YI ZHAO, YING ZHAO, and INSEONG SONG*

Recent studies in marketing have consistently shown that all customers are not equally profitable. In the credit card business, all customers are not equally risky. When a customer misses one payment on a credit card bill, a signal is sent to the credit card company. It is important for the card issuer to interpret the signal and to identify whether the customer is a low-risk one, who will eventually pay back the debt and contribute to the card issuer's profits by paying interest on the overdue balance, or a high-risk one, who will not pay back the debt. The issuer can then customize its policies to deal with these different consumer types. This article develops a dynamic model for debt repayment behavior of new customers in the credit card market that makes it possible to differentiate between low-risk, delinquent customers and high-risk customers. The authors apply the model to a data set of new consumers' monthly spending and repayment records.

Keywords: state space model, risk prediction, Bayesian estimation, market segmentation, customer relationship management

Predicting New Customers' Risk Type in the Credit Card Market

Credit cards are the second most popular noncash payment instrument in the United States, and their popularity is growing around the world. They are a convenient payment method whereby consumers can purchase the product first and pay later. However, credit card lending is risky for card issuers because the loans are usually not secured by any assets. Furthermore, unlike traditional loans, which are discrete, typically involve an individual analysis of credit risk, and have a specific maturity date, credit cards invite a continuous flow of borrowing with limited subsequent checks of financial status after the initial issuance of the card. There is also information asymmetry in the credit card market in the sense that the borrowers know their own ability and willingness to repay the debt better than the card issuers. Given the risk associated with credit card lending, it is important for card issuers to identify consumer risk types as early as possible to prevent risky consumers from borrowing too much before default occurs and to customize their marketing strategies to different customer groups. This article provides a model that predicts the risk type of new customers using their initial card usage data.

The credit card industry categorizes cardholders into three segments. First, approximately 40% of cardholders pay their balances in full each month; these cardholders contribute to the revenue of the card issuer in the form of interchange fees and merchant discounts that the issuers charge to the merchants. These fees and discounts accounted for 15% of revenue for the average issuer in 2001 (Evans and Schmalensee 2005). Second, there are consumers whose accounts are "charged off"; at this point, the card company writes off the card balance as bad debt. The bank card industry average for charge-offs fluctuates, with a low of approximately 4% of receivables in May 2000 and a peak of nearly double that in March 2002 (www.standardandpoors.com). Third, the largest segment—approximately 60%—is known as revolvers (http://www.cardweb.com/cardlearn/stat.html); these cardholders typically carry a balance on their cards. This segment includes consumers with a high risk of default, who exhibit payment behaviors suggestive of someone struggling with debt. The credit card industry's best customers are among the revolvers. They borrow at high interest rates, but they eventually (in most cases) repay their loans. A card customer is usually required to pay only 5% of the outstanding amount, and the card issuer charges interest on the amount that is rolled over to the subsequent periods. Interest on

*Yi Zhao is a Fellow of the Center for Excellence in Brand & Customer Management, J. Mack Robinson College of Business, Georgia State University (e-mail: mktyizhao@gsu.edu). Ying Zhao is Assistant Professor of Marketing (e-mail: mkyzhao@ust.hk), and Inseong Song is Associate Professor of Marketing (e-mail: mksong@ust.hk), Department of Marketing, Hong Kong University of Science and Technology. Ying Zhao and Inseong Song acknowledge support from the Research Grants Council of the Hong Kong Special Administrative Region, Hong Kong, China (Project No. HKUST 6461/05H). Sunil Gupta served as associate editor for this article.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

these rollovers contributes the maximum revenue slice. Indeed, on average, finance charges account for 70% of card revenue (Evans and Schmalensee 2005). Consumers who fully intend to borrow on their credit card accounts are not ideal customers for the card company. They have bad credit risk, borrow large sums, and often default.

Recent studies in marketing have consistently found that not all customers are equally profitable. According to the literature on customer management and lifetime value analysis, for firms to be profitable in the long run, they must either convert unprofitable customers to a profitable status or "fire" them (Blattberg, Getz, and Thomas 2001; Gupta et al. 2006; Gupta and Lehmann 2002, 2005; Rust, Lemon, and Zeithaml 2004). However, a better customer management strategy would require the firm to identify the more profitable customers from the less desirable ones. In this article, we develop a model that enables card issuers to identify the risk types of their customers by studying their repayment behaviors using a data set that includes the inception of the consumers' credit histories with the card issuers. Risk type identification is an important managerial issue for the credit providers. For example, the subprime mortgage problem has gone beyond the financing firms to have a significant impact on the entire U.S. economy and the world economy as well. Credit card debt is no less severe than mortgage debt if the risk of credit debt is not well controlled. Our approach enables card issuers to identify low-risk and profitable customers separately from high-risk customers so that they can customize their marketing strategies to encourage low-risk customers to borrow while screening out high-risk customers at the earliest stage possible to reduce losses.

A major issue in this article in terms of identifying consumer types is that low-risk consumer can be delinquent as a result of non-risk-related reasons, such as consumer oversight. The low-risk but occasionally delinquent consumer segment is a good source of revenue for credit card companies because these consumers pay the interest on the overdue amount and will eventually pay off their debts. Mislabeling these low-risk and profitable consumers as part of the high-risk group and imposing an unfavorable credit policy on them simply because of occasional delinquency would decrease the credit card company's revenues. Therefore, a model that provides a way to differentiate between low-risk but occasionally delinquent consumers and high-risk consumers would help credit card companies improve their profits.

To identify high-risk consumers at the earliest stage possible, it is important for the card issuer to use the spending and repayment data from the first month when a consumer opens a credit card account with the company. An issue that we need to consider is that it takes some time for a new customer to become familiar with a new card, and as a result, his or her behavior may evolve over time.[1] The marketing literature suggests that familiarity, defined as the number of product-related experiences the consumer has accumulated, affects consumer behavior. For example, simple repetition improves task performance by reducing cognitive effort. A consumer's ability to analyze information and elaborate on given information improves as familiarity increases (Alba and Hutchinson 1987). In the context of credit cards, we expect that consumers become better at processing information related to the usage of the new cards and perceive less uncertainty and risk in using the new cards compared with other familiar payment instruments, such as cash, checks, and other existing credit cards, as they accumulate more experience with the new card. That is, a new customer's spending and repayment behaviors are likely to evolve gradually over time as a result of his or her increased knowledge of the features of the new card and the possibility that this learning process will require experience with spending and repayment with the card over time. Thus, we need a dynamic model that captures such evolutionary behaviors. Given the possible survival bias of examining data from a stationary period only, we need to model consumer behavior from the inception of card use.

To attain this goal, we incorporate time-varying parameters using a state space specification in a Type II Tobit model. State space models have been used in various marketing modeling areas, such as brand choice modeling (Akcura, Gonul, and Petrova 2004; Kim, Menzefricke, and Feinberg 2005; Lachaab et al. 2006), market structure modeling (Van Heerde, Mela, and Manchanda 2004), sales forecasting (Neelamegham and Chintagunta 2004), and advertising effect analysis (Bass et al. 2007). A distinctive feature of our model is that the observation equations are built on a Tobit II structure. We treat the discrete variable of whether a consumer is delinquent in each month and the continuous variable of the amount paid in each month, conditional on not being delinquent, as two separate but possibly correlated observations, and we jointly model the two aspects of debt repayment behavior using the Type II Tobit model framework. Furthermore, unlike the typical Tobit model, our model separates the repayment amount, conditional on consumers being delinquent, into two groups: an actual zero resulting from consumers' inability to pay back the debt and a censored zero resulting from other factors, such as consumer oversight. This unique feature of the model enables us to capture the possibility that an observed consumer nonpayment can be due to either consumer risk or other factors, such as consumer oversight.

We organize the rest of this article as follows: Next, we describe the data used in this study. Then, we propose a state space model of consumer debt repayment behavior in the credit card market. We then discuss the estimation results. Finally, we conclude with managerial implications and suggestions for further research.

THE DATA

A Hong Kong bank provided the data set used in this study. Our sample includes information from credit applications and monthly statements of 1500 cardholders from January 2000 through August 2002. All the cardholders in our sample opened their accounts in or after January 2000, ranging from January 2000 to June 2001. Although the data do not start and stop for every customer at the same time, our data include the inception of each consumer's history with the company. All the cardholders in the sample have

[1]In Hong Kong, where the data set for this study is from, there was no central credit bureau through which the credit card companies share the data. Credit card companies did not share any data for the data period. Thus, the managers do not have a clear way of knowing whether the customer is new to the category or only to the company.

only one account with the company and do not hold credit cards with the five other major card companies.[2] The data set records cardholder delinquency, purchases, cash advances and repayment histories, credit limits, and interest rates for late payments and cash advances. In addition, the data set provides information on consumers' demographics, such as income, educational level, years of employment, and residential status, at the time they filled out the application form. Following Steckel and Vanhonacker (1993), we used two-thirds of the sample for the estimation and the remaining one-third as a holdout sample for validation purposes. That is, the calibration sample contains 1000 consumers, and the validation sample includes 500 consumers. We used the observations from the first 12 months of the calibration sample's history to estimate the model.

We present summary statistics of the consumers' demographic variables and the credit card–related variables in Table 1, Panel A. Consumer monthly spending and repayment variables appear in Table 1, Panel B. The average annual income for the customers in our sample was approximately HK$209,000 (HK$1 = US$.13), 48% of the customers owned their residence, the average number of years of employment was 6.87 years, and the majority of the cardholders did not have a college degree. The average credit limit is HK$36,720, and the average credit card annual interest rate is 27%. The delinquency cases accounted for 9% of the sample. The average monthly total balance for the accounts in our sample was HK$7,820, and the mean monthly expense was HK$2,330. In addition, the average amount of monthly cash advance was HK$310, and the average amount of monthly debt repayment was HK$2,080. Note that in the model part, we use the ratio instead of the absolute values of the variables to remove the possible heteroskadesticity. The average ratio of the monthly debt repayment to the total balance was .74, with the maximum ratio value of 2.84. On average, the total balance on an account was 23% of the available credit limit. The average percentage of the total balance in a month that was due to the expense occurring in that month was 75%. On average, 4% of the monthly expense was due to a cash advance. Table 2 shows how the cardholders cleared their balances. In our sample, 27.32% of the cardholders cleared their balances each month, and another 21.45% cleared their balances most of the time, though they carried a balance over less than 10% of the months. The sum of the size of these balance-clearing consumer groups is 48.77%, which is comparable to the U.S. figure of 40%. We also observe that rollovers are evenly distributed, though there is a large (17.13%) group of frequent, heavy rollovers; these cardholders carried balances over almost every month.

[2]This information was not available to the card issuer during the data period. We verified it by cross-checking data from several banks.

Table 1
SUMMARY STATISTICS

A: Consumer- and Card-Specific Variables				
	M	*SD*	*Minimum*	*Maximum*
Education (0 for secondary or below and 1 for otherwise)	.22	.41	.00	1.00
Residence status (1 for owned housing and 0 for others)	.48	.50	.00	1.00
Years of employment	6.87	6.17	1.00	29.00
Annual income (HK$100,000)	2.09	1.18	.44	8.51
Credit limit (HK$1,000)	36.72	20.50	5.00	99.00
Card age (months)	12.45	8.14	1.00	32.00
Interest rate	.27	.03	.24	.30

B: Spending and Repayment Variables				
	M	*SD*	*Minimum*	*Maximum*
Delinquency (1 if delinquent and 0 if otherwise)	.09	.28	.00	1.00
Total balance (HK$1,000)	7.82	13.32	–7.54	94.35
Expense (HK$1,000)	2.33	4.35	.00	82.87
Cash advance (HK$1,000)	.31	2.22	.00	80.00
Actual payment (HK$1,000)	2.08	4.05	.00	103.36
Actual payment/total balance	.74	.41	.00	2.84
Total balance/credit limit	.23	.34	–.48	1.18
Expense/total balance	.75	.40	.00	1.00
Cash advance/expense	.04	.17	.00	.97

THE MODEL

Our model consists of two equations: the observation equations and the state equations. The observation equations specify a discrete observation of whether a customer is delinquent in a given month and a continuous observation of how much is repaid, conditional on the customer not being delinquent. If a customer does not meet the minimum payment requirement, he or she is considered delinquent. Both outcome variables are specified as functions of explanatory variables and individual-specific, time-varying parameters. The state equations describe the nature of the dynamics of the parameters in the observation equations.

Observation Equations

The delinquency equation. We define a discrete variable, y_{1it}, for cardholder i in month t that is equal to 1 if cardholder i fails to meet the minimum payment requirement in month t and is equal to 0 if otherwise. We assume that the discrete variable, y_{1it}, is a function of a continuous latent variable, y^*_{1it}, as follows:

$$y^*_{1it} = X_{1it}\gamma_{1it} + \varepsilon_{1it}; \tag{1}$$

$$y_{1it} = 1 \text{ if } y^*_{1it} > 0, \text{ and } \quad y_{1it} = 0 \text{ if } y^*_{1it} \leq 0,$$

where X_{1it} is a row vector ($1 \times k_1$) of the explanatory variables associated with y^*_{1it}, γ_{1it} is a column vector ($k_1 \times$

Table 2
DISTRIBUTION OF BALANCE-CLEARING TENDENCY

Ratio	*Distribution (%)*
.00	27.32
.00~.10	21.45
.10~.20	10.54
.20~.30	4.50
.30~.40	2.89
.40~.50	2.49
.50~.60	2.20
.60~.70	2.41
.70~.80	2.46
.80~.90	3.63
.90~.95	2.98
.95~1.00	17.13

Notes: Ratio = number of months not clearing balance/number of total months.

1) of the cardholder's specific and time-varying parameters, and ε_{1it} is a random term that captures information unobserved by the researchers.

The repayment equation. The amount that a consumer repays in month t is a positive number if the consumer is not delinquent in month t. If a consumer is delinquent in month t, the failure to meet the minimum repayment requirement could be due to the high risk of the consumer or to something else, such as the oversight of a low-risk consumer who would have made the repayment if he or she had remembered to pay the bill. To capture this, we allow for the possibility that a cardholder is capable of paying back the debt but misses the payment because of non-risk-related reasons, and we explicitly model this.

As we discussed previously, we use the ratio of repayment in our model to remove the possible heteroskedasticity. We denote $y_{2it} = AP_{it}/TB_{it}$, where AP_{it} is the actual repaid amount by consumer i in month t and TB_{it} is the total balance of consumer i at time t. However, as we discuss subsequently, we also include the total balance information in the set of explanatory variables. We denote y^*_{2it} as the latent payment index and y_{2it} as the observed payment ratio. We model the latent repayment index y^*_{2it} as a function of the explanatory variables, X_{2it}. Specifically,

$$y^*_{2it} = X_{2it}\gamma_{2it} + \varepsilon_{2it}; \tag{2}$$

if $y^*_{2it} \leq 0$, then $y_{2it} = 0$ or, equivalently, $y_{1it} = 1$; if $y^*_{2it} > 0$, then $y_{2it} = y^*_{2it}$, with a probability of $1 - \delta$; and $y_{2it} = 0$ or, equivalently, $y_{1it} = 1$, with a probability of δ, where δ is the probability that the delinquency comes from a consumer who has the ability but forgets to pay, X_{2it} is row vector $(1 \times k_2)$ of the explanatory variables associated with y_{2it}, γ_{2it} is a column vector $(k_2 \times 1)$ of the cardholder's specific time-varying parameters, and ε_{2it} is a random term that captures information unobserved by the researchers. Note that y_2 is observed when $y_1 = 0$ in our model, unlike typical Tobit specifications. We use this particular specification to denote y_1, the delinquency outcome. In addition, unlike the typical Tobit model, which would treat the repayment of all the delinquent consumers as censored data, the proposed model assumes that the repayment ratio conditional on consumers being delinquent can be categorized into two groups: an actual zero resulting from consumers' inability to pay back the debt and a censored zero resulting from other factors, such as consumer oversight. We assume that the error terms of Equations 1 and 2 are normally distributed and possibly correlated. That is,

$$\begin{pmatrix} \varepsilon_{1it} \\ \varepsilon_{2it} \end{pmatrix} \sim N(0, \Omega).$$

For identification purposes, the variance of ε_{1it} is set to 1 because the first component deals with the latent variable in a binary choice. Therefore, we model Ω as follows:

$$\Omega = \begin{pmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

The set of covariates we use in both equations are INT, $TB_{it}/CRLIMIT_{it}$, $EXPENSE_{it}/TB_{it}$, and $CSHADV_{it}/EXPENSE_{it}$, where INT is the intercept, TB_{it} is the total balance for consumer i in month t, $CRLIMIT_{it}$ is consumer i's credit limit in month t, $EXPENSE_{it}$ is the expense for consumer i in month t, and $CSHADV_{it}$ is the amount of cash advance for consumer i in month t. These variables represent all the information a credit card company observes about its customers' monthly activity, and credit card companies typically use these variables to conduct customer risk analysis.

There is a large dispersion of each of the aforementioned variables across observations. For example, the coefficients of variation for TB, EXPENSE, and CSHADV are 1.70, 1.87, and 7.16, respectively. To remove the possible heteroskadesticity, we use the ratio instead of the absolute values of the variables. Furthermore, the ratio variables are commonly used in the finance and accounting literature streams because these ratios reflect a customer's risk level to some degree.[3] For example, the ratio of $TB_{it}/CRLIMIT_{it}$ is typically used as an indicator of the risk level of a cardholder, such that a value approaching 1 suggests that the customer is more likely to be high risk. The ratio of $EXPENSE_{it}/TB_{it}$ indicates the percentage of the total balance in a given month that is due to the expense occurring in that month for a consumer, and $1 - EXPENSE_{it}/TB_{it}$ is the percentage of the total balance that is due to finance charges, interest, or other penalties from failing to clear the balance from the previous month, which could be closely related to the customer's debt repayment ability. The ratio of $CSHADV_{it}/EXPENSE_{it}$ reflects the percentage of the monthly expense that is due to a customer's cash advance in a given month. In summary, these ratios capture a customer's financial risk status to some degree. In this study, we examine how these ratios are correlated with the debt repayment behavior of cardholders.

State Equation

The state equation describes how the parameters of the observation equations evolve over time. We use the following state equation to model the evolution of consumer specific parameters:

$$\gamma_{it} = \theta_i\gamma_{i,t-1} + \omega z_{it} + h_i + \xi_{it}, \tag{3}$$

where γ_{it} is a column vector that combines γ_{1it} and γ_{2it} and θ_i is a diagonal matrix (i.e., $\theta_i = \text{diag}[\theta_{i1}, ..., \theta_{ik}]$). We use a multivariate normal distribution to model consumer heterogeneity as follows:

$$(\theta_{i1}, \ldots, \theta_{ik})^T \sim N(\overline{\theta}, V_\theta),\ \gamma_{i1} \sim N(a_1, p_1),\ h_i \sim N(\overline{h}, V_h), \text{ and}$$

$$\xi_{it} \sim N(0, \Sigma),$$

where $\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_k)$ and $k = k_1 + k_2$.[4] In the model, θ_{ik} is an individual-specific autoregressive parameter that represents the amount of carryover effects in γ_k for each customer; γ_{i1} is a column vector that represents the initial state of the parameters; z_{it} is a p-dimensional vector, including consumer-specific and card-specific variables; and ω is a $k \times p$ matrix of parameters that captures observed heterogeneity and the effects of the possibly time-varying credit terms. The consumer-specific variables that

[3]See, for example, Altman (1968), Zmijewski (1984), and Domowitz and Sartain (1999).

[4]We also estimated a model in which Σ was a full matrix. However, we find that the model with the full matrix of Σ does not improve model fit and prediction.

enter the state equation are the cardholders' income, years of employment, educational level, and residential status (if they own their current home).[5] We include these variables to capture the observed consumer heterogeneity in their sensitivity to the X variables. The card-specific variables we use in the estimation are the age, interest rate, and credit limit of the card.[6] The term h_i is a k-dimensional column vector that captures the unobserved heterogeneity in the steady-state mean. The inclusion of the consumer- and time-specific random term, ξ_{it}, allows the evolution process to be probabilistic rather than deterministic.

The proposed model also provides a flexible structure to accommodate the possibility of parameter variation over time, as well as observed and unobserved consumer heterogeneity. In our model specification, when θ is significantly different from zero and/or when the standard deviation of ξ is significantly different from zero, our model is a dynamic model that captures parameter variation over time. When θ equals zero and the variance of ξ_{it} equals zero, our model becomes a Tobit II model.

Model Estimation

We estimate the model parameters using a hierarchical Bayesian approach. We use the Markov chain Monte Carlo method to generate the posterior distributions. A total of 40,000 draws were made from the series of the full conditionals by Gibbs sampling. We checked convergence by monitoring the time series of the draws. We discarded the initial "burn-in" of 30,000 draws and kept the last 10,000 draws to make our inferences. We provide detailed descriptions on our priors and full conditionals in the Web Appendix (http://www.marketingpower.com/jmraug09).

EMPIRICAL RESULTS AND DISCUSSIONS

In this section, we discuss some of the estimation results of the proposed model and compare the model fits and the in-sample and out-of-sample predictive ability of the proposed model with three benchmark models. We also compare the proposed model with a naive industry model in terms of their respective abilities to identify consumer risk types. We then discuss the potential benefits of our model to the credit card issuers.

Parameter Estimates

The estimation results appear in Table 3. We present the averages and the standard deviations of simulated draws from the posterior distributions. Our estimation results support the evolutionary aspect of consumer repayment behavior. That is, we find evidence that parameters indeed change over time. In our model, there are three sources for parameters to change over time. First, there are decaying carryover effects, which are represented by the autoregressive term, θ_i. If this term is not equal to one, the consumer-specific coefficients at time t can be systematically different from their previous values. Second, there are effects of time-varying variables, which are captured by the term ω. Although the demographic variables are constant over time for a given consumer, the credit terms can be changed by the card issuer. Third, there is a consumer- and time-specific random term, ξ. A significant variance for ξ represents the portion of the dynamics that is not captured by the covariates or by the decaying carryover.

Table 3, Panel A, reports the impact of time-varying card-specific variables and the consumer-specific demographic variables on the consumer-specific coefficients in the delinquency and the repayment equations. Among the demographic variables, the number of years of employment is negatively related to the intercept in the delinquency equation. Income is positively related to the effect of TB/CRLIMIT on delinquency. An increase in TB/CRLIMIT (or, equivalently, an increase in the proportion of total balance in the credit limit) would increase the delinquency probability for high-income customers more than for low-income customers.

For the credit card–specific variables, we find that only the credit limit variable is related to the coefficients for delinquency outcomes. The effect of credit limit on the coefficient for CSHADV/EXPENSE is positive. An increase in CSHADV/EXPENSE would imply larger probabilities of delinquency for customers with large credit limits than for customers with small credit limits. This suggests that an unusually high proportion of cash advance in the credit card bill could be a premonition of delinquency, particularly for customers with large credit limits. This result is intuitive. A consumer with a large credit limit tends to have high income and to be less likely to resort to a credit card to withdraw cash. Therefore, a large proportion of cash advance over the total expense in a month in the credit card bill is more likely to be a stronger indicator of a risky financial status for a consumer with a large credit limit than for a consumer with a small credit limit.

When it comes to the coefficients in the repayment equation, we find no significant effect of the residence status and the number of years of employment on the customer-specific coefficients. Among demographic variables, we find that more educated people and/or people with higher incomes tend to have larger repayment ratios. Income is significantly related to the coefficient of EXPENSE/TB. When EXPENSE/TB is large, people with higher incomes tend to repay less than others. In other words, a larger proportion of financial charge in the credit card bill (or a smaller value of EXPENSE/TB) is related to a higher repayment ratio for customers with higher incomes. Education is significantly related to the coefficient of CSHADV/EXPENSE, suggesting that when CSHADV/EXPENSE is high, people with higher education levels tend to repay more than those with lower education levels. Among the card-specific variables, card age is the only variable that has significant correlation with the repayment ratio. The result shows that the longer a customer is with the bank, the higher the repayment ratio is. This is consistent with previous findings in the literature that accounts may become less likely to default/delinquent as they age (Gross and Souleles 2002).

[5]The Equal Credit Opportunity Act (15 U.S.C. § 1691 et seq.; implemented by the Federal Reserve Board's Regulation B) prohibits creditors from discriminating in any aspect of a credit transaction on the basis of an applicant's race, color, country of origin, gender, marital status, or age. Therefore, we do not use data on the consumers' age, gender, and marital status in our model estimation.

[6]The card age at time t is specified as the number of months a card has been held by the consumer at time t.

Table 3
PARAMETER ESTIMATES

A: The Coefficient Matrix ω

	Delinquency Equation				*Repayment Equation*			
	INT	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*	*INT*	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*
EDU	–.1498	.0544	.1275	–.0307	.0402	.0318	–.0160	.1209
	(.0947)	(.1126)	(.1004)	(.1118)	(.0167)	(.0342)	(.0141)	(.0529)
RESID	–.0005	.0745	–.0102	.0282	.0162	.0153	.0057	.0184
	(.0514)	(.0530)	(.0634)	(.1059)	(.0164)	(.0278)	(.0127)	(.0324)
YRE	–.0330	.0243	.0208	–.0013	–.0020	.0006	.0026	.0006
	(.0122)	(.0129)	(.0134)	(.0155)	(.0017)	(.0025)	(.0016)	(.0030)
INCOME	–.0019	.1040	–.0172	–.0014	.0220	–.0002	–.0208	–.0207
	(.0423)	(.0493)	(.0434)	(.0632)	(.0083)	(.0147)	(.0075)	(.0192)
CRLIMIT	.0054	.0041	–.0161	.0408	.0019	–.0018	–.0016	–.0008
	(.0093)	(.0118)	(.0103)	(.0128)	(.0015)	(.0024)	(.0013)	(.0040)
CARDAGE	–.0997	–.0563	–.2476	–.0098	.1354	–.1132	.0526	–.2152
	(.1320)	(.1747)	(.1985)	(.1979)	(.0555)	(.2529)	(.0429)	(.1548)
INTEREST	–.0027	.0015	.0016	–.0011	–.0006	–.0014	.0004	.0016
	(.0039)	(.0050)	(.0042)	(.0041)	(.0005)	(.0009)	(.0005)	(.0012)

B: The Autoregressive Parameter θ and the Variance of Error Term in the State Equation

	Delinquency Equation				*Repayment Equation*			
	INT	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*	*INT*	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*
$\bar{\theta}$	.4449	.3525	.4116	.3988	.1516	.4819	.2283	.3510
	(.0313)	(.0535)	(.0505)	(.0512)	(.0118)	(.0166)	(.0182)	(.0467)
V_θ	.0250	.0168	.0220	.0208	.0031	.0372	.0078	.0171
	(.0034)	(.0047)	(.0050)	(.0051)	(.0004)	(.0029)	(.0013)	(.0046)
λ	.0256	.0319	.0460	.0393	.0029	.0951	.0020	.0424
	(.0116)	(.0133)	(.0233)	(.0248)	(.0002)	(.0071)	(.0001)	(.0066)

C: The Heterogeneity Term and the Variance of Error Term in the State Equation

	Delinquency Equation				*Repayment Equation*			
	INT	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*	*INT*	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*
$\bar{h}$	–.5957	–.0371	–.2070	–.0344	.6540	–.1811	.0895	–.0470
	(.0781)	(.0498)	(.0701)	(.0541)	(.0214)	(.0816)	(.0143)	(.0454)
V_h	.0894	.0152	.0363	.0111	.0130	.0469	.0039	.0122
	(.0210)	(.0104)	(.0160)	(.0066)	(.0011)	(.0073)	(.0005)	(.0032)

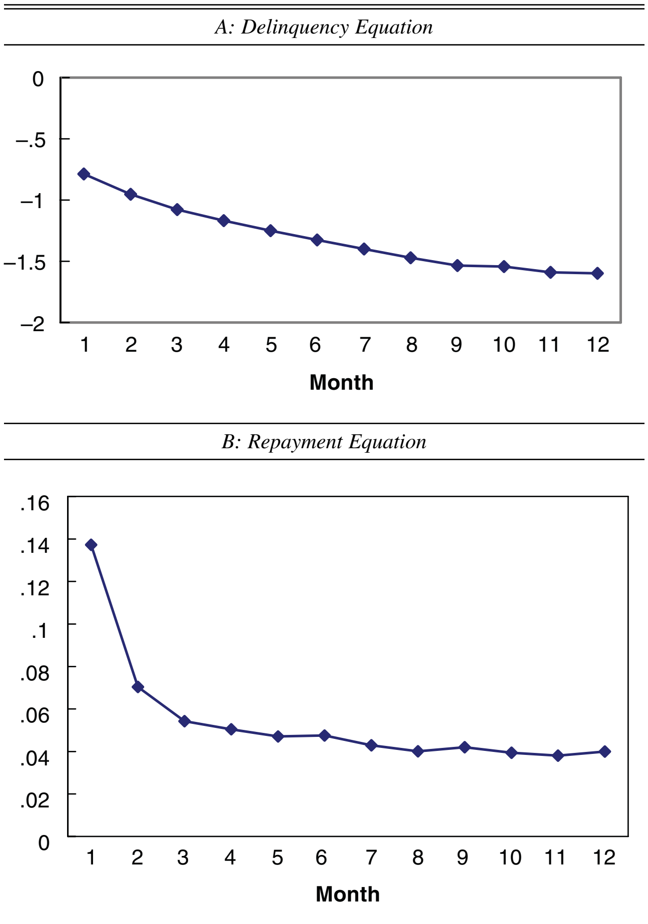
D: The Initial States

	Delinquency Equation				*Repayment Equation*			
	INT	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*	*INT*	*TB/CRLIMIT*	*EXPENSE/TB*	*CSHADV/EXPENSE*
α_1	–1.2287	.2281	–.5357	.4229	.6935	–.5859	.3090	–.5398
	(.2810)	(.1740)	(.2828)	(.1996)	(.0339)	(.0609)	(.0337)	(.0672)
p_1	.3609	.1573	.1816	.1829	.0223	.5549	.0187	.1671
	(.1489)	(.1787)	(.1131)	(.1209)	(.0018)	(.0943)	(.0017)	(.0396)

In Table 3, Panel B, we present the estimation results for the autoregressive parameter, θ, that describes the extent of carryover effects in the consumer-specific coefficients, γ. A smaller value of θ indicates carryover decay, suggesting that the value of γ in the current period would be different from its previous value and that the γ in the current period is more easily affected by the outside variables than by the γ in the previous period. We find that the values of the autoregressive parameters are smaller than one for all coefficients in both the delinquency and the repayment equations, in support of the time-varying nature of the parameters. We also find that consumers are heterogeneous in terms of the carryover parameters from the estimates of V_θ. Finally, we find that the carryover effect can differ across variables. For example, the carryover parameter in the coefficient for EXPENSE/TB in the repayment equation is clearly different from the others.

In Figure 1, we plot a cardholder's sensitivity to the EXPENSE/TB variable in the delinquency equation and in the repayment ratio equation as an illustration. We observe that the parameters indeed change over time and that the magnitude of the changes is significant. We also observe that the effect of the EXPENSE/TB variable on the repayment ratio quickly converges to a stable level but that the convergence occurs more slowly for that parameter in the delinquency equation. Such a difference in convergence speed can be explained by the difference in the carryover effect parameter, θ.

Figure 1
CARDHOLDERS' SENSITIVITY TO THE EXPENSE/TB VARIABLE



The last source for parameter dynamics among the three aforementioned sources for parameters turns out to be a significant source for parameter dynamics. As we show in Table 3, Panel B, most of the variances of the ξ terms are significant, implying that there are some unobserved within-consumer time-varying shocks contributing to the parameter dynamics, which are not explained by the changes in card terms over time.

In Table 3, Panels C and D, we report the estimates of the mean and the variance of consumer-specific level shifters (h_i) and the estimates of the distribution of the initial state in the state equation. The significance of the variance estimate (V_h) indicates that there are unobserved across-consumer heterogeneity in addition to the unobserved heterogeneity explained by demographics and card terms. We also find that consumers are heterogeneous in the initial condition from the significant estimates of p_1. The probability (δ) of having a missed payment as a result of factors such as consumer oversight is estimated to be .05. This suggests that in approximately 5% of the cases, an observed nonpayment is due to factors such as consumer oversight. Finally, we report the estimate of the variance of error terms in the observation equation. Because the delinquency equation deals with a binary variable, we normalize the variance of ε_{1it} to one. Thus, we report the variance of ε_{2it} and the correlation between ε_{1it} and ε_{2it}. The correlation of the error terms between the delinquency equation and the repayment equation is significantly negative. This suggests that the higher the repayment ratio, the lower is the probability that an account will be delinquent, which makes intuitive sense.

Prediction and Model Comparison

To assess the importance of accounting for consumer heterogeneity and parameter dynamics in fitting the data, we compare the fit of the proposed model with those from three benchmark models. These benchmark models are (1) a model with neither consumer heterogeneity nor dynamics (Benchmark Model 1), which is a simple Type II Tobit model; (2) a model with consumer heterogeneity but without dynamics ($\theta = 0$, and $\Sigma_\xi = 0$) (Benchmark Model 2); and (3) a model with dynamics but without consumer heterogeneity (Benchmark Model 3).

We compare the in-sample fit and the holdout prediction ability. We use three sets of holdout samples: (1) same people, later months; (2) different people, same months; and (3) different people, later months. As we noted previously, the estimation data consist of the first 12 months of observations from 1000 consumers. Thus, the first holdout set consists of the observations from the 13th to 15th month for the 1000 consumers whose first 12 months of observations are used for parameter estimation. The second holdout set consists of another 500 consumers' first 12 months of observations. Finally, the third set consists of the next three months of observations of the 500 consumers. By having various sets of holdout samples, we can determine how dynamics/heterogeneity is related to the predictive ability in different settings.

Using the parameter estimates, we predict delinquency probability and the repayment ratio in month t using consumers' demographic information, consumers' repayment information from month 1 to month t – 1, and consumers' spending information from month 1 to month t for the holdout sample. For the same-people/later-months holdout sample and the different-people/later-months holdout sample, we made predictions about the delinquency probability and the repayment ratio for the next consecutive three months. To make predictions for the different-people/same-months holdout sample, we calculated the delinquency probability and the repayment ratio for the 500 customers in the holdout sample for their first 12 months, starting from inception.

In Table 4, we present the in-sample fit along with the holdout fits both from the proposed model and from the benchmark models. The mean square errors suggest that the proposed model, which takes into account both consumer heterogeneity and dynamics, performs the best in predicting the delinquency probability and repayment ratios across all the samples, followed by the model with dynamics but without consumer heterogeneity and then by the model with consumer heterogeneity but without dynamics. The model with neither dynamics nor consumer heterogeneity is the worst at making predictions. This suggests that both the consumer heterogeneity and the dynamic components of the model contribute to the better prediction of the proposed model compared with the null models, which include either or none of the two components.

The finding that the benchmark model with dynamics and without heterogeneity performs consistently better than the benchmark model with heterogeneity but without

Table 4

FIT COMPARISON BETWEEN THE BENCHMARK MODELS AND THE PROPOSED MODEL (SUM OF THE SQUARED ERRORS)

			Benchmark Model 1	*Benchmark Model 2*	*Benchmark Model 3*	*Proposed Model*
Same people, same times		MSE1	.1050	.0785	.0706	.0583
		MSE2	.0434	.0296	.0265	.0015
Same people, later times	Next 1 stage	MSE1	.1072	.1113	.0717	.0704
		MSE2	.0325	.0317	.0211	.0185
	Next 2 stage	MSE1	.1062	.1115	.0685	.0679
		MSE2	.0361	.0324	.0241	.0218
	Next 3 stage	MSE1	.1050	.1143	.0681	.0669
		MSE2	.0350	.0322	.0243	.0228
Different people, same times		MSE1	.1132	.0841	.0770	.0679
		MSE2	.0484	.0349	.0392	.0195
Different people, later times	Next 1 stage	MSE1	.1015	.1183	.0940	.0872
		MSE2	.0337	.0326	.0257	.0241
	Next 2 stage	MSE1	.1001	.1212	.0913	.0886
		MSE2	.0335	.0384	.0297	.0275
	Next 3 stage	MSE1	.1125	.1176	.0865	.0838
		MSE2	.0341	.0422	.0298	.0286

Notes: $\text{MSE1} = \frac{1}{n}\sum_{i=1}^{n}|\,y_{i1} - \Phi(\hat{y}_{1i})\,|^2$, and $\text{MSE2} = \frac{1}{n}\sum_{i=1}^{n}|\,\exp(y_{2i}) - \exp(\hat{y}_{2i})\,|^2$.

dynamics in making predictions suggests that the dynamic component of the model plays a more important role than the heterogeneity part. This is due to the flexible nature of the models that allow for parameter dynamics, which can pick up the time-varying factors that affect a consumer's behavior during the data period, whereas the models that do not allow parameters to change over time pick up only the average effect, which is averaged across the periods for which the calibration samples are observed. If cardholder behaviors indeed change over time, such averaging across the estimation periods—the first 12 months in our case—will result in poor predictions for the holdout periods because of the systematic difference between the parameter estimates and the true future parameters. In particular, the proposed model first predicts the consumer-specific parameters for the holdout periods using state equation (Equation 3) and then predicts the behavioral outcomes based on the new parameters using the observation equations (Equations 1 and 2). Incorporating dynamics is particularly important for predicting future behaviors when consumer behaviors indeed change over time, and our results also indicate that the dynamic aspect needs to be incorporated into our data set. Given the nature of credit card lending management, managers must identify consumer risk types as early as possible and thus need to use data from the inception of a cardholder's history with the company. Such a situation requires managers to take into account time-varying behaviors because consumers are likely to show transient behaviors in early periods. Because the model that takes into account both consumer heterogeneity and the time-varying parameters performs better than the two baseline models, we focus our subsequent discussion on the proposed dynamic model only.

Identifying Consumer Risk Type

We now explain how our model can be used to identify low-risk consumers among delinquent consumers. A nice feature of the proposed model is that we can predict the cardholder's repayment ratio ($\hat{y}_{2t}$) given the cardholder's credit card repayment information from month 1 to month t – 1 and his or her spending information from month 1 to month t.[7] In our Type II Tobit model, the value of y_2 for a delinquent consumer captures the repayment a consumer would have made if he or she had chosen not to be delinquent. Our modeling approach enables us to recover this value. We claim that consumers with a low value of $\hat{y}_2$ are riskier than consumers with a high value of $\hat{y}_2$ if both types of consumers are delinquent. In other words, a delinquent consumer who has a low $\hat{y}_{2t}$ is more likely to remain delinquent until t + 2 than a delinquent consumer who has a high $\hat{y}_{2t}$. To verify our claim, we conduct the following exercise using the holdout sample of 500 consumers: Assuming that the time span for a consumer in the data is T months, we calculate $\hat{y}_{2,\,T-2}$ using the consumer's repayment data from month 1 to month T – 3 and the spending data from month 1 to month T – 2. We then calculate the mean of $\hat{y}_{2,\,T-2}$ for three subgroups of consumers: (1) the consumers who are not delinquent in month T – 2, (2) those who are delinquent in month T – 2 but are not delinquent in both month T – 1 and month T, and (3) those who are delinquent for the three months consecutively. We also make predictions on the nondelinquency probability. As a benchmark, we estimate a naive model that is commonly used for predicting consumer risk in the credit card industry. This naive model is a logistic regression model that accounts for consumer heterogeneity to model the probability of delinquency.[8] We also calculate the predicted nondelinquent probability for the three customer groups using the estimation results from the naive model for the holdout sample.

In Table 5, Panel A, we present the predicted repayment ratio ($\hat{y}_{2,\,T-2}$) and the predicted nondelinquency probability ($\hat{y}_{1,\,T-2}$) based on the results from our proposed model

[7]Note that given the observed consumer spending and repayment history and the estimated model parameters, we can predict the repayment ratio ($\hat{y}_{2t}$) for each consumer at each month, thus allowing consumers' risk profiles to evolve over time.

[8]Discriminant analysis such as linear/logistic regression is by far one of the most common quantitative techniques in credit management. It typically uses information on delinquency, account activity, account balances, amount overdue, and age of the account (Rosenberg and Gleit 1994).

Table 5

COMPARISON OF THE IDENTIFICATION ABILITY OF THE INDUSTRY NAIVE MODEL AND THE PROPOSED MODELS

A: Proposed Model 1

		Proposed Model (Repayment Ratio, y_2)		*Proposed Model (Nondelinquent Probability, y_1)*		*Naive Model (Nondelinquent Probability)*	
Type of Consumers	*Number of Consumers*	*M*	*SD*	*M*	*SD*	*M*	*SD*
No delinquency	404	.78	.39	.93	.13	.93	.11
Short-term delinquency (one month)	66	.49	.50	.75	.23	.76	.17
Long-term delinquency (three months)	30	.02	.21	.65	.28	.71	.22

B: Proposed Model 2

		Type I Error			
		10.00%	*20.00%*	*30.00%*	*40.00%*
Type II error	Proposed model	36.37%	33.33%	31.82%	28.79%
	Naive model	81.30%	68.74%	53.57%	46.38%

along with the predicted nondelinquent probability based on the results from the naive model for the three consumer groups. We find that the consumers who are delinquent for three months up to T have the lowest mean of $\hat{y}_{2,\,T-2}$, which is .02. The mean of $\hat{y}_{2,\,T-2}$, is the highest for the consumers who are not delinquent in month T – 2, which is .78. The mean for those who are delinquent at T – 2 but are not delinquent at both T – 1 and T is .49. This suggests that $\hat{y}_2$ is negatively correlated with the risk level of an account that demonstrates long-term delinquency. It also indicates that the delinquent consumers with a large value of $\hat{y}_{2,\,T-2}$ are more likely to be low-risk types because they meet the minimum payment requirement in the subsequent months, while the delinquent consumers with a smaller value of $\hat{y}_{2,\,T-2}$ are more likely to be high-risk types because they tend to become long-term delinquents. Note also that the means of $\hat{y}_{2,\,T-2}$ for the consumers who are not delinquent in month T – 2 and the consumers who are delinquent in month T – 2 but are not delinquent in both month T – 1 and month T are above .5 and are close together, while the mean of $\hat{y}_{2,\,T-2}$ for those who are delinquent for three months consecutively is significantly smaller than .5, and there is a noticeable gap between the means of $\hat{y}_{2,\,T-2}$ from this last group of consumers (those who are delinquent for three consecutive months) and the low-risk consumers (those who are not delinquent in month T – 2, or those who are delinquent in month T – 2 but are not delinquent in both month T – 1 and month T).

What is determined by the industry naive model? Using the results from the industry naive model, we calculated the mean predicted probability of nondelinquency for the three consumer groups. Note that the naive model uses discrete delinquency outcomes, so the prediction can be made for the delinquency probability only. The results show no clear differentiation in the predicted nondelinquency probabilities between the high-risk and the low-risk consumers (.71 versus .76), suggesting that our proposed model performs better than the naive model in identifying customer risk groups. We also calculated the mean predicted probability of nondelinquency using the proposed model for the three consumer types. The means of the nondelinquency probability were .93, .75, and .65, respectively, for the first, second, and third groups of consumers. The difference (.65 versus .75) in the predicted probabilities of nondelinquency between the high-risk and the low-risk consumers based on the proposed model is larger than the difference (.71 versus .76) based on the naive model. However, such a difference in the predicted nondelinquency probabilities is not as noticeable as the difference (.02 versus .49) in the mean predicted repayment ratios ($\hat{y}_{2,\,T-2}$) based on the proposed model. This indicates the importance of using the delinquency and repayment data jointly in predicting consumer risk types.

Subsequently, we further demonstrate the incremental gain of the proposed model in predicting consumer risk types over the industry naive model using the holdout sample. In this exercise, we use a consumer's spending and repayment data from time 1 to time T – 3 and the consumer's spending information at time T – 2 to calculate the predicted repayment ratio, $\hat{y}_{2,\,T-2}$, based on the estimation results from the proposed model, and we use $\hat{y}_{2,\,T-2}$ to classify consumers into high- and low-risk groups. We also use the predicted probability of nondelinquency for each consumer based on the estimation results from the industry naive model to classify consumers into high- and low-risk groups. What cutoffs have been used to classify consumers? Because the scales of the predicted quantities are not comparable between the two models (i.e., the repayment ratio for the proposed model versus the nondelinquency probability for the naive model), we cannot come up with a common cutoff value. Instead, we impose a common level of classification error. For a given cutoff for classification, we can compute the size of the classification error by cross-validating the prediction-based classification with the actual outcome in the data we observe up to time T. For example, it is possible that a consumer is classified as a long-term delinquent consumer by the model but is not delinquent according to the data. Such a case is recorded as a classification error. Note that there are two types of errors. In one case, researchers may misclassify a high-risk customer as a low-risk customer. In another, researchers may misclassify a low-risk customer as a high-risk customer. We label the first error as Type I and the second error as Type II.

For any model, at a particular cutoff value, it is possible to compute the amount of Type I error, the sample probability of misclassifying a high-risk consumer as a low-risk consumer. In other words, at a given value of Type I error, it is possible to find the respective corresponding cutoffs for the proposed model and for the naive model. To make a comparison between two models, we impose a certain level of Type I error. At a particular value of Type I error, we compute the two cutoffs, one for the proposed model and one for the naive model. From the cutoffs, we compute the amount of Type II error, the sample probability of misclassifying a low-risk consumer as a high-risk consumer, for each model. The results appear in Table 5, Panel B. For example, for the cutoffs at which the Type I error is 10%, the amount of Type II error in the proposed model is 36.37%, and that in the naive model is 81.3%. The table shows that the proposed model produces consistently smaller Type II errors at all the levels of Type I error, which suggests that the proposed model outperforms the industry naive model in correctly identifying customer types.

Simulation Exercise

Using $\hat{y}_2$ as a leading indicator for consumer risk levels, managers of credit card companies can develop better card policies to improve profit. Given the revenue structure of card issuers, it is profitable not to block the accounts of delinquent consumers if those consumers are likely to pay the balance in the near future. We present a simulation exercise to show how a credit card issuer can increase its revenue by developing a targeted strategy to its delinquent customers using the information on $\hat{y}_2$. We focus on the potential risky consumers (the delinquent consumers). Consider the following three strategies managers can adopt when a consumer is delinquent: (1) blocking the consumer from using the credit card until he or she meets the minimum payment requirement (the blocking strategy), (2) leaving the consumer alone without doing anything (the nonblocking strategy), and (3) blocking only the high-risk consumers from using their credit cards until they meet the minimum payment requirement (the targeting strategy). The second strategy is what card issuers currently do. Card issuers do nothing unless a consumer fails to meet the minimum requirement for three consecutive months. Our simulation exercise explores the possibility for the card issuer to improve profit by identifying consumer risk types earlier, namely, at the first delinquency.

In this exercise, we first make assumptions about the cash flow structure. We begin with an assumption on a consumer's lifetime contribution to the card issuer's cash flow as follows:

$$\mathrm{LV_i} \equiv \frac{(\mathrm{mexpense_i} \times \text{merchant rate})/.15}{\text{discount rate}},$$

where merchant rate is the rate the credit card company charges the merchants when the company's card is used in the transaction, $\mathrm{mexpense_i}$ is the expected per-month expense for consumer i, and discount rate is the discount rate the credit card company uses to calculate the present value of the future cash flow. The term $\mathrm{mexpense_i} \times$ merchant rate is the expected interchange fee income for the consumer per month. On the basis of the historical figure in the U.S. credit card industry (Evans and Schmalensee 2005), we assume that the interchange fee accounts for 15% of revenue for the issuer. So, the total cash flow for a month from consumer i would be ($\mathrm{mexpense_i} \times$ merchant rate)/.15, and thus the lifetime contribution would be given by $\mathrm{LV_i}$.

When an account is blocked, a high-risk consumer will leave the company without paying back the debt.[9] We assume that a low-risk consumer who happens to be delinquent at time t will pay the full balance at time t + 1. Even if a low-risk consumer eventually repays the full balance, the blocking action might lower the consumer's preference for that card and make him or her stop using the card. Suppose that a low-risk consumer's probability of choosing to stay with the company even when the account is blocked is *p*.

Assuming that consumer i is delinquent at time t but will repay the minimum required amount in the subsequent period (low-risk consumers), the payoff if the account is blocked is as follows:

$$\mathrm{payoff_{1i}} = [\mathrm{TB_i} \times \mathrm{r_i} + \mathrm{TB_i}(1 + \mathrm{r_i}) \times \mathrm{r_i} + \text{late fee}_\mathrm{i} + \mathrm{p} \times \mathrm{LV_i}]/ (1 + \text{discount rate}),$$

where $\mathrm{TB_i}$ is the total amount that consumer i owes at the time of delinquency, $\mathrm{r_i}$ is consumer i's monthly interest rate for the overdue amount, and late $\mathrm{fee_i}$ is the late fee charged for the consumer's delinquency. The last term in the brackets is the expected future lifetime value of the consumer under the blocking action. If the account is not blocked, the payoff for a low-risk consumer is as follows:

$$\mathrm{payoff_{2i}} = [\mathrm{TB_i} \times \mathrm{r_i} + \mathrm{TB_i}(1 + \mathrm{r_i}) \times \mathrm{r_i} + \text{late fee}_\mathrm{i} + \mathrm{LV_i}]/ (1 + \text{discount rate}).$$

The difference between $\mathrm{payoff_{1i}}$ and $\mathrm{payoff_{2i}}$ comes from the chance that the customer will not use the company's card any more if the account is blocked. If consumer i is indeed a high-risk consumer who will eventually default, the payoff from blocking the account is as follows:

$$\mathrm{payoff_{3i}} = -\mathrm{TB_i}.$$

That is, the write-off of the outstanding balance would be the loss. However, if the card issuer does not block the high-risk consumer's account and instead waits for two more months until the consumer is eventually classified as a three-month delinquent consumer, there will be further loss because the consumer will keep using the card for the next two months. Therefore, the payoff for not blocking a high-risk consumer's account is as follows:

$$\mathrm{payoff_{4i}} = -\mathrm{TB_i} - \frac{\mathrm{mexpense_i}(1 - \text{merchant rate})}{(1 + \text{discount rate})} - \frac{\mathrm{mexpense_i}(1 - \text{merchant rate})}{(1 + \text{discount rate})^2}.$$

Suppose that $\mathrm{n_1}$ is the set of consumers who are delinquent only temporarily and that $\mathrm{n_2}$ is the set of consumers

[9]The current setting of the simulation exercise does not allow for the possibility that account blocking can lead to changes of the repayment behavior of the high-risk consumers. We acknowledge this as a possible limitation.

who are long-term delinquents. If the credit card company blocks all these accounts, the revenue is as follows:

$$\text{revenue}_{\text{block}} = \sum_{i \in n_1} \text{payoff}_{1i} + \sum_{i \in n_2} \text{payoff}_{3i}.$$

If the credit card company does not block any of these accounts, its revenue is as follows:

$$\text{revenue}_{\text{nonblock}} = \sum_{i \in n_1} \text{payoff}_{2i} + \sum_{i \in n_2} \text{payoff}_{4i}.$$

Finally, when the credit card company uses a targeted strategy by classifying consumers into low- and high-risk categories using our model, it would block the accounts of the customers who are in the high-risk category until they meet the minimum payment requirement to reduce the risk of further loss and would still generate revenue from the low-risk customers. The credit card company's revenue under the targeted strategy is as follows:

$$\text{revenue}_3 = \sum_{i \in n_3} \text{payoff}_{1i} + \sum_{i \in n_4} \text{payoff}_{2i} + \sum_{i \in n_5} \text{payoff}_{3i} + \sum_{i \in n_6} \text{payoff}_{4i},$$

where n_3 is the set of low-risk consumers who are wrongfully identified as high-risk consumers, n_4 is the set of low-risk consumers who are identified as low-risk consumers, n_5 is the set of high-risk consumers who are identified as high-risk consumers, and n_6 is the set of high-risk consumers who are wrongfully identified as low-risk consumers.

For consumers who are delinquent at T – 2, we first identify the risk type of each consumer using $\hat{y}_2$. We use various cutoff values to classify the consumer type. For each consumer, the values of TB_i, r_i, and late fee_i are available in the data. We also compute the average monthly expense to approximate mexpense_i. We assume that the merchant rate is 3% and that the discount rate for the credit card company is .013. We experiment with several values for *p*, the probability that a low-risk consumer will continue to use a credit card even after he or she experiences account blocking. We compute the revenues for each of the three strategies. We also calculate the revenue under the assumption that the company is able to identify the consumer type with perfect foresight. The revenue under this assumption provides the upper bound of revenue that the firm can obtain.

The results based on the holdout sample of 500 consumers appear in Table 6. The revenue from not blocking any account is HK$639,070. The revenue from blocking all delinquent accounts decreases as the likelihood that a consumer whose account is blocked will stay with the company decreases. The revenue from the targeting strategy is jointly determined by the value of *p* and by the cutoff point at which we choose to classify consumers into high- and low-risk groups. Because the payoff of the targeting strategy is a random variable, we run 5000 simulations to estimate the mean payoff and the standard error. The results suggest that at large values of *p* (in most of the cases), the targeted strategy generates higher revenues for the credit card company than the total blocking and the nonblocking strategies.[10] We find that the three benchmark models also do a creditable job in identifying risky prospects and in capturing the profitability implications of different policies in the simulation. The results suggest that the knowledge of the predicted y_2 is useful in helping credit card companies design targeted strategies.[11]

CONCLUSION

Consumer delinquency rates in the credit card market are consistently higher than those in the other parts of the loan market. Unlike traditional loans, credit card lending is not secured by any assets. It is important for a credit card company to monitor and predict its consumers' debt repayment behaviors and to identify consumer segments to develop targeted marketing strategies. We propose a dynamic Type II Tobit model to predict consumer debt repayment behavior in the credit card market. In our modeling framework, we treat the discrete variable of whether a consumer is delinquent in each month and the continuous variable of the amount paid in each month, conditional on not being delinquent, as two separate but possibly correlated observations to identify low-risk consumers even when they are delinquent. Furthermore, our model separates the repayment amount, conditional on consumers being delinquent, into two groups: an actual zero resulting from consumers' inability to pay back the debt and a censored zero resulting from other factors, such as consumer oversight. We allow the model parameters to vary over time to account for a consumer's possible learning about card usage and the

[10]As Table 6 shows, the proposed model performs well if the cost of misclassification is low (a high *p* value). We acknowledge this as a limitation. However, our numerous conversations with the senior mangers from the major credit card companies and an in-depth interview with a major credit card company's vice president in Asia suggest that the probability that a low-risk consumer will pay back the debt is high after he or she receives reminders from the bank, possibly because of switching costs and loyalty programs offered by the banks.

[11]Note that in practice, credit card managers should monitor the risk type of their delinquent customers and adjust their strategies constantly over time because consumers' risk type may change over time.

Table 6

RESULTS OF THE STRATEGY SIMULATION BASED ON THE PROPOSED MODEL

	Nonblocking Strategy	*Targeting Strategy at Different Cutoff Points (y_2)*				*Blocking Strategy*	*Perfect Foresight*
		.20	*.40*	*.60*	*.80*		
p = .95	639.07	(729.32, 797.30)	(735.59, 839.62)	(780.84, 834.26)	(802.68, 814.76)	779.84	875.99
p = .90	639.07	(679.83, 749.89)	(678.50, 781.98)	(720.44, 774.14)	(737.53, 753.99)	690.36	875.99
p = .85	639.07	(629.73, 703.07)	(621.24, 725.56)	(659.95, 714.09)	(671.84, 693.76)	600.88	875.99
p = .80	639.07	(599.10, 676.80)	(585.80, 691.32)	(629.36, 684.16)	(605.94, 633.74)	511.4	875.99

Notes: All the numbers are in HK$1,000, and the numbers in parentheses are the 95th percentile interval for the revenues from targeting strategy.

change of card policies over time. We also account for consumer heterogeneity.

We apply our proposed model to a unique data set that includes the inception of a consumer's credit card history of monthly spending and repayment. Our estimation results show that the consumers' repayment behavior changes over time. We also find that a consumer with a high predicted repayment ratio is likely to be a profitable consumer for the credit card company, even though the consumer sometimes misses monthly payments.

Managerially, our study can benefit the credit card companies by better controlling for credit risk. Our approach enables credit card companies to identify high-risk consumers and potentially profitable consumers. As a result, the credit card companies can be more flexible and effective in their credit supply policies. Using the approach we developed in this study, credit card companies can develop effective marketing strategies by targeting specific segments of their consumers. The study also helps the credit card companies understand the effects of credit policies on consumer delinquency in terms of different consumer segments. Our simulation exercise shows how the credit company can use our approach to develop targeted card policies to control consumer risk and reduce consumer default.

There are several possible extensions to the proposed study. The focus of the study was to develop a managerially relevant prediction model of consumer repayment behavior using information on a consumer's past repayment and spending behavior.[12] A possible extension is to develop a structural model that explicitly models consumers' spending and repayment decisions simultaneously to study the process underlying their credit card spending and repayment behaviors. Because most consumers typically own multiple credit cards, our study can be extended to the case of several credit cards to study the correlation of consumers' repayment behaviors across credit cards subject to data availability. One limitation of the article is that we assume that credit card terms, such as credit limits and interest rates, are exogenous variables, though in reality, credit limits and annual percentage rates can be adjusted over time by the banks according to the needs and past spending and repayment behaviors of the consumers. Because of our data limitation, there could be omitted variable bias in our estimation. Further research could jointly model the card issuer's decision on credit card terms and the consumer's decision on monthly spending and repayment simultaneously. Finally, consumers' long-term spending and repayment behaviors could evolve as a result of changes in their socioeconomic characteristics over time. The current research can be further extended by accounting for this dynamic in the model, subject to data availability.

[12]A model to formulate a dynamic multiperiod problem in which consumers solve their own allocation problem over time and firms maximize the long-term profit stream would be ideal. However, we are not able to do this given the current data limitation, and we acknowledge this as a limitation of the article.

REFERENCES

Akcura, M.T., F.F. Gonul, and E. Petrova (2004), "Consumer Learning and Brand Valuation: An Application on Over-the-Counter Drugs," *Marketing Science*, 23 (1), 156–69.

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (March), 411–54.

Altman, Edward I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23 (4), 589–609.

Bass, Frank M., Norris Bruce, Sumit Majumdar, and B.P.S. Murthi (2007), "Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising–Sales Relationship," *Marketing Science*, 26 (2), 179–95.

Blattberg, Robert C., Gary Getz, and Jacquelyn S. Thomas (2001), *Customer Equity: Building and Managing Relationships as Valuable Assets*. Boston: Harvard Business School Press.

Domowitz, Ian and Robert L. Sartain (1999), "Determinants of the Consumer Bankruptcy Decision," *Journal of Finance*, 54 (1), 403–420.

Evans, David S. and Richard Schmalensee (2005), *Paying with Plastic*. Cambridge, MA: MIT Press.

Gross, David B. and Nicholas S. Souleles (2002), "An Empirical Analysis of Personal Bankruptcy and Delinquency," *Review of Financial Studies*, 15 (1), 319–47.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V. Kumar, Nathaniel Lin, et al. (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139–55.

——— and Donald R. Lehmann (2002), "Customers as Assets," *Journal of Interactive Marketing*, 17 (1), 9–24.

——— and ——— (2005), *Managing Customers as Investments*. Philadelphia: Wharton School Publishing.

Kim, Jin Gyo, Ulrich Menzefricke, and Fred M. Feinberg (2005), "Modeling Parametric Evolution in a Random Utility Framework," *Journal of Business and Economic Statistics*, 23 (3), 282–94.

Lachaab, Mohamed, Asim Ansari, Kamel Jedidi, and Abdelwahed Trabelsi (2006), "Modeling Preference Evolution in Discrete Choice Models: A Bayesian State Space Approach," *Quantitative Marketing and Economics*, 4 (1), 57–81.

Neelamegham, Ramya and Pradeep K. Chintagunta (2004), "Modeling and Forecasting the Sales of Technology Products," *Quantitative Marketing and Economics*, 2 (3), 195–232.

Rosenberg, Eric and Alan Gleit (1994), "Quantitative Methods in Credit Management: A Survey," *Operations Research*, 42 (4), 589–613.

Rust, Roland T., Katherine N. Lemon, and Valarie Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (January), 109–126.

Steckel Joel H. and Wilfried R. Vanhonacker (1993), "Cross-Validating Regression Models in Marketing Research," *Marketing Science*, 12 (4), 415–27.

Van Heerde, Harald J., Carl F. Mela, and Puneet Manchanda (2004), "The Dynamic Effect of Innovation on Market Structure," *Journal of Marketing Research*, 41 (May), 166–83.

Zmijewski, Mark E. (1984), "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, 22 (Supplement), 59–82.