Yong Cao & Thomas S. Gruca

Reducing Adverse Selection Through Customer Relationship Management

Adverse selection is an important problem for marketers. To reduce the chances of acquiring an unprofitable customer, companies may screen prospects who respond to a marketing offer. Prospects who respond are often not approved. At the same time, prospects who are likely to be approved are unlikely to respond to a given marketing offer. Using data from a firm's customer relationship management system, the authors show how to target prospects who are likely to respond and be approved. This approach increases the number of customers who are approved and reduces the number of applicants who may defect after being turned down. This method can be extended to new customer acquisition and more effective targeting of costly promotions to migrate customers to higher levels of lifetime value.

> In many businesses, the customers most likely to sign on are precisely the worst customers you could possibly find.
> —Reichheld (1996, p. 76)

> The customers you want to attract don't respond, and the ones you don't want to attract do.
> —Richard E. Mirman, Chief Marketing Officer, Harrah's Entertainment, quoted in Levey (2002, p. 1)

> Most of the people applying for a card with, say, a 12 percent APR, were the last people issuers would approve.
> —Lazarony (2000)

Successful customer relationship management (CRM) begins with acquisition of the right customers. Many writers on CRM issues focus on the identification, valuation, and retention of good customers (e.g., Dowling 2002; Rigby, Reichheld, and Schefter 2002; Verhoef 2003; Winer 2001). A firm should put a great deal of emphasis on discovering who its best customers are and how to find new customers who will be similarly loyal and profitable. At the same time, many companies would benefit immensely from avoiding customers at the other end of the value spectrum (i.e., bad customers).

Bad customers are those who buy only when deep discounts are offered, buy in much smaller quantities than normal, or are otherwise much more costly to serve. In other cases, they defect to another company before their acquisition and other up-front costs are recovered. Unfortunately, in firms' attempts to attract new customers, they often elicit responses from the very type of customer they wish to avoid.

Yong Cao is Assistant Professor of Marketing, Department of Business Administration, College of Business and Public Policy, University of Alaska (e-mail: afyc@uaa.alaska.edu). Thomas S. Gruca is Associate Professor of Marketing, Department of Marketing, Henry B. Tippie College of Business, University of Iowa (e-mail: thomas-gruca@uiowa.edu). The authors thank an anonymous company for providing the data used in this study. They also thank Wagner Kamakura, Bruce Klemz, Lopo Rego, and Gary Russell for their comments on this article. Author names are listed in alphabetical order. Professor Gruca acknowledges the support of the Lloyd J. and Thelma W. Palmer Research Fellowship.

This predicament, known as adverse selection, is most closely linked with the marketing of products such as loans or insurance. However, the experiences of many failed Internet retailers are case studies in how to attract and reward bad customers. A leading online grocer found that 75% of its customers were "price butterflies," bargain hunters chasing deep discounts from one Web site to the next (Reichheld and Schefter 2000). Regrettably, this problem is not limited to online neophytes. Every company has the potential to attract customers whose loyalty and value to the firm are suspect. Long-distance carriers spent billions of dollars courting disloyalty by sending large checks to tempt competitors' customers into switching (Naik 1995). A multi-industry study by McKinsey found that bad customers may account for 30%–40% of a typical company's revenue (Leszinski et al. 1995). It is clear that many firms do a poor job screening out bad customers (Reichheld 1996).

In markets for risk products, the purchase of the product imposes risks on the seller rather than on the buyer (Mitchel 2002). In such situations, a company's survival depends on its decisions about which customers to seek and acquire. These firms use a screening process to determine whether a customer responding to an offer should be approved. This step reduces the likelihood of acquiring an unprofitable customer. Ultimately, many people who respond to a favorable marketing offer will not be approved. At the same time, because the most attractive prospects for any offer are likely to have many alternatives from which to choose, they are less likely to respond.

The problems of adverse customer selection are exacerbated for firms attempting to cross-sell (or up-sell) a product to their existing customers. Several recent articles have tried to understand why customers would be interested in buying more from the same company. Some determinants of cross-buying that have been studied thus far include ownership of other products from the same firm (Kamakura, Ramaswami, and Srivastava 1991), satisfaction with those

ISSN: 0022-2429 (print), 1547-7185 (electronic)

products (Wangenheim 2004) and their prices (Verhoef, Franses, and Hoekstra 2001), and environmental factors, such as technological volatility (Stremersch et al. 2003).

The core of any relationship is loyalty and respect (Fournier, Dobscha, and Mick 1997). By soliciting a response to an offer from an existing customer and then rejecting that customer's response, the firm shows contempt for its relationship with that customer. It is likely that such consumers may reduce their commitment to the firm or defect completely. Such a reaction by consumers has a real economic cost in lost future revenues and other benefits, such as positive word-of-mouth advertising and referrals of new customers (Reichheld 1996).

Fortunately, for firms implementing CRM, there is a solution to the twin problems of adverse selection and costly screening. In this article, we develop a modeling framework to identify the customers who are more likely to respond to an offer and to become more profitable customers (i.e., by being approved for the product). Using data from a major financial institution, we show that our approach to prospect selection results in more approvals and fewer rejected customers. Overall, these effects result in increased profitability.

We provide a practical method for effectively implementing cross-selling activities, a cornerstone to increasing profitability through CRM (Dowling 2002). Specifically, we show how CRM enables a firm to target its marketing efforts better to current customers (Rigby, Reichheld, and Schefter 2002). Furthermore, this modeling framework can be easily extended to parallel cross-selling situations, such as those involving costly sampling of nonfinancial products.

Modeling Framework

In this article, we focus on the situation in which a firm tries to cross-sell a product to its existing base of customers through direct marketing. This is a key situation in which CRM technology can fulfill its promise to improve firm performance (Rigby, Reichheld, and Schefter 2002, p. 6). Because we are focusing on marketing a risk product, we assume that the firm incurs substantial costs to screen its customers and that not all of the customers who respond to the offer will be approved. (We discuss the extension of this framework to the situation of marketing nonrisk products subsequently.)

Typically, a direct marketing campaign has two phases. In the first phase, a relatively small sample of the population is selected to evaluate and respond to a particular offer from the firm. The results of this test (i.e., who responded and who did not) are combined with the firm's data about these customers to develop a prospect selection model. The prospect selection model links observed behavior with household characteristics, such as prior purchase behavior and geographic, demographic, and lifestyle variables, to predict response probability. By ranking the remaining customers from the most likely to the least likely to respond to an offer, the firm can select the most promising set of customers for a larger solicitation. During the second phase, the company contacts only those prospects who are most likely to respond.

Prospect Selection Under Adverse Selection

Under adverse selection, there are limitations to the traditional approach of selecting prospects on the basis of their propensity to respond. Although the firm would elicit many responses, most of these prospects are unlikely to be approved. Therefore, the number of rejected customers will be high, and it is possible that this rejection could damage the firm's relationships with these customers.

An alternative is to use screening criteria to identify prospects that would be approved if they responded to the offer. Financial institutions use this approach when they use third-party credit data to solicit customers with offers of preapproved credit cards. However, the best customers are unlikely to respond to any particular offer because they have more options than do less desirable customers. To illustrate this point, consider the response to a major bank's mailing of approximately two million preapproved credit card applications. At the end of the campaign, the bank issued only 15,000 new credit cards, a response rate of less than 1% (Ausubel 1999). The response rate was low despite the bank's offering various levels of favorable introductory rates and time periods.

Adverse selection necessitates the costly screening of customers who respond to a firm's offer. The customers who are deemed acceptable will be approved, and the others will be rejected. For a firm to identify the best prospects from its remaining customers, it must understand the factors that drive response and approval. This entails the development of a model that enables the firm to identify prospects who are likely to respond and to be approved. We illustrate this process in Figure 1.

Using this simultaneous approach, we can identify prospects who are likely to respond and to be approved. As a result, the number of approvals should be higher than either alternative we previously discussed. This is due to the lower number of rejected applications compared with the approach in which prospects are selected solely for their propensity to respond. We also expect a greater number of responses from using our simultaneous approach than from selecting prospects on the basis of their likelihood to be approved. Thus:

> H_1: Under adverse selection, identifying prospects using a simultaneous model of response and approval likelihoods results in (a) a greater number of approvals than selecting prospects on the basis of their likelihood to be approved and (b) fewer rejections than selecting prospects on the basis of their likelihood to respond.

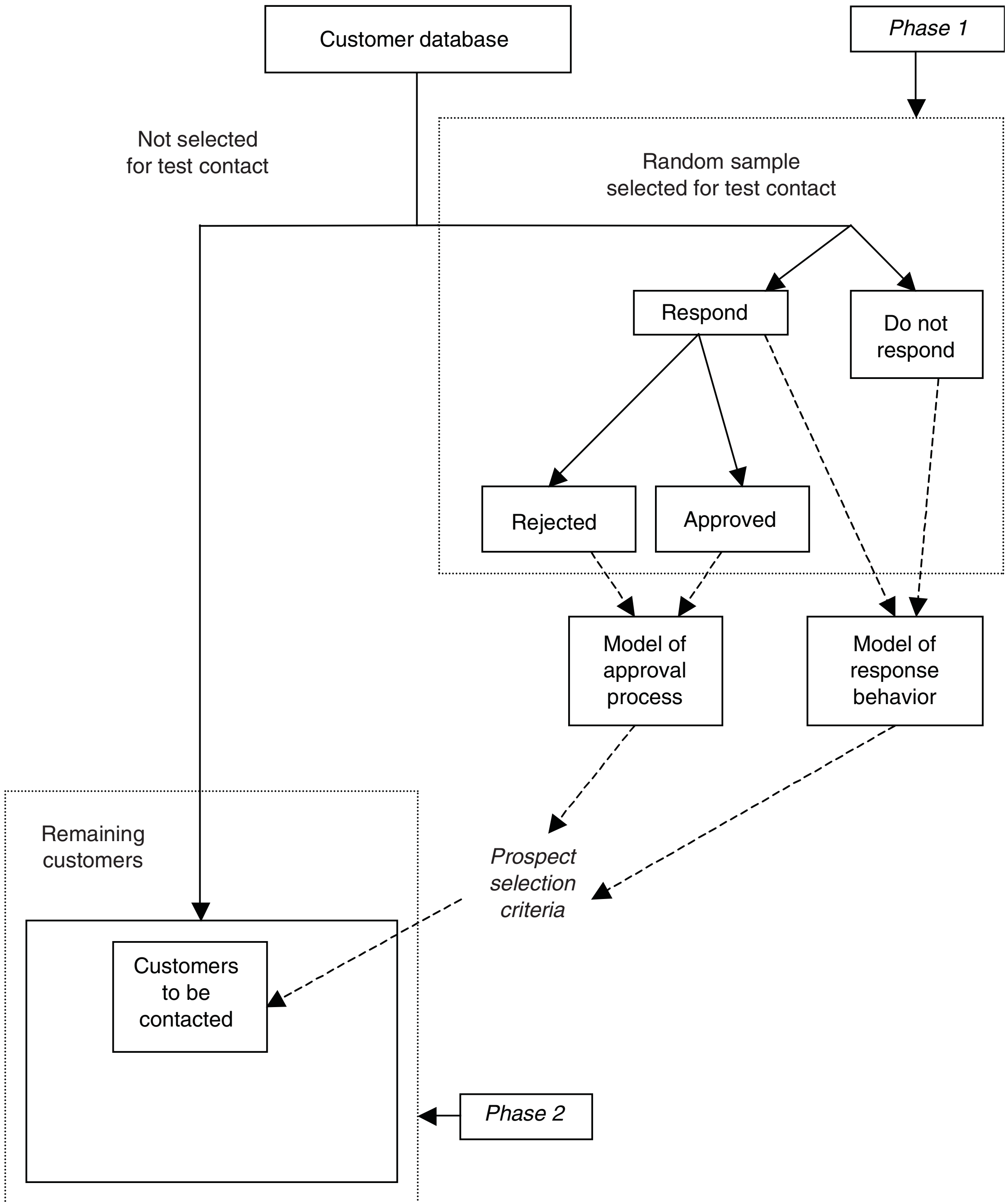
In the next section, we present the results of an empirical study that demonstrates the use of a simultaneous model for prospect selection under adverse selection, and we provide a comparative evaluation of the financial impact of using this modeling framework.

Empirical Study

Setting and Sample

Our data were provided by a major financial institution that was considering a large mailing for a new secured loan product. We were provided with a random sample of 11,710

FIGURE 1
Prospect Selection Under Adverse Selection and Costly Screening



records from a larger test mailing that used the bank's database. To test our hypothesis, we divided the entire sample into an estimation (two-thirds) and holdout (one-third) sample. The estimation sample consisted of 7854 households, 3844 of which responded and 1414 of which were approved. The holdout sample included 3856 households, 1870 of which responded and 690 of which approved.

The company that provided the data for this study was concerned about revealing the true response rates for this test mailing; this was information the company considered confidential. Therefore, rather than using the entire sample of nonrespondents, they provided a random sample of nonrespondents that was approximately the same size as the number of respondents (i.e., a synthetic retrospective sample; see Mantel 1973). This sampling method affects only the estimation of the intercept term, not the other parameters in a binary response model. This bias in the intercept term does not affect the use of a model estimated on data of this type for ranking purposes, because the intercept is the same for every observation by definition.

Household Information on Credit History and Experience

For this sample of customers, the company's CRM system provided measures of each household's general credit history and experience. We describe these items in Table 1. The credit status and history variables supply a great deal of detailed but redundant information about the household. To reduce the overlap across these indicators, we used principal components analysis to reduce this set of variables to a more manageable number of factors (i.e., related linear combinations of the original measures).

From the original set of 24 measures, we identified six factors that accounted for 72% of the total variance in our sample. We present a Varimax rotation of the factor loadings in Table 2. The first factor is associated with the borrower's use of credit compared with current limits. We designate this factor "credit limits." A higher score on this factor indicates that a borrower is using a high percentage of the credit available to him or her. This reduces the attractiveness of the borrower for many lenders (Delaney 1997). The second factor is related to recent approvals for bank cards, including premium bank cards. We call this factor "new bank cards." Borrowers with a high score on the third factor, which we designate "payment problems," have a less desirable payment history for unsecured debt. The fourth factor is related to the number of finance company accounts specifically and the total number of accounts in general. We designate this factor "finance company accounts." Both the second and fourth factors are related to the breadth of credit the borrower uses. The fifth factor measures the length of credit history and is so named. Only one measure, the number of inquiries, has a high loading on the sixth factor (i.e., "inquiries").

The factors identified in our sample are consistent with the categories of inputs that commercial suppliers of credit scores use (Delaney 1997). In our empirical study, we use the factor scores as indicators of creditworthiness. In addition to these measures of general creditworthiness, we were also provided with a measure of recent payment history with respect to mortgage payments. The credit-scoring firm Fair Isaac considers such data important in evaluating secured home loans (Delaney 1997).

TABLE 1
Measures of Credit History and Experience

Measure	Sample Mean
Average number of months that accounts are on file	78.68
Months on file	142.70
Months since most recent account opened	20.95
Months since most recent bank card account opened	29.42
Number of accounts	15.03
Number of accounts ever 30 or more days past due	2.46
Number of accounts 30 or more days past due in 24 months	1.22
Number of bank card accounts open in 6 months	.15
Number of bank card accounts open in 24 months	.80
Number of bank cards, finance, personal revolving accounts	4.51
Number of currently active personal finance accounts	.29
Number of different subscribers	12.37
Number of finance accounts verified in 12 months	1.44
Number of inquiries	2.97
Number of premium bank card accounts open in 24 months	.18
Number of public record and account line derogatory remarks	1.00
Number of satisfactory finance company revolving accounts	.84
Percentage of accounts never delinquent	78.45
Percentage of active accounts with positive balance	67.27
Percentage of bank card accounts greater than 50% of limit	29.00
Ratio: balance to high credit	44.26
Ratio: bank card balance to high credit	2.94
Ratio: retail balance to high credit	16.24
Ratio: revolving balance to high credit	33.39

Lifestyle Determinants of Demand

The theory of adverse selection suggests that people who respond to an offer for a secured loan product are likely to be greater credit risks than nonrespondents, but creditworthiness is not the only determinant of demand. Canner, Durkin, and Luckett (1998) find that households that have home equity loans or lines of credit tend to have higher incomes and higher levels of education. Most of these borrowers are between the ages of 35 and 64 years. These households are also more likely to have substantially more equity in their homes. In addition, data from the 2002 American Home Survey show that married households are also more likely to have a home equity loan.

The financial institution augmented its database of credit data with measures indicating income, age, marital status, presence of children in the home, and occupation. The income data (INCOME) was estimated by means of microlevel geodemographic data (e.g., census tract averages). We did not have data beforehand about a prospect's actual level of home equity. Therefore, we used a proxy measure, an estimate of the average mortgage amount (MORTGAGE) from a small area sample. All else being equal, households with higher mortgages might be less interested in a home equity loan because they may have recently refinanced their first mortgage, leading to a higher current payment (Canner, Durkin, and Luckett 1998). We included an indicator for households headed by a single adult (SINGLE) because we expect such households to be less likely to use this type of product. Because education expenses are a major use for home equity loans (Canner, Durkin, and Luckett 1998), we also included an indicator for the presence of children in the home (CHILDREN). Although we did not have access to data on the borrower's level of education, we used occupation as a proxy. We constructed an indicator (MANAGER) for the households that owned their own business, were professionals or executives,

TABLE 2
Factor Loadings for Measures of Credit History and Experience

Measures	Credit Limits	New Bank Cards	Payment Problems	Finance Company Accounts	Length of Credit History	Inquiries
Ratio: bank card balance to high credit	.903	.163	.045	.107	.075	–.062
Percentage of bank card accounts greater than 50% of limit	.893	.144	.038	.088	.065	–.059
Ratio: revolving balance to high credit	.866	–.055	.165	.074	–.098	.028
Ratio: balance to high credit	.653	–.192	.202	.062	–.317	.134
Percentage of active accounts with positive balance	.631	–.230	.088	.112	–.288	.195
Ratio: retail balance to high credit	.575	–.184	.226	.051	–.132	.081
Number of bank card accounts open in 24 months	.050	.833	–.136	.109	–.212	.038
Number of bank cards, finance, personal revolving accounts	–.060	.752	–.098	.409	.278	–.090
Number of premium bank card accounts open in 24 months	–.195	.740	–.196	.107	.005	–.004
Number of bank card accounts open in 6 months	–.006	.669	–.155	–.055	–.239	.106
Number of accounts ever 30 or more days past due	.251	–.104	.908	.100	–.021	.053
Number of accounts 30 or more days past due in 24 months	.240	–.085	.839	.102	–.062	–.006
Number of public record and account line derogatory remarks	.013	–.129	.817	–.151	–.114	.128
Percentage of accounts never delinquent	–.136	.295	–.778	.274	.210	–.044
Number of finance accounts verified in 12 months	.224	–.045	–.021	.836	–.231	.076
Number of satisfactory finance company revolving accounts	.074	.129	–.135	.806	–.032	–.074
Number of accounts	.060	.536	.051	.751	.168	.075
Number of different subscribers	.026	.625	–.002	.678	.239	.011
Number of currently active personal finance accounts	.272	–.311	.168	.398	–.296	.225
Average number of months accounts on file	–.210	.036	–.131	–.078	.862	–.005
Months since most recent bank card account opened	.011	–.163	–.176	.062	.708	–.112
Months on file	–.126	.142	–.072	.116	.616	.591
Months since most recent account opened	–.152	–.226	.059	–.401	.540	–.103
Number of inquiries	.166	.044	.187	.007	–.171	.819
Eigenvalue (unrotated solution)	5.9	4.8	2.1	2.0	1.5	1.1

Notes: We used principal components analysis with Varimax rotation.

or held middle-management positions. For age, we created separate indicators for the 35–44 age group (35T44), the 45–54 age group (45T54), the 55–64 age group (55T64), and the 65 and over age group (65PLUS). We expected higher levels of response from the 35–55 age groups. We were also provided with a variable that indicated whether this household had responded to a mail solicitation from a financial services firm in the recent past (MAILBUY). We list these variables in Table 3.

To test our hypotheses, we compared the prospects identified using the proposed simultaneous response and the approval model with prospects selected solely on the basis of their response probability and the prospects with the highest likelihood of being approved. We began by dividing the entire sample into an estimation sample and a holdout sample as we previously discussed.

We used the model parameters from the estimation sample to identify prospects in the holdout sample using the same criteria. For example, using the estimated parameters of a model of response likelihood, we computed the probability that a prospect in the holdout sample would respond to the offer. This information provided a rank ordering of all elements of the holdout sample with respect to their attractiveness as a prospect.[1]

[1]The response and approval rates for the validation sample will be overstated compared with actual results, given the artificially high response rate in our synthetic retrospective sample. However,

TABLE 3
Lifestyle Determinants of Demand for Secured Loan Products

Measure	Description	Sample Mean
INCOME	Estimated income/$10,000	46.30
MORTGAGE	Estimated mortgage	64,601.09
SINGLE	Head of household is single	.07
CHILDREN	Presence of children under the age of 18	.16
MANAGER	Borrower is a business owner, professional, executive, or middle manager	.04
35T44	Borrower is between the ages of 35 and 44 years	.38
45T54	Borrower is between the ages of 45 and 54 years	.27
55T64	Borrower is between the ages of 55 and 64 years	.13
65PLUS	Borrower is 65 years of age or older	.04
MAILBUY	Household responded to mail solicitation from a financial company recently	.08

To proxy the selection of the best prospects from the remaining customers (e.g., for the simultaneous model of response and approval, see Phase 2 in Figure 1), we compared the expected results for the top 30% of the prospects for each alternative method. The differences between the methods are more pronounced for smaller subsets of the holdout sample (10% or 20%). However, we chose this larger subset to avoid appearing as if we capitalized on chance variations in smaller samples (e.g., there are fewer than 60 approvals in the first decile of prospects chosen for their response likelihood). We discuss these empirical results subsequently.

Prospect Selection Based on Response Likelihood

The first model focused on the probability that a household would respond, given the household's creditworthiness and other lifestyle characteristics that we previously discussed. Using a binary probit formulation, we modeled the responses on the 7854 households in the estimation sample, 3844 of which responded to the offer. The resultant coefficient estimations appear in Table 4 (Column 2). The overall chi-square fit statistic for the model is significant at the $p < .01$ level.

Households with lower incomes and lower mortgages and applicants between 35 and 64 years of age were more likely to respond. Previous response to direct mail was also positively associated with a response to this offer. Neither marital status (SINGLE) nor the presence of children in the home (CHILDREN) was a significant determinant of the propensity to respond to the offer. Business owners and middle managers (MANAGER) were also less likely to respond. A possible explanation for this is that occupation is a poor choice for a proxy variable in this sample. Alternatively, because these customers probably have higher incomes, they might be better able to afford refinancing to extract equity or to combine various types of debt (e.g., first mortgages, credit card debt) into a new mortgage.

The impact of the creditworthiness factor scores provides clear evidence of adverse selection. Poor mortgage payment performance, a higher ratio of credit used to available credit, a greater number of inquiries, a shorter credit history, and more indicators of problems with prior payment history were all positively associated with an increased probability of responding to this offer. The respondents were also less likely to have been recently approved for a new bank card and had less broad credit experience.

the results are comparable across models, which is the key to our analyses.

TABLE 4
Estimation Results

Variable	Response-Oriented Selection Model	Approval-Oriented Selection Model	Simultaneous Response and Approval Selection Model (Bivariate Probit Model)	
	Coefficient	Coefficient	Response Coefficient	Approval Coefficient
Constant	.1459**	–.8363**	.1453*	–.8727**
Credit limits	.1985**	.0092	.1983**	.0163
New bank cards	–.1891**	.0377	–.1890**	.0309
Payment problems	.2983**	–.2961**	.2984**	–.2860**
Finance company accounts	–.0897**	.0823**	–.0899**	.0791**
Length of credit history	–.1915**	.0025	–.1911**	–.0034
Inquiries	.1262**	–.0572**	.1260**	–.0533*
INCOME	–.0032**	.0128**	–.0031**	.0126**
MORTGAGE	–3.80E–06**	8.22E–07	–3.80E–06**	6.67E–07
Number of 30 day or worse ratings on mortgage accounts	.0411**	–.0557**	.0410**	–.0540**
SINGLE	.0810	—	.0838	—
CHILDREN	.0403	—	.0419	—
MANAGER	–.1585*	—	–.1578*	—
MAILBUY	.2169**	—	.2184**	—
35T44	.1994**	—	.1992**	—
45T54	.2085**	—	.2078**	—
55T64	.1734**	—	.1702**	—
65PLUS	.1229	—	.1172	—
Chi-square (model)	1618.9**	352.7**	.54	ρ (error correlation) .062

*$p < .05$.
**$p < .01$.

Using these results, we predicted the probability of response for the holdout sample of 3856 households. We then ranked the results and assigned each household to one of ten (nearly) equally sized segments (i.e., deciles). We expected that the households in the highest decile would be the most likely to respond to the offer.

Because we have the actual response data (and approval data) for these households, we can examine how well the response-oriented selection model would have performed if it were used to score the profiles of the holdout sample. The average response rate for the top three deciles is 72%, compared with 38% for the remaining deciles. However, the approval rate for these same households was only 26%, compared with the rate of 46% for the remaining prospects. Therefore, under adverse selection, selecting prospects on the basis of response likelihood results in a predictably high response rate but a low approval rate.

Prospect Selection Based on Approval Likelihood

Selecting prospects using approval as the dependent variable involves a change in the sample to accompany the change in the dependent variable. We can only model the probability of approval using data from that subset of customers who responded to the offer. This means restricting ourselves to the 3844 observations for which we have data on the outcome of the approval decision. To predict approval, we relied on the key income-, credit-, and mortgage-related variables.

The modeling of approval as a probabilistic process might seem unnecessary in today's environment of automated credit scoring and online credit card approvals. However, in this particular application, important variables such as current income and size of mortgage are unknown at the time of prospect selection, and verification of such information is an important component of the costs of screening applicants.

Using a binary probit model, we estimated the approval likelihood for the households in the estimation sample that had responded to the offer. The resultant coefficient estimations appear in Table 4 (Column 3). The overall chi-square fit statistic for the model is significant at the $p < .01$ level.

Applicants with higher incomes, good records of keeping current with mortgage payments, and fewer inquiries into their credit history are likely to be approved. In addition, applicants with a recent approval for a bank credit card are likely to be approved, whereas applicants with poor records of payment on unsecured debts are less likely to be approved. We used the results from this second model to estimate the approval probabilities for all 1870 households in the holdout sample. We then ranked the households and assigned them to deciles; again, those in the highest decile were the most likely to be approved.

Using the actual approval data for the validation sample, we observe how well this approach would have performed had it been used to identify prospects. If we consider only the customers who responded, the approval rate for a mailing to the top three deciles of this sample (62%) is greater than the approval rate (32%) for the remaining applicants. However, the response rate for the top three deciles is much lower (28% versus 57% for others in sample).

Consistent with the executives' experiences reflected in the opening quotations, households that are likely to be approved for the offer are the same ones that are less likely to respond. This is because good customers (e.g., customers with more favorable credit ratings) receive a large number of offers, including those from companies trying to acquire them as a new customer. Thus, their propensity to respond to any one particular offer is predictably quite low.

Prospect Selection Based on Response and Approval Likelihoods

To identify prospects who are likely to both respond to our offer and be approved, we need to consider two dependent variables: response and approval. If we follow the probit model formulations, we have the following system of equations to estimate:

$$\text{(1)} \qquad \begin{cases} R = \beta_1 X_1 + \mu \\ A = \beta_2 X_2 + \nu \end{cases},$$

where X_1 and X_2 are descriptors of the individual household. These include the measures of creditworthiness, income, mortgage, and lifestyle characteristics that we previously discussed. The dependent variable R is the observed response to the offer (0/1). The dependent variable A is the observed binary approval outcome (0/1). Note that the dependent variable A is observed only when the other dependent variable R (response) has a value of 1. We further assume that the error terms μ and ν follow a standard bivariate normal distribution and that the correlation between the error terms (μ, ν) equals ρ.

That the dependent variable A (approval) is observed only when the other dependent variable R (response) has a value of 1 introduces the possibility of bias in the estimates of the approval prediction equation. One source of this bias may be an omitted variable (or a set of correlated variables) that explains both response and approval. To estimate such a system, we used a bivariate probit model with sample selection (Meng and Schmidt 1985). This formulation is often used in the credit-scoring literature (Banasik, Crook, and Thomas 2003; Boyes, Hoffman, and Low 1989; Jacobson and Roszbach 2003).

The simultaneous estimation of the response and approval models is only the first step. We must then combine the resulting information to identify households that are likely to respond and to be approved. To identify such prospects, we estimate the probability that both R = 1 and A = 1. This probability is given by the following:

$$\text{(2)} \qquad P(R = 1 \text{ and } A = 1) = \Phi 2(B_1 X_1, B_2 X_2, \rho),$$

where $\Phi 2$ is the cumulative bivariate standard normal distribution, B_1 and B_2 are the parameters estimated in the system given in Equation 1, and ρ is the correlation between the error terms. The terms X_1 and X_2 are the respective variables used to predict response and approval, as we previously noted.

For our empirical study, we estimated the system of equations (Equation 1) using the bivariate probit with the selection procedure in LIMDEP. The results appear in

Columns 4 and 5 of Table 4. Because we observe the responses of the entire estimation sample, there should be little difference between the coefficients of the response model in the bivariate probit system (Table 4, Column 4) and those in the preceding response model (Table 4, Column 2). The differences are negligible.

Contrary to our expectations, we find that the correlation between the error terms in the two equation system is not statistically significantly different from zero ($\rho = .062$, $p < .83$).[2] This implies that the estimates of the coefficients of the approval model are not significantly biased because of their being estimated with only the subset of respondents. This result is confirmed by two other findings. First, in comparing the coefficients for the stand-alone approval model (Table 4, Column 3) and the coefficients estimated for the approval model in this simultaneous system (Table 4, Column 5), we find few, if any, differences. Second, we find that the improvement in the fit of the bivariate probit model over the separate response and approval probit models is not significantly different from zero ($\chi^2 = .54$, $p < .46$, degree of freedom = 1).

The estimated coefficients of the response and approval models in the simultaneous model (Equation 1) do not differ significantly from those estimated by the separate models. However, the prospects we identified using the formula in Equation 2 yield different results. By avoiding elicitation of responses from prospects who will ultimately be rejected and by reducing the number of prospects unlikely to respond, the simultaneous approach to identifying prospects should result in better outcomes for the company.

To identify prospects who are likely to respond and to be approved, we used the estimated coefficients from the system in Equation 1 as inputs into Equation 2.[3] Using the independent variables for each household in the holdout sample, we computed the probability that a given household would respond and be approved. As before, we ranked the entire holdout sample and separated the households into deciles. The households in the highest decile are the most likely to respond and to be approved.

Using the actual response and approval results for the holdout sample, we determined the expected number of responses and approvals for each decile. Concentrating on the first three deciles again, we observe the superiority of this simultaneous approach to prospect selection. The response rate is lower (60%) than that for the approach that focuses on selecting prospects on the basis of their response likelihood (72%). The approval rate (41%) is lower than that for the approval-focused selection approach (62%). However, this combined approach results in more approved customers overall (284 versus 198 for approval-based selection). This finding supports H_{1a}. In addition, the simultaneous approach results in fewer rejected customers (405 versus 624 for approval-based selection), in support of H_{1b}.

It is worthwhile to compare the number of approvals for each method of identifying prospects with random selection. A random selection of 30% should yield 30% of the approvals. The top three deciles of prospects based on response likelihood yielded 214 (31%) approvals, which is slightly better than random selection. The top three deciles of prospects based on approval likelihood yielded only 198 (29%) approvals, which is actually worse than random selection. This is because so few of the households in the most highly rated deciles actually respond to the offer. In contrast, the prospects identified by our simultaneous model yielded 284 (41%) approvals, a 37% increase over random selection.

Financial Impact

In this section, we attempt to assess the financial impact of selecting prospects using a model focused simultaneously on response and approval probabilities. This entails computing the relative profits, including approval costs for the firm and the potential costs of losing customers who respond to an offer only to be turned down.

We make several simplifying but realistic assumptions in this analysis. First, we restrict ourselves to the top three deciles (30% of the holdout sample) for each method of identifying prospects. Second, we assume that the unit costs of contacting ($2) and approving ($15) are the same for all customers. For the value of the offer to the firm, we continue with our example of a secured loan product. Currently, the average home equity loan is approximately $30,000. Some loans are resold to the secondary market, and others are retained in-house. Without further information, we estimate the value of an approval at 1% of the origination value (i.e., $300).

To estimate the potential value lost if a current customer terminates his or her relationship with the financial institution as a result of being turned down for a secured loan, we assume that the financial institution is using its database of its credit card customers. Reichheld (1996) estimates the average profits from a credit card customer to be approximately $100 per year. A 10% discount rate yields a lifetime value of $1,000.

We do not know the probability that a credit card customer who is turned down for a loan by a bank will drop the credit card associated with that bank. Therefore, we consider two possibilities, a .5% defection rate and a 2% defection rate. In Table 5, we compare the outcomes of mailing the offer to the top 30% of prospects, using each of the alternative prospect selection approaches.

By using the usual direct marketing metrics of response rates and cost per response, the optimal approach would be to emphasize response likelihood in the selection of prospects. In contrast, selecting prospects on the basis of approval likelihood is more efficient because this approach yields the lowest screening cost per approval and the high-

[2]This result is largely due to the wealth of information we have about each household from the firm's CRM system. For example, if we were unable to obtain estimates for income or mortgage size, the correlation between the error terms for the two equations would be large ($\rho = .46$) and significant ($p < .02$).

[3]In this particular case, the estimates from the separately estimated models would also perform well because the coefficients in the approval model do not suffer from a great deal of bias. However, this situation is likely to be unique to this data set. Managers interested in this method should use the bivariate probit formulation to avoid the usual problems with censored data.

TABLE 5
Financial Comparisons

	Prospect Selection Based on Response Likelihood	Prospect Selection Based on Approval Likelihood	Prospect Selection Based on Likelihood of Response and Approval
Mailing size (top three deciles)	1157	1157	1157
Total contact costs ($2 per household)	($2,314)	($2,314)	($2,314)
Number of responses	835	319	689
Response rate	72%	28%	60%
Contact cost per response	($2.77)	($7.25)	($3.36)
Total screening costs ($15 per response)	($12,525)	($4,785)	($10,335)
Number of approvals	214	198	284
Screening cost per approval	($59)	($24)	($36)
Approval rate	26%	62%	41%
Contact and screening cost per approval	($69)	($36)	($45)
Total revenue ($300 per approval)	$64,200	$59,400	$86,200
Gross profits (without defection costs)	$49,361	$52,301	$72,551
Expected losses due to rejections (½%/2% defection rate and $1,000 lifetime value)	($3,105)/($12,420)	($605)/($2,420)	($2,025)/($8,100)
Net proceeds (½%/2% defection rate)	$46,256/$36,941	$51,696/$49,881	$70,526/$64,451

est approval rate. The simultaneous emphasis on response and approval in selecting prospects is dominated by the other strategies with respect to these functionally oriented metrics.

With respect to the overall profitability of the campaign, the results are different. Selecting prospects on the basis of both their likelihood of response and approval is the most profitable approach. When we incorporate the possible losses of future revenue from the defection of customers who were rejected, our proposed method for identifying prospects remains superior.[4]

Discussion

This article makes an important contribution to CRM research by showing how to reduce the negative effects of adverse selection and costly screening using CRM data. In an empirical study of cross-selling a secured loan product, we show that the negative effects of adverse selection are considerable. They can have a detrimental effect on the profitability of a marketing campaign in the short term. Furthermore, because of rejected customers' potential defections, adverse selection can have an important negative influence on the future profitability of the firm. Our model for selecting prospects on the basis of their likelihood of responding to an offer and being approved relies on the CRM data that most financial firms already possess. This approach is superior to alternative methods for identifying prospects within a database of current customers.

In addition to being more profitable, this modeling approach is more consistent with the goals of CRM insofar as it is sensitive to customers' needs. There is a monetary value in treating all customers, not just the best customers, with respect. For the customers who are unlikely to respond, firms should not bother them with an offer. For those who are unlikely to be approved, firms should not set them up for disappointment. Our proposed modeling framework results in a win-win situation for the firm and all its customers.

This study emphasizes the importance of the method that companies use to identify prospects for cross-selling to their own customers. The selection of a method to identify such prospects depends on two factors. The first of these is the relationship between a household's propensity to respond and the likelihood that it will become a more valuable customer (e.g., be approved). Under adverse selection, these factors have an inverse relationship. The most attractive customers are unlikely to respond, whereas potentially unprofitable customers respond in droves. If the relationship is positive or zero, the second factor, "screening costs," becomes important. The firm must consider the costs of determining whether a prospect will become a more valuable customer. In our empirical study, these were the screening costs. If these screening costs are relatively large, the firm should identify prospects on the basis of their likelihood of becoming a more valuable customer; this increases the chances that costs will be recouped. If these costs are comparatively small, the firm should target prospects who are likely to respond to the offer, thereby increasing the expected number of customers whose value to the company is increased over time. Although these guidelines should hold for companies selling products or services outside the typical boundaries of risk products, more research is necessary to evaluate their effectiveness in practice. In the next sections, we discuss how this modeling framework can be adapted to assist managers with similar marketing problems.

[4]As with any simulated results, the numbers in Table 5 depend on the assumptions about contact costs, screening costs, and so forth.

Cross-Selling in Nonfinancial Markets

Although our motivating example is a secured loan product, the application of this modeling framework to cross-selling to current customers in associated markets, such as credit cards, lines of credit, home improvement loans, and so forth, is readily apparent. However, it is also worthwhile to consider the extension to nonfinancial markets.

Marketers often face a problem that is parallel to that of costly screening due to adverse selection. Many companies use costly sampling programs to increase a consumer's lifetime value. Consider a cable company that is interested in upgrading customers to a digital service that would necessitate the in-home installation of a decoder box or digital video recorder. Often, the cable company will offer free installation of the equipment and a limited period of reduced fees. This cost is incurred in the expectation that responding households will be sufficiently satisfied to keep the equipment and continue to pay full price for the upgraded service after the expiration of the trial offer.

Using their CRM database, the cable company could determine the characteristics of the type of households that respond to such offers and identify which households tend to continue with the higher-priced services (at least until the installation costs are recovered). By incorporating such information into a bivariate probit model such as the one we previously described, the cable company can more selectively target their promotions to households that are more likely to be profitable over time.

Many firms use costly promotions to move customers from one level of consumption to a higher, more profitable level. By combining offer testing and the data in their CRM systems, such companies can use this modeling framework more profitably to identify good prospects for costly upselling promotions. Such a modeling approach represents an important departure from previous research on cross-buying behavior (e.g., Kamakura, Ramaswami, and Srivastava 1991; Verhoef, Franses, and Hoekstra 2001). In these studies, there is a tacit assumption that firms should try to increase cross-buying and that this is a universally good thing for the firm. However, under adverse selection, firms must also consider whether attempts to broaden their relationship with some customers might be ultimately detrimental to profitability.

Prospecting for New Customers

With some adaptation, this modeling framework can also be used to identify high-quality prospects among noncustomers. For example, when a firm uses an established list to find new customers, it can identify which households responded to a promotion and which ones turned into profitable customers. The next step would be to append identifying information onto this purchasing and profitability data from third-party data vendors.

This augmentation of lifetime value data would enable the firm to build a model to predict which households are likely to respond and which of those become profitable. The model can then be used to guide the firm's selection of future prospects to approach with similar offers. A variation of this method is currently being used by some insurance companies to identify good automobile and home insurance prospects using credit bureau data (Simon 2002).

Limitations and Directions for Further Research

The marketing executives quoted at the beginning of this article suggest that adverse selection is a serious problem for marketers. The customers who are likely to be profitable acquisitions are unlikely to respond, whereas prospects with less favorable future values to the firm are more likely to respond. This article provides the first empirical evidence we know of that supports this widely held view of who responds and who does not respond to marketing offers.[5]

We have shown the impact of adverse selection for an important but limited case of cross-selling risk products. Further research is necessary on the magnitude and impact of adverse selection in the response to promotions, especially costly ones. Whether adverse selection is as significant a problem for new customer acquisition or for companies selling nonrisk products is an important area for further research.

In this study, the firm used a fairly stable model to approve the applications of responding prospects. With nonrisk products, the situation may change. In these applications of our framework, approval is replaced by a customer moving to a new, more profitable level of consumption. How firms should model these transitions and how soon they can identify those who will become highly valued customers is an important extension of our study and a key problem in CRM research in general.

Conclusions

A positive by-product of our approach to customer selection is the inherent coordination between the marketing and the credit functions of the firm that is selling risk products. Selecting prospects on the basis of their likelihood to respond and to be approved removes a major source of conflict. Such interfunctional coordination is a major goal of CRM as a whole (Rigby, Reichheld, and Schefter 2002). Furthermore, our proposed approach to prospect selection addresses an important problem in the customer relationship life cycle, namely, the identification of the right prospects. Its narrow focus and modest goals are consistent with the successes that are now being experienced in CRM practice (Rigby and Ledingham 2004).

[5]All of the prospects who responded in Aubusel's (1999) study were preapproved for the product being promoted. His results demonstrate adverse selection within a set of prospects who were already judged to be attractive by the firm (i.e., bank approval preceded customer response).

REFERENCES

Ausubel, Lawrence M. (1999), “Adverse Selection in the Credit Card Market,” working paper, Department of Economics, University of Maryland, (accessed May 1, 2005), [available at http://www.bsos.umd.edu/econ/bankruptcy/adverse.pdf].

Banasik, Jonathan, John Crook, and Lyn Thomas (2003), “Sample Selection Bias in Credit Scoring Models,” *Journal of the Operational Research Society*, 54 (8), 822–32.

Boyes, William J., Dennis L. Hoffman, and Stuart A. Low (1989), “An Econometric Analysis of the Bank Credit Scoring Problem,” *Journal of Econometrics*, 40 (1), 3–14.

Canner, Glen B., Thomas A. Durkin, and Charles A. Luckett (1998), “Recent Developments in Home Equity Lending,” *Federal Reserve Bulletin*, 84 (April), 241–51.

Delaney, Kevin J. (1997), “Knowing the Score: Your Credit History Is an Open Book—and One You Should Read Carefully,” *The Wall Street Journal*, (December 1), R15.

Dowling, Grahame (2002), “Customer Relationship Management: In B2C Markets, Often Less Is More,” *California Management Review*, 44 (3), 87–104.

Fournier, Susan, Susan Dobscha, and David Glen Mick (1997), “Preventing the Premature Death of Relationship Marketing,” *Harvard Business Review*, 75 (January–February), 2–8.

Jacobson, Tor and Kasper Roszbach (2003), “Bank Lending Policy, Credit Scoring and Value at Risk,” *Journal of Banking and Finance*, 27 (4), 615–33.

Kamakura, Wagner, Sridhar N. Ramaswami, and Rajendra K. Srivastava (1991), “Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services,” *International Journal of Research in Marketing*, 8 (November), 329–50.

Lazarony, Lucy (2000), “What's the Rate on that Card? You Won't Know Until It Arrives,” (August 30), (accessed June 1, 2005), [available at http://www.bankrate.com/ibd/news/cc/20000830.asp].

Leszinski, Ralf, Felix A. Weber, Roberto Raganoni, and Thomas Baumgartner (1995), “Profits in Your Backyard,” *McKinsey Quarterly*, 4, 118–27.

Levey, Richard H. (2002), “New Shooter at Harrah’s,” *Direct*, 14 (14), 1.

Mantel, N. (1973), “Synthetic Retrospective Studies and Related Topics,” *Biometrics*, 29 (3), 479–86.

Meng, Chun-Lo and Peter Schmidt (1985), “On the Cost of Partial Observability in the Bivariate Probit Model,” *International Economic Review*, 26 (1), 71–86.

Mitchel, James O. (2002), “Comparison of Sales Performance Across Financial Service Sectors,” *Journal of Financial Services Marketing*, 7 (2), 178–85.

Naik, Gautam (1995), “AT&T Seems to Call Truce in Price Wars,” *The Wall Street Journal*, (August 19), A3.

Reichheld, Frederick F. (1996), *The Loyalty Effect*. Cambridge, MA: Harvard Business School Press.

——— and Phil Schefter (2000), “E-Loyalty: Your Secret Weapon on the Web,” *Harvard Business Review*, 78 (July–August), 105–113.

Rigby, Darrell K. and Diane Ledingham (2004), “CRM Done Right,” *Harvard Business Review*, 84 (November), 118–27.

———, Frederick F. Reichheld, and Phil Schefter (2002), “Avoid the Four Perils of CRM,” *Harvard Business Review*, 80 (February), 5–11.

Simon, Ruth (2002), “Numbers Game: Looking to Improve Your ‘Credit Score’? Fair Isaac Can Help—That’s Because It’s the Leader In Calculating the Figure, Widely Used by Lenders—Why Insurers Like Them Too,” *The Wall Street Journal*, (March 19), A1.

Stremersch, Stefan, Alan M. Weiss, Benedict G. Dellaert, and Ruud T. Frambach (2003), “Buying Modular Systems in Technology Intensive Markets,” *Journal of Marketing Research*, 40 (August), 335–50.

Verhoef, Peter C. (2003), “Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development,” *Journal of Marketing*, 67 (October), 30–45.

———, Philip Hans Franses, and Janny C. Hoekstra (2001), “The Impact of Satisfaction and Payment Equity on Cross-Buying: A Dynamic Model for a Multi-Service Provider,” *Journal of Retailing*, 77 (Fall), 359–78.

Wangenheim, Florian V. (2004), “Predicting Usage Intensity and Upgrading Behavior of Service Customers: A Model for Lifetime Value Estimation at Early Relationship Stages,” working paper, School of Business Administration, University of Dortmund.

Winer, Russell S. (2001), “A Framework for Customer Relationship Management,” *California Management Review*, 43 (Summer), 89–105.