This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Risk Elements in Consumer Instalment Financing, Technical Edition

Volume Author/Editor: David Durand

Volume Publisher: NBER

Volume ISBN:

Volume URL: http://www.nber.org/books/dura41-1

Publication Date: 1941

Chapter Title: Appendix B: Application of the Method of Discriminant Functions to the Good- and Bad-Loan Samples

Chapter Authors: David Durand

Chapter URL: http://www.nber.org/chapters/c12956

Chapter pages in book: (125 - 142)

Appendix B

Application of the Method of Discriminant Functions to the Good- and Bad-Loan Samples

CONSIDERING the fundamental assumption of a dichotomous classification of loans, the problem of analysis is to discover differences between the good-loan and bad-loan distributions. The factors analyzed in this report fall into two rough categories: the qualitative attributes like occupation and marital status, and the quantitatively measurable variates like income and number of years at present address. Analysis of the qualitative attributes may be made by comparing the proportion of good loans in a given occupational group, for example, with the proportion of bad loans. Analysis of the quantitatively measurable factors can, of course, be carried out by the same process. The proportion of good loans in any income class can be compared with the proportion of bad loans; but one further step in the analysis is generally desirable and possible. A difference between the income distributions of the good and bad loans usually can be translated into a difference in mean or average income, a difference in the standard deviation about the mean, a difference in skewness, a difference in kurtosis, etc.

In the non-technical sections of this report, the distributions of all factors, quantitative and qualitative alike, are shown on the same basis; in all cases the percentage of good and bad loans in each of a small number of class intervals is determined; and for the quantitative factors no attempt is made to measure mean value, variance, skewness, kurtosis, etc. Nevertheless, a difference in mean values is frequently obvious. For example, the good-loan samples in Table 11 undoubtedly have a longer average tenure of employment than the bad-loan samples, although the amount of the difference is not readily apparent. Differences in other measures, such as variance or skewness or kurtosis, are much more obscure; and the difficulty of analyzing these differences is often great. On the whole, the analysis of the quantitative factors thus far has consisted of a rough attempt to determine differences between the means of the good and bad loans.

The analysis in Chapter 3 consists of a set of individual treatments of separate factors. The differences that were discovered between the samples of good and bad loans related only to separate factors—income distribution, occupational distribution, etc. This individualistic approach has its shortcomings, however. A more satisfactory approach would be to consider each of the samples as a single distribution in a number of variates. Any difference between two distributions could be used for the purpose of differentiation; for example, the correlation coefficient between tenure of residence and tenure of occupation might be one value for the good loans and another for the bad. In practice, however, differences between means are the most obvious and by far the easiest to handle, i.e., when quantitatively measurable factors are concerned. For this purpose the use of discriminant functions, described in Appendix A, has two distinct advantages; it provides a means by which a number of credit factors can be weighted and combined into an index of credit risk; and it helps to indicate when individual analyses may be specious because of correlation between factors.

The method of linear discriminant functions is the ideal method of analysis when the two populations have multivariate normal distributions with equal variances and covariances but differing means. In the good- and bad-loan samples, where the assumed conditions are not actually met, this method is no longer ideal, but it may be a useful approximation.

An experiment with discriminant functions was carried out for the used-car sample. Four factors were singled out for analysis: cash price, actual down payment in dollars, purchaser's monthly income, and length of contract. These factors were chosen because they are fundamentals from which a number of other factors can be derived. From the ratio of down payment to price, the percent down payment is derived. The difference between price and down payment is the unpaid balance, which is usually a fair approximation to the amount of the note; the ratio of the unpaid balance to contract length is an approximation of the amount of the monthly payment; and the ratio of this last factor to income is an index of the burden of the debt upon the borrower's purchasing power. Instead of a separate investigation of all these derivative factors, a single discriminant function analysis of the four basic factors appears to be more systematic and more expedient.

The four selected factors have already received separate analysis. The distribution of cases was presented in Tables 4, 6, 7, and 8. A summary of these analyses is presented in Table B-1,

TABLE B-1

MEANS AND STANDARD DEVIATIONS OF NON-REPOSSESSED AND REPOSSESSED USED-CAR SAMPLE, BY PRICE, DOWN PAYMENT, INCOME, AND MATURITY

	Price	*Down Payment*	*Income*	*Maturity*
Mean (non-repossessions)	\$410	\$166	\$172	$13\frac{1}{2}$ mos.
Mean (repossessions)	\$344	\$119	\$148	$13\frac{1}{2}$ mos.
Difference	\$66	\$47	\$24	0 mos.
Standard deviation (both samples)	\$195	\$89	\$93	3.4 mos.
Ratio $\frac{\text{mean difference}}{\text{standard deviation}}$	.34	.53	.26	.00
Theoretical efficiency index[a]	13	21	10	0

[a] This index was not determined from the actual distribution of loans; it was computed from the ratio of mean difference to standard deviations by means of a table of the normal curve.

which presents means and standard deviations instead of percentage distributions. These values have been determined from the entire used-car sample of 484 non-repossessions and 485 repossessions of which 439 of the non-repossessions and 448 of the repossessions reported full data on price, contract length, down payment, and income.

This tabulation suggests that the first three variates are related to risk; that the order of importance is down payment, price, income; and that the last variate, contract length, is not related. As we have pointed out before, these conclusions may be specious if the correlation between the variates is high; and the correlation coefficients in Table B-2 indicate considerable correlation between some of the variates.

TABLE B-2

CORRELATION COEFFICIENTS FOR SELECTED FACTORS

Factor	*Price*	*Down Payment*	*Income*
Down payment	.87		
Income	.33	.29	
Length of contract	.62	.47	.05[a]

[a] Does not differ significantly from zero.

The discriminant function for these four factors was found to be $Z = d - .174p + .124i - 6.45m$, where d is the down payment in dollars, p is the price in dollars, i is the monthly income in dollars, and m is the length of contract in months.

The effectiveness of the function Z can be measured by the ratio of the difference between its two means (the mean for the good sample and the mean for the bad sample) to its standard deviation. The value of this ratio can be estimated without the actual computation and tabulation of the value of Z for each loan. The ratio is .63, which is an appreciable though not startling increase over the value of .53 for down payment alone; the corresponding efficiency indices computed from a table of the normal curve are 25 and 21. This increase is not striking;

if the factors had been independent, the ratio would have been .68, and the efficiency index would have been 27.

On the basis of these data we can now show that the individual analyses and their indications of the relative importance of factors are sometimes misleading. In the individual analysis, length of contract does not appear to be related to risk, for the good- and bad-loan samples have the same mean value. In the discriminant function Z, however, relation between contract length and risk does appear. Owing to the correlation between factors, the coefficient for length of contract is -6.45, which indicates that risks tend to improve as length becomes shorter. This inconsistency, as we have explained earlier, is attributable to the fact that few of the lower-priced used cars are financed on contracts of more than 12 months; for cars of the same price, the short terms are distinctly superior.

In the individual analysis, a high price appears to indicate good risk; but in the discriminant function, the price coefficient, $-.174$, indicates exactly the opposite. This apparent inconsistency can be explained by the high correlation between price and down payment. High price indicates good risk as long as it is accompanied by a high down payment, which is usually the case; but when down payments remain constant, the higher prices indicate poorer risks.

Nevertheless, the coefficients of the various factors are not entirely reliable as indices of the relative importance of the various factors. If the function Z is transformed to express the measurement of each variate in units of one standard deviation—a process analogous to the computation of the Beta-coefficients in multiple correlation—the transformed coefficients are somewhat more reliable, but they are not yet ideal. Transformed to units of one standard deviation, the discriminant function found above becomes $Z' = d - .382p + .131i - .246m$.

One possible way of measuring the relative importance of the factors is to determine discriminant functions for a number of combinations based on fewer than four factors. For combina-

TABLE B-3

CORRELATION COEFFICIENTS, MEAN DIFFERENCES, AND STANDARD DEVIATIONS, FOR SEVEN RISK FACTORS, COMPUTED FROM A COMMERCIAL BANK SUBSAMPLE OF 191 GOOD LOANS AND 190 BAD LOANS[a]

	Sex[b]	Real Estate[c]	Number of Years at Present Address	Nature of Occupation[d]	Number of Years at Occupation	Bank Account[e]	Life Insurance[f]
Correlation coefficients							
Real estate	−.09[g]						
	.03[g]						
Number of years at present address	.17	.16					
	−.03[g]	.30					
Nature of occupation	.56	.09[g]	.08[g]				
	.11[g]	.05[g]	−.01[g]				
Number of years at occupation	.00[g]	.24	.12[g]	.09[g]			
	.03[g]	.22	.15	.07[g]			
Bank account	−.04[g]	.26	.09[g]	.21	.25		
	.11[g]	.21	.02[g]	.15	.01[g]		
Life insurance	.20	.05[g]	−.04[g]	.11[g]	.13[g]	−.13[g]	
	.03[g]	−.13[g]	−.12[g]	.06[g]	−.20	.12[g]	
Mean	.241	.241	5.937	1.094	8.785	.492	.235
	.090	.090	5.160	.695	5.884	.205	.305
Difference	.151	.151	.777	.399	2.901	.287	.070
Standard Deviation	.303	.303	5.15	.562	5.81	.353	.300
	.202	.202	5.40	.503	4.88	.285	.326

Footnotes will be found on page 133.

tions of down payment with each of the other factors, the results are as follows:

$$Z_1 = d + .214i$$
$$Z_2 = d - .232p$$
$$Z_3 = d - 12.5m$$

The theoretical ratios of mean difference to standard deviation were determined; they are .54, .58, and .60, respectively, and naturally enough, they lie between .53 for down payment alone and .63 for all four factors. Of these three combinations, that containing contract length has the highest ratio, but it is not strikingly better than the one containing price.

A second attempt with discriminant functions was made with the commercial bank sample. Here the number of available cases was so large that the drawing of a random subsample of 191 good loans and 190 bad seemed expedient. The computations were made entirely on the basis of this subsample.

Seven factors were selected for analysis: sex, stability of residence, stability of occupation, nature of occupation, bank account, life insurance, and real estate. Five of these factors are merely qualitative attributes incapable of quantitative measurement. As such, they are not directly subject to the discriminant function analysis; however, by assigning arbitrary numerical values to the qualitative categories the mechanical process of computing a discriminant function can still be followed. Thus women were given a value of 1, and men a value of 0; occupations were divided into three groups, with the poorest risk group having a value of 0, the middle group a value of 1, and the best a value of 2; cases with bank account were given a value of 1 compared with 0 for those with no bank account; and a similar process was used for life insurance and real estate. The means of the two samples, the standard deviations, and the correlation coefficients, are shown in Table B-3.

The form of the discriminant function obtained has been shown on page 86. This formula agrees well with the result of the individual analyses except in regard to stability of residence; the negative weight given to stability of residence suggests that risk increases as residence becomes more stable, which is a direct contradiction of the individual analysis. This discrepancy seems to be traceable, however, to a substantial sampling error in the subsample.[1]

Values of Z were computed and tabulated for all loans in the commercial bank sample; and with slight modifications, the process was then extended to the industrial banking company sample. The efficiency index based on combined factors was in each case noticeably higher than that for any one of the individual factors. Since these efficiency indices were obtained from actual distributions and not from theoretical estimates, they are particularly important. The assumptions underlying the classical discriminant function approach were sadly lacking, and the function itself was determined from a relatively small subsample of the total available cases. Despite these serious drawbacks, the method produced concrete results.

SHORT-CUT METHODS FOR COMPUTATION, ON THE ASSUMPTION OF INDEPENDENCE

Ordinarily the process of computing a discriminant function is arduous; but when the factors in question are independent, the process is simplified. If the distributions are normal or approximately normal, a mere simplification of the standard procedure is appropriate. The equations (see p. 111)

$$s_{11}l_1 + s_{12}l_2 + \ldots = a_1$$
$$s_{12}l_1 + s_{22}l_2 + \ldots = a_2$$
$$\ldots\ldots\ldots\ldots$$

become

$$s_{11}l_1 = a_1$$
$$s_{22}l_2 = a_2$$
$$\ldots\ldots$$

in case of complete mutual independence. If a state approaching independence is suspected, the l's can be computed directly from the mean differences and the variances; and the resulting function will probably be a good approximation. If, however, the distributions depart markedly from normality, an alternative procedure may be preferable. This second short-cut method is based on the simple principle that the probability of two or more events may be computed, in the case of independence, merely by multiplying together the individual probabilities of the occurrence of the events.

Suppose that as far as factor A is concerned, the good and bad loans are distributed among p discrete classes. Let a'_i represent the percentage of good loans in the A_i class ($i = 1 \ldots p$); let a''_i represent the percentage of bad loans; then $\frac{a''_i}{a'_i}$ is the bad-loan relative. Similarly for factor B with q discrete classes, b'_j and b''_j represent the percentage of good and bad loans in

[1] The following percentage distribution of loans in the subsample, with an efficiency index of 8.3, is distinctly at variance with the corresponding distribution in the total sample, with an index of 13.8. The difference, which is not excessive in a sample of this size, is large enough to affect the discriminant function considerably.

	0–2 years	2–6 years	6–10 years	10 years and over
Subsample				
Good	29.8	34.0	7.9	28.3
Bad	32.6	36.9	10.5	20.0
Total sample				
Good	28.0	34.8	10.1	27.1
Bad	40.4	36.2	7.2	16.2

Footnotes for Table B-3 on page 130

[a] The upper figure of each pair refers to the good-loan sample; the lower figure refers to the bad-loan sample. The correlation coefficients and standard deviations can be appropriately averaged by pairs to obtain a pooled estimate of the supposedly equal value for both distributions. Since the numbers of cases in each sample are virtually equal, an unweighted arithmetic average will suffice.
[b] Males given a value of 0, females of 1.
[c] Non-owners of real estate given a value of 0, owners of 1.
[d] Better than average given a value of 2, average of 1, worse than average of 0.
[e] Those with bank accounts given a value of 1, those without of 0.
[f] Those with life insurance given a value of 1, those without of 0.
[g] Does not differ significantly from 0. See also footnote 3, p. 135.

class B_i, and $\frac{b''_j}{b'_j}$ is the bad-loan relative. On the assumption of independence, the expected percentages of loans belonging to both class A_i and class B_j are $a'_i b'_j$ and $a''_i b''_j$ with a bad-loan relative of $\frac{a''_i b''_j}{a'_j b'_i}$. The result can be generalized to any number of factors.

The generalized bad-loan relative $\frac{a''_i b''_j c''_k \ldots}{a'_i b'_j c'_k \ldots}$ will serve as a sort of discriminant function; if it is greater than one, it signifies a worse-than-average loan, and conversely. In actual practice, modifications of this procedure will be found convenient. The logarithm of the reciprocal of the bad-loan relative, which equals

$$\log \frac{a''_i}{a'_i} + \log \frac{a''_j}{a'_j} + \log \frac{a''_k}{a'_k} + \ldots$$

is probably the most fundamental. This function is positive for better-than-average loans and negative for worse-than-average.

This short-cut method may be combined with the classical method of discriminant functions. Suppose three variates a, b, c are normally distributed and highly correlated. A discriminant function

$$z = L_a a + L_b b + L_c c$$

would be determined. There would be two normal distributions, one for the good loans and one for the bad. A transformation[2] can be made so that these distributions take the form

$$\frac{1}{\sigma_z\sqrt{2\pi}}\, e^{-\frac{(z+Dz/2)}{2\sigma_z^2}} dz, \quad \frac{1}{\sigma_z\sqrt{2\pi}}\, e^{-\frac{(z-Dz/2)^2}{2\sigma_z^2}} dz$$

where Dz is the mean difference and 0 is the dividing line between better-than-average and worse-than-average cases. The bad-loan relative for any particular case is the ratio of the two

$$\frac{e^{-\frac{z^2+zDz+Dz^2/4}{2\sigma_z^2}}}{e^{-\frac{z^2-zDz+Dz^2/4}{2\sigma_z^2}}} = e^{-\frac{zDz}{\sigma_z^2}}$$

The natural logarithm of the reciprocal of this is $\frac{zDz}{\sigma_z^2}$; it will be positive for better-than-average and negative for worse-than-average loans. If some additional factors D, E, . . . are not correlated, the discriminant function for all factors will be

$$\frac{zDz}{\sigma_z^2} + \log_e \frac{d'_1}{d''_1} + \log_e \frac{e'_i}{e''_i} + \ldots$$

APPLICATION OF THE SECOND SHORT-CUT METHOD TO COMMERCIAL BANK SAMPLE

The evidence obtained from the available samples indicates that the factors under investigation are not entirely independent, but the degree of interdependence is surprisingly small. In Table B-3, which refers to the subsample of 191 good and 190 bad loans, the highest correlation coefficient (.56) is between occupation and sex in the good-loan sample, and the next highest (.30) is between stability of address and ownership of real estate in the bad-loan sample. These particular coefficients are more than large enough to be statistically significant, but most of the others are not.[3] Even the significant coefficients, however,

[2] Probably the most convenient transformation is of the form

$$A = a - \frac{\bar{a}' + \bar{a}''}{2},\ B = b - \frac{\bar{b}' + \bar{b}''}{2}, \text{ etc.}$$

where $\bar{a}'$ is the a-mean of the good loans and $\bar{a}''$ is the a-mean of the bad loans, etc. The effect is to make the origin the midpoint between the means. Any transformation, however, that makes

$$L_a\bar{A}' + L_b\bar{B}' + L_c\bar{C}' = \frac{Dz}{2}$$

and

$$L_a\bar{A}'' + L_b\bar{B}'' + L_c\bar{C}'' = -\frac{Dz}{2}$$

will suffice.

[3] On the assumption of true independence in the parent universe, the standard error of the correlation coefficient is $\frac{1}{\sqrt{189}} = .073$ in a sample of 190 cases. Since the 5 percent significance level is .143 (.073 × 1.96, where 1.96 is the 5 percent value of t for the normal curve), all values of .14 or less for the coefficient may be considered non-significant.

are not sufficiently high to suggest a particularly close relationship; hence a situation approximating independence may, perhaps, be indicated.

Further evidence on independence is obtained from a series of 21 2 x 2 breakdowns of the commercial bank loans, one for each of the 21 possible pairs of the seven factors shown in Table B-3. For each of these factors the entire sample may be divided into two parts. In the case of some factors, like ownership of bank account, only two classifications are possible; for others like occupation, an arbitrary division is made so that the better risks are included in one classification and all the rest in another. For each pair of factors a two-way distribution may then be arranged by distributing all loans among four classes. Table B-4, which presents these data, will require some explanation. The first column is a percentage distribution of both good and bad loans by sex and real estate. The top figure (4.14) is the percent of females owning real estate among the good loans; beneath this figure is a similar percent (1.52) for the bad loans, followed by the bad-loan relative (.37). The next group of three figures (14.85; 6.99; .47) gives the percent of females not owning real estate and the bad-loan relative; the third group refers to males owning real estate; and the fourth refers to males not owning real estate. The second column gives the situation that would exist in a state of complete independence. The top figure (5.20) represents the expected proportion of females owning real estate among the good loans. This figure is determined by multiplying the total proportion of females among the good loans (4.14 + 14.85 = 18.99) by the total proportion of all persons owning real estate (4.14 + 23.23 = 27.37). Below this top figure is the expected proportion of females owning real estate among the bad loans (1.21), followed by the expected bad-loan relative (1.21 ÷ 5.20 = .23). All these expected figures can be calculated easily from the summary totals at the end of Table B-4.

This table permits comparison of the actual proportion of good or bad loans in any class with the proportion that would be expected in case of complete independence; it also permits comparison of the actual bad-loan relative with the expected bad-loan relative. This last comparison is important; for as long as the actual and expected relatives are approximately equal, the second short-cut method of computing the discriminant function can be used with assurance.

In Table B-4 the expected and actual values of the bad-loan relatives are surprisingly similar in most cases. The four most noticeable exceptions are for females owning real estate, females in the bad occupations, owners of real estate not owning life insurance, and persons having both bank account and real estate. Interestingly enough, the first three of these four cases include only a small proportion of all borrowers.

Although the evidence indicates that complete independence does not exist in the good- and bad-loan samples, we feel that the use of the second short-cut method is amply warranted in the case of the commercial bank sample. The standard discriminant function approach, which accounts for correlations between variates, is based on assumptions of normality that are not supported by the available evidence. The second short-cut method, which assumes independence but makes no assumption of normality, may be quite as realistic as the standard approach.

When the second short-cut method was tried for the commercial bank sample, the two factors age and income were added to the seven used in the previous experiment. The formula resulting from the experiment appears on page 85; and the distribution of loans is shown in Table 18. To illustrate the computation procedure, we shall show how some of the terms of this formula were computed.

The bad-loan relative for persons having bank accounts is .5 (see summary of Table B-4); the reciprocal is 2.0; and the common logarithm of the reciprocal is .30. For persons not having bank accounts, the relative is 1.4; the reciprocal, .715; the logarithm, $\bar{1}.85$ or $-.15$. At this point two alternative proce-

TABLE B-4

Percentage Distributions for the Commercial Bank Sample, Showing Interdependence Among the Following Seven Credit Factors: Sex, Possession of Life Insurance, Ownership of Real Estate, Possession of Bank Account, Stability of Residence, Stability of Occupation, and Nature of Occupation[a]

Classification[b]	SEX: FEMALE+, MALE−						LIFE INSURANCE: OWNED+, NOT OWNED−							
	Real Estate		*Occupation*[c]		*Bank Account*		*Real Estate*		*Occupation*[c]		*Sex*		*Stab. of Res.*	
	Owned X Not Owned O		Good X Bad O		Owned X Not Owned O		Owned X Not Owned O		Good X Bad O		Female X Male O		3 Yrs–Up X 0–3 Yrs[d] O	
	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.
+X														
Good loans	4.14	5.20	16.92	12.28	8.29	8.51	23.61	22.40	53.34	52.92	14.07	15.54	49.63	47.48
Bad loans	1.52	1.21	6.08	3.82	2.43	1.91	12.47	10.64	32.42	33.42	4.46	6.34	33.84	33.20
Relative	.37	.23	.36	.31	.29	.22	.53	.48	.61	.63	.32	.41	.68	.70
+O														
Good loans	14.85	13.79	2.07	6.71	10.70	10.48	58.22	59.43	28.49	28.91	67.76	66.29	32.20	34.35
Bad loans	6.99	7.30	2.43	4.69	6.08	6.60	62.00	63.83	42.05	41.05	70.01	68.13	40.63	41.27
Relative	.47	.53	1.17	.70	.57	.63	1.06	1.07	1.48	1.42	1.03	1.03	1.26	1.20
−X														
Good loans	23.23	22.17	47.75	52.39	36.53	36.31	3.76	4.97	11.33	11.75	4.92	3.45	8.39	10.54
Bad loans	12.77	13.08	38.80	41.06	20.06	20.58	1.82	3.65	12.46	11.46	4.05	2.17	10.74	11.38
Relative	.55	.59	.81	.78	.55	.57	.48	.73	1.10	.98	.82	.63	1.28	1.08
−O														
Good loans	57.78	58.84	33.26	28.62	44.48	44.70	14.41	13.20	6.84	6.42	13.25	14.72	9.78	7.63
Bad loans	78.72	78.41	52.69	50.43	71.43	70.91	23.71	21.88	13.07	14.07	21.48	23.36	14.79	14.15
Relative	1.36	1.33	1.58	1.76	1.61	1.59	1.65	1.66	1.91	2.19	1.62	1.59	1.51	1.85

(continued on next page)

TABLE B-4

Percentage Distributions for the Commercial Bank Sample[a] (*continued*)

Classification[b]	REAL ESTATE: OWNED+, NOT OWNED−						BANK ACCOUNT: OWNED+, NOT OWNED−							
	Stab. of Res.		*Occupation*[c]		*Stab. of Occup.*		*Real Estate*		*Stab. of Res.*		*Occupation*[c]		*Life Insurance*	
	3 Yrs–Up X 0–3 Yrs[d] O		Good X Bad O		6 Yrs–Up X 0–6 Yrs[d] O		Owned X Not Owned O		3 Yrs–Up X 0–3 Yrs[d] O		Good X Bad O		Owned X Not Owned O	
	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.
+X														
Good loans	19.81	15.88	17.16	17.70	19.90	16.30	15.66	12.27	28.14	26.00	30.07	28.99	39.04	36.68
Bad loans	9.73	6.37	7.80	6.41	8.21	5.76	5.47	3.21	10.74	10.03	12.77	10.09	18.34	16.75
Relative	.49	.40	.45	.36	.41	.35	.35	.26	.38	.39	.42	.35	.47	.46
+O														
Good loans	7.56	11.49	10.21	9.67	7.47	11.07	29.16	32.55	16.68	18.82	14.75	15.83	5.78	8.14
Bad loans	4.56	7.92	6.49	7.88	6.08	8.53	17.02	19.28	11.75	12.46	9.72	12.40	4.15	5.74
Relative	.60	.69	.64	.81	.81	.77	.58	.59	.70	.66	.66	.78	.72	.71
−X														
Good loans	38.21	42.14	47.51	46.97	39.67	43.27	11.71	15.10	29.88	32.02	34.60	35.68	42.79	45.15
Bad loans	34.85	38.21	37.08	38.47	32.11	34.56	8.82	11.08	33.84	34.55	32.11	34.79	56.13	57.72
Relative	.91	.91	.78	.82	.81	.80	.75	.73	1.13	1.08	.93	.98	1.31	1.28
−O														
Good loans	34.42	30.49	25.12	25.66	32.96	29.36	43.47	40.08	25.30	23.16	20.58	19.50	12.39	10.03
Bad loans	50.86	47.50	48.63	47.24	53.60	51.15	68.69	66.43	43.67	42.96	45.40	42.72	21.38	19.79
Relative	1.48	1.56	1.94	1.84	1.63	1.74	1.58	1.66	1.73	1.85	2.21	2.19	1.73	1.97

(continued on next page)

TABLE B-4

PERCENTAGE DISTRIBUTIONS FOR THE COMMERCIAL BANK SAMPLE[a] (*continued*)

Classification[b]	STAB. OF RES.: 3 YRS–UP+, 0–3 YRS–						STABILITY OF OCCUPATION: 6 YRS–UP+, 0–6 YRS–							
	Occupation[c]		*Stab. of Occup.*		*Sex*		*Occupation*[c]		*Sex*		*Life Insurance*		*Bank Account*	
	Good X / Bad O		6 Yrs–Up X / 0–6 Yrs[d] O		Female X / Male O		Good X / Bad O		Female X / Male O		Owned X / Not Owned O		Owned X / Not Owned O	
	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.	Actual	Exptd.
+X														
Good loans	38.50	37.52	37.40	34.56	11.42	11.02	39.27	38.52	11.76	11.31	51.04	48.75	27.86	26.70
Bad loans	21.58	20.01	21.38	17.97	4.46	3.79	19.45	18.09	3.75	3.43	32.11	30.03	9.02	9.07
Relative	.56	.53	.57	.52	.39	.34	.50	.47	.32	.30	.63	.62	.32	.34
+O														
Good loans	19.52	20.50	20.62	23.46	46.60	47.00	20.30	21.05	47.81	48.26	8.53	10.82	31.71	32.87
Bad loans	23.00	24.57	23.20	26.61	40.12	40.79	20.87	22.23	36.57	36.89	8.21	10.29	31.30	31.25
Relative	1.18	1.20	1.13	1.13	.86	.87	1.03	1.06	.76	.76	.96	.95	.99	.95
–X														
Good loans	26.17	27.15	22.17	25.01	7.57	7.97	25.40	26.15	7.23	7.68	30.79	33.08	16.96	18.12
Bad loans	23.30	24.87	18.94	22.35	4.05	4.72	25.43	26.79	4.76	5.08	42.36	44.44	13.47	13.42
Relative	.89	.92	.85	.89	.54	.59	1.00	1.02	.66	.66	1.38	1.34	.79	.74
–O														
Good loans	15.81	14.83	19.81	16.97	34.41	34.01	15.03	14.28	33.20	32.75	9.64	7.35	23.47	22.31
Bad loans	32.12	30.55	36.48	33.07	51.37	50.70	34.25	32.89	54.92	54.60	17.32	15.24	46.21	46.26
Relative	2.03	2.06	1.84	1.95	1.49	1.49	2.28	2.30	1.65	1.67	1.80	2.07	1.97	2.07

(*concluded on next page*)

TABLE B-4

PERCENTAGE DISTRIBUTIONS FOR THE COMMERCIAL BANK SAMPLE[a] (*continued*)

Classification	SUMMARY													
	Sex		*Life Insurance*		*Real Estate*		*Bank Account*		*Stab. of Res.*		*Stab. of Occup.*		*Occupation*[c]	
	Male	Female	Owned	Not Owned	Owned	Not Owned	Owned	Not Owned	3 Yrs –Up	0–3 Years[d]	6 Yrs –Up	0–6 Years[d]	Good	Bad
Good loans	81.01	18.99	81.83	18.17	27.37	72.63	44.82	55.18	58.02	41.98	59.57	40.43	64.67	35.33
Bad loans	91.49	8.51	74.47	25.53	14.29	85.71	22.49	77.51	44.58	55.42	40.32	59.68	44.88	55.12
Relative	1.13	.45	.91	1.41	.52	1.18	.50	1.40	.77	1.32	.68	1.48	.69	1.56

[a] This table is based on 1,179 good loans and 987 bad loans, all of which reported complete information for all seven factors. For explanation of table, see text, p. 136.

[b] The meanings of the symbols in this column are indicated by corresponding symbols in the captions over the columns of percentages.

[c] The following occupational groups of Table 13, pp. 70–71, were considered good: professional (1a and 1b); clerical, except outside salesmen and commercial representatives (2a, 2b, 2d); policemen and firemen (3); and proprietors (4). All others were classed as bad.

[d] Upper limit excluded from this class interval.

dures are possible. One is to add .30 to the score of all cases with bank account and to subtract .15 from those without; the other is to add the difference, .45, to those having bank accounts and to subtract nothing from the others. With the first scheme, the point 0 is the dividing line between the better-than-average and worse-than-average cases; with the second, the point .15 is the dividing line. The second scheme, which may be a little easier for computing actual scores, was used here. The dividing line for the entire scoring system was 1.25, the sum of .15 for bank account plus eight similar quantities for the other factors.

A rough job of curve fitting was done in the case of stability of residence. The bad-loan relative is 1.6 for the class of less than one year; and it decreases more or less regularly to .6 for the class of 10 years or over (Table 12). The common logarithm or the reciprocals increase from −.20 to .22 so that the difference between the extremes is .42. For each year up to 10 at present address the loan was rated one-tenth of .42 or .042. Since the class of 10 years and over was not subdivided, we have no evidence to show whether the bad-loan relatives continue to fall as the length of residence increases above 10 years. For this reason the total score was limited to .42 no matter how long the tenure of residence. Some readers may take exception to this conservative policy; they may feel that an additional score of .042 should be added for each year over 10. While this point of view may be justified, we merely suggest that such a policy may give too high a rating to the young person of 25 who has never been away from home.

APPENDIX C

Tests of Significance and Sampling Errors