2

How Risks Can Be Studied

THE analysis of risk factors presented in this volume is the result of a statistical sampling procedure based on a categorical classification of all loans into two mutually exclusive classes, "good" loans and "bad" loans. Theoretically, a good loan is distinguished from a bad loan by the fact that the gross profit on a good loan is sufficient to cover all expenses including possible losses; but in practice the distinction is much less precise. Many, perhaps most loans are repaid in full and on time, and are therefore considered by lenders to be good loans. Some loan accounts become delinquent, however, and sooner or later the lender begins to take action; follow-up letters and calls by collectors usually come first; later comes legal action, which includes seizure and sale of collateral as well as the garnishment of wages; and finally, if all efforts appear fruitless, the loan may be charged off. Although no lender can determine precisely when a loan ceases to be profitable and begins to become unprofitable, many lenders draw some qualitative distinction between their worst loans and the others. Some lenders, for example, set up a Grade A class of borrowers, comprised of those who have repaid promptly and in full, a Grade B class consisting of those who have repaid but with occasional delinquency, and a Grade C class including those who have shown serious delinquency leading to court action, charge-off, or repossession. Such a distinction may be very useful in determining which borrowers merit additional loans in the future.

Since most lenders' files are arranged with some sort of separation between good and bad loans, separate analyses of

these two classifications, rather than a single analysis of all loans, are made in the present study. The characteristics of the borrowers in each class—average age, occupational distribution, percentage of persons having bank accounts, etc.—are compared. The analysis consists, then, of a study of the important differences in borrowers' characteristics between good loans and bad loans. Each lender who contributed material was requested to provide a sample of good loans and an approximately equal sample of bad loans. The process of making such a selection, while appearing simple, involves a number of serious complications, which are discussed later in this chapter.

ILLUSTRATIVE ANALYSIS

The procedure of analysis adopted for this study may be described by illustrating its use in a specific case—for example, in the analysis of samples of 100 good loans and 100 bad loans obtained from the personal loan department of a New England commercial bank. We know from the questionnaires described in Chapter 1 that lenders consider stability of occupation an important credit factor, and we wish to determine whether or not the samples bear this out. The data requested from the bank include the number of years the borrower had been engaged in the occupation in which he was employed at time of application. We have used this information as the basis for a measure of stability, although admittedly a measure based on previous employment as well as present employment would be more satisfactory. Among the cases submitted, the borrowers' present employment records were reported for all the bad loans and all but one of the good loans. This fact is important, for if the information had not been reported for a substantial number of cases, the results would have been questionable if not entirely invalid. In most of the tables accompanying this report,

the number of cases not reporting information requested is given in addition to the number of cases reporting; when the number not reporting seems sufficiently high to discredit the result, attention is called to this fact.

One possible method of showing whether stability is related to risk is to compute the means of the employment records of the two samples. In this illustrative case the mean of the good-loan sample is 10.76 years, and that of the bad is 7.16 years.[1] If these averages are reliable, they indicate that satisfactory borrowers in the past have been persons with occupations more stable than those of the unsatisfactory borrowers. Most people will be willing to infer that future applicants with stable employment records are likely to be better risks than those with unstable records.

The next point to consider is whether or not the averages are reliable. A skeptic might object: "I believe that if you took sufficiently large samples, you would find no difference between the means of the good loans and of the bad loans; I believe that the apparent difference in the stability of employment in these two groups of loans is a pure coincidence entirely attributable to sampling errors, which are bound to occur in inadequate samples." Such a coincidence is of course possible, but extremely unlikely. A standard test[2]

[1] After reading a preliminary draft of this study, one of our critics reported that these averages are considerably higher than his experience would indicate. Upon investigation, we discovered that the occupational tenures reported by this bank are among the longest reported by any of the contributing banks. This fact we attribute either to selection on the part of the bank officials or to the possibility that the community served by this bank may be a particularly stable one. In any case, the sample is satisfactory for illustrative purposes. Furthermore, it is typical of all other samples in that the average tenure for the good loans is greater than the average tenure for the bad loans.

[2] A description of this test, called the t-test, particularly its application to small samples, will be found in R. A. Fisher, *Statistical Methods for Research Workers* (London and Edinburgh, 6th edition, 1936) Chapter 5 (in particular sec. 25.1); G. Udney Yule and M. G. Kendall, *An Introduction to the Theory of Statistics* (London, 11th edition, 1937) Chapters 20 and 23; George W. Snedecor, *Statistical Methods Applied to Experiments in Agriculture and Biology* (Ames, Iowa, 1937) Chapters 2, 3, 4.

of statistical significance indicates that there is not one chance in a hundred that such a coincidence could have occurred. Some grounds therefore exist for believing that the results are reliable. If, however, the test of significance had indicated that the chance of a sampling coincidence was considerably more than one in a hundred—say 10 in 100, or 1 in 10—we should have dismissed the evidence as unreliable.

That tests of significance demonstrate reliability only in a limited sense should be emphasized. Such tests actually show whether or not the sample is large enough to be reliable. If the test of significance indicates that the sample is not large enough, no further evidence is necessary to demonstrate unreliability. But if the sample is large enough to be reliable, it may still be unreliable for a number of other reasons. For example, borrowers may have made false or misleading statements on their applications, and the prevalence of falsehood may be lower among the good loans than among the bad; errors of transcription or tabulation may have been made, and these may for some reason affect the good and bad loans differently. Errors of this sort, however, can only be eliminated at their source, by systematic credit investigation and by careful checking of statistical transcriptions and computations.

Table 3, giving percentage distributions of the good and bad loans according to the borrowers' stability of occupation, illustrates an alternative method of sample analysis, used as the standard throughout this report. In this type of analysis we are no longer interested in the average number of years of tenure of occupation for each sample, but in the difference between the percentage of good and the percentage of bad loans for any particular group of borrowers. In the example in Table 3, 30.0 percent of the bad loans show tenure of less than three years, but only 22.2 percent of the good loans are in this same class. Similar discrepancies for the other class intervals will be noted.

TABLE 3

THE RELATION BETWEEN BAD-LOAN EXPERIENCE AND STABILITY OF OCCUPATION, AS SHOWN BY THE GOOD-LOAN AND BAD-LOAN SAMPLES SUBMITTED BY ONE COMMERCIAL BANK[a]

Number of Years at Present Occupation[b]	*Percentage Distribution*		*Ratio of Bad Loans to Good*
	Good Loans	Bad Loans	
0– 3	22.2	30.0	1.4
3– 6	19.2	30.0	1.6
6–10	13.1	18.0	1.4
10 and over	45.5	22.0	.5

Remarks: The discrepancy between the samples is statistically significant. The efficiency index is 23.5; for description of efficiency index, see text, pp. 28-31.

[a] The good-loan sample consisted of 100 cases, of which 1 did not report information, and the bad-loan sample of 100 cases, all reporting.
[b] Upper limit of class interval excluded.

These distributional differences are not explainable as sampling coincidences any more than the average differences discussed above; an appropriate test for this arrangement[3] indicates that there is not one chance in a hundred that these results could have occurred as a sampling coincidence. This fact is indicated in Table 3 under "remarks," which include a statement to the effect that the results are significant. Most of the other tables accompanying this report also contain remarks indicating whether the evidence is significant, questionably significant, or not significant.[4] Significance refers, of course, to statistical significance, which only means that the sample is of sufficient size to justify drawing conclusions.

[3] The Chi-square test. Cf. R. A. Fisher, *op. cit.*, Chapter 4; Frederick C. Mills, *Statistical Methods* (New York, revised, 1938) pp. 618-36; George W. Snedecor, *op. cit.*, Chapters 1 and 9.
[4] Results are considered significant if they satisfy the 1 percent criterion; they are considered questionably significant if they meet only the 5 percent criterion; and otherwise they are considered not significant.

INDEX OF BAD-LOAN EXPERIENCE

Table 3 also gives ratios of the percent of bad loans in any class interval to the percent of good loans in that class interval. This ratio, called the bad-loan relative, may be used as an index of bad-loan experience for the cases in that interval. Since the ratio or index for all classes combined is 1 (100 percent to 100 percent), a ratio of 1, when it occurs, indicates average experience; a ratio greater than 1 indicates worse-than-average risk; and a ratio smaller than 1 indicates better-than-average risk. Thus for the interval of fewer than three years in Table 3 the ratio of 30.0 percent to 22.2 percent, or 1.4, indicates worse-than-average experience; and for the interval of 10 years and over the ratio .5 indicates better-than-average experience. In samples of only 100 good and 100 bad loans, the bad-loan relative is subject to a large sampling error; about all the relative can indicate is whether a particular class interval, or group of borrowers, is better than average, roughly average, or worse than average. In much larger samples, however—samples of several thousand would be necessary—the relative takes on more precise significance.[5]

When a sufficiently large unselected sample is obtained—i.e., a sample that represents the true relative importance of the good and bad loans—the bad-loan relative can be supplanted by the ratio of the number of bad loans in any class interval to the number of all loans handled in that class interval, which is obviously preferable to the relative. Of course, the bad-loan relative can be used to estimate the desired ratio for a particular class interval if the over-all ratio of the number of bad loans in all classes to the number of all loans handled is known. The process may be illustrated by the following example. Suppose the banker who sub-

[5] See section on size of sample, pp. 35-37 below, and also Appendix C.

mitted the sample of Table 3 discovered from past experience that 2 percent of all loans made were bad loans. If he wanted to know the ratio for borrowers with less than 3 years' employment tenure, he could obtain an estimate by multiplying 2 percent by 1.4, i.e., by multiplying the over-all bad-loan ratio by the bad-loan relative for the class interval in question.

THE EFFICIENCY INDEX

An abstract interpretation of the result of this sample experiment can be given easily. The questionnaire results reviewed in Chapter 1 show that lenders believe that stability of occupation is an important indicator of creditworthiness, and the sample data bear out this belief. This conclusion is not of much use, however, in the formulation of loan policy. Although loan policy can be satisfactorily discussed only in terms of operating cost—as we shall show later—a concrete example of the type of problem involved can be obtained immediately by reference to Table 3. In this table, three class intervals, comprising all borrowers with tenure of employment of less than 10 years, are worse than average. On the basis of this evidence, however, a loan officer is not likely to reject all future applications from applicants with occupation records of less than 10 years; Table 3 suggests that by setting up a 10-year minimum tenure standard a lender will lose more than half his present business, which he probably will not wish to lose even if it is worse than average. Before making any minimum requirements, a lender will want to make sure that the borrowers thus eliminated are so much worse than average that they are absolutely unprofitable.

A factor, to be really effective as a credit indicator, must provide some means whereby a substantial number of bad accounts can be eliminated without appreciable rejection of

good business. In this connection a simple though rough measure of the effectiveness of various factors can be computed. To illustrate: The three worse-than-average class intervals in Table 3, including all borrowers with tenures of less than 10 years, contain 78.0 percent of the bad loans but only 54.5 percent of the good; the difference between these two percentages is 23.5 percent. The one better-than-average class, that with tenures of 10 years and over, contains 45.5 percent of the good loans and only 22.0 percent of the bad; and again the difference is 23.5 percent. Conceivably this difference can vary all the way from 0 to 100. When it is 0, the distributions of good and bad loans are identical; therefore, if any class of borrower is rejected, the same percentages of good and bad loans will be eliminated. If the difference should ever be 100, the better-than-average classes would contain all the good loans, and the worse-than-average groups would contain all the bad loans; hence, all bad loans could be eliminated without the loss of any of the good loans. Thus, the larger differences between 0 and 100 generally indicate greater opportunities for eliminating bad risks without undue elimination of good risks. This difference, which we shall call the efficiency index, provides the desired measure of the usefulness of any factor (in our illustration, the particular factor is tenure of occupation) as a means of credit control.[6]

In the course of this report, the efficiency index will receive considerable emphasis; its function is to separate the more effective credit factors from the less effective. The highest

[6] The efficiency index for normal distributions is an easily determined function of the ratio of the mean difference between the two samples to the standard deviation. (See Appendix A, pp. 106-8.) In most technical discussions, this ratio is a more fundamental concept than the efficiency index. The efficiency index has the advantage, however, of being determinate for qualitative attributes, such as occupation, where there is no ratio of mean difference to standard deviation.

index discovered in the entire analysis is 46 for percent of down payment in the new-car sample.[7] From this maximum, the efficiency indices for other factors range down to almost zero, and most of them are below 20. Indices of less than 10 may usually be considered practically equivalent to zero; this matter will be amplified in Chapters 4 and 5. A tabulation of the efficiency indices for the more important credit factors appears in Table 17, Chapter 3.

Discussion of the efficiency index introduces a major problem in interpreting results. The index is offered as a measure of the effectiveness of a factor as a risk selector; what it really measures, however, is not the inherent effectiveness of a factor, but its effectiveness in future selection only. When the sample analysis of a factor shows no significant difference between good and bad loans, or when the efficiency index is small, the most natural interpretation is that the factor is unrelated to risk. This interpretation would be the only correct one if it were based on samples of totally unselected loans, but the fact that all loans have been carefully selected permits another interpretation. When, in the granting of loans, considerable emphasis is laid on a given factor, and when these loans are used as a basis for sample analysis, a low efficiency index for the factor—even an important factor—is likely to result. The low index merely means that further emphasis on this factor is undesirable; it does not mean that less emphasis is desirable.

Lenders who wish to make studies of their own loan experience should not consider results yielding an efficiency index of less than 15.0 as significant. This precaution, used along with two others to be recommended later[8] (a minimum sample of 200 good and 200 bad cases, and a minimum total of 30 good and bad cases in each class interval), may suffice

[7] See Table 9, p. 61.
[8] See pp. 35-36.

as a rule-of-thumb substitute for a precise test of statistical significance. This substitute rule, while not infallible, will aid in securing sample reliability. We recommend, however, that investigators acquaint themselves with the standard sampling methods, especially if they intend to make very extensive investigations.

SELECTION OF SAMPLES

The specific case used for illustrative purposes above was based on an analysis of 100 good loans and 100 bad loans. The objection may be raised that an analysis based on samples of equal size gives undue weight to the bad loans, which are considerably less important numerically than the good loans. This objection can arise only from a misconception of the purpose of the equal sample method and of the principles of modern statistical sampling theory. The analysis of bad-loan experience may be considered in two distinct parts. The first part is the measurement of the relative importance of the two groups of loans—i.e., the ratio of bad loans to good, or of bad loans to total number of cases handled—and for this purpose equal samples are obviously useless. The second part is the portrayal of characteristic differences between the good and bad loans; and for this one purpose the equal sample approach is admirable, for it provides maximum reliability with a minimum number of cases. We have found that a total sample of 200 cases is often large enough to determine some of the differences between the two groups if the sample is equally divided between good and bad loans. But a sample of 190 good and 10 bad loans, which is the sort of distribution that would truly represent the relative importance of good and bad, would be inadequate because of the small number of bad loans. To obtain reliable results, a sample of some 2000 cases containing perhaps 1900 good

and 100 bad loans would be necessary. No one will deny that such a representative sample of 2000 cases is preferable to an equally divided, selected sample of 200 cases. But if the cost of obtaining 2000 cases is prohibitive, a sample of 100 good and 100 bad loans may be better than no sample at all. The essential point is to obtain a sufficiently large sample to be statistically reliable for each of the two categories.

Another objection to equal samples arises from the popular belief that the reliability of a sample is determined by its coverage, i.e., the percentage of all cases represented by the sample. Modern sampling theory rarely finds the concept of coverage very useful. Except in special cases—and the analysis of loan experience is not one of them—a sample is not thought of as a finite percentage of a finite population (i.e., the total group from which the sample is drawn), but rather as an infinitesimal part of an indefinitely large population—a hypothetical infinite universe, so called. Although this view of sampling may seem radical, it is actually the most conservative possible. For example, if a sample of 250 cases is large enough to represent reliably an infinite universe, it will represent better a finite population of 1000 cases, and still better, one of 300 cases. The important fact in sampling is not coverage but the attainment of a sample large enough to represent faithfully an infinite universe. This policy is followed in the present analysis.

RANDOM SAMPLING TECHNIQUE

In statistical investigations of the kind outlined in this volume, correct random sampling procedure is extremely important; it is also one of the most difficult problems encountered in loan sample analysis. A standard satisfactory method cannot be formulated because the design of a suitable method often depends upon the nature of the problem

at hand. All we can do in this study is illustrate good sampling procedure in the following rather simplified imaginary situation.

A lender has on record 237 particularly unsatisfactory loans made during 1938 and 1939. He also has some 15,000 other loans made during the same period; these other loans are generally satisfactory, containing nothing worse than cases of minor delinquency. For his study the lender decides that the 2-year period is sufficiently homogeneous and sufficiently short so that selection of cases by chronological distribution is not necessary. He also decides to take the entire 237 cases for a bad-loan sample and to draw a random sample of approximately 237 cases from the 15,000 satisfactory cases, believing that for his study the additional accuracy obtainable by using more than 237 good cases does not justify the additional work involved. The only difficulty is the problem of drawing the random sample of good cases.

Several simple methods of drawing are possible. One is to take 237 cases haphazardly from the filing cabinets; another is to take some letter in the alphabet that will provide about 237 cases; and a third is to count out the loans and take every 63rd one. All of these methods, however, are frowned on by some statisticians. A more acceptable method is to make out a control card for each loan and to shuffle the cards in a mechanical shuffler, but this procedure is extremely cumbrous. An acceptable and at the same time practical method, which can be used if the loans to be sampled are numbered consecutively, may be found in a table of random numbers.[9]

Suppose the 15,000 loans are numbered consecutively from 10,000 to 25,000. The loans are probably arranged in chronological order, but that is of no consequence. A sample of 237

[9] One table of random numbers appears in *Tracts for Computers, No. 15, Random Sampling Numbers,* compiled by L. H. C. Tippets (London, 1927). Another appears in R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (London and Edinburgh, 1938), Table XXXIII, pp. 82 ff.

cases can be drawn easily from a table consisting of columns of random digits as follows:

8091	9271	1473	....	....
0818	4452	0627	....	
2314	5748	3108		
0550	5465	9463		
1351	1788	2406		

A column of five digits may be marked off, and from this column all numbers between 10,000 and 25,000 may be selected. In the above sample table we can take the first five-digit column (the first four-digit column plus the first digit in the second column); the third number in this column, 23145, is within the required range; so is the fifth, 13511. In this way 237 random numbers can be obtained, and the loans with the corresponding numbers can then be secured from the file. If a few of the numbers are missing, additional numbers can be drawn until the sample reaches the required size.

Usually, however, the sampling problem is not nearly so simple. The loans may not be filed consecutively by number, or a selected chronological distribution of loans may be considered necessary. In such cases proper random sampling can be accomplished by means of shuffling, or the loans can be specially numbered to permit the use of a table of random numbers, but the mechanical difficulty of either process will probably induce many to use less acceptable but simpler methods.

SIZE OF SAMPLE REQUIRED

We assume throughout this report that the best samples to use are approximately equal samples of good and bad loans. This assumption, of course, is true only when good and bad cases are equally easy to obtain and tabulate; when they are not, very unequal samples may be utilized. For example,

a lender with a good punch card system and with records available for 10,000 good loans and 200 bad loans may find the tabulation of the entire 10,000 good loans as simple as the drawing of a random sample of 200 good loans to match the 200 bad; he will probably decide that the small additional expenditure of effort is more than repaid by the additional precision inherent in the larger sample. But in many other cases little is gained by using unequal samples, and this study is primarily concerned with such cases.

The optimum size of sample depends upon a number of things: the cost or difficulty of obtaining cases, the degree of precision desired, the nature of the questions to be answered, and so on. The required number will be small if the problem at hand is merely to answer some such question as the following: Is there a significant difference between good and bad loans in respect to stability of occupation? One possible approach, already pointed out, is to determine the means and standard deviations of the two samples and then to apply an appropriate test of significance.[10] Our experience indicates that samples based on as small a number as 55 good loans and 55 bad loans may give significant results. The alternative method, that of making frequency distributions and then employing another test of significance,[11] probably requires a somewhat larger number of cases, but actual experience indicates that samples of 100 good and 100 bad loans are frequently large enough to demonstrate significance.

Although significant results can be obtained with samples of 100 good and 100 bad loans (or even fewer), we do not recommend that lenders wishing to make analyses of their own loans use such small samples. Those who understand the use of tests of significance can form their own judgment concerning the size of the sample, but those who do not understand these tests should use samples of at least 200 good

[10] See footnote 2, p. 24.
[11] See footnote 3, p. 26.

loans and 200 bad loans. Furthermore, they should arrange the distributions so that the minimum total of good and bad loans in any class interval is 30. This last restriction, which is important, will either limit the number of class intervals to a few, or it will necessitate samples larger than 200 of each type of loan. For example, if loans are to be classified into 20 occupational classes, a total of 30 good and bad loans in each class cannot be obtained with fewer than 300 good loans and 300 bad, and in all likelihood many more will be required.

The above estimates of sample size are all made on the supposition that no great degree of precision is required. We have already mentioned that the bad-loan relative (the percentage of bad loans in any class to the percentage of good loans in that class) is not very accurate for small samples; if the bad-loan relative is to be determined precisely, much larger samples will be required. A degree of precision can be defined, for illustration, as follows: suppose that for some class interval or group of loans the true, or so-called population value of the bad-loan relative is 1.0 (this is the value that would be approached if the sample were indefinitely large), and suppose that the sample must be large enough to insure, within reasonable likelihood, that the sample value of the bad-loan relative shall be between .9 and 1.1; then the minimum number of cases in each sample is about 7200 if there are to be 10 class intervals, and about 15,200 if there are to be 20 intervals.[12] Samples of this size would be appropriate in studies of bad-loan experience by occupation, for example. For such a study the relative merits of the different occupational groups would have to be accurately determined. Furthermore, to be satisfactory, the occupational classification would have to be fairly detailed; 20 different

[12] The range of error from .9 to 1.1 is taken to represent a sampling variation of two standard errors on either side of the true value, 1.0. Thus there is only one chance in twenty that any given sample estimate will lie outside the specified range. See Appendix C, pp. 151-53.

occupational groups would not be too many, and even 50 occupational groups might be desirable.

CONSOLIDATION AND CONSISTENCY OF INDIVIDUAL SAMPLES

Since many lenders contributed samples, a separate analysis of each contribution is not presented in this study, but all the available samples have been consolidated into six general groups, as follows: commercial banks, industrial banks, personal finance companies, appliance finance companies, new-car transactions, and used-car transactions. In the process of consolidation most samples were merely added together, but the commercial bank samples were specifically weighted to compensate for the effect of samples containing an unequal number of good and bad loans.[13] A consolidation of sam-

[13] For commercial banks, the distributions presented throughout this study are weighted averages of the percentage distributions of the 12 component samples. These averages were computed because different banks contributed different proportions of good and bad loans; some contributed twice as many good loans as bad, whereas others contributed an equal number. If all these available samples had been merely added together, the good-loan experience of the banks submitting twice as many good loans would have been overrepresented; and if any variation had existed in the loan experience of the different banks, a source of error would have been introduced. To avoid this source of error, a weight was given to each bank sample, and the same weight was applied to both the good- and the bad-loan distributions of that bank sample. The weight was determined by the total number of loans in the smaller of the two samples; if the bad-loan sample was the smaller, the number in that sample was taken as the weight, and conversely. The sum of the weights was, in most cases, 1294, which we have termed the effective number of cases. This is a fictitious number used for the purpose of making tests of significance, and does not refer to the actual number of loan schedules, which was 1468 good and 1297 bad loans. A measure of statistical significance based on 1294 will slightly underestimate the true significance.

In many of the distributions shown here, information was not reported for some of the cases. In such instances the effective number of cases was reduced in accordance with the number for which data were not reported. For all the other types of lending institutions submitting samples, the number of good and bad loans was approximately equal; consequently no process of weighting seemed necessary, and all component samples were merely added together.

ples, even samples from the same general type of institution, has serious drawbacks, however. When samples drawn independently from different lenders' loan portfolios are haphazardly collected and consolidated, the net result is not a sample of any particular homogeneous universe. The combined samples represent a diversity of influences: they represent no standard degree of goodness or badness; they represent lenders operating in different geographical locations and employing different credit policies; and they cover an undetermined period of time, during which lending conditions and credit experience may have varied considerably.

Although a serious attempt was made to secure uniformity in the goodness and badness of the loans submitted for analysis, the loan samples received were anything but uniform. For example, commercial bankers were requested to distinguish bad loans by one of the following criteria: loan was more than 90 days delinquent; comaker paid all or part of loan after demand by bank; legal action was taken; loan was charged off. But upon analysis, the samples submitted were found to vary surprisingly. In one sample the proportion of cases that were excessively delinquent without receiving further action by the bank was only 2 percent; in another sample, it was 90 percent.[14] The banker who submitted the second sample wrote by way of explanation that he had a dearth of really bad loans to choose from; that many of the cases submitted were delinquencies of less than the specified 90 days; and that in many cases these so-called bad loans were not bad enough to prevent the borrowers from obtaining other loans in the future. In the auto finance samples, bad loans were supposed to contain only repossessions, and

[14] For a description of the composition of the bad-loan samples submitted by the various contributing commercial banks, see National Bureau of Economic Research (Financial Research Program), *Commercial Banks and Consumer Instalment Credit,* by John M. Chapman and Associates (1940) Table B-1, p. 275.

good loans were to contain only paid out accounts; but one large contributor had trouble obtaining enough paid out accounts because lack of storage space prevented retention of the records. Consequently this company was forced to provide a good-loan sample consisting partly of paid out accounts and partly of current accounts that had not yet become bad. In short, neither bad loans nor good loans in the available samples are a clearly defined species. The selection of good or of bad loans depended largely upon the judgment of the contributing lender and upon the quality of the material he had readily available. In spite of these difficulties, we feel confident that the repayment experience represented by the good-loan samples is clearly and substantially superior to that represented by the bad-loan samples; and as long as this is true, these samples will suffice for the sort of analysis we are trying to make.

Because of the possibility that bad-loan experience might vary considerably from lender to lender, the loan samples submitted by each contributor were analyzed separately if they were large enough to assure reliability; otherwise they were combined with other similar small samples until sufficiently large units were obtained. Thus 10 of the 21 commercial bank samples obtained were analyzed separately, and the other 11 were combined and analyzed as 2 separate units; 2 of the 10 industrial bank samples were treated separately, and the other 8 were combined into one unit; the 2 personal finance company samples and the one appliance finance company sample were each treated separately; and finally 2 of the 3 automobile finance company samples were analyzed separately, and the other was broken down into 2 units representing the operations of 2 branch offices of the same company. The individual tabulations are not reproduced in this study, but in most of the tables of composite experience, remarks will be found indicating the degree of consistency

observed among the components.[15] No objective test is used herein for judging consistency. While an objective test is undoubtedly desirable, the construction of one that would not entail an exorbitant expenditure of labor seems impossible. The only feasible procedure, therefore, is to examine each component superficially and subjectively to see whether or not it is consistent with the composite. Since consistency may be taken in more than one sense, its meaning should be clarified. A good-loan sample received from a New York City lender indicates that 16 percent of all cases report ownership of real estate, whereas a similar sample from Los Angeles indicates 40 percent. While there is no consistency between the 16 percent and the 40 percent reporting ownership, there is consistency of bad-loan experience because the real estate owners appear to be definitely good risks in both samples. The latter meaning of consistency—consistency of bad-loan experience—is the only one used in this report.

Since the time element may cause considerable variation in risk experience, some method of control is desirable. One possible method is to select a number of short, homogeneous time periods, and to make separate analyses of the loans made in each of these periods; a sample of good loans made in the first half of 1936 could be compared with a similar sample of bad loans. Carried far enough, this process might eventually result in a description of secular and cyclical changes in risk experience. An alternative method is to choose a longer period of time and to select the chronological distribution of the good and bad loans so that they are approximately identical; that is, if 25 percent of the good sample is selected from loans made in the first half of 1936, about the same proportion of the bad sample should cover the same period. Lenders who contributed to this study were asked to select their samples by the latter method; they were requested to select their bad-loan sample first, and then to select the good-loan sample, with approximately the same distribution. On the whole, we do not have information concerning either the accuracy with which they were able to follow this procedure or the sort of chronological distribution that resulted, but we presume that most of the loans in the samples were made during the period from 1935 through 1938. One of the industrial banking company samples, it is true, was carefully broken down to show experience in three successive years; in this form the sample failed to show any significant variation, but this failure may well be attributable to the fact that the number of cases in the sample was smaller than one would wish. Obviously this study does not throw any light on the effect of time on risk experience, and the results should be considered as averages related to a rather undefined period of about 4 years' duration in the near past.

[15] The results of some of these analyses have appeared elsewhere. For actual tabulation of the component commercial bank samples see John M. Chapman and Associates, *op. cit.*, Appendix B. A tabulation of the industrial bank components will be found in National Bureau of Economic Research (Financial Research Program), *Industrial Banking Companies and Their Credit Practices,* by Raymond J. Saulnier (1940) Chapter 6.

SUMMARY OF PROCEDURE

The following summary lists the more important steps to be taken and the more serious difficulties likely to be encountered in an analysis of risk experience based upon sampling procedure.

Determination of the quality of loans to be included in both the good-loan sample and the bad-loan sample is the first problem of risk analysis. The bad-loan cases should, if feasible, contain all types of clearly unsatisfactory repayment experience, and nothing else. In some cases, however, the mechanical process of selecting loans from the files will be greatly simplified if the bad loans are limited to some specific class, such as repossessions or charge-offs; in other cases, the number of clearly bad loans may be so small that the inclu-

sion of borderline cases may be necessary to obtain a sample of adequate size, i.e., a sample that includes at least 200 cases. Good loans can be variously defined, depending on the desires of the analyst and the type of filing system from which the loans are drawn; they can be defined as clearly exemplary cases, or as cases not classified as bad loans.

The number of cases chosen will depend on several considerations: for example, the nature of the specific task to be performed, the amount of labor time available, and the degree of precision desired. In general, 200 good loans and 200 bad loans represent the absolute minimum on which a sample should be based, although trained statisticians may frequently see opportunities for solving special problems with considerably smaller numbers. Even 200 cases, however, will probably be insufficient for a satisfactory study of occupation or other factors requiring detailed analysis; a thousand cases is probably desirable here, and even more may be required if particular detail or great accuracy is necessary.

The mechanical process of drawing cases out of the loan file is one that must be devised to fit the individual case. The first requisite is that the drawing should be properly random in order to eliminate all conscious or unconscious personal bias as well as other undesirable biases that sometimes result from non-random sampling; the use of a table of random numbers is definitely advantageous. The second requisite is economy of effort, and in this connection, a little ingenuity on the part of the analyst may save considerable work.

The effect of changes in time on risk experience can be avoided in three ways: the study can be limited to a short and rather homogeneous period; the selection of loans can be so arranged that the chronological distribution of the good loans is approximately identical with that of the bad; and a number of separate studies can be made of several short, homogeneous periods.

An illustration of the method by which samples can be

tabulated appears in Table 3. Limitation of the number of class intervals is important in making such a tabulation; moreover, no class interval should contain fewer than 30 loans, good and bad combined. As soon as the percentage distributions among the various class intervals have been computed, the bad-loan relatives and the efficiency index can be computed. The bad-loan relative, which is the percentage of bad loans in any class interval divided by the percentage of good loans, will indicate the classes that represent particularly good or particularly bad risks; and the efficiency index, which has been described above, will permit comparison of the effectiveness of different factors as indicators. The differences observed between the good- and bad-loan distributions based on a sample of only 200 cases, however, may not be genuine. While the reliability of the results should be examined by use of one of the standard tests (see footnotes 2 and 3, pages 24 and 26), the efficiency index can be used as a poor substitute. If all results yielding an efficiency index of 15 are rejected, a number of false conclusions will be avoided. Of course, if a result obtained by the procedure outlined above is rejected as unreliable, further evidence may be sought to establish reliability. Whether to discard a result or to seek additional information is usually a question that must be decided in relation to circumstances.