financial statement fraud detection: An analysis of statistical and machine learning algorithms," *Auditing*, vol. 30, no. 2, pp. 19–50, 2011.

[22] G. O. Ginn, J. J. Shen, and C. B. Moseley, "Hospital financial position and the adoption of electronic health records," *J. Healthcare Manage.*, vol. 56, no. 5, pp. 337–350, 2011.

[23] B. Shen, T. M. Choi, Y. Wang, and C. K. Y. Lo, "The coordination of fashion supply chains with a risk averse supplier under the markdown money policy," *IEEE Trans. Syst., Man, Cybern. Syst.,* vol. 43, no. 2, pp. 266–276, Mar. 2013.

[24] D. N. Joanes, "Reject inference applied to logistic regression for credit scoring," *IMA J. Math. Appl. Bus. Industry*, vol. 5, no. 1, pp. 35–43, 1993.

[25] S. Kotsiantis, E. Koumanakos, D. Tzelepis, and V. Tampakas, "Forecasting fraudulent financial statements using data mining," *Int. J. Comput. Intell.*, vol. 3, no. 2, pp. 104–110, 2006.

[26] T. M. Choi, C. H. Chiu, and P. L. Fu, "Periodic review multi-period inventory control under a mean-variance optimization objective," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 8, pp. 678–682, Jul. 2011.

[27] J. W. Lin, M. J. Hwang, and J. D. Becker, "A fuzzy neural network for assessing the risk of fraudulent financial reporting," *Manage. Audit. J.*, vol. 18, no. 8, pp. 657–665, 2003.

[28] S. Viaene, G. Dedene, and R. A. Derrig, "Auto claim fraud detection using Bayesian learning neural networks," *Expert Syst. Appl.*, vol. 29, no. 3, pp. 653–666, 2005.

[29] C. Ghidini, C. Francescomarino, M. Rospocher, P. Tonella, and L. Serafini, "Semantics-based aspect-oriented management of exceptional flows in business processes," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 1, pp. 25–37, Jan. 2012.

[30] T. M. Choi, C. H. Chiu, and P. L. Fu, "Periodic review multi-period inventory control under a mean-variance optimization objective," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 4, pp. 678–682, Jul. 2011.

[31] J. Sobehart, S. Keenan, and R. Stein, *Benchmarking Quantitative Default Risk Models: A Validation Methodology*. New York, NY, USA: Moody's Investors Service, 2000.

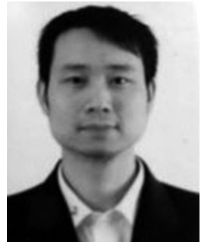
[32] H. Akaike, "Likelihood of a model and information criteria," *J. Econ.*, vol. 16, no. 1, pp. 3–14, 1981.

[33] G. E. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[34] P. McCullagh and J. Nelder, *Generalized Linear Models*. London, U.K.: Chapman and Hall, 1989.

[35] D. W. Hosmer and S. Lemeshow, "A goodness-of-fit test for the multiple logistic regression model," *Commun. Statist.*, vol. 9, no. 10, pp. 1043–1069, 1980.

[36] S. Zeng, I. Boier-martin, P. Melville, C. Murphy, and C. A. Lang, "Predictive modeling for collections of accounts receivable," in *Proc. 2007 Int. Workshop DDDM*, New York, NY, USA, pp. 43–48.

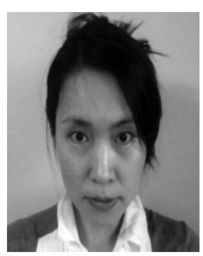


Desheng D. Wu is an Affiliate Professor and Managing Director at the RiskLab of University of Toronto, Toronto, ON, Canada. He has published over 80 papers in refereed journals including *Production and Operations Management, Decision Sciences, Risk Analysis*, and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. His current research interests include enterprise risk management in operations, performance evaluation in financial industry, and decision sciences.

Prof. Wu is the Editor of Springer book series titled *Computational Risk Management*. He has served as an Editorial Board Members/Guest Editors in journals including the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *Annals of Operations Research*, *Computers, and Operations Research*, *International Journal of Production Economics*, and *Omega*.

David L. Olson is the James & H.K. Stuart Professor at MIS and Chancellor's Professor at the University of Nebraska-Lincoln, Lincoln, NE, USA. He has published over 100 refereed journal articles, primarily on the topic of multiple objective decision-making and information technology. He is an Associate Editor of *Service Business* and Co-Editor-in-Chief of *International Journal of Services Sciences*. He has made over 100 presentations at international and national conferences on research topics. He has authored 17 books, including *Decision Aids for Selection Problems*, *Introduction to Information Systems Project Management*, and *Managerial Issues of Enterprise Resource Planning Systems,* as well as coauthored the books *Introduction to Business Data Mining*, *Enterprise Risk Management*, *Advanced Data Mining Techniques*, *New Frontiers in Enterprise Risk Management*, *Enterprise Information Systems*, and *Enterprise Risk Management Models*. His current research interests include management information systems, management science, and operations management areas.

Mr. Olson is a Member of the Association for Information Systems, the Decision Sciences Institute, the Institute for Operations Research and Management Sciences, and the Multiple Criteria Decision Making Society, and a Fellow of the Decision Sciences Institute. He was named the Best Enterprise Information Systems Educator by IFIP.



Cuicui Luo received the B.Sc. degree in actuarial science and statistics, and the M.Sc. degree in mathematics finance at University of Toronto, Toronto, ON, Canada. She is currently pursuing the Ph.D. degree in mathematics finance and is a Research Associate with RiskLab at the same university.

Her current research interests include financial market analysis and decision analysis.