International Journal of Forecasting 19 (2003) 71–85
www.elsevier.com/locate/ijforecast

Forecasting consumer credit card adoption: what can we learn about the utility function?

Min Qi[a,*], Sha Yang[b,1]

[a]Department of Economics, College of Business Administration, Kent State University, Kent, OH 44242, USA
[b]Gary Anderson Graduate School of Management, University of California, Riverside, Riverside, CA 92521-0203, USA

Abstract

How to accurately predict customers' adoption behavior is becoming more important and challenging to many credit card marketers as competition increases. This calls for more knowledge about the consumer utility function and the corresponding decision behavior. In this study, we challenge the commonly used logit model which implies linear utility function and constant marginal rate of substitution (MRS) with a neural network model that can accommodate nonlinear utility function and changing MRS between card attributes. Using the data from a national survey of credit card usage, we find that the neural network model significantly outperforms the logit in predicting consumer card adoption decisions. Our results indicate that consumers do not make linear tradeoffs between card attributes and the MRS between card features does not remain constant even within the same demographic group.

Keywords: Neural networks; Forecasting; Credit card adoption; Utility function; Marginal rate of substitution

1. Introduction

For direct marketers of credit cards, it is very important to understand how consumers make their card adoption decisions. When a consumer receives a credit card offer, it is natural for the consumer to compare the attributes, such as APR (annual percentage rate), annual fee, credit limit, and cash rebate, etc., of the card offered to the cards one already holds. Depending on the primary purpose of using credit cards various attributes may affect the decision very differently. For people who use credit cards primarily for convenience, they may never carry a credit card balance, thus APR might be irrelevant in their decision on whether to accept a credit card offer. Instead, they will compare annual fee, credit limit, and cash rebate, etc. However, for people who use credit cards primarily for borrowing, APR might be the key element in their decision making. They will adopt a card that offers lower APR though it may have a higher annual fee and no cash rebate.

These behaviors suggest that within certain ranges, card attributes such as annual fee and cash rebate

*Tel.: +1-330-672-1088; fax: +1-330-672-9808.
E-mail addresses: mqi@bsa3.kent.edu (M. Qi), sha.yang@ucr.edu (S. Yang), shayang@stern.nyu.edu (S. Yang).
[1]Present address: Marketing Department, Stern School of Business, New York University, 44 West 4th Street, KMC 9-77, New York, NY 10012-1126. Tel.: +1-212-998-0500; fax: +1-212-995-4006.

0169-2070/02/$ – see front matter
PII: S0169-2070(01)00118-2

might not be compensatory to APR, and vice versa. Other card features such as annual fee, credit limit, cash rebate, etc, however, might be compensatory. For example, a consumer may be willing to pay a not-so-small annual fee if a card offers a lucrative cash rebate. Note, however, that the normally noncompensatory attributes may become compensatory in certain extreme ranges. For example, if the APR becomes so low that the credit card becomes an attractive way of financing, then even convenience seekers may start to make tradeoffs between APR and other card features. Furthermore, even when the compensatory weighted-additive rule is applicable, the marginal rate of substitution (MRS) between card attributes is unlikely to be constant. For example, when the credit limit on a card is low, the consumer may be willing to pay a higher fee to raise his or her credit limit, say a $10 fee hike for each $1,000 increase in credit limit. However, when the credit limit has been raised higher and higher, the consumer will be willing to pay less and less for each additional $1000 raise in credit limit. Therefore, it is reasonable to believe that the MRS between fee and credit limit diminishes as the credit limit increases.

From the above discussion, one can easily see that whether a consumer will accept or reject a credit card offer is a complex decision making behavior. The consumer utility function may be noncompensatory. Even if it is compensatory in certain ranges and between certain features, the utility function is unlikely to be linear and the MRS between various features is unlikely to be constant. Therefore, for different types of consumers, for different card features, and at different values of card attributes, the MRS between features may not be the same, and it may range anywhere from 0 or infinity (indicating noncompensatory decision making) to certain nonzero values (indicating compensatory decision making). As such, the standard linear utility function and constant MRS implied by the commonly used statistical models, such as probit and logit, will fail to provide an accurate prediction of consumer credit card adoption decisions. This calls for models that can accommodate nonlinear and/or noncompensatory consumer decision making.

Artificial neural networks (NN) have drawn considerable attention in many disciplines that involve pattern recognition and forecasting. This rich class of flexible nonlinear models can approximate any function (linear or nonlinear) arbitrarily well (see Hornik, Stinchcombe, & White 1989; White, 1990, among others). While NN have been widely studied in various applications,[2] they have also been found useful in several marketing and consumer behavior studies. For example, Bentz and Merunka (2000) used NN as a diagnostic and specification tool for multinomial logit (MNL) in modeling brand choice decisions. In a simulation study and a study on consumer patronage behavior, West, Brocket, and Golden (1997) find that NN can offer significant improvement over traditional linear models such as discriminant analysis and logistic regression because NN can capture nonlinear relationships associated with the use of noncompensatory decision rules. Agrawal and Schorling (1996) apply NN to forecast brand shares in grocery product categories, and find that NN is better able to handle nonlinearities in the data than the commonly used multinomial logit model. Kumar, Rao, and Soni (1995) compare NN with logistic regression in modeling the decision of a supermarket chain whether to carry new products, and find that NN are parsimonious, produce better classification, handle complex underlying relationships better, and are stronger at interpolation. These and other studies[3] show that NN are promising for classification and multiple criteria decision making in terms of predictive accuracy, adaptability, and robustness.

Despite its practical and theoretical importance, very few studies have been conducted on consumer credit card usage. Previous empirical work on the credit card market predominantly employs probit/logit models. For example, Canner and Cyrnak (1986) use a logistic regression model to understand

[2]See, for example, option pricing (Hutchinson, Lo, & Poggio, 1994; Garcia & Gencay, 2000), time series prediction (Swanson & White, 1995, 1997a,b; Balkin & Ord, 2000; Darbellay & Slama, 2000; Tkacz, 2001), stock market prediction (Gencay, 1998; Qi, 1999; Qi & Maddala, 1999), exchange rate forecasting (Gencay, 1999), and student performance prediction (Gorr, Nagin, & Szczypula, 1994).

[3]Additional studies comparing NN to traditional statistical models in classification and multiple criteria decision making include Malakooti and Zhou (1994), Dasgupta, Dispensa, and Ghose (1994), Hruschka (1993), Archer and Wang (1993), Tam and Kiang (1992), and Salchenberger, Cinar, and Lash (1992) etc.

the determinants of credit card usage for convenience or credit revolving. Calem and Mester (1995) investigate imperfect competition in the credit card market via a probit model.

The main purpose of the present paper is to fill the gap in the studies of consumer credit card adoption behavior by investigating the relevance of NN in meeting the afore-mentioned challenges. Specifically, we use the logit model as a benchmark to see if nonlinearity and noncompensatability modeled by NN help to improve prediction accuracy.

The rest of the paper is organized as follows. In Section 2, we provide a description of NN and benchmark statistical models, as well as the performance measures and test statistics. Section 3 describes the data, and the experimental design. Section 4 reports the empirical results. Managerial implications are provided in Section 5. Finally, conclusions and directions for future research are provided in Section 6.

2. Model description, performance measures, and test statistics

2.1. Model description

Assume that there is a utility attached to a credit card which is modeled as a function of card attributes X (such as APR, or annual fee), an associated parameter vector β, and some unobserved random term ε:

$$U = f(X, \beta) + \varepsilon \tag{1}$$

We assume that a consumer will accept the card with highest utility.

In our study, card adoption is modeled as a process of comparing the new offers with existing wallet cards, similar to the approach used by Yang and Allenby (2000). Specifically, we adopt the ideal reference point[4] approach which assumes that a person will accept an offer if the difference between the utility of the new offer (U^{new}) and the utility of a hypothetical card that features the best attributes of each of the wallet cards (U^{wallet}) exceeds some threshold value ($\delta > 0$):

$$U^{\text{new}} - U^{\text{wallet}} > \delta \tag{2}$$

or

$$[f(X^{\text{new}}, \beta) + \varepsilon^{\text{new}}] - [f(X^{\text{wallet}}, \beta) + \varepsilon^{\text{wallet}}] > \delta \tag{3}$$

where X^{new} contains the attributes of the offered card, and X^{wallet} is measured as containing the best value across the wallet cards for each card feature. For example, person h has 5 active cards in his wallet, and the best APR is 5.75%, the highest credit limit is \$5000, and the lowest annual fee is \$0 across these five cards. In this case, $X^{\text{wallet}} = (5.75\%, \$5000, \$0)$ though these attributes might not belong to one specific card. This is a reasonable assumption for people who basically want a card that is good in all features. We adopt this decision rule based on the fact that there is only a small proportion of credit card offers that were accepted in the data, which gives a strong signal that consumers are very selective in acquiring additional credit cards in the current marketplace.

Eq. (3) can be rewritten as:

$$-\eta < U^* \tag{4}$$

where $U^* = V - \delta$, $V = f(X^{\text{new}}, \beta) - f(X^{\text{wallet}}, \beta)$, and $\eta = \varepsilon^{\text{new}} - \varepsilon^{\text{wallet}}$. U^* is the increase in utility in excess of the threshold which we name as surplus utility, the higher the surplus utility, the more likely a credit card offer will be accepted. If we let the cumulative density function of η in Eq. (4) be a logistic distribution, denoted as $g(a) = 1/(1 + e^{-a})$, then we get a binary logit model where

$$\Pr(\text{a card is adopted, or } y = 1) = \Pr(-\eta < U^*) = g(U^*) \tag{5}$$

$$\Pr(\text{a card is not adopted, or } y = 0) = 1 - g(U^*) \tag{6}$$

We have two x variables: difference in interest rates and difference in fees, the theoretical basis for which is provided by Yang and Qi (2000) in a two-period utility maximization problem via budget

[4]Ideal reference point has been documented in the marketing literature. For example, Rajendran and Tellis (1994) show that consumers are more likely to use the best price paid previously as the reference price, rather than the worst or the average paid. Parasuraman, Berry, and Zeithaml (1988) deals with consumer perceptions of service quality.

allotment expansion. Canner and Cyrnak (1986) find that the usage pattern of credit cards is related to consumer demographic profile. It is also of interest in our study to identify how demographic information can aid in predicting consumer sensitivity to card attributes which further leads to decisions on card adoption. This information can help credit card firms better understand the preferences of consumers and optimally customize their offers. Firms can also do a better job keeping old customers by understanding their interests and concerns. Therefore, we will combine all these demographic variables to the two card feature variables to explain the consumers' decision whether to accept a card offer or not. We also add interaction terms of card attribute characteristics (X) and individual respondent's demographic variables (Z) in order to capture person-specific threshold value.

In the traditional logit model, the utility function is linear additive. Thus in Eq. (4), V takes the form $V = (X^{\text{new}} - X^{\text{wallet}})'\beta$. For the jth card offered to the hth individual, the model thus can be written as

$$V_{hj} = (X_{hj}^{\text{new}} - X_{hj}^{\text{wallet}})'\beta \tag{7}$$

$$\delta_{hj} = Z'_h\alpha + [(X_{hj}^{\text{new}} - X_{hj}^{\text{wallet}}) \otimes Z_h]'\gamma \tag{8}$$

where $h = 1, \ldots, H$ (H is the total number of individuals in the sample); and $j = 1, \ldots, J_h$ (J_h is the total number of observations from the hth individual); vector Z_h contains five individual demographic variables including age, income, education, gender, and a dummy variable indicating whether the respondent uses the credit card for convenience or as a source of revolving debt; and α and γ are parameter vectors associated with the consumer demographic profiles and their interaction with the card features, respectively. As is obvious from Eqs. (4), (7) and (8), when demographics and interaction between demographics and card attributes are brought into the utility function, the MRS between the two card feature variables will be[5]

$$\frac{MU_1^*}{MU_2^*} = \frac{\beta_1 - Z'_h\gamma_1}{\beta_2 - Z'_h\gamma_2} \tag{9}$$

[5]See, for example, Varian (1993) for the definition of MRS. For convenience, we omitted the negative sign at the end.

Therefore, the MRS tends to vary across different groups of people with different profiles. However, for people with the same demographic profile, the MRS is constant according to the linear utility specification.

The logit model described above implies that the credit card adoption decision is guided by a compensatory linear weighted-additive utility function. As we discussed in the introduction, for convenience seekers, APR may be irrelevant and thus may not be compensatory to other card attributes, such as annual fee and credit limit. The opposite may be true for credit revolvers. Furthermore, even if some card features may be compensatory in certain ranges, the MRS is unlikely to remain constant within the group of people with the same demographic profile. These pose severe challenges to the traditional statistical models such as probit and logit that imply linear utility function, compensatory decision making rule, and constant MRS.

NN are a class of flexible nonlinear models inspired by the way in which the human brain processes information. Given an appropriate number of hidden-layer units and sufficient data, NN can approximate any nonlinear (or linear) function to an arbitrary degree of accuracy through the composition of a network of relatively simple functions. The flexibility and simplicity of NNs have made them a popular modeling and forecasting tool across different research areas in recent years. Here we briefly describe the NN model we used to model consumer credit card adoption decision. See Ripley (1994) for a more detailed introduction to NN models and a survey on neural networks for classification, Kuan and White (1994) for an econometric perspective, Gorr (1994) for a forecasting perspective, Qi (1996) for financial applications, and Zhang, Patuwo and Hu (1998) for a more recent survey.

A variety of different NN models have thus been developed, among which the three-layer feedforward network is the most widely used and is adopted in the present study. Let f be the unknown utility function (linear or nonlinear) through which a vector of explanatory variables $X = (x_1, x_2, \ldots, x_k)' = [(X^{\text{new}} - X^{\text{wallet}})', Z', (X^{\text{new}} - X^{\text{wallet}})' \otimes Z']'$ relates to the surplus utility U^*, i.e., $U^* = f(X)$. Then f can be approximated by a three-layer NN model.

The NN model can be written as:

$$U^* = f(X) = \alpha_0 + \sum_{j=1}^{n} \alpha_j g\left(\sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j}\right) \qquad (10)$$

where n is the number of units in the hidden layer, k is the number of explanatory variables (including card attributes, demographic information and the interaction terms, which are exactly the same as those used in the benchmark logit model), $\{\alpha_j,\ j = 0,1,\ldots,n\}$ represents a vector of parameters from the hidden to the output-layer units, and $\{\beta_{ij},\ i = 0, 1,\ldots,k,\ j = 0, 1,\ldots,n\}$ denotes a matrix of parameters from the input to the hidden-layer units. If n is too large, the NN may overfit in which case the in-sample errors can be made very small but the out-of-sample errors may be large. The choice of n depends on the number of explanatory variables and the nature of the underlying relationship. In the present study, we estimate the NN model of different hidden-layer units using the training data, and the Bayesian Information Criterion (BIC) is employed to select the best model[6]. The best model is then utilized to generate out-of-sample forecast results.

From Eq. (10), it is obvious that NN yield a flexible nonlinear utility function and unlike the linear utility function, the MRS between card features modeled by NN will be allowed to differ even among people with the same demographic profile. Therefore, the issues in modeling the complex consumer credit card adoption decision making process may be solved by an NN model.

2.2. Performance measures and test statistics

The prediction accuracy of the two alternative models are measured by MAD (mean absolute deviations) and RMSE (root mean square errors), as well as BIC. Though MAD and RMSE are common measures of prediction accuracy, they provide no information on the percentage of correct predictions of adoption and rejection decisions. The latter is much more relevant for direct marketing of credit card. Therefore, the percentage of correct predictions (PC) is also reported for the out-of-sample predictions. In calculating PC, the cutoff value 0.5 is first used, i.e., if the predicted probability, $\hat{y}$, is greater or equal to 0.5, then the consumer is predicted to accept the offer. Otherwise, he is predicted not to adopt the card. We compute the percentage of correct predictions for the entire holdout sample, only the offers that are accepted, and only the offers that are not adopted. We also report the percentage of correct predictions using a cutoff value of 0.32 (the sample probability of accepting a card offer) to see how sensitive the performance of each model is with the choice of cutoff probabilities.

The comparative analysis between the NN and logit model outlined in the present study has practical and theoretical importance. If the NN model outperforms the logit model, it indicates that NN not only can be used as a more accurate prediction model, but also can help us better understand the consumer utility function and credit card adoption decision behavior. It suggests that consumer utility function is nonlinear, the MRS is changing, and the credit card adoption decision is more likely based on noncompensatory decision rule rather than constant MRS linear compensatory decision rule as implied by the logit model. Therefore, it is important to test whether the difference in the prediction accuracy between the two alternative models is statistically significant. For this purpose, we use both Diebold and Mariano (1995) (DM) and Wilcoxon's signed-ranks (SR) tests for the significance of the difference between the squared forecast errors of the alternative models. The p-values of both tests will be reported along with MAD, RMSE, and BIC.

3. Data

3.1. Data description

Data were obtained from a national survey of credit card usage that started from April 1994 and ran through December 1996. Respondents were asked to record attributes of their current portfolio of credit cards, including annual percentage rates (APR), annual fees, credit limits, card issuer and other card type measures such as gold, platinum,

[6]We thank Reviewer 2 for suggesting BIC to us. BIC selected the neural network with 4 hidden layer units. In a previous version of the paper, AIC was used which selected the NN with 20 hidden layer units. The more parsimonious NN model selected by BIC performs better in out-of-sample prediction.

affinity card, etc. Respondents also reported their usage rates and balance on each of their wallet cards. Other information was provided for a variety of socio-demographic data including respondents' age, gender, income, education, revolving status (whether the person usually revolves balance or pays back immediately), and other variables.

When a new offer was accepted, the person would specifically list when the card was added and features associated with it. This provides us information about people's adoption behavior. We also need to determine the card features of those offers that were not accepted by respondents. We imputed this information by using the mean feature values of those cards that have been adopted. More specifically, since we know the distribution of card features for new cards adopted in each period, we can simply use mean values of these features as explanatory variables when a person does not accept a card offer during that same period. Based on the fact that there is a high frequency of credit card offers in the current marketplace, we further assume that people make a card adoption decision every quarter. The assumptions utilized here are crucial and necessary in some sense for analyzing people's card adoption behavior, though more accurate inferences can be made if respondents also recorded the features of those cards that ended up being rejected in each period of time. Unfortunately, this information is not available.

Although credit limit might be an important card attribute that could affect consumer card adoption decision for both convenience seekers and credit revolvers, there are a lot of missing values in this variable in the survey. Including it into the sample will significantly reduce our sample size, therefore, we decided to leave it out. The total number of respondents that are included in our sample is 271. Since there may be multiple observations on each individual in our sample, the total number of observations is 1804. The variable definitions and sample statistics are listed in Table 1. From Table 1, about 77% of the individuals in our sample are older than 35, only 12% of the people have annual income greater than $75,000. About 40% of the respondents are college graduates or are college students, 80% of them use credit cards for revolving debt, and 52% of the sample consists of males. Finally, in our sample 32% of the card offers are accepted.

3.2. Experimental design

Model construction and predictions are made in two different ways: within and cross-individual prediction. This is made possible by dividing the training and forecasting samples in corresponding ways. Normally, there are multiple observations on each individual in our sample. To generate within-individual prediction, the last observation of each individual with 2 or more observations is drawn to form the testing sample, and the remaining observations are pooled together to estimate model parameters. This way the training sample consists of 1551 observations, and the testing sample 253 observa-

Table 1
Variable definition and sample statistics

Variable	Mean (Std.[a])
Choice (1 = adopted, 0 = otherwise)	0.32
Card attributes	
Difference in APR (annual percentage rate %)	2.01 (4.49)
Difference in fee (annual fee $)	2.73 (10.45)
Demographics	
Age (1 = age > 35, 0 = otherwise)	0.77
Income (1 = annual income > $75 K, 0 = otherwise)	0.12
Education (1 = college or above, 0 = otherwise)	0.40
Revolve (1 = revolver, 0 = transactor)	0.80
Gender (1 = male, 0 = female)	0.52

[a] Standard errors are reported only for continuous variables.

tions for within-individual modeling and forecasting. The cross-individual prediction is generated differently in that all observations from a randomly drawn group of individuals are used as the testing sample, and all observations from the remaining individuals are used to estimate the model parameters. We have 1321 observations in the training and 483 in the testing sample for cross-individual prediction.

The within and cross-individual predictions serve slightly different purposes. While the former is concerned with forecasting a consumer decision on a new card offer when the demographic information of that consumer has already been used in estimating the model parameters, the latter is just the opposite. A model with good cross-person prediction will be particularly useful for a company trying to attract new customers, or customers who have not yet had a credit card, although we anticipate that cross-individual prediction will be more difficult.

4. Empirical results

4.1. Prediction accuracy

The empirical results on various performance measures of the within and cross-individual predictions are reported in Tables 2–4. In Tables 2 and 3, both the in and out-of-sample MAD, RMSE, and BIC are reported for the NN and the logit models. The DM and SR test results are also given. From Table 2, it is obvious that for the within-individual prediction, NN are significantly better than the logit both in and out of sample. The in and out-of-sample MAD and RMSE of NN are just fractions of those of the logit. The BICs of NN are much smaller than those of the logit both in and out of sample. Both the DM and SR tests indicate that the difference in the squared out-of-sample prediction errors of the two

Table 4
Percentage of correct prediction on the card adoption behavior

	Over all	Adopted	Not adopted
Panel A. Cutoff Prob. = 0.5			
Within-individual			
NN	100.00	100.00	100.00
Logit	83.40	25.81	91.44
Cross-individual			
NN	97.31	94.65	98.99
Logit	64.80	29.41	87.16
Panel B. Cutoff Prob. = 0.32			
Within-individual			
NN	99.60	100.00	99.55
Logit	66.80	77.42	65.32
Cross-individual			
NN	97.72	95.72	98.99
Logit	63.77	64.17	63.51

Table 2
Within-individual prediction of credit card adoption decision

	In-sample			Out-of-sample		
	MAD	RMSE	BIC	MAD	RMSE	BIC
NN	0.0106	0.0681	−5.0100	0.0056	0.0291	−5.3882
Logit	0.3954	0.4430	−1.5432	0.3229	0.3674	−1.6091
	DM Test			SR Test		
Logit vs. NN	14.1478 (0.0000)			13.7808 (0.0000)		

Table 3
Cross-individual prediction of credit card adoption decision

	In-sample			Out-of-sample		
	MAD	RMSE	BIC	MAD	RMSE	BIC
NN	0.0174	0.0559	−4.7308	0.0402	0.1445	−1.4244
Logit	0.3521	0.4172	−1.6506	0.4088	0.4755	−1.2567
	DM Test			SR Test		
Logit vs. NN	11.3300 (0.0000)			17.9717 (0.0000)		

models are highly significant (*P*-values are 0). Similar patterns are observed for the cross-individual prediction reported in Table 3.

Table 4 provides the out-of-sample percentage of correct predictions for both models. Panels A and B report the results when the cutoff probabilities are 0.5 and 0.32, respectively. In Panel A, for within-individual predictions, while NN can correctly predict all the credit card offers (both adopted and not adopted), the logit model can only correctly predict 25.81% of the adopted, 91.44% of the not adopted offers, and 83.40% overall. For cross-individual predictions, NN can correctly predict 94.65% of the adopted, 98.99% of the not adopted, and 97.31% overall, yet, these percentages for the logit model are 29.41%, 87.16%, and 64.80%, respectively. From Panel A, one can easily see that the under performance of the logit model is primarily due to its inability to correctly predict the adoption decisions. Approximately only 1 out of 4 adoption-decisions are correctly predicted by the logit model despite that it can correctly predict most of the non-adoption decisions.

As the cutoff probability lowers to 0.32 (Panel B), the percentage of correct predictions of the NN model largely remain the same among the adopted, not adopted, and over all for both within and cross-individual predictions. Interestingly, although the logit model shows large improvement in predicting the adopted offers, it shows large deterioration in predicting the not adopted. Since the majority of the offers are not adopted, the overall percentages drop.

Another observation from Tables 2–4 is that by all performance measures (MAD, RMSE, BIC, and percentage of correction prediction), the within-individual predictions are slightly more accurate than the cross-individual predictions, and this pattern exists for both the NN and the logit model. This observation confirms our prediction that the cross-individual prediction will be less accurate because it deals with forecasting a consumer decision on a new card offer when the demographic information of that consumer has not been used in model estimation.

As discussed in the introduction, while the logit model implies constant MRS linear compensatory decision rule, the NN model can capture changing MRS nonlinear decision behavior. The superior performance of NN to the logit model reported in this section suggests that when consumers make a decision on whether or not to adopt a card offered to them, the card features are not linearly compensatory and thus are not considered to be linear tradeoffs. Furthermore, the MRS may be different even for the same person but at different values of card feature variables. These will be demonstrated further in the next subsection.

4.2. Model interpretation

To show that NN have the comparative advantage of modeling flexible nonlinear utility functions and allow changing MRS between card attributes, we compare the utility function and MRS obtained from the NN and the logit models. We try to address the following questions: Does demographic profile matter? How large is the variation in the MRS? How nonlinear is the utility function modeled by NN? How different are the estimated marginal effects on the adoption probability from each model?

We take the logit and NN models estimated for the within-person prediction to demonstrate the model interpretations, since the most observations are used in the estimation (1551). The parameter estimates of the logit are reported in Table 5[7]. Among the two card feature variables, APR is significantly negatively related to the adoption probability. Among the five profile variables, three are significant. The older the consumer, the higher the income, the less likely a card offer will be accepted. This is reasonable because this consumer group will be more financially established when they are older and have higher income, thus they tend to be very selective about the card to adopt. On the contrary, people who tend to use credit card to revolve their debt are much more likely to adopt a new card. Among the ten interaction terms, only APR×Age, APR×Gender, and Fee×Revolve are significant.

As shown in Section 2, the MRS implied by logit will be the same within the same demographic group

[7] Note that there may exist error serial correlation between observations that are drawn from the same household, which may invalidate the statistical significance of the logit model reported in Table 5.

Table 5
Parameter estimates of the logit model

Variable	Estimate	Std.	Chi-Square	P-value
Intercept	−0.626*	0.231	7.357	0.007
APR	−0.189*	0.056	11.504	0.001
Fee	−0.026	0.025	1.069	0.301
Age	−0.330**	0.154	4.605	0.032
Income	−0.502*	0.190	6.996	0.008
Education	−0.062	0.129	0.230	0.631
Revolve	0.585*	0.183	10.261	0.001
Gender	0.135	0.124	1.179	0.278
APR×Age	0.079**	0.038	4.426	0.035
APR×Income	0.055	0.046	1.453	0.228
APR×Education	−0.024	0.031	0.581	0.446
APR×Revolve	−0.006	0.046	0.015	0.902
APR×Gender	−0.052***	0.030	3.061	0.080
Fee×Age	−0.009	0.014	0.406	0.524
Fee×Income	0.011	0.016	0.484	0.487
Fee×Education	0.004	0.011	0.152	0.697
Fee×Revolve	0.036***	0.020	3.167	0.075
Fee×Gender	0.001	0.011	0.009	0.924

Note: Superscripts *, **, and *** indicate significant at 1%, 5%, and 10% respectively.

and different across, and for NN, the MRS can be different even within the same demographic group. We thus compute the MRS for each demographic group in our sample for both the logit and NN models, and the results are provided in Table 6. Columns (1)–(6) in Table 6 list all possible demographic groups, the total number of which is $2^5 = 32$, and Column (6) shows the number of observations belonging to each particular group. Column (0) gives a number (Group Code) to represent the group. While Column (7) gives the MRS from the logit, Columns (8) to (11) provides the descriptive statistics of the MRS from the NN since they are different even within the same group. Note that from Column (6) there are no observations for 6 out of 32 groups, thus we do not need to report the MRS for these 6 groups. Among the remaining 26 groups, the one that consists of older, poorer, and less educated female revolvers has the most observations (331, Group 19), and the one that consists of older, poorer, and less educated male revolvers has the second most observations (247, Group 20).

From Table 6, the MRS modeled by logit ranges from −80 to 23 across different groups, indicating the importance of demographic information. The range of the MRS is much larger for NN, from −90 to 164.[8] The extreme values of MRS can be interpreted as indications of noncompensatory decision behavior of the consumers. Take, for example, the maximum MRS of 164 in Group 14 (consists of younger, richer, better educated male non-revolvers) indicates that this individual is much more sensitive to APR than to Fee. Therefore, for this particular consumer, a low fee but high APR card offer is not going to be attractive. For consumers with extremely high MRS, a profit-seeking marketer should offer them cards with medium or low APR but high fees. Group 26 (consists of older, richer, less educated male non-revolvers) is just the opposite. The MRS of all five consumers in that group are zero, suggesting they are not sensitive to APR at all (noncompensatory decision making). For this cohort, a credit card company should not offer them lower APR but higher fee cards. The MRS's implied by the NN model are plotted in Fig. 1 for the largest two

[8]For one observation in each of the Groups 14, 17, and 22, the marginal utility with respect to both APR and Fee are all zero, rendering the MRS of that individual undefined.

Table 6
MRS between APR and fee for logit and NN

(0) Group code	(1) Demographic profile Age	(2) Inc.	(3) Edu.	(4) Rev.	(5) Gen.	(6) Count	(7) Logit	(8) NN Mean	(9) Std.	(10) Min	(11) Max
1	0	0	0	0	0	0	–	–	–	–	–
2	0	0	0	0	1	8	10	2	3	0	6
3	0	0	0	1	0	74	−19	0	4	−16	20
4	0	0	0	1	1	77	−22	1	1	−2	1
5	0	0	1	0	0	8	10	6	11	−7	30
6	0	0	1	0	1	15	13	2	4	−2	7
7	0	0	1	1	0	74	−15	0	1	−7	1
8	0	0	1	1	1	55	−17	0	1	−2	0
9	0	1	0	0	0	0	–	–	–	–	–
10	0	1	0	0	1	0	–	–	–	–	–
11	0	1	0	1	0	0	–	–	–	–	–
12	0	1	0	1	1	5	−9	1	1	0	2
13	0	1	1	0	0	0	–	–	–	–	–
14*	0	1	1	0	1	7	23	NaN	NaN	10	164
15	0	1	1	1	0	0	–	–	–	–	–
16	0	1	1	1	1	3	−8	2	0	2	2
17*	1	0	0	0	0	45	3	NaN	NaN	−90	59
18	1	0	0	0	1	42	5	2	6	−6	21
19	1	0	0	1	0	331	−80	1	1	−4	3
20	1	0	0	1	1	247	−67	0	1	−2	1
21	1	0	1	0	0	11	4	4	6	−11	8
22*	1	0	1	0	1	32	6	NaN	NaN	−14	44
23	1	0	1	1	0	74	−24	2	2	1	8
24	1	0	1	1	1	239	−28	0	0	−2	0
25	1	1	0	0	0	9	2	6	6	−4	10
26	1	1	0	0	1	5	5	0	0	0	0
27	1	1	0	1	0	39	−5	2	2	−3	3
28	1	1	0	1	1	14	−8	2	0	2	2
29	1	1	1	0	0	18	4	9	7	−4	20
30	1	1	1	0	1	32	7	5	10	−29	19
31	1	1	1	1	0	40	−5	5	11	−1	66
32	1	1	1	1	1	47	−7	3	2	−2	4

Note: *In Groups 14, 17, and 22, for one observation the marginal utility with respect to both APR and Fee are all zero, rendering the MRS of that individual undefined.

demographic groups: the upper panel shows the 331 observations in the 19th group and the lower panel shows the 247 observations in the 20th demographic group.

To visualize the utility function modeled by the NN and the logit models, we plot the surplus utility against APR in Fig. 2. The upper panel shows the 331 observations in the 19th group and the lower panel shows the 247 observations in the 20th demographic group. The surplus utility is computed for both models using card feature variables (APR and Fee), demographic variables (Age, Income, Education, Revolve, and Gender, which are the same within each group), and their interactions. Then the calculated surplus utility is plotted against APR, with observations sorted by an ascending order of APR. As expected, in both panels, the surplus utility is negatively related to APR in a linear fashion for the logit and in a nonlinear fashion for the NN. Moreover, although each plot in Fig. 2 represents the same demographic group, while for logit, only one value of the surplus utility is corresponding to any particular value of APR, but for NN, the surplus utility may still vary depending on the Fee. Again, this

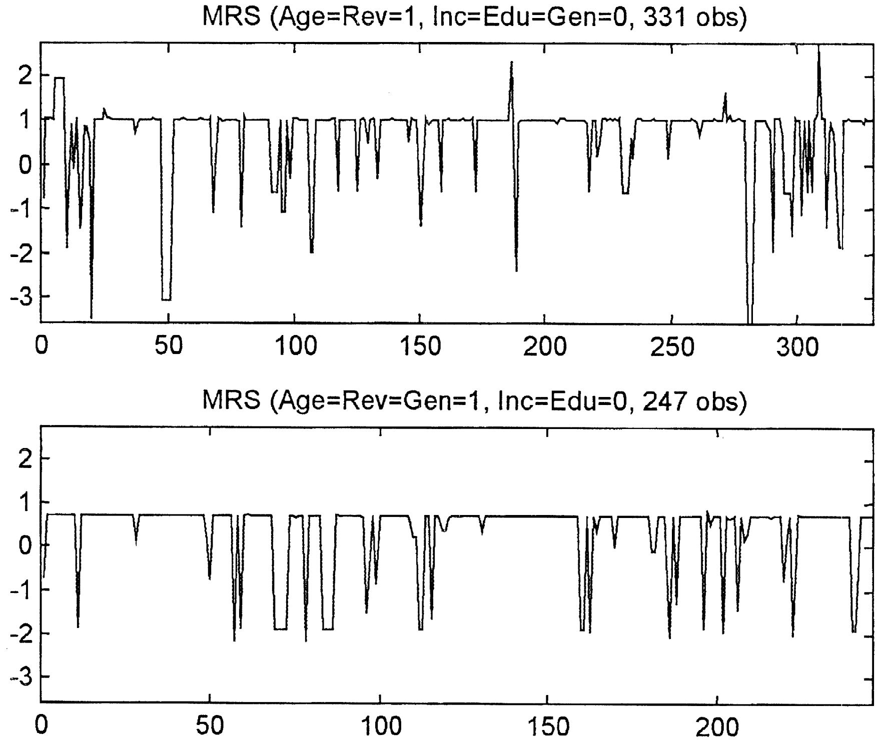

Fig. 1. MRS between APR and Fee implied by the NN model.

implies varying MRS even within the same demographic group.

Finally, Table 7 reports the descriptive statistics on the marginal effect of each card feature variable on the adoption probability for both NN and logit models for all the 1551 in-sample observations used in the estimation of the within-person prediction model. Although the signs of the mean marginal effect of both features agree among the two models, the marginal effects of the NN model have a much larger range than those of the logit with a more than doubled standard error for both features.

5. Managerial implications

Our findings have important managerial implications for credit card marketers. First, targeting the right group of people is crucial for cost savings and profit maximization. The more accurate prediction of NN can help credit card companies identify customers who are most likely to adopt the card they offer. To illustrate this point, we conduct a simple profit analysis on the forecasts generated by the NN and the logit model.

Assume that the life-time earnings that can be generated from a credit card is m (this can be calculated roughly by the annual profit from an average card times the average number of years in which the card is active). Also assume the cost of targeting a potential consumer is c (for purchasing information of potential customers, mailing and contacting, and other promotions). A credit card offer will be made to a potential customer if a model predicts an adoption probability of 0.5 or higher. Depending on whether to make an offer or not, and whether the offer is accepted or not, there are four scenarios. If an offer is made and accepted, then a profit of $(m - c)$ will be generated; if an offer is made but not accepted, a cost of $(-c)$ will occur; if an offer is not made, no matter whether the person targeted will accept or not, there will be no profit or loss. The results of profit analysis are reported in Table 8 in which the profits are given under three

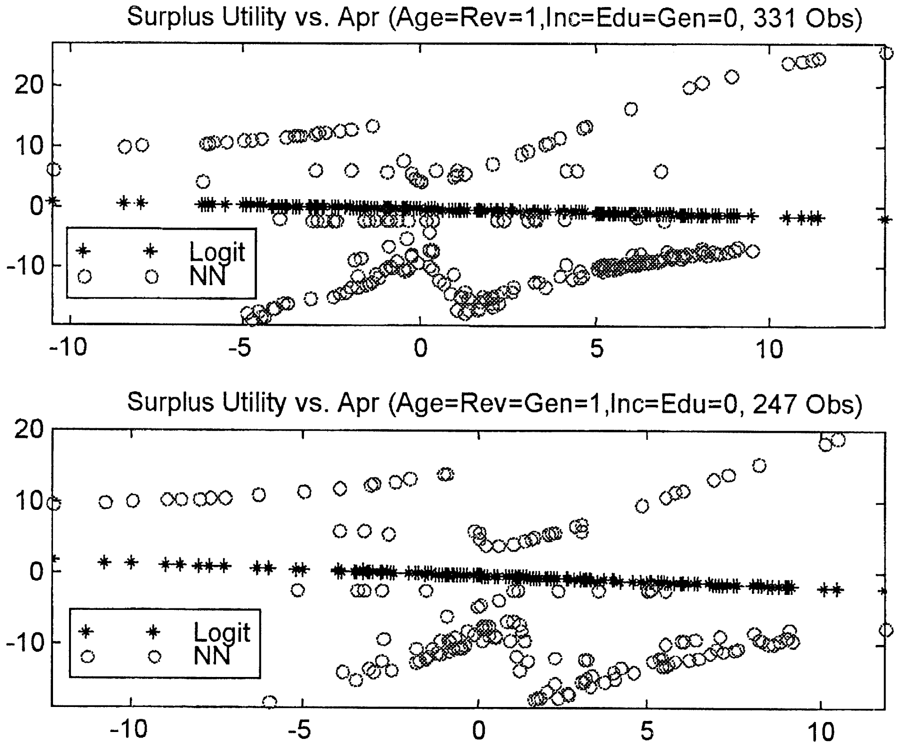

Fig. 2. Plot of surplus utility versus APR modeled by NN and logit.

Table 7
Descriptive statistics of the marginal effect on the probability of adoption

	Mean	Std.	Min	Max
Logit				
APR	−0.0319	0.0124	−0.0675	−0.0049
Fee	0.0005	0.0024	−0.0086	0.0067
NN				
APR	−0.0090	0.0328	−0.5991	0.4499
Fee	0.0024	0.0092	−0.0116	0.1165

possible values of m ($100, $300, and $500), and two possible marketing costs c per offer ($10, $20).

It is clear from Table 8, the NN generates much larger profits than the logit model under all six m and c combinations for both within or cross-individual predictions. In general, for within-individual predictions, the profit from NN and logit are $31(m-c)$ and $(8m-27c)$, respectively, which imply that as long as $m/c \geq 0.17$ NN will be more profitable than logit. For cross-individual predictions, the profit from NN and logit are $(177m-180c)$ and $(55m-93c)$, respectively. This implies that as long as $m/c \geq 0.71$, NN will be more profitable than logit. In reality, the m/c almost always far exceeds 0.17 or 0.71, thus a company using a NN prediction model will almost always have a higher profit than a similar company who uses a logit model.

To see whether the profit results are sensitive to the choice of cutoff probabilities, we also report in Table 8 the results of a cutoff value of 0.32, which is the probability of accepting an offer in our sample. At this lower cutoff value, while the NN predicts almost the same number of adoptions as with the 0.5 cutoff, the logit predicts a lot more adoptions. However, for the logit model the reduction in Type I error is at the cost of a significant increase in Type II error. As a result, for both within and cross-individual predictions, the NN model still generates much higher profits than the logit model under all m and c combinations. In fact, as long as $m/c \geq -9.86$ (or -0.78), the NN will be more profitable than the logit for within (or cross) individual predictions.

Table 8
Within and cross-individual profit analysis

	Within-individual		Cross-individual	
(m, c)	NN	Logit	NN	Logit
Panel A. Cutoff Prob. = 0.5				
(100, 10)	\$2,790	\$530	\$15,900	\$4,570
(300, 10)	\$8,990	\$2,130	\$51,300	\$15,570
(500, 10)	\$15,190	\$3,730	\$86,700	\$26,570
(100, 20)	\$2,480	\$260	\$14,100	\$3,640
(300, 20)	\$8,680	\$1,860	\$49,500	\$14,640
(500, 20)	\$14,880	\$3,460	\$84,900	\$25,640
Profit formula	$31(m-c)$	$8m-27c$	$177m-180c$	$55m-93c$
Panel B. Cutoff Prob. = 0.32				
(100, 10)	\$2,780	\$1,390	\$16,080	\$9,720
(300, 10)	\$8,980	\$6,190	\$51,880	\$33,720
(500, 10)	\$15,180	\$10,990	\$87,680	\$57,720
(100, 20)	\$2,460	\$380	\$14,260	\$7,440
(300, 20)	\$8,660	\$5,180	\$50,060	\$31,440
(500, 20)	\$14,860	\$9,980	\$85,860	\$55,440
Profit formula	$31m-32c$	$24m-101c$	$179m-182c$	$120m-228c$

Second, understanding that consumer credit card decision is based on a nonlinear and noncompensatory rule can help credit card companies to design new cards that offer the right features to the right group of people. For example, for consumers who use credit cards primarily for convenience, low APR will not be attractive. Credit card companies should entice a convenience seeker with other features, such as high credit limit or cash rebate. For debt revolvers, however, low APR and high credit limit are the key features that affect their adoption decision. Other attributes such as low annual fee and cash rebate may not be necessary.

6. Conclusions and discussions

We investigated the relevance of NN that can model nonlinear utility functions and changing MRS in consumer credit card adoption behavior. The results are compared to the widely used logit model. We find that NN predict much more accurately than the logit model in terms of smaller mean absolute errors and root mean square errors and higher percentage of correct predictions. The same pattern carries through both within and cross-individual predictions. Our model interpretation indicates that the credit card adoption by consumers is likely to be based on a nonlinear utility function that implies changing MRS, and consumers do not make constant linear tradeoffs among card attributes. Furthermore, the MRS may vary even within the same demographic group.

Though the results of our study have demonstrated some clear conclusions based on our sample, more work can be done in the following area. First, since we are constrained by a fairly small sample size and limited number of card attributes, we could have overstated the issue of nonlinearity and noncompensatability. As the sample size increases, we may be able to detect more useful findings between demographic profile and the attribute sensitivity. This will directly guide firms to better customizing their card offers and increasing card adoption rate.

Second, in the current study, we focus on the issue of nonlinear utility function and changing MRS and their impact on consumer decision making. However, our NN model is aggregate in nature which only captures the observed heterogeneity (different response rate for people with different demographics or purchase history) while ignoring unobserved heterogeneity (information that is not revealed to a researcher, e.g., motivation, Allenby & Rossi, 1999). Therefore, one future direction is to explore models that can account for both heterogeneity and nonlinearity. Finally, it is a fairly common practice among the credit card companies to partially customize card offers to potential customers based on their profile. Therefore, it is reasonable to view the card features as endogenously determined. It will be interesting to investigate what effect this might have on the conclusions and managerial implications of the study. We leave this and other issues for future research.

Acknowledgements

We would like to thank Michael Butler, Michael Y. Hu, Chung-Ming Kuan, Lawrence Marsh, and Randall Mount for helpful conversations and comments. We are indebted to Wilpen L. Gorr (the editor), V. Kumar (the associate editor) and the anonymous referees whose comments have led to an improvement of the paper. Min Qi acknowledges with thanks the financial support from Kent State University's Research Council and College of Business Administration. Any remaining errors are our own responsibility.

References

Agrawal, D., & Schorling, C. (1996). Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model. *Journal of Retailing*, *72*, 383–407.

Allenby, G. M., & Rossi, P. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, *89*, 57–78.

Archer, N. P., & Wang, S. (1993). Application of the back propagation neural network algorithm with monotonicity constraints for two-group classification. *Decision Science*, *24*, 60–75.

Balkin, S. D., & Ord, J. K. (2000). Automatic neural network modeling for univariate time series. *International Journal of Forecasting*, *16*, 509–515.

Bentz, Y., & Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modelling: a hybrid approach. *Journal of Forecasting*, *19*, 177–200.

Calem, P. S., & Mester, L. J. (1995). Consumer behavior and the stickiness of credit-card interest rates. *The American Economic Review*, *85*, 1327–1336.

Canner, G. B., & Cyrnak, A. W. (1986). Determinants of consumer credit card usage patterns. *Journal of Retail Banking Services*, *8*, 9–18.

Darbellay, G. A., & Slama, M. (2000). Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, *16*, 71–83.

Dasgupta, C. G., Dispensa, G. S., & Ghose, S. (1994). Comparing the predictive performance of a neural network model with some traditional market response models. *International Journal of Forecasting*, *10*, 235–244.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Garcia, R., & Gencay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, *94*, 93–115.

Gencay, R. (1998). The predictability of security returns with simple technical trading rules. *Journal of Empirical Finance*, *5*, 347–359.

Gencay, R. (1999). Linear, nonlinear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, *47*, 91–107.

Gorr, W. L. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*, *10*, 1–4.

Gorr, W. L., Nagin, D., & Szczypula, J. (1994). Comparative study of neural network and statistical models for predicting professional student grade point averages. *International Journal of Forecasting*, *10*, 17–34.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359–366.

Hruschka, H. (1993). Determining market response functions by neural network modeling: a comparison to econometric techniques. *European Journal of Operational Research*, *66*, 27–35.

Hutchinson, J., Lo, A., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, *49*, 851–889.

Kumar, A., Rao, V. R., & Soni, H. (1995). An empirical comparison of neural network and logistic regression models. *Marketing Letters*, *6*, 251–263.

Kuan, C., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Review*, *13*, 1–91.

Malakooti, B., & Zhou, Y. Q. (1994). Feedforward artificial neural networks for solving discrete multiple criteria decision making problems. *Management Science*, *40*, 1542–1561.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Marketing*, *64*, 12–40.

Qi, M. (1996). Financial applications of artificial neural networks. In Maddala, G. S., & Rao, C. R. (Eds.), *Handbook of statistics*, vol. 14: Statistical methods in finance. Amsterdam: Elsevier Science.

Qi, M. (1999). Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business & Economic Statistics*, *17*, 419–429.

Qi, M., & Maddala, G. S. (1999). Economic factors and the stock market: A new perspective. *Journal of Forecasting*, *18*, 151–166.

Rajendran, K. N., & Tellis, G. J. (1994). Contextual and temporal components of reference price. *Journal of Marketing Research*, *58*, 22–34.

Ripley, B. D. (1994). Neural networks and related methods for classification. *Journal of Royal Statistical Society B,* 409–456.

Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Science*, *23*, 899–916.

Swanson, N. R., & White, H. (1995). A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, *13*, 265–275.

Swanson, N. R., & White, H. (1997a). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, *79*, 540–550.

Swanson, N. R., & White, H. (1997b). Forecasting economic time series using flyexible vs. fixed specification and linear vs. nonlinear economic models. *International Journal of Forecasting*, *13*, 439–461.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, *38*, 926–947.

Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*, *17*, 57–69.

Varian, H. R. (1993). *Intermediate microeconomics: a modern approach*, 3rd ed. New York: Norton.

West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, *16*, 370–391.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, *3*, 535–549.

Yang, S., & Allenby, G. M. (2000). A model for observation, structural, and household heterogeneity in panel data. *Marketing Letters*, *11*, 137–149.

Yang, S., & Qi, M. (2000). Heterogeneous consumer preferences for instant and future benefits: an application to credit card adoption. Technical Report, A. Gary Anderson Graduate School of Management, University of California, Riverside.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35–62.

Biographies: Min QI is an assistant professor of economics at Kent State University. She obtained her Ph.D. from the Ohio State University. Her research interests are applied econometrics, financial econometrics, computational methods, forecasting, and neural networks. Her work has appeared in *Journal of Business & Economic Statistics, Journal of Forecasting, IEEE Transactions on Neural Networks, Journal of Empirical Finance, International Journal of Forecasting, Handbook of Statistics* (North-Holland, 1996), and *Computational Finance* (MIT Press, 2000), etc. Her profile has been published in *Who's Who in America* and *Who's Who in Finance and Industry*.

Sha YANG is assistant professor of Marketing at the Stern School of Business, New York University. She obtained her Ph.D. from the Ohio State University. Her research interests include modeling consumer choice and panel data analysis, consumer heterogeneity and market segmentation, and Bayesian methods. Her recent work has appeared in *Marketing Science* and *Marketing Letters*.