The Profiles of Late-Paying Consumer Loan Borrowers: An Exploratory Study

A Note by James S. Ang, Jess H. Chua, and Clinton H. Bowling

1. Introduction

Empirically derived credit-scoring models have been extensively examined in the literature, e.g. [1]. The attention received by the models will probably increase since the Equal Credit Opportunity Act (ECOA) recognizes these credit-scoring models as a basis for nondiscriminatory loan granting decisions, but stipulates that the models must be "demonstrably and statistically sound."

One commonly acknowledged problem with existing credit-scoring models is the treatment of late-paying loans. Some models have simply ignored late-paying loans and excluded them from the empirical samples. Others have arbitrarily classified late-paying loans as good loans. Therefore, they have either failed to account for the existence of late-paying loans or failed to recognize the considerable costs involved in late-payment notices, follow-ups, and collections. Since late-paying loans are more commonly encountered than outright defaults, it might be argued that existing credit-scoring models have misplaced their emphasis as far as credit evaluation is concerned.

Credit-scoring models that can also discriminate between late-paying and good loans are needed. However, before we plunge into the process of constructing credit-scoring systems, it is worthwhile to see if statistically significant differences can be found to differentiate late-paying loans. This is the purpose of the present study. The result sought is an understanding of the profiles of late-paying borrowers rather than a predictive model. It is hoped that such an understanding will help provide a list of the relevant variables that must be included in a predictive model. It may also indicate the type of predictive model that would be most appropriate.

2. Data

A random sample of 180 late-paying loans was drawn from the file of three branches of a major California bank. They are all consumer automobile loans

JAMES S. ANG *is professor of finance, Oklahoma State University.* JESS H. CHUA *is assistant professor of finance, Oklahoma State University.* CLINTON H. BOWLING *is loan officer, First National Bank of Oklahoma City.*

0022-2879/79/0579-0222$00.50/0
JOURNAL OF MONEY, CREDIT, AND BANKING, vol. 11, no. 2 (May 1979)

granted in the period 1970–73. The personal characteristics and payment records used were constructed from computer printouts, individual credit files, and loan applications.

The purpose of the study is to characterize late-paying loans. Therefore, a measure of "lateness" is necessary. Since the loans in the sample involve different total numbers of loan payments, the number of delinquent payments is not an appropriate measure. Instead, the percentage of total number of payments that were late was used.

The variables available in the data files to characterize late-paying loans were: gross amount of loan, age, sex, marital status, number of dependents, years lived at residence, monthly take-home pay, monthly take-home pay of spouse, own or rent residence, other monthly income, total monthly payments on all debts, type of bank accounts, number of credit references listed, years on job, total family income per month, debt to income ratio, total number of payments on the loan, and annual percentage interest on the loan.

3. *Methodology*

Conventional clustering techniques were not used because of several possible complications. First, many of the variables are categorical (marital status, sex, rent or own residence) or measured in discrete and unequal intervals (under $400 or over $2,000 in monthly take-home pay). Second, there may be nonlinear relationships. For example, there may be "threshold-type" relationships. Late payments may take a quantum jump when certain values for some of the variables are obtained. Third, there may be considerable interaction effects. Certain variables may be predictors of late payment only in the presence of other specific variables. To allow for these possibilities, the Automatic Interaction Detector (AID) [2] was used.

4. *Results*

The AID technique was applied to the 180 borrowers sampled. The results are presented in Figure 1. The late-paying loans were classified into fourteen groups from the lowest average percentage payments late (7 percent) to the highest (66 percent). The technique found seven variables relevant in characterizing late-paying loans. These are: residence ownership, debt/income ratio, age of borrower, number of payments, number of years at residence, total monthly income, and number of years employed at present job. Altogether, 39.3 percent of the total variance is explained by these seven variables. The F-value of 6.06 indicated that the predictors are significant at the 0.001 level. A breakdown of the portion of variance explained by each variable is presented in Table 1.

The characteristics of each final group can be obtained by tracing the relevant branch of the tree backward. For instance, the borrowers with the lowest late payment rate (group 1) had the following profile: A loan with over 31 payments, over 29 years of age, on the job over 5 years, and lived at the same residence for over 10 years. On the other hand, the borrowers with the highest percentage of late payments

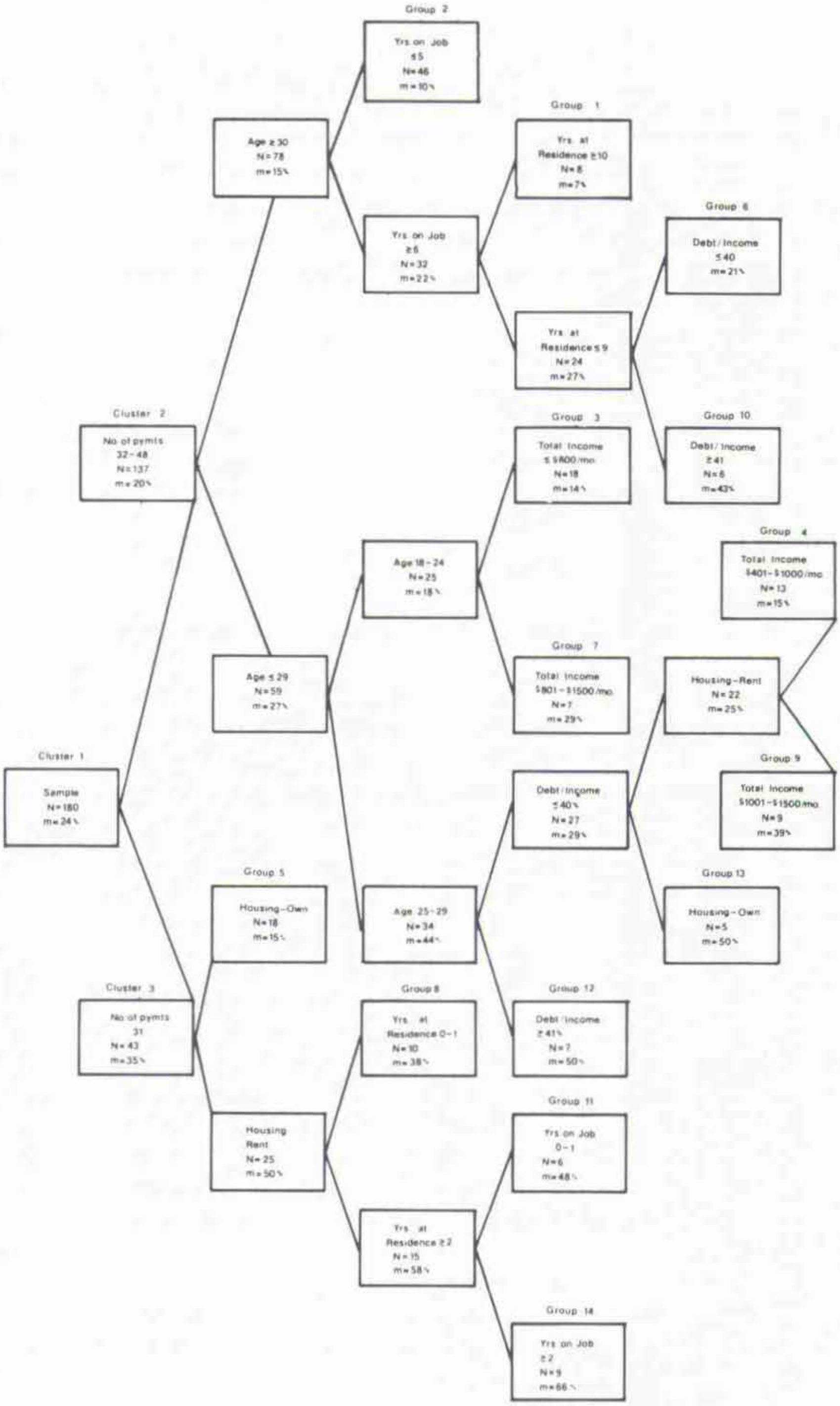

Fig. 1. Borrower Profiles as Predictors for Repayment. (N = number of loan; m = average percent of late payment in the sample).

TABLE 1

THE PROPORTION OF VARIATION IN THE PERCENTAGE OF LATE PAYMENT EXPLAINED BY THE PREDICTORS (Sample Size = 180)

Predictors	Portion of Total Variance Explained
Residence ownership (own or rent)	0.119
Debt to income ratio	0.063
Age of borrower	0.061
Number of payments	0.058
Number of years at residence	0.036
Total monthly family income	0.032
Number of years employed at present job	0.026
r^2	0.393

(group 14) had been on their jobs over 2 years, lived at their present residences over 2 years, rented their homes, and had a loan of 31 or fewer payments.

Closer examination of the results reveals nonlinear and interactive relationships concerning every predictor. We shall discuss those of age in detail because of their implications for the ECOA. The ECOA allows age to be used as a predictor in credit-scoring models as long as it is based on experience and statistical evidence. However, it prohibits the assignment of adverse scores to applicants over 62 years old on the basis of their ages. This policy would be rational only if late-payment and default rates exhibit nonlinear relationships with age; i.e., within a certain age range, age is relevant, but outside this range, it is not.

First, observe that, when loans involve 31 or less payments, age is not a relevant variable at all. When loans involve more then 31 payments, borrowers were partitioned into those aged 30 or over and those less than 30. The older ones were shown to have a lower average late-payment percentage. The younger group was further split into 18–24 years old and 25–29 years old. Note the surprising result, showing that the 25–29 age group actually had a higher frequency of late payments. More importantly, the technique did not further partition the older group according to age again, as it did the younger group. This shows that as long as the borrowers are 30 or over, differences in late-payment rates are due to factors other than age. Therefore, at least in this sample, the ECOA stipulations concerning age are rational and not restrictive.

Other results that have implications for the ECOA are those on sex and marital status. Neither one was a relevant variable for characterizing delinquent borrowers. Since data on race, color, and religion were not collected, there was no way the study could determine their relevance.

5. *Suggestions for Further Research*

The results of AID analysis have shown that the relationships between late payments and predictor variables are highly nonlinear. The existence of interactions cause combinations of variables to produce irregular late-payment patterns. This implies that linear credit-scoring models will probably be inadequate in accounting for the differences in borrower quality. We would suggest the following approach.

Since linear credit-scoring models are still the most convenient to construct, they may still be used. However, several models will have to be used to account for the nonlinearities. For example, at least two models (one for loans with over 31 payments and one for loans with 31 or less payments) are indicated by the results here. The marginal improvement in predictive ability of two or more linear models over one universal model should be measured.

LITERATURE CITED

1. Orgler, Yair E. "A Credit-Scoring Model for Commercial Loans." *Journal of Money, Credit, and Banking*, 2 (November 1970), 435–45.

2. Sonquist, John A., and James N. Morgan. *The Detection of Interaction Effects*. Ann Arbor: Institute for Social Research, 1964.