Credit Scoring and the Equal Credit Opportunity Act

David C. Hsia

Follow this and additional works at: https://repository.uchastings.edu/hastings_law_journal

Part of the Law Commons

Recommended Citation
Available at: https://repository.uchastings.edu/hastings_law_journal/vol30/iss2/5

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Law Journal by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.
Credit Scoring and the Equal Credit Opportunity Act

By David C. Hsia*

Credit scoring is an empirical technique that uses statistical methodology to predict the probability of repayment by credit applicants. A typical system evaluates certain financial and nonfinancial characteristics of each applicant on a scoring table (See Figure 1 on page 375). The table derives from the creditor's past experience with similarly situated applicants and assigns a varying number of points to each of the characteristics, depending upon the applicant's responses. The creditor adds up the applicant's points to obtain a total score. If the applicant scores higher than the predetermined cutoff, the creditor approves the credit request.¹

Recently, much political and legal controversy has surrounded the role of scoring systems in analyzing creditworthiness. There is particular concern over the inherent potential of such systems for discrimination against federally protected classes of applicants.² Observers also disagree about the extent to which federal law applies to credit analy-

---


sis. To date, this conflict has yielded considerably more questions than answers.

This Article attempts to resolve some common misunderstandings as to the nature and function of credit scoring, and to allay the confusion regarding the application of the law to scoring systems. It also seeks to illuminate those areas of the law whose impact on credit scoring remains uncertain and to indicate the ways in which empirical systems of credit analysis may tend to discriminate illegally. Sections I and II detail the practical reasons and scientific principles underlying credit scoring, and Section III delineates the process by which scoring systems are constructed. These sections are offered as introductory material for those unfamiliar with credit scoring. The remaining sections discuss legal problems associated with scoring.

The Reasons for Credit Scoring

Judgmental systems

Because the creditor seldom knows the applicant personally, and because success or failure depends on extending credit to those who will repay, the creditor needs some means of analyzing creditworthiness. Traditional credit analysis uses human judgment to evaluate creditworthiness. Credit officers analyze incoming applications in light of their own prior experience and their employer's institutional guidelines. If, for example, plumbers or elderly applicants traditionally


Suppose that someone walks into a bank or store. That person wants to obtain certain goods or services, but does not have sufficient funds to do so. The prospective customer therefore requests credit. The creditor receiving the request does not know whether the credit applicant will honor the promise to repay the loan. The creditor therefore needs credit analysis.

Credit analysis is a method of determining whether an applicant constitutes a good or bad credit risk. If the applicant repays the loan as agreed, the creditor will profit from the interest it charges. Conversely, if the applicant defaults, the creditor stands to lose whatever amount remains unpaid. Consequently, a creditor's ability to distinguish creditworthy from noncreditworthy applicants is crucial to its success in the credit business.

5. See BANK ADMINISTRATION INSTITUTE, JOB DESCRIPTIONS FOR BANK PERSONNEL 133-37 (1975); Speech by Dr. Edward M. Lewis, Consumer Bankers Association, Central
have had trouble repaying, the credit officer will summarily reject future applications from these groups. In theory, the institution directs its credit officers to look for applicants with the (1) ability and (2) willingness to repay. These two attributes would appear to cover every possible situation. An applicant with no means of repaying the credit constitutes a poor risk, however good his or her intentions. Similarly, an applicant who does not intend to repay should not receive credit, regardless of personal wealth or income.

In practice, however, these judgmental systems of credit analysis suffer from several critical flaws. First, the credit officer may have an imperfect recollection of past experience, or one very troublesome incident may distort the officer's view of a particular group.

In addition, the judgmental system fails to react well to changes in the composition and creditworthiness of the pool of credit applicants. Conceivably, tree surgeons, for example, may have been poor credit risks at one time. However, as years pass tree surgeons may improve substantially as credit risks because there may be growing public interest in proper care of trees and restrictive government licensing of new tree surgeons. The credit officer might recall the previous unfortunate credit experience with a tree surgeon and steadfastly refuse to extend any tree surgeon credit. As a result, the creditor will never discover that the creditworthiness of tree surgeons has improved.

Western Section Annual Meeting (May 17, 1977); Interview with Mr. Edward Berger, Montgomery Ward, by the Federal Reserve Board staff, in Washington, D.C. (Dec. 19, 1974).


Judgmental systems also lack accuracy in identifying creditworthy applicants because of unrealizable institutional guidelines. These guidelines serve as a substitute for the experience that a new credit officer may lack. Unfortunately, many of these so-called credit policies arise without empirical verification. Thus, the creditor's senior management may adopt a traditional slogan, such as the three “P’s” or the three “B’s,” on little more than an intuitive notion that the principle makes sense. In fact, a quantitative study might show that some people in these discredited categories actually make excellent credit risks. Without controlled research, the observer cannot ascertain the validity of institutional guidelines.

Finally, judgmental systems may discriminate illegally. The credit officer has complete discretion when weighing all the pertinent information mentally. Personal appearance unquestionably plays a role in the credit granting decision. Studies indicate that factors such as the applicant's race, sex, marital status, and age may also receive consideration. However, federal law now makes consideration of these factors illegal discrimination.

Despite these potential defects, judgmental systems of credit anal-

over $25,000 may not properly reflect the effects of rapid inflation. Many of the new applicants with sufficient income to qualify under the old standard may in fact lack creditworthiness because their income has failed to keep pace with the rising costs of living.

9. “Never lend to preachers, plumbers, or prostitutes.” Main, supra note 1; cf. M. IRVING, THE BANK BOOK (1973) (suggesting other occupations as falling within the three “P’s”). However, do not confuse this with more general rules such as the three “C’s,” New Math, supra note 6, at 2.

10. “Never lend to beauticians, bartenders, or barbers.” Main, supra note 1.

11. See S. REP. No. 94-589, supra note 2, at 6. Because little controlled research of this nature apparently exists, most institutional guidelines probably have limited validity. E. FIEDELER, MEASURES OF CREDIT RISK AND EXPERIENCE (1971).

Even the more abstract criteria such as “ability and willingness to repay” may prove unreliable. What constitutes ability to repay? Experienced credit officers can recount tales of unemployed fathers who nevertheless managed to keep making payments on their mortgages month after month, or of convicted felons with no declared income who continued both to use and pay off their credit cards while incarcerated. An applicant's willingness to repay is even more difficult to ascertain than his or her ability.


ysis are still widely used.\textsuperscript{14} Indeed, despite the possible failings enumerated above, judgmental systems provide their users with a relatively effective approach to credit analysis. Many judgmental systems enjoy extremely accurate predictions of creditworthiness, especially if operated by an experienced group of officers using well-tried institutional guidelines.\textsuperscript{15}

Scoring systems

Credit scoring has developed as a recent alternative to judgmental credit analysis. A scoring system typically consists of two main components: a scoring table (Figure 1) developed from analysis of the creditor's past applicants and a repayment probability table (Figure 2) which relates total points scored on the scoring table with a percentage probability of default or repayment.\textsuperscript{16}

Figure 1—Hypothetical Credit Scoring Table*

<table>
<thead>
<tr>
<th>Applicant's Age</th>
<th>under 25 years</th>
<th>25-29</th>
<th>30-37 or no answer</th>
<th>38-46</th>
<th>47-50</th>
<th>51-61</th>
<th>62 and up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time at Present Address</td>
<td>under 1 year</td>
<td>1-2</td>
<td>2-3</td>
<td>3-5</td>
<td>5-9</td>
<td>10 or over</td>
<td>no answer</td>
</tr>
<tr>
<td>Age of Apto</td>
<td>None</td>
<td>2</td>
<td>3-4</td>
<td>5-7</td>
<td>8 or over</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Housing Cost</td>
<td>$125 or more</td>
<td>$125-270</td>
<td>$271 or more</td>
<td>Owns Clear</td>
<td>Relatives</td>
<td>24</td>
<td>no answer</td>
</tr>
<tr>
<td>Finance Company References</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Credit Card</td>
<td>None</td>
<td>1-5%</td>
<td>2-10</td>
<td>3-15</td>
<td>4 and up</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Debt to Income Ratio</td>
<td>No Debts</td>
<td>1-5%</td>
<td>6-15%</td>
<td>16% or over</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of Downpayment</td>
<td>None</td>
<td>1-15%</td>
<td>16-50%</td>
<td>50-85%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This table does not reflect any creditor's actual experience, but is offered as an illustration only.


The scoring table typically contains six to twelve characteristics.

On the hypothetical scoring table in Figure 1, the characteristics include: age, time at present address, age of auto, monthly housing cost, bank accounts, finance company references, major credit cards, debt to income ratio, and amount of downpayment. Each characteristic contains two or more attributes. Thus, the attributes: under 25 years, 25-29, 30-37, 38-46, 47-50, 51-62, and 62 and up comprise the characteristic "applicant's age" on the hypothetical table. The attributes therefore constitute the various possible answers to the credit questions asked of the applicant. The scoring system uses the answers to predict creditworthiness. An attribute may encompass a single possible response, such as no, yes, or no answer; or it may span a series of possible responses, such as 2-5 years. The intervals spanned by the latter type of attributes may change regularly, as with 0-5, 6-10, 11-15, etc.; or irregularly, as with 1-2, 3-7, 8-11, etc. An applicant is assigned only one attribute for each characteristic.

Finally, each attribute shares its cell with a numerical score. This number expresses the degree to which the creditor's past experience statistically associates creditworthiness with that attribute when analyzed in conjunction with the other characteristics in the system. Thus, a high score does not indicate that past applicants with that attribute all repay their loans on schedule, but that applicants combining this attribute with other high score attributes make superior credit risks. Bor-
rowers behave in consistent, but nonprogressional ways which accounts for the scores' tendency to change erratically from cell to cell rather than always to ascend in a smooth curve.

The scoring system processes credit applications by noting the responses disclosed on the application and comparing them with the scoring sheet. The creditor totals the scores for each of the applicant's attributes. The total score either meets or fails to meet a predetermined cutoff. The cutoff score determines whether the creditor rejects the applicant, refers the application to a credit officer for further review, obtains a more detailed credit report, or accepts the applicant immediately. The cutoff score derives from the creditor's business evaluation of the repayment probability table and the acceptable default rate.

Public interest in credit scoring

The importance of credit scoring to the public is principally a result of its increased utilization by large credit grantors. Several thousand scoring systems currently analyze creditworthiness. While the number of applications processed by each system varies considerably, their users include many of the country's largest lenders. Thus, scoring systems process an increasing fraction of the 200 million applications submitted to creditors each year.

The uniformity imposed upon credit analysis by a single scoring system also makes this technology increasingly important to the public. For all practical purposes, under a judgmental system each individual credit officer forms a separate system. The officer employs subjective standards for analysis and creditworthiness, balancing all the pertinent considerations mentally. Two officers of the same judgmental creditor may easily disagree about the acceptability of the same application. In addition, even the most zealous of judgmental credit officers can evaluate only a limited number of applications per day. In contrast, a

---

18. Concepts of Scoring, supra note 17; Main, supra note 1; Weingartner, supra note 16.
20. Section III describes the origin of the repayment probability table.
21. FICO Interview, supra note 7.
large scoring system can process an almost unlimited volume of applications, depending only upon the availability of clerks to apply the system. Thus, a single scoring system can have a much greater impact upon credit applicants than can any judgmental system. If the scoring system works well, most applicants will receive fair treatment based on their actual creditworthiness. However, if it works poorly, the scoring system may have a disproportionately harsh impact upon large numbers of creditworthy applicants.

The public interest in credit scoring systems also derives from the increased attention Congress has devoted to scoring technology. For example, in addition to considerations already prohibited, in its 1976 amendments to the Equal Credit Opportunity Act, Congress extended its ban on sex and marital status discrimination to also prohibit discrimination on the basis of age in any aspect of a credit transaction. Congress provided, however, that a creditor did not commit age discrimination if it used "any empirically derived credit system which considers age if such system is demonstrably and statistically sound in accordance with regulations of the [Federal Reserve] Board, except that in the operation of such system the age of an elderly applicant may not be assigned a negative factor or value." This means that while judgmental systems of analysis may no longer consider age as a predictor of creditworthiness, properly constructed scoring systems may continue to do so.

---

25. MDS Systems Approach, supra note 7; Executive Summary, supra note 7, at 1.
26. 15 U.S.C. § 1691(a)(1) (1976). Although a credit scoring system may not consider the other prohibited bases of the Equal Credit Opportunity Act, such as race, color, religion, national origin, sex, marital status, receipt of public assistance, and good-faith exercise of Consumer Credit Protection Act rights, certain existing systems apparently use the applicant's address as a predictive variable. See text accompanying note 212 infra. This treatment may derive from actual evidence. D. Beebe, Retail Credits and Collections 20 (1919) [hereinafter cited as Beebe]; Myers & Forgy, supra note 14, at 802. Such practices have understandably given rise to allegations of indirect discrimination and redlining. See authorities cited note 3 supra; Capon, Credit Ratings and Rights, Washington Post, Dec. 17, 1977, at A19, col. 1; Address by Commissioner Elizabeth Hanford Dole, Nat'l. Ass'n of Women Lawyers Annual Meeting, in Chicago (Aug. 6, 1977), in FTC Investigates Big Credit Card Issuers for Possible Bias Against Applicants, Wall Street Journal, Aug. 8, 1977, at 11, col. 2. These charges have resulted, in turn, in the introduction of additional legislation to further regulate the use of credit scoring systems. H.R. 8451, H.R. 8464, H.R. 8510, H.R. 8580, H.R. 9196, 95th Cong., 1st Sess. (1977).
Private interest in credit scoring

Many business considerations have prompted creditors to adopt scoring on a large scale. The primary advantage is scoring's more accurate credit decisions through use of the institution's actual credit experience. Credit scoring avoids limitations of individual remembrance utilized by judgmental systems by statistically analyzing the institution's actual experience.29 A computer isolates the various predictive characteristics and accurately counts the actual repayment experience associated with each characteristic. This purely analytical methodology should reduce the opportunity for human bias and thereby increase the accuracy of credit analysis.30 A sound scoring system, therefore, confers the multiple advantages of fair treatment of individual applicants based on actual creditworthiness, improved quality of the creditor's loan portfolio, and increased profitability.

Credit scoring enjoys the second business advantage of providing the exact probability of repayment associated with a particular score. A judgmental system requires the credit officer intuitively to estimate the probability of repayment based upon his or her prior experience. Credit scoring eliminates this guesswork by comparing the applicant's characteristics with a sample of the creditor's past applicants with similar characteristics. Statistical analysis determines the exact percentage of those applicants who repaid their credit in a satisfactory manner. Knowledge of the exact probability of repayment permits the creditor's management to select rationally the degree of risk it wishes to accept when extending credit.31

29. BANK ADMINISTRATION INSTITUTE, SEVENTH BIENNIAL SURVEY OF BANK PERSONNEL POLICIES AND PRACTICES 6 (1974); Myers & Forgy, supra note 14, at 799. Cf. C. PHELPS, RETAIL CREDIT FUNDAMENTALS 73 (4th ed. 1963) (credit analysts should give attention to changes in credit conditions) [hereinafter cited as PHELPS]. A judgmental system uses the memories of its credit officers to synopsize the institution's past experience. If the officer associates certain attributes with defaulting borrowers, the officer will use this association to reject future applicants having those attributes. This method of using the institution's prior credit experience suffers from the drawback of incomplete or inaccurate human recollection. The institution's credit officers may turn over rapidly. Experienced credit officers may be unduly impressed by unusual events, such as defaults, and consequently exclude satisfactory borrowers with identical attributes.


Credit scoring also has the advantage of incurring very low operating costs. A judgmental system requires that skilled credit officers give extensive personal scrutiny to each application.\(^\text{32}\) Although scoring entails very high startup costs for development and installation, thereafter clerical personnel, supported by a high degree of automation, can process applications.\(^\text{33}\)

Scoring may also reduce the cost of credit analysis by reducing the use of credit bureau reports. The credit report costs the creditor several dollars, while the creditor already has the "scorable" application in hand.\(^\text{34}\)

For a high volume creditor, the cost of obtaining credit reports may run into millions of dollars each year. Credit reports serve two principal functions, confirmation of information disclosed on the application and investigation of the applicant's past credit history.\(^\text{35}\) Credit scoring requires neither of these functions. Judgmental credit analysis operates on the theory that certain facts about credit applicants, such as their income or job status, are positively or negatively correlated with the probability of repayment on schedule. The credit officer therefore needs to confirm the veracity of each applicant's statements about these pertinent elements of creditworthiness. Credit scoring relies upon an alternative theory, that large populations behave consistently.\(^\text{36}\) Accordingly, a scoring system processes the fact that an applicant says his income amounts to $18,000, not that the income actually amounts to $18,000.


\(^{33}\) See Biborsch, Credit Scoring Systems Have Built In Bonuses, Bankers Monthly, Mar. 1967, at 40; Buel & Lewis, Credit Scoring—and Beyond, Banking, Feb. 1969, at 42 [hereinafter cited as Buel & Lewis]; Credit Scoring Can Mean Efficiency, Credit Union Magazine, June 1974, at 39; Executive Summary, supra note 7, at 1; Presby & Simon, Credit Scoring Can Save Money and Improve Credit Granting Too, Stores, Oct. 1969, at 17; Weingartner, supra note 16.

\(^{34}\) See authorities cited note 38 infra; Hearings on S. 823, infra note 35, at 334; Chandler & Ewert, The Value of Credit Reports Versus Their Costs (1976) (unpublished paper, Georgia State University); Chandler & Coffman, supra note 7, at 1; Roy, supra note 7, at 27.

\(^{35}\) National Comm'n Report, supra note 1, at 212; Hearings on S. 823 Before the Subcomm. on Financial Institutions of the Senate Comm. on Banking and Currency, 91st Cong., 1st Sess., at 236 (1969) [hereinafter cited as Hearings on S. 823].

\(^{36}\) C. Moore & P. Klein, The Quality of Consumer Installment Credit 19, 83 (1967); Weingartner, supra note 20; National Comm'n Report, supra note 1; Hearings on S. 823, supra note 35.
that much. Indeed, some scoring systems score the statements from credit reports just as they do applications, on the basis of their having been made rather than their veracity.\textsuperscript{37}

Credit reports also disclose an applicant’s prior credit performance which theoretically indicates the borrower’s willingness to repay.\textsuperscript{38} Critics have charged, however, that credit reports frequently prove inaccurate.\textsuperscript{39} They may serve as no more than a probable indicator of the applicant’s actual “character.” Scoring the application alone can act as an equally reliable predictor of character. It merely uses the applicant’s statements on the application, rather than the credit bureau’s statements about the applicant, as predictors.\textsuperscript{40} Both sources of information present potential for inaccuracy. Indeed, the applicant’s score, as calculated from the application, may prove a more accurate and less expensive predictor of character than a credit report. The creditor understands both the individual score and the scoring system and controls all aspects of their use, whereas the credit report emanates from a credit bureau whose operations are beyond the control of the creditor.

Scoring has the fourth advantage of giving corporate management precise control over the amount of credit its officers extend. The relative attractiveness of consumer credit as an investment varies as interest rates respond to changes in the economy. By adjusting the cutoff score, management can increase or decrease the amount of consumer loans in the creditor’s portfolio. Statistical analysis of the institution’s past applicants also provides some advance indication of the exact degree to which changes in the cutoff score will affect its volume of consumer credit. Under a judgmental system, when management desires to reduce credit volume it can only instruct its credit officers to “tighten up” credit standards. In a large bank or finance company, the effect of such instructions will vary from branch to branch. Management cannot fine tune the credit volume.

Scoring has the final business advantage of providing a mechanism to detect changes in the characteristics which predict creditworthiness and in the composition of the population of creditworthy applicants. As the economy meanders through its normal business cycles, the composition of a creditor’s applicants will alter. New types of

\textsuperscript{37} See, e.g., Concepts of Scoring, supra note 17, at 73.
\textsuperscript{38} BEEBE, supra note 26, at 77, 109; PHELPS, supra note 29, at 62, 74.
\textsuperscript{40} See, e.g., Concepts of scoring, supra note 17, at 73; MDS Systems Approach, supra note 7.
people move into an area. Local conditions change. Inflation occurs. As these events transpire, the accuracy of an older scoring system will eventually decline. When predictive ability declines to unacceptable levels, the creditor simply undertakes renewed statistical analysis of more recent applicants and develops a new scoring system which takes into account changes in the characteristics which predict creditworthiness. The predictive ability of a judgmental system can decay too, but the system does not detect changes as routinely, and there exists no obvious remedy for its decline.

Scientific principles of credit scoring

Before proceeding with a description of the construction of a scoring system and the statutory constraints on its use, the "scientific" principles from which scoring derives should be examined. The following discussion simplifies the actual nature of the methodology, but seeks to convey a general grasp of the underlying concepts.

The future will resemble the past

All forms of credit analysis necessarily start with the assumption that future events will resemble past events. Thus, future creditworthy applicants will tend to resemble past creditworthy applicants. Scoring merely refines the identification of personal characteristics associated in the past with satisfactory credit risk.

The reader should note two corollaries to this principle. First, the characteristics used by a scoring system to predict creditworthiness do not cause creditworthiness. The scoring system models the behavior

---


42. Roy & Lewis, supra note 7, at 18.

43. The assumption that events, precipitated by human behavior, follow consistent patterns may seem unwarranted. Its validity persists, however, because credit analysis has not yet developed any superior substitute for past experience as a predictor of future experience. In addition, this assumption has worked reasonably well, obviating the need to find a substitute. Cf. T. Kuhn, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970) (paradigm shift in the sciences occurs suddenly and in response to substantial breakdown of previously accepted models).

44. Address by William R. Fair, Fair, Isaac Technical Seminar on Credit Scoring (April 14, 1977); Weingartner, supra note 20; cf. authorities cited note 76 infra.
of borrowers; it does not analyze the reasons for their behavior. The system developer determines what characteristics are positively or negatively correlated with prompt repayment, but not why such a correlation occurs. Therefore, many characteristics that may have high predictability may incidentally create a disproportionate impact on certain minority groups.\textsuperscript{45} Statistical methods for investigating causal forces do exist, but scientists have not yet deployed these techniques in the credit analysis field.\textsuperscript{46} Present scoring neither assumes nor investigates the existence of causal linkages.

Second, the assumption that the future will resemble the past does not require that credit applicants tell the truth.\textsuperscript{47} It merely requires that they respond consistently. The system relies upon the applicant's assertion of a certain attribute and scores that attribute's creditworthiness, rather than relying on the fact that the applicant actually possesses the asserted attribute, as a predictor.\textsuperscript{48}

Groups of individuals have predictable behavior

The theories of probability and statistics depend upon the existence of a large number of similar events for analysis.\textsuperscript{49} A single event has little meaning to the statistician.\textsuperscript{50} However, a series of events per-
mits numerous conclusions and projections about what will happen “in the long run.”

Accordingly, scoring depends upon the availability of a large amount of credit experience. A group of past applications forms the development data base. A large volume of new applications assures frequent future comparisons to the profile compiled from the development data base. The system predicts the performance of these new applicants in the aggregate. The overall credit performance of many new applicants will approximate that of the data base. Credit granted to one applicant may or may not result in repayment. Credit granted to many applicants, however, entails predictable levels of risk. On the average, the same percentage of these applicants will repay as in the equivalent portion of the development data base.

That their validity is based on an average does not necessarily mean that scoring systems must function more impersonally than judgmental systems. In fact, scoring treats the applicant just as individually as a judgmental system. Both approaches to credit analysis compare the applicant to the mass of past, similarly situated applicants. The scoring approach tends to have more scientific accoutrements and therefore gives the impression of treating the applicant as a standardized component. The judgmental system treats the applicant in exactly the same way, while giving the impression of being more personal.

A set of measurements can distinguish between two groups

The population of credit applicants consists of two nonoverlapping subpopulations, creditworthy applicants and noncreditworthy applicants. An applicant must belong to one, and only one, of these subpopulations. Accordingly, the creditor has great interest in knowing


53. Chandler & Coffman, supra note 7, at 7; New Math, supra note 6, at 2; see Roy & Sanderson, supra note 17, at 27.

Any given population may consist of two or more nonoverlapping subpopulations. Examples of such populations include males and females, left and righthanded people, and persons with IQs of over 100 and under 100. Each member of the population belongs to one, and only one, of the subpopulations.
what characteristics distinguish the two subpopulations. For example, consider an *interval scale* characteristic such as the size of a credit applicant's proposed downpayment.54

Figure 3—Hypothetical frequency distribution (average separation)

As illustrated in Figure 3, plotting the number of creditworthy and noncreditworthy applicants at each point of the scale results in a *frequency distribution*.55 The frequency distribution indicates the degree to which this characteristic distinguishes the two subpopulations. If the overlap is considerable, as in Figure 4, “size of downpayment” does a

Figure 4 - Hypothetical frequency distribution (poor separation)

54. An interval scale has a unit of measurement that serves as a common standard for all possible attributes. Thus, all incomes appear somewhere on an interval scale running from zero dollars to infinite dollars, and all ages appear on an interval scale of years. The interval scale contrasts with the nominal scale. A characteristic forming a nominal scale has several different attribute categories that neither overlap with each other nor share a common unit of measurement. Thus, the characteristic “marital status” offers only two possible alternatives, married or unmarried. These attributes, although mutually exclusive, do not share an underlying unit of measurement. Similarly, the responses to “place of residence” might include owns, rents, parents, or other. See Blalock, supra note 46, at 14.

55. An ordinary frequency distribution counts the number of cases or events (each applicant, in the case of credit scoring) occurring at each point on an interval scale. It displays this “frequency count” vertically as a height over each point on the horizontal scale. See generally Blalock, supra note 46, at 41-53.
poor job of predicting credit performance. If, on the other hand, the overlap is insignificant, the characteristic effectively separates the subpopulations.

At some points on the scale, the subpopulations do not overlap. Future applicants whose attributes fall in such nonoverlapping spans are assigned to the appropriate subpopulation. As the width of nonoverlapping spans increases, more applicants are correctly classified and the characteristic better predicts creditworthiness. Even where overlap does occur, the relative heights of each subpopulation curve at that point on the frequency distribution still indicates the ratio of creditworthy to noncreditworthy applicants in the total population. In distributions of limited overlap, the creditor can therefore use the predictive characteristic to grant credit to applicants on some overlapping portions of the frequency distribution because the percentage of noncreditworthy applicants granted credit remains at predictable and acceptable levels.

The creditor should clearly prefer to use the available characteristic that maximizes the separation between the two subpopulations. Credit scoring elaborates on the foregoing principle by relying on combinations of distinguishing characteristics to further increase the distinctiveness of the subpopulations. High income and home ownership may both indicate creditworthiness on their own frequency distributions. However, a combination of both attributes may mean extremely high creditworthiness.

A sophisticated statistical technique known as discriminant analysis refines this intuitive process of combining several predictive characteristics. Basically, it subjects creditworthiness data to mathematical calculations in order to assemble an index of creditworthiness. This

---


\[ Z = L_1 X_1 + L_2 X_2 + L_3 X_3 + \ldots + L_p X_p \]

where \( X \) represents the characteristics available to distinguish the subpopulations and \( L \) represents the linear coefficient that maximizes the subpopulations' divergence. Fisher, The Use of Multiple Measurements in Taxonomic Problems, 7 Annals of Eugenics 179 (1936); Fisher, The Statistical Utilization of Multiple Measurements, 8 Annals of Eugenics 376 (1938).

57. Chaterjee & Barcum, A Nonparametric Approach to Credit Screening, 65 Am. Stat. Assn. J. 150 (1970); see Durand, supra note 6, at 22-28; Myers & Forgy, supra note 14, at
index includes the available characteristics that, taken together, best predict creditworthiness. Discriminant analysis not only selects these characteristics but also takes into account the scaling of their attributes and weighs each characteristic so as to express properly its relative contribution to the final index of creditworthiness. Plotting the subpopulations over the index produces a frequency distribution with minimum overlap with respect to the data available.

Figure 5—Hypothetical frequency distribution
(excellent separation)

The characteristics that distinguish best are incorporated into the index and will appear on the scoring sheet. An applicant’s place on the index is determined by the total score. Charting the scores and performance of a sample of applicants results in a frequency distribution similar to that produced in Figure 5 for a single-interval characteristic. Because the horizontal axis\(^58\) graphs the combined effect of several characteristics, instead of a single predictive characteristic, the subpopulations overlap less. The user can therefore make more accurate predictions about the subpopulation to which a new applicant belongs.

**System Construction**

As a prerequisite to the full comprehension of how the law affects credit scoring, the reader must understand the technical construction, operation, and nature of scoring systems. This section therefore attempts to describe how a developer goes about building a typical scoring system. This discussion is designed only to familiarize the reader

\[ \frac{(Uc - Un)^2}{1/2 (Dc^2 + Dn^2)} \]

where \(Uc\) represents the mean score of the creditworthy subpopulation, \(Un\) represents the mean score of the noncreditworthy subpopulation, \(Dc\) represents standard deviation of the creditworthy subpopulation and \(Dn\) represents the standard deviation of the noncreditworthy subpopulation.

\(^58\) The horizontal axis is also called the X axis, the abscissa, or the independent axis in a plane Cartesian coordinate system. Other names for the vertical axis include the Y axis, ordinate, and dependent axis.
with the nomenclature and processes of system development. Other procedures for system construction exist besides the one described here. However, most professional developers generally use the following model.

Definition of system scope

The creditor first decides upon the coverage of the system, namely, what applicants it will process. Some nationwide creditors use a single system for the entire country. Others use a different system for each branch or store. Other systems process only applicants under age twenty-five from the Midwest. Full service creditors offering a variety of credit plans may use scoring for some plans but not for others.

The choice of system scope depends principally upon the nature of

59. For example, with a large enough sample, the developer could simply divide the applicant population along arbitrarily selected characteristics and attributes. Each of the thousands of cells of this multidimensional matrix would have a repayment rate calculated on the basis of the applicants falling into that cell. The creditor would select the overall repayment rate it found acceptable and the developer would use this cutoff rate to designate all cells as either acceptable or unacceptable. An incoming application would be assigned to one of the cells in the matrix according to its respective attributes. The cell's designation would determine whether the creditor approved the request. However, compiling enough data to compute a reliable repayment rate for each of the thousand-odd cells would entail a development sample having more than a million cases. Logistical and cost considerations preclude such a mathematically inefficient approach to system construction.

More realistically, one professional developer uses a technique known as dynamic programming to construct its scoring systems, instead of the more conventional combination of regression analysis and discriminant analysis described above. Dynamic programming solves the problem of analyzing multiple relationships by dividing the problem into "decision stages," working backwards from the stated objectives, solving the simplest stages first, and assembling the individual stages into a complete system only after independently solving all of the intermediate stages. Mathtech, MATHSCORE-A Computer System for Numerical Credit Scoring (Nov. 15, 1977) (publication of Mathtech, Princeton Station Office Park, P.O. Box 2392, Princeton NJ 08540) [hereinafter cited as MATHSCORE]; Mathtech, Mathscore Weight Development Algorithms (Oct. 1, 1977) (publication of Mathtech, Princeton Station Office Park, P.O. Box 2392, Princeton, NJ 08540); Interview with Richard D. Koelsch, Director of Business Services, Mathtech, Inc., in Washington, D.C. (Jan. 5, 1978); cf. F. Hillier & G. Lieberman, INTRODUCTION TO OPERATIONS RESEARCH 248-79 (2d ed. 1974).

60. This discussion will not enable the reader to rush out and immediately develop a scoring sheet and repayment table. Indeed, very few creditors have the staff expertise to develop their own scoring systems. Instead, they usually hire outside consultants to examine their files and assemble the credit information into one or more scoring systems. Surprisingly, few businesses offer system development services. Only half a dozen such consultants have come to the author's attention.

the creditor’s business. Although a system requires, at least, several thousand applications as a data base, too inclusive a development population may result in unnecessary loss of predictive power. The creditor, using a single system for the whole country, may find itself unable to extend any credit in Florida because of the unusual demographic composition of the state’s population. The creditor exercises business judgment, rather than scientific doctrine, when selecting whom the system will cover.

Classification of accounts

The creditor then defines what it considers a satisfactory account and an unsatisfactory account. Using hindsight, it classifies past borrowers as creditworthy or noncreditworthy. This definition establishes a nominal scale for the characteristic creditworthiness. This classification process serves as the starting point from which the two subpopulations which the creditor wishes to separate will derive.

Sampling

The initiative now passes to the system developer, who uses the creditor’s records to draw samples of satisfactory and unsatisfactory accounts and rejected applicants. Each of these three samples will, if properly drawn, statistically represent the characteristics of the underlying applicant population in an unbiased fashion. Selection of appropriate sampling techniques comprises a specialized and arcane subject. Accordingly, this subsection can mention only a few sam-

62. See note 51 & accompanying text supra.
63. Churchill, Nevin & Watson, Credit Scoring—How Many Systems Do We Need?, THE CREDIT WORLD, Nov. 1977, at 6; see text accompanying note 217 infra.
64. Chandler & Coffman, supra note 7, at 7; New Math, supra note 6, at 2; FICO Position Paper 2, supra note 61, at 1; MDS Systems Approach, Sec. III, supra note 9; Roy, supra note 7, at 56-58.
67. See generally W. Deming, SAMPLE DESIGN IN BUSINESS RESEARCH (1960); W. Deming, SOME THEORY OF SAMPLING (1950); M. Hansen, W. Hurwitz & W. Madow, SAMPLE SURVEY METHODS AND THEORY (1953); W. Hendricks, THE MATHEMATICAL
pling issues of particular interest to scoring system development.

Most creditors have far more satisfactory accounts than they have unsatisfactory accounts.68 A normal, random sample therefore cannot fully represent the attributes of the subpopulation of unsatisfactory accounts without pulling far more satisfactory accounts than necessary. Thus, most developers elect to use stratified sampling, a technique which draws an equal number of cases from each of several subpopulations of disproportionate size.69 This technique gives a more efficient representation of the attributes of the smaller subpopulation.70 Thus, the developer might select equal samples of 1,000 cases each from unequal subpopulations of 7,000 satisfactory borrowers, 1,000 unsatisfactory borrowers, and 2,500 rejected applicants.

In addition, sampling of existing accounts does not give proper consideration to the attributes of rejected applicants.71 The scoring system must model the population consisting of all the creditor's applicants, not merely the creditor's borrowers. To do otherwise would predict the performance of only part of the population, whose attributes may differ greatly from those of the overall population it will process. Methods for integrating rejected applicants into the borrower population will receive further treatment below.72

A proper development sample also requires that its accounts have sufficient "age." Like any other product, credit accounts have life-cycles. A closed-end account73 goes through a sequence of steps: appli-
CREDIT SCORING

cation, credit analysis, approval, disbursement of proceeds, a series of repayments by the borrower, possible default, and eventual full payment or writeoff. A creditor's portfolio consists of many accounts in different stages of their lifecycles. In order to compare the attributes of satisfactory, as opposed to unsatisfactory borrowers, the accounts used in the development sample must have had the opportunity to develop both default and repayment experience. A sample of new accounts imparts no credit information because none of the borrowers has yet had time to default. Frequently, only accounts opened several years previously have accumulated sufficient unsatisfactory experience to become useful to the developer. The accepted and rejected applications in the sample must have occurred at about the same time in order to permit meaningful comparisons. This need for vintage accounts also implies a constant gap between the attributes and performance of current applicants and of the development sample.

Coding

The developer next utilizes every available, pertinent item of information about each individual in the three samples drawn as described above. The creditor's loan files and payment records constitute the developer's main source of applicant information. The loan file may include the original application, verifying information, credit reports, the loan officer's personal evaluation, cash flow analysis, appraisals, approval sheets, the contract, disclosure statements, and collection materials. Payment records, however, frequently exist only on a computer. After determining the extent of available information, the developer designs a coding scheme for transcribing all information into

---


75. Cf. Blalock, supra note 46, at 510-12 (certain deficiencies also exist in noncurrent data). However, some developers attempt to improve the quality of the "aged" development sample by adjusting it to better represent the attributes of more recent applicants. This approach assumes that while only older accounts have sufficient credit experience to permit their classification as creditworthy or noncreditworthy, a sample composed of such accounts fails to allow for changes in the demographic, financial, and creditworthiness makeup of more recent applicants. Accordingly, this latter approach draws two sets of samples, one of aged accounts having credit experience and one of recent applicants. It then compares the attributes of each set of samples and uses quantitative techniques to reweight the "aged" sample to conform to the composition of the recent applicant samples. Thus, if the new sample shows more car owners than the "aged" sample, those few car owners in the "aged" sample will receive more weight in the final development sample.

computer-usable form. After conversion to machine-readable format, the data go into the computer and its electronic manipulation commences. These data typically consist of several hundred attributes for each of several thousand applicants.

Attribute classing

The computer’s first task consists of deciding how to divide each characteristic having an interval scale into attribute cells. Recall that attributes may form either nominal or interval scales. Nominal scales usually convert readily into attribute cells on a scoring sheet. For a characteristic such as checking account, the responses yes and no comprise all of the possible answers. An interval scale, however, imposes no special limits on the number or width of the attribute cells of the interval characteristic. Whether the first cell should span 0-6 months or 0-8 months depends instead on the particular data base.

At this point in the development process, the data for an interval characteristic exists in the computer’s memory in exactly the form in which the applicant submitted it to the creditor. No rounding, refining, or condensing of the numbers has occurred. Thus, income might appear in one dollar increments, or years on job by one month intervals.

In order to make the attributes more comprehensible and more susceptible to statistical manipulation, the ideal development process converts the repayment-rate curve into a series of fixed steps over the characteristic’s interval as illustrated in Figure 6. When the repayment rate remains stable over a range of attribute values, this span is collapsed into a single attribute cell, as in Figure 7.

---

77. See note 54 supra.
This conversion process largely defines the perimeters of the attribute cells that will comprise the scoring sheet's columns. Note that the final attributes selected by the computer's conversion of raw data can span irregular intervals. The characteristic "number of credit references" might have the following attributes: 0, 1-2, 3-5, 6-12, 13-17, 17-19, over 19, depending upon the data's actual plateaus. Nominal scale attributes also are assigned to common cells where they exhibit similar repayment rates.

Rejects

As described above, a scoring system models the behavior of the population consisting of all a creditor's applicants. It seeks to divide the applicant population into two, nonoverlapping subpopulations, creditworthy applicants and noncreditworthy applicants. By identifying the attributes which distinguish the subpopulations, the system can properly classify future applicants of unknown creditworthiness with a high degree of accuracy.

However, the creditor's records do not fully characterize all its applicants as either creditworthy or noncreditworthy on the basis of experience. Some applicants are rejected. Because the creditor grants no credit to rejected applicants, it does not generate any default or repayment experience for them. It cannot automatically classify them in either subpopulation.

The mathematics of discriminant analysis evaluates the separation of only two subpopulations. The next step in system development, therefore, consists of collapsing three sample groups into two develop-

---

78. See text accompanying notes 53-58 supra. However, certain special systems intentionally analyze the quality of loans only after the approval and disbursement. See Orgler, supra note 19.
ment subpopulations: satisfactory accounts (creditworthy applicants), unsatisfactory accounts (noncreditworthy applicants), and rejects (creditworthy and noncreditworthy applicants).

In the actual development process reclassification of rejects occurs prior to selection of predictive characteristics and calculation of scores. Various methods of accounting for rejected applicants exist.

Do nothing

The developer may ignore the presence of rejects. This method assumes that the attributes of the accepted population largely represent those of the applicant population.79

This approach suffers from the potentially faulty assumption that past credit analysis resulted in random acceptances, whereas rejected applicants usually have lower quality economic and social attributes. Otherwise, the judgmental creditor would not have rejected them in the first place. Accordingly, the accepted population usually has considerably different attributes from the applicant population. Failure to allow for the rejects results in a development sample which measures only the better quality applicants.

Assume rejects are noncreditworthy

The developer may assume that all rejects would have defaulted and are therefore noncreditworthy.80 This method justifies its derogatory treatment of rejects by focusing upon their relatively low economic attributes.

Other analyses of rejects do not, however, confirm this presumption.81 While the repayment rate may decline as attribute quality declines, the rate never reaches zero. Although the repayment rate among rejects may become too low for profitability, it still includes a high percentage of creditworthy applicants. The creditor rejects these applicants only because it lacks a practical means for distinguishing them from the balance of the noncreditworthy subpopulation.82

---

79. See text accompanying note 78 supra.
80. The author is not acquainted with any professional developers who use this approach.
81. This would mean that the existing mode of credit analysis, usually the human judgment of a credit officer, would have almost perfect accuracy. This system would reject only noncreditworthy applicants and accept all the creditworthy applicants along with a few noncreditworthy ones. Any system performing this well certainly does not need replacement.
82. The repayment probability table mentioned in Section III, part N, for example, illustrates that credit standards tend to be imperfect. Tightening the standards can change
The assignment of all rejects to the noncreditworthy category therefore misclassifies many rejected applicants. This error means that the two subpopulations used to develop the scoring system will not accurately reflect the attributes of the true underlying subpopulations of creditworthy and noncreditworthy applicants. The predictive power of the final system will suffer because of the numerous classification errors caused by this approach.

**Buy experience**

A third and perhaps best way of measuring the total applicant population involves an actual check of the creditworthiness of the "through the door" population. The creditor may approve all incoming applications, up to a certain number regardless of the information disclosed by the applicant. It then studies the borrowers' performance. This method makes the accepted population identical to the applicant population. With no rejects, the satisfactory and unsatisfactory accounts, respectively, automatically become the sole source of information about creditworthy and noncreditworthy applicants.

Obviously, economic factors constrain the use of this method of gleaning information about rejects. Presumably the default rate rises considerably in the complete absence of any credit analysis. The creditor purchases credit experience with higher loan losses. The creditor can to some extent control unnecessary losses by randomly granting credit to only a sample of current applicants or by strictly limiting the time period over which it grants credit to all applicants. However, the creditor should avoid scrutinizing the unanalyzed slow loans with special care or resorting to special collection efforts because this would produce abnormal data about the default rate.

**Use old credit reports**

A creditor's application files frequently include a credit report, the ratio of satisfactory to unsatisfactory accounts, but can result in the rejection of all unsatisfactory accounts only in the extreme case in which it declines virtually all applications. This observation equally applies to judgmental analysis. Raising the required income from $12,000 to $14,000 and increasing the quality of the required credit references will reject proportionately more noncreditworthy applicants, but will also reject large numbers of creditworthy applicants.

83. See text accompanying notes 5-15 supra.

obtained at the time of the credit request which summarizes the applicant's past credit experience.\textsuperscript{85} Using the general credit analysis assumption that future behavior will resemble past behavior, the developer assigns rejects to the creditworthy or noncreditworthy subpopulation on the basis of the credit report.\textsuperscript{86}

This procedure has the advantage of low cost, but it has the disadvantage of not reacting to subsequent changes in applicant creditworthiness. The scoring system seeks to separate applicants on the basis of their performance after the credit request, not before. The main advantage of using "aged" accounts lies in knowing their actual, rather than estimated, credit performance.\textsuperscript{87} Estimates based on outdated credit reports, however, increase the unreliability of the samples. In addition, an applicant's credit performance may vary from creditor to creditor, and a creditor very well may have reports on only a few of the total rejects if the creditor has relied on scoring.

\textit{Use new credit reports}

A similar method of determining the performance of rejects uses new credit reports.\textsuperscript{88} This approach assumes that the rejected applicant eventually obtained credit from some other source and that a current credit report will reflect the history of the substitute credit. The developer can thereby determine which rejected applicants warrant classification as creditworthy in the development sample.

The Fair Credit Reporting Act\textsuperscript{89} presents a potential obstacle to the use of this technique for assigning rejects. The Act provides that a consumer reporting agency (a credit bureau) may supply a consumer report only for a "permissible purpose" and enumerates the uses that

\textsuperscript{86} This process increases the opportunity for errors in classification. The consequences of such misclassification can include development of an inaccurate scoring formula. See generally R. Eisenbeis & R. Avery, Discriminant Analysis and Classification Procedures (1972); Piffalls, supra note 65, at 893.
\textsuperscript{87} Creditors frequently have considerable difficulty classifying accounts as satisfactory or unsatisfactory even when they have complete information about account performance. See note 65 & accompanying text supra. However, most credit reports are even less helpful in classifying applicants.
\textsuperscript{88} The creditor obtains an updated credit report on everyone in the development sample and classifies each person as satisfactory or unsatisfactory on the basis of this report.
constitute permissible purposes. These do not include construction of scoring systems.

**Augmentation**

Most professional developers of scoring systems use an augmentation technique to account for the attributes of rejects. This approach capitalizes on the ability of discriminant analysis to separate any two subpopulations by using a composite of predictive characteristics. In credit scoring the developer wants to distinguish between creditworthy and noncreditworthy applicants, and does so by selecting characteristics and weighting their attributes so as to maximize the divergence between these two subpopulations.

![Figure 8—Augmentation](image)

For augmentation, the developer uses different subpopulations. Accordingly, a scoring system which differentiates between accepted and rejected applicants instead of creditworthy and noncreditworthy applicants is developed. To construct this system, the developer uses the same selection and weighting techniques he will use for the final scoring system. A scoring sheet and odds table, similar to the one in Figure 8, results.

93. Discriminant analysis can process any data consisting of two discrete subpopulations that do not overlap and that have distinguishing characteristics. The seminal work on discriminant analysis involved classification of flowers into different species using measurements of their stems and leaves as indicators. Fisher, *The Use of Multiple Measurements in Taxonomic Problems*, 7 ANNALS OF EUGENICS 179 (1936). The same statistical methodology can develop formulas for good and bad commercial loans, recidivist and nonrecidivist parolees, evaders and nonevaders among taxpayers, or successful and unsuccessful students or employees. D. MORRISON, MULTIVARIATE STATISTICAL METHODS 230 (1967).
This augmentation system models the behavior of the creditor rather than that of its applicants. By comparing the heights of the two normal curves at any point along the accept/reject score line, the developer knows the probability of the creditor accepting an application attaining that score. The computer reweights the accepted applications at each score to account for the probability of their acceptance. Thus, for example, where the curves intersect, each accepted application counts double to allow for an equal number of rejected applications.

The developer has creditworthiness information about the reweighted accepted applications. By using the reweighted sample in system development calculations, the accepted applications indirectly represent the attributes of all applicants. After this threshold screening, the accept/reject scoring system has no further function in the development of the final scoring system.

First characteristic selection

The process next selects the single characteristic that gives the most information about creditworthiness. It does this by performing a statistical test that expresses a characteristic's predictive power as a single figure. The higher the figure, the better its predictive power (i.e. it correctly classifies more creditworthy and noncreditworthy accounts). The characteristic with the highest figure, out of the hundreds of available characteristics, becomes the scoring sheet's first characteristic.

First score calculation

Having mathematically selected the single most useful characteristic, the process next sets the scores for each attribute of that characteristic. Because a one characteristic system does not have to account for the correlation effect of any other characteristics, the developer simply

---

94. See authorities cited note 91 supra; Roy & Sanderson, supra note 17.
95. More sophisticated statistical techniques that can achieve an accepted population representative of the whole applicant population are not described here because of their advanced, and sometimes proprietary nature. See Pitfalls, supra note 65, at 894.
96. See Chandler & Coffman, supra note 7, at 16; Concepts of Scoring, supra note 17, at 44; MDS Systems Approach, Sec. III, pt. 4, supra note 17.
97. This is typically accomplished by the Chi-square, Kendall's Tau, or Phi tests. Each of these measures the association between nominal scale variables, but in slightly different ways. See BLALOCK, supra note 46, at 275-77, 295-98, 418-26.
calculates the ratio of creditworthy to noncreditworthy applicants for each attribute, while controlling for the relative sample sizes. The computation may also logarithmically rescale this ratio to convert a multiplication process into an addition process.

Second characteristic selection

The developer next selects the second characteristic by pairing the first characteristic selected as described above, with each of the remaining characteristics in the database. Each two-characteristic combination forms a primitive index of creditworthiness having an interval scale. The machine therefore calculates the overall predictive power of each two-characteristic combination. This calculation allows for correlation between the several variables. For example, age and income may each predict creditworthiness independently, but may function poorly as a two-characteristic system because older people have higher incomes. Each characteristic tends to measure the same subgroup of creditworthy applicants and therefore adds little useful information to the other’s predictions. The process selects as the second characteristic of the final scoring system, the characteristic that adds the most marginal improvement to the system’s predictive ability.

Second score calculation

Having selected the second most predictive characteristic, the developer next derives the scores to be assigned to each attribute of that characteristic. For example, suppose that the process has selected “time at address” as the first characteristic, calculated its attributes and trial scores, and selected “size of monthly car payment” as having the next most predictive information in relation to time at address. Suppose this characteristic has four attributes, no car, under $80, $80-100, over $100, derived as suggested. The process must still determine how many points each attribute should receive.

Developers use several methods to make this assignment. The most conventional approach calculates the points for the second and subsequent characteristics from the results of the multivariate, statistical routines. Another approach uses an iterative, trial and error proce-

---


99. Concepts of Scoring, supra note 17, at 50; Control Systems, supra note 92, at 57, 59.

100. See text accompanying notes 96-97 supra.
dure, sometimes called a "guided direct search," to derive the optimal assignment of points experimentally. 101

Subsequent characteristic selection and score calculation

The computer keeps repeating the process detailed in the preced-

101. Under the latter procedure, the computer sets each attribute's score at zero as illustrated in Figure A.

Figure A—Hypothetical initiation of score calculation

<table>
<thead>
<tr>
<th>characteristics</th>
<th>monthly car payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>attributes</td>
<td>none</td>
</tr>
<tr>
<td>scores</td>
<td>.00</td>
</tr>
</tbody>
</table>

The computer then increases and decreases the first attribute's score slightly, perhaps to .01 and -.01. It calculates the subpopulation separation for each combination of attribute scores. The combinations having poorer separatory power are thrown out.

Figure B—Hypothetical results of first pass

<table>
<thead>
<tr>
<th>none</th>
<th>1-79</th>
<th>80-100</th>
<th>100+</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Holding the results of the first pass constant, the machine repeats this process in turn for each succeeding attribute cell. It changes the zero score in the cell slightly, compares the separatory power of each combination, and retains the best predictor.

Figure C—Hypothetical results of first cycle

| .01 | -.01 | .00 | .01 |

After doing this once for each attribute cell, the machine has completed a cycle. The attribute scores span the range .01 to -.01 in the combination that maximizes the separation of the creditworthy and noncreditworthy subpopulations. The machine then performs a second cycle. It changes the first cycle scores slightly, testing successive combinations for predictive ability, and retaining the superior combination.

Figure D—Hypothetical results of second cycle

| .01 | -.02 | -.01 | .02 |

This incrementation process continues for cycle after cycle. During this process scores rise and fall. Eventually the process reaches a point at which marginal changes in the attribute scores do not change the overall separation of the two subpopulations of interest, thus completing the trial score calculation for the second characteristic.
ing subsections, retaining those characteristics already selected and sequentially assembling these characteristics into trial systems in combination with each of the remaining characteristics in the data base. Note that the trial scores assigned to the characteristics retained from previous steps may change under the influence of characteristics subsequently added to the system.

Eventually, this process reaches the data's reliability limit.

Figure E-hypothetical conclusion of score calculation

<table>
<thead>
<tr>
<th>monthly car payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
</tr>
<tr>
<td>$1-79</td>
</tr>
<tr>
<td>$80-100</td>
</tr>
<tr>
<td>$101+</td>
</tr>
<tr>
<td>.27</td>
</tr>
<tr>
<td>-.78</td>
</tr>
<tr>
<td>-.19</td>
</tr>
<tr>
<td>.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Separation</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of applicants</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>O</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>X-creditworthy</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>O</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>O- noncreditworthy</td>
</tr>
<tr>
<td>O</td>
</tr>
<tr>
<td>X</td>
</tr>
</tbody>
</table>

This process also uses statistical tests that express a two (or more) characteristic system's predictive power as a single number. This time, however, the process employs tests that allow for the intercorrelation between the predictive characteristics. Thus, income and years of education might independently predict creditworthiness. However, if most applicants with high income also have many years of education, and vice-versa, then using both characteristics will not improve on the predictions made by using only one of them. Multivariate statistics constitutes a complex and esoteric field of scholarship. See generally Blalock, supra note 66, at 429-70; W. Cooley & P. Lohnes, Multivariate Procedures for the Behavioral Sciences (1962); N. Draper & H. Smith, Applied Regression Analysis (1966); M. Ezekiel & K. Fox, Methods of Correlation and Regression Analysis (3d ed. 1959); J. Johnston, Econometric Methods (2d ed. 1972); F. Kerlinger & E. Pedhazer, Multiple Regression in Behavioral Research (1973); S. Roy, Some Aspects of Multivariate Analysis (1957); J. Van de Geer, Introduction to Multivariate Analysis for the Social Sciences (1971); Chandler & Coffman, supra note 7, at 17.

102. Concepts of Scoring, supra note 17, at 58. The developer sometimes artificially restricts certain perimeters of the guided direct search process. For example, the developer might specify that no single score may contribute more than 25% to the total score in order to prevent a single characteristic from dominating the final system, regardless of its power as a predictor of creditworthiness. Alternately, the development process might use the credit report only as a last resort because the creditor must pay for the credit report, whereas the information from the application costs nothing. Statisticians have not yet decided whether such constraints can affect the scientific validity of the final system.

103. This statement is true at least for a process using either a guided direct search or dynamic programming. Developers do not presently know whether reliability limit can
Neither adding more characteristics nor adjusting the trial scores will produce more subpopulation separation. At this point the system typically consists of some six to twelve characteristics. The degree of separation of the two subpopulations at this point corresponds to the system's final predictive ability.

Score transformation

In the preceding calculation using the iterative approach, scores arbitrarily spanned a limited range. Summing small numbers while keeping track of signs and decimal points presents considerable opportunity for arithmetic error. Therefore, the developer makes an algebraic transformation of the calculated scores. This process does not affect the relative relationship of the scores or the total separation conferred by their numerical values. It merely changes the scores in the cells to convenient whole numbers. Essentially, the final scoring sheet results from this transformation.

Validation

Both scientific principles and federal laws compel the developer's next step, which is system validation. All statistical analysis assumes a small but finite probability of error. Frequently, this error results from the ill fortune of drawing a skewed sample for use in development, or from sloppy data-transcription procedures which result in classification errors with respect to some attributes. As a result of

104. The range spanned depends upon the constants used by the score calculation algorithm. With slight changes in computation limits, the calculation could span -1 to +1, -100 to +100, or 0 to 1000.

105. E.g., Figure 1 supra; deKerchove, Reject Reasons and Scoring, Viewpoints, Winter 1977, at 3; Main, supra note 1; A Simplified Look at Credit Scoring, Consumer Month, May 1977, at 4.


107. See W. Cochran, Sampling Techniques 359 (3d ed. 1977); Blumenfeld, Godbey
such errors, the development sample would not fully represent the attributes of the underlying population of applicants when taken as whole. A scoring system developed by means of an erroneous sample would do a poorer than expected job of predicting creditworthiness. Therefore, the developer must validate the system to estimate its future error rate prior to its actual use for credit evaluation.

Numerous methods exist for assessing the performance of systems. The simplest approach involves taking the original development sample, running it through the proposed system, and observing whether it predicts as well as expected. This method has the drawback of producing unduly optimistic predictions of the system's future performance because the biased validation data still does not reflect the attributes of the true population.

Accordingly, most developers validate by means of a "holdout sample." After drawing the original development sample, the developer "holds out" about a third of the cases and does not use them in system development. The holdout sample therefore has an independent opportunity to represent the actual population of applicants. The system processes the new data from the holdout sample and the developer compares its creditworthiness prediction with the holdout applicant's actual performance in order to obtain an independent estimate of the system's validity.

Compilation of the repayment probability table

Prior to the initial use of the system, the developer will compile its other component, the repayment-probability table. The developer reweights the development sample to represent the applicant population's actual ratio of creditworthy to noncreditworthy persons. The sys-


108. See Frank, Massy & Morrison, Bias in Multiple Discriminant Analysis, 2 J. MARKETING RESEARCH 250 (1965).


110. Other error-estimation methods exist but are rarely used to validate scoring sys-

See generally Pitfalls, supra note 65, at 894.
tem processes the reweighted sample to calculate a score for each applicant. The computer counts the number of creditworthy and non-creditworthy applicants attaining each score and calculates the percentage of applicants who repaid at each score.111

By cumulating the number of creditworthy and noncreditworthy applicants at or above a certain score, the developer can calculate the overall repayment rate for that cutoff level. This percentage expresses the quality of the loan portfolio and estimates the number of applications accepted. Conversely, cumulation by ascending score describes the number of applications not accepted and the projected repayment rate among the rejects.

Theoretically, the creditor uses the repayment-probability table to set its cutoff score.112 By setting it high, the creditor obtains a small, but quality, loan portfolio. If it lowers the cutoff, the marginal default rate increases, but so does the portfolio’s size. The profitability tradeoff between volume and quality will vary from creditor to creditor. Ideally, the creditor balances the cost of a default against the profit on a repayment, and sets the cutoff score at the marginal profitability rate.113 In practice, creditors rarely know the breakdown of their costs and profits. Consequently when portfolio performance seems unsatisfactory, the creditor raises the cutoff. When greater volume is desired, it lowers its cutoff.

ECOA and Scoring

Until 1977, credit scoring comprised a technology largely unregulated by the government. Designers could construct systems using any criteria they wanted. Creditors could make any use of a system they saw fit. Use of scoring increased solely because of its business advantages.

In 1976 Congress passed certain amendments to the Equal Credit Opportunity Act (ECOA).114 Effective March 23, 1977, these amendments added race, color, religion, national origin, age, receipt of public assistance, or good faith exercise of any Consumer Credit Protection

111. Concepts of Scoring, supra note 17, at 68; Management Tool, supra note 1, at 22.
112. New Math, supra note 6, at 5; Management Tool, supra note 1, at 22.
Act right as prohibited bases, to the 1975 version of the Act which prohibited credit discrimination on the basis of sex or marital status.\textsuperscript{115}

Almost as an aside, however, the Act noted that a creditor does not discriminate if it uses “any empirically derived credit system which considers age if such system is demonstrably and statistically sound in accordance with regulations of the [Federal Reserve] Board, except that in the operation of such system the age of an elderly applicant may not be assigned a negative factor or value . . . .”\textsuperscript{116} This section investigates the effect of the federal legislation upon credit scoring.

Demonstrably and statistically sound, empirically derived credit system

With the advent of ECOA, the Federal Reserve Board\textsuperscript{117} assumed the duty of interpreting Congress’ prohibition against credit discrimination and its impact upon scoring technology. In writing ECOA regulations the Board faced several problems of system coverage and definition, such as what constitutes a demonstrably and statistically sound, empirically derived credit system, or when does an elderly applicant receive a negative factor or value for age.

The Board’s Regulation B responds to these quandries by dividing all credit analysis into two halves: (1) judgmental systems and (2) demonstrably and statistically sound, empirically derived credit systems. Only systems that conform to detailed regulatory standards fall into the latter category.\textsuperscript{118} All other approaches to credit analysis comprise judgmental systems.\textsuperscript{119} Judgmental systems therefore include the vast majority of credit analysis techniques, including procedures not actually involving a credit officer’s personal interview and evaluation of the applicant.

In compliance with its Congressional mandate to define what constitutes a demonstrable and statistically sound, empirically derived credit system, the Board adopted a two-stage approach. To be “empirically derived” a credit system must operate in certain general ways. To be “demonstrably and statistically sound,” as well as empirically derived, the scoring system must pass a series of additional tests. In particular, an empirically derived credit system must evaluate creditworthiness primarily by allocating points to applicant attributes.

\textsuperscript{117} Hereinafter called the “Board.”
\textsuperscript{118} Regulation B, 12 C.F.R. § 202.2(p) (1978).
\textsuperscript{119} Regulation B, 12 C.F.R. § 202.2(t) (1978).
The points must derive from empirical comparison of the creditworthy and noncreditworthy applicant population of the creditor. The creditor must use the overall score to determine applicant creditworthiness, but may also consider other information.

Once these threshold requirements are met, the empirically derived system is deemed demonstrably and statistically sound if it meets three statistical standards. These Regulation B tests impose only minimal statistical standards for predictive accuracy. Any professional developer of scoring systems would, for mathematical and business reasons, insist that a system pass a much more rigorous battery of tests before releasing it for actual use. Any system with a modicum of predictive power should easily pass government specifications.

The Board's first test of soundness seeks to assure accurate development data. In particular, the data used to develop the system must consist of either the whole applicant population or else a proper sample drawn from the whole applicant population. The applicant population includes, of course, both rejected and accepted applicants. Any development sample which properly integrates rejected applicants into the system will therefore accurately represent the attributes of the actual, underlying population. A statistically biased sample could result in inaccurate predictions of creditworthiness or unnecessary harm to a class of applicants protected by ECOA. The creditor's business interest in reliable credit analysis parallels any regulatory compulsion to use proper sampling.

As the second standard, Regulation B requires validation of a newly developed system prior to its regular use on incoming applicants. Upon validation the system must separate creditworthy and noncreditworthy applicants at a statistically significant rate. This requirement apparently means that the system must have some predictive accuracy. That is, it should make credit decisions at least slightly better than a completely random process. The regulation does not specify the level by which the system's performance must exceed zero. Courts in employment discrimination cases may well, however, accept as statistically significant correlations that could happen randomly only five times in 100 trials (a ninety-five percent significance level).

For business reasons, most professional developers employ more rigorous tests of accuracy than is required by the regulations. A

---

holdout sample, for example, compares the system’s predictions with the known performance of a different group of applicants. A comparison of the system’s results to a hypothetical system having perfect predictive ability might constitute an even more rigorous performance requirement.

The Board’s third test for system soundness applies only after the system becomes operational. As time passes, the nature of a creditor’s applicants will change and the predictive power of its scoring system will inevitably decay. The creditor must therefore “revalidate” its system at appropriate intervals.

A revalidation would employ the same statistical tests as the initial validation. However, the revalidation sample consists of recent credit applicants who have since developed representative credit experience. This recent sample replaces the applicant sample used in the initial validation, whose credit performance has since become outdated. Comparison of the system’s predictions to the revalidation sample’s actual credit performance yields a statistic expressing the system’s continuing predictive power.

Again, Regulation B imposes a less stringent standard for revalidation than a professional developer would employ. The economics of credit granting usually require a much higher degree of accuracy than merely “nonrandom.” A creditor makes only a moderate profit from a satisfactory account, while it can lose a large amount of the credit’s principal balance and incur substantial collection costs on an unsatisfactory account. Accordingly, there exist strong business reasons for abandoning a decaying system long before it fails the minimum regulatory test for revalidation.

Regulation B leaves certain details of revalidation within the creditor’s discretion. For example, it does not specify the tests that the developer must use or the significance level that the revalidation of the system’s accuracy must attain. As noted above, a Phi coefficient and a ninety-five percent significance level, a medium standard for error, are probably acceptable.

Regulation B also omits explicit guidance as to how frequently revalidation must occur. Since the national population’s financial

---

123. See text accompanying note 111-15 supra.
125. See authorities cited note 31 supra.
characteristics change with the major business cycle, this presumably would be the maximum period between revalidations. Economists, however, do not wholly agree as to the length of the American business cycle. Accordingly, the creditor may find that its scoring system loses the minimum accuracy necessary to do business profitably long before Regulation B compels revalidation.

Finally, the third Regulation B test neglects to articulate the consequences of a failure to revalidate. By negative implication, the system would lose its designation as a “demonstrably and statistically sound” system. It might, however, retain its appellation as an “empirically derived” system because its method of operation and its origin continue to conform to that portion of the Regulation B definition. Curiously, an empirically derived credit system that no longer meets the tests for demonstrable and statistical soundness, would apparently constitute a judgmental system. The creditor could, therefore, still legally use the failed system, so long as it does not score the age of applicants. Nevertheless, the creditor presumably would replace such a system because of the need for accurate creditworthiness predictions.

Two competing considerations may have affected the Board’s decision as to how rigorous to make the standards for verifying the statistical validity of scoring systems. First, many creditors use poorly developed, inaccurate scoring systems. If a group of credit officers meets to select characteristics and scores on the basis of their anecdotal experience, the resultant “scoring” system will predict creditworthiness less reliably than a mathematically rigorous system. The nonempirically derived system may reflect the prejudices of its creators and thereby unnecessarily disadvantage some protected classes of applicants. Imposition of strict statistical standards assures that all scoring systems in use measure actual creditworthiness, rather than perceived creditworthiness.

The second competing consideration concerns the high potential for discrimination in judgmental systems of credit analysis. A properly

128. Regulation B defines a judgmental system as “any system for evaluating the creditworthiness of an applicant other than a demonstrably and statistically sound, empirically derived credit system.” Regulation B, 12 C.F.R. § 202.2(t) (1978).
developed scoring system should measure only creditworthiness. It may still confer less credit on some protected classes of applicants, but it does so because of their past credit performance and socioeconomic characteristics, rather than because of intentional or accidental discrimination. In contrast, credit analysis based on human evaluation of non-quantifiable factors such as the applicant's "character" offers considerable opportunity for the introduction of human prejudice.

Too rigorous a validation standard for scoring systems would disqualify many poorly developed systems from inclusion within the definition of demonstrably and statistically sound, empirically derived credit systems. They would comprise judgmental systems, however, which cannot consider age as a characteristic. This would encourage creditors to revert to purely judgmental systems with their high potential for discrimination.

The Board, therefore, had to balance the discriminatory potential of inaccurate scoring systems against that of judgmental systems, when deciding how rigorous to make validation standards. The Board may have concluded that encouraging empirical techniques would result in the least discrimination. Regulation B therefore has relatively low statistical standards. The desire for mathematical purity defers to the desire for the development of scoring techniques. In addition, mathematicians do not fully agree about what "validation" means in a purely theoretical context.

Negative factor or value

Having defined what makes a scoring system demonstrably and statistically sound, as well as empirically derived, the Board faced the further task of defining what actions constitute assigning a "negative factor of value" to the age of an elderly applicant. Regulation B defines "negative factor or value" as meaning "in relation to the age of elderly applicants, utilizing a factor, value, or weight that is less favorable regarding elderly applicants than the creditor's experience warrants or is less favorable than the factor, value, or weight assigned to the class of applicants that are not classified as elderly applicants and are most favored by a creditor on the basis of age." A creditor can score the applicant's age directly. Normally an elderly applicant

---

gets the number of points for age that the development process assigns on the basis of age-related default rates. If, however, one of the nonelderly attribute cells earns a higher age score because of better credit performance, then the elderly applicants receive at least that many points for age, regardless of their actual credit performance.

Several reasonable interpretations of "negative factor or value" have been suggested. This provision might have prohibited the direct consideration of age by a scoring system, even if demonstrably and statistically sound. Under this interpretation, however, a system could use characteristics such as time on the job or time of local residence, that have a high correlation with age. Some commentators believe that a complete prohibition of direct consideration of age best protects the elderly from unwarranted discrimination, and at least one state has embraced this approach.132

This interpretation ignores the language of ECOA, that provides in relevant part: "It shall not constitute discrimination . . . for a creditor . . . to use any empirically derived credit system which considers age . . . ."133 Congress would not have inserted this language expressly allowing age consideration in scoring systems if it intended an absolute prohibition. In the first place, the Act's general prohibition on age discrimination alone should have curtailed all consideration of age. In addition, Congress knew that the elderly comprise the age category presenting the best credit risks.134 Permitting consideration of age would increase the amount of credit the elderly will receive.

Concluding that Congress intended to permit credit scoring systems to consider an applicant's age directly, Regulation B then confronts the question of how a system can score age. At least four alternatives present themselves. It could permit scoring systems to consider age in any manner consistent with statistical principles. Alternatively, it could prohibit a scoring system from specially disadvantaging elderly applicants by requiring that it allocate to the age-attribute cell for elderly applicants at least as many points as to the nonelderly age-attribute cell scoring the least number of points. It could also require that elderly applicants receive at least as many points for their age as the class of nonelderly applicants most favored on the basis of age. Finally, it could mandate that elderly applicants receive some median number of points for their age, relative to the points allocated to other

Again, two competing considerations affect the choice among these possible interpretations of "negative factor or value." First, because the elderly have the best credit performance of any age group, minimum government interference with scoring would generally result in their receiving the most points of any age class. Second, because of the impact on large numbers of elderly applicants that would result from improper scoring, extensive government intervention, and a broad interpretation of "negative factor or value," may be justified.

The interpretation finally adopted in Regulation B succeeds in simultaneously satisfying both competing considerations. Elderly applicants must receive the higher of (1) the score warranted on the basis of that creditor's experience with elderly applicants or (2) the score assigned to the class of nonelderly applicants most favored on the basis of age. Given that elderly applicants usually demonstrate the highest creditworthiness, for most creditors this means that the elderly receive the same number of points for age that they would have if age scores were assigned on purely statistical bases. For those very few creditors whose population of elderly applicants does not demonstrate the maximum creditworthiness, Regulation B requires that they receive the advantage of the same score as the most creditworthy age group. This bifurcated treatment of age confers the maximum possible amount of credit on the elderly while doing minimal damage to the predictive accuracy of the scoring system and the quality of the creditor's portfolio. It also minimizes the need for continuing government intervention.

A related controversy concerns the definition of the age at which an applicant becomes "elderly." Congress wanted to protect the elderly, but never articulated a precise definition of the term. The Board initially elected to emulate this legislative omission, and therefore every applicant became elderly with respect to all younger applicants. If the Board defined "negative factor or value" as meaning that elderly applicants must receive no fewer points than the lowest nonelderly age score, then every age category would have to receive at least as many points as the next youngest age category. The scores for age would

137. See text accompanying note 34 supra.
then form a monotonically increasing function.\textsuperscript{140} In certain cases, this would result in excessively favorable treatment for middle-aged applicants, a group whose creditworthiness typically drops, while not particularly aiding older applicants.\textsuperscript{141}

To avoid such unrealistic outcomes, Regulation B does define elderly. Obviously, such a decision involves certain arbitrary elements. As a policy matter, by choosing a low age such as fifty or fifty-five, more applicants would receive the benefit of coverage by the "negative factor or value" prohibition. By selecting a very high figure such as seventy or seventy-five, fewer applicants would fall into the elderly age category, but the category would receive a much larger number of points because creditworthiness usually continues to rise sharply with advancing age. Regulation B's definition of elderly as ages sixty-two and up constitutes something of a compromise, and is consistent with the Social Security laws.\textsuperscript{142}

**Reasons for Adverse Action**

The Equal Credit Opportunity Act also requires that when a creditor declines to grant a request for credit, it must give the applicant a statement of the reasons for the adverse action.\textsuperscript{143} Congress stated that the reasons given must be specific, but did not illustrate what constitute specific reasons for credit refusal by a scoring system. A scoring sheet presents certain obstacles to arriving at any meaningful selection of reasons for an application's refusal. Each applicant's final score derives from the cumulative effect of his attributes. The creditor rejects the applicant because the total score falls below the preselected cutoff. All of the submaximum characteristics therefore contribute jointly to the applicant's failure to get the credit. In addition, creditors adjust their cutoff score from time to time in response to changing business conditions. Had the applicant been submitted at a different time,

\begin{itemize}
\item \textsuperscript{140} A monotonically increasing function is a mathematical sequence whose successive members either increase or stay the same, but do not oscillate in relative value. For the ECOA this would mean that as applicant age increases, the points assigned for that age must either become larger or remain constant, but never decline. Figure 6 depicts a monotonically increasing function, although that graph does not relate to age and does not increase because of a statutory requirement.
\item \textsuperscript{141} Letter from Richard F. Kerr, Federated Department Stores, to Neil Butler, Federal Reserve Board staff (Dec. 9, 1976) (on file with the Federal Reserve Board under Docket No. R-0031).
\item \textsuperscript{142} 42 U.S.C. § 402(a)(2), (b)(1)(B), (c)(1)(B) (1970).
\end{itemize}
when the creditor had a lower cutoff, the applicant might have obtained the credit.

Regulation B therefore confronts the problem of determining what constitutes a set of specific reasons for adverse action by a credit scoring system. Users of scoring systems argue that the specificity requirement is satisfied simply by informing the rejected applicant that he failed to score enough points to pass the cutoff score.144 This standard would also substantially simplify the logistics of generating the adverse-action notices required by ECOA. A single preprinted notice giving the same reason, "insufficient score," would go to all applicants rejected by scoring.

Regulation B does not embrace this approach. Users of scoring systems must give "specific" reasons.145 "Insufficient points" states a conclusion, not the underlying reasons that caused the failure to attain the needed score.

When enacting ECOA, Congress wanted applicants to receive specific reasons for adverse action so that they could use this information to take remedial action.146 If the creditor bases its decision on erroneous assumptions then the applicant can take steps to rectify this misapprehension. For example, when the statement of reasons discloses "no credit file found" as the reason for denial, the applicant can submit additional information concerning prior use of credit and possibly obtain favorable reconsideration of the application. Similarly, when the denial results from a mutable characteristic, the applicant will know to reapply if conditions change. Thus, an applicant denied credit because of "too short a period of local residence" would know that an application submitted at a later date might have a better chance of approval. The applicant needs to know what characteristics will produce additional points and how much improvement is necessary to obtain the desired credit.

Regulation B therefore leaves the scoring-system creditor with the problem of how to select specific reasons for rejection. Creditors have considerable regulatory discretion in dealing with this selection question. To date no clearly superior method, on either legal or scientific

grounds, has emerged from among the several available alternatives. Perhaps the most widely used approach to selection of specific reasons for denial under a credit scoring system involves judgmental selection by a credit officer. In the judgmental system, the credit officer balances all of the relevant information about the applicant mentally to arrive at an adverse conclusion. It therefore seems entirely reasonable that the same individual who made the adverse decision can arrive at a meaningful selection of specific reasons for that conclusion. Where credit was declined because of scoring, the credit officer performs only the second function, perusing the application and determining the reasons for the scoring system's adverse action.\textsuperscript{147}

No clear consensus presently exists on the related question of whether judgmental selection of reasons may include characteristics not among those on the scoring sheet. Proponents argue that ECOA limitations on scoring apply only to the credit decision. The law imposes no limits on the process by which the creditor then selects reasons for the decision. Moreover, certain events, such as bankruptcy, occur too infrequently to warrant inclusion in the scoring sheet. However, human scrutiny of the whole application would certainly find bankruptcy to be the most meaningful reason for the applicant's lack of creditworthiness. It merely confirms the scoring system's conclusion.

Opponents of external reason selection argue that only the characteristics that contribute to the total score constitute meaningful reasons for the applicant's failure to receive the credit. Disclosure of a reason not scored misleads the applicant. Subsequent changes in the applicant's status with respect to the disclosed reason for rejection will not result in credit, because the overall score remains the same. Conversely, the scoring system might approve a subsequent application disclosing superior attributes, notwithstanding the stated reason for the prior adverse action.

The other approach to disclosing reasons for adverse action involves mechanical selection of the "specific reasons" to be disclosed. When a computer does the credit analysis, it can also generate the statement of reasons for adverse action. The reasons the machine gives the applicant resemble those produced by a judgmental process, but

\textsuperscript{147} In the matter of Alden's, Inc., 43 Fed. Reg. 6,622, 6,624 (1978) (FTC consent agreement); Letter from Anne Geary, Federal Reserve Board staff, to Lloyd Hackler, President, American Retail Federation (Oct. 27, 1977) (on file under Docket No. EC-497); Letter from Neil Butler, Federal Reserve Board staff, to Glenn W. Hampton (May 3, 1977) (on file under Docket No. EC-178).
their actual selection involves an entirely automated procedure. The machine compares the scored characteristics to see where the applicant lost the most points and prints out a standard form or letter. The computer necessarily limits its selection to scored reasons.148

Any single applicant receiving a mechanically generated statement of reasons gets about as much meaningful information from this process as is received from a judgmental process. However, examination of the mix of reasons given to the population of all rejected applicants may reveal a deficiency in the aggregate information disclosed. If the scoring sheet gives an extra-large number of points to one particular attribute cell, then a majority of rejected applicants may be informed that deficiencies with respect to this characteristic caused their rejection. As the disclosed reasons become increasingly uniform, they also lose specificity. If the machine selects “lack of home ownership” as the reason for virtually all adverse action, this disclosure may then have no more meaning than “insufficient points.”

Some developers have considered remedying this potential loss of specificity by substituting other algorithms for simple subtraction when calculating the characteristics that result in rejection. Possible alternative methods include selection based on the greatest percentage difference between points scored and points possible, the greatest difference between points scored and average or median points scored by the whole applicant population, or disclosure of the combination of characteristics that would, with the change of one cell, give the applicant enough points to pass the cutoff. To date, no clearly superior method of achieving the greatest specificity has emerged.149

A related problem in selection of reasons for adverse action involves disclosure of reasons that appear contrary to common sense. Suppose that the system development process selects time of local residence as a characteristic, sets the attribute intervals at 0-1 year, 1-4 years, 4-19 years, and over 19+ years; and assigns scores of 12, 3, 7, and 0 to the respective attributes. Further suppose that the creditor’s method of choosing reasons for adverse action results in the selection of this characteristic for a certain applicant with twenty-one years of local residency. The creditor therefore informs the applicant that it denied the credit because of “excessive local residency” or “you have lived in your present house too long.” One may question whether this disclo-

149. ECO, supra note 14, at 103.
sure furnishes the rejected applicant with a meaningful reason for the adverse action.

Most judgmental systems associate increasing creditworthiness with longer local residence. The credit officer reasons that prolonged local residence means better applicant stability. Persons who have lived in the neighborhood for many years seem unlikely to abandon their established lifestyles by immediately absconding with the creditor's money. In addition, they tend to enjoy reliable sources of income with which to repay the credit.

The hypothetical scoring system appears to contradict this traditional view. This discrepancy could occur for several reasons. First, credit applicants do not necessarily comprise a representative cross-section of local residents of varying seniority. People may seek credit most frequently when they have just moved into the area. If many of them moved because they received promotions or took new jobs at higher salaries, then people with local residence of short duration would have unusually high creditworthiness. Second, many of the most predictive characteristics may correlate positively with each other, as well as with creditworthiness. Use of more than one of them would not add to the system's basic predictive power, so the system-development process selects only the one characteristic that predicts best for this purpose. The system may utilize the less predictive, correlated characteristics to refine the basic prediction to conform to the particular idiosyncrasies of the applicant population. Accordingly, while the credit officer may have the correct view of the average effect of a characteristic such as time of local residence, the multivariate analysis used in system development considers its marginal, rather than average, contribution to the creditworthiness prediction.

Nevertheless, given the way the scoring sheet processes applications, the credit system rejects the applicant because of "too long a period of local residence." The peculiar makeup of the system user's creditworthy applicants or the system's particular use of the characteristic does not make the attribute any less the reason for the adverse decision. With proper explanation, giving this attribute as the reason for rejection can convey just as much meaningful information as any other properly selected reason. The ECOA mandates selection of the "real" reason for adverse action, regardless of whether it confirms or disabuses the applicant's notions of how credit analysis should work.¹⁵⁰

Effects test

When Congress amended the Equal Credit Opportunity Act in 1976, it expressed its intent that the effects test, first developed in relation to employment discrimination, apply to credit discrimination.\textsuperscript{151} Regulation B incorporates this legislative intent both by reference and in its limitation of certain credit practices.\textsuperscript{152} To date, little judicial interpretation of the credit effects test has occurred.\textsuperscript{153} Accordingly, the credit effects test may or may not resemble its employment predecessor closely.

The effects test originated through judicial interpretation of the Equal Employment Opportunity Act which prohibits certain discriminatory employment practices.\textsuperscript{154} The Supreme Court in \textit{Griggs v. Duke Power Company}\textsuperscript{155} held that unintentional, as well as intentional, discrimination may violate this prohibition.\textsuperscript{156} Employment practices that have unequal impact on a protected class may constitute employment discrimination even though the employer did not intend to discriminate.

The Supreme Court subsequently expanded on this theme in \textit{McDonnell Douglas Corp. v. Green}\textsuperscript{157} and \textit{Albemarle Paper Co. v. Moody}.\textsuperscript{158} These cases establish a three part standard of proof for employment cases using the "effects test," that shifts the burden of proof back and forth between complainant-employee and defendant-employer:\textsuperscript{159}

1. The employee has the initial burden of establishing a prima facie case. To do so, the employee must show that an employment practice has a discriminatory effect on a protected class.

2. The employer may rebut the prima facie case by showing that

\begin{itemize}
  \item \textsuperscript{151} S. REP. No. 94-589, \textit{supra} note 2, at 4; H.R. REP. No. 210, 94th Cong., 1st Sess. 5 (1975).
  \item \textsuperscript{152} Proposed Regulation B, § 202.6(a), 42 Fed. Reg. 1,242, 1,255 n.7 (1977).
  \item \textsuperscript{153} \textit{Contra}, Carroll v. Exxon, 434 F. Supp. 557 (E.D. La. 1977).
  \item \textsuperscript{155} 401 U.S. 424 (1971).
  \item \textsuperscript{156} \textit{Id}. at 430.
  \item \textsuperscript{157} 411 U.S. 792 (1973).
  \item \textsuperscript{158} 422 U.S. 405, 425 (1975).
\end{itemize}
the employment practice has a manifest relationship to the job in question.

3. It then remains open to the employee to show the availability of an alternative employment practice that would have a less discriminatory effect on the protected class and that serves the employer's legitimate business needs at least as well as the disputed practice.

An employment practice that violates this "effects test" constitutes illegal discrimination even though the employer may not have intended such a result. Title VII cases measure discrimination by the effect of an employer's policies, as well as by its motives when establishing them. Courts use the effects test to measure the discriminatory impact of intelligence tests, seniority systems, and transfer and assignment policies.160

The effects test presents numerous interesting implications for credit scoring systems. A scoring system may violate the effects test by having disproportionate impact on a protected class of applicants. A restatement of the employment effects test, as adapted to credit scoring, might read as follows:

1. A credit applicant establishes a prima facie effects test case by showing that the scoring system has a disproportionate impact on a protected class of applicants.

2. The system user can rebut this prima facie case by showing that the system predicts creditworthiness effectively.

3. The applicant can still prevail by showing the existence of another scoring system, method of credit analysis, or adjustment to the existing scoring system that has a less discriminatory impact on the protected class and predicts creditworthiness at least as well as the present system.

Because of the paucity of ECOA case law, the veracity, meaning, and application of this restatement are yet unclear.161 The credit effects test could conform closely to the precepts of its employment progenitor or it could formulate its own standards for measuring unintentional discrimination. When the courts apply the ECOA effects test to credit scoring, they will presumably use the Albemarle formulation of a three-step shifting burden of persuasion as at least a starting point.162 However, a number of considerations distinguish measurement of credit discrimination from measurement of employment discrimination.


161. Hsia, supra note 159, at 790-95.

First, there remain the evidentiary difficulties of determining the demographic composition of the applicant population.\textsuperscript{163} Regulation B prohibits creditors from asking about an applicant's sex or race in connection with a request for an extension of credit unless the transaction involves the purchase of residential real property.\textsuperscript{164} Accordingly, for most applications the system user does not have reliable information about the demographic makeup of its applicant population. Without statistical data, proof of disparate effect, manifest relationship, or alternative practice becomes most difficult. At most, some system users will have data on the marital status and age of their applicant populations. Reconstruction of prohibited bases such as sex and race would require indirect inferences from available information. For example, the system user can guess at an applicant's sex with a fair degree of reliability by using the first name disclosed on application forms,\textsuperscript{165} and an applicant's home address may be used to estimate racial demographics by comparing them to census tract data.\textsuperscript{166}

Second, users of scoring systems can measure both the credit performance of individual borrowers and the predictive performance of the scoring system with much greater precision than an employer can measure either job performance or the ability of employment practices to predict such performance. Most of the leading employment effects-test cases involved unskilled work. In \textit{Griggs}, the job in question involved shoveling coal from one end of a courtyard and feeding it into a slot at the other end of the yard.\textsuperscript{167} Determining how well a person shovels coal necessarily presents certain quantification difficulties. Basically, most people of a given strength can shovel coal adequately. Choosing among available candidates, therefore, entails employer discretion, rather than screening based on personal qualifications. This element of uncertainty may cause courts to examine claims of "manifest relationship" closely. Scoring suffers from no such handicap of ad-

\textsuperscript{163} Employment cases rely heavily upon statistical comparisons of the percentages of members of protected classes hired by the employer and available for employment in the surrounding area. Robinson v. Union Carbide Corp., 538 F.2d 652, 660 (5th Cir. 1976); Green v. Missouri Pacific R.R., 523 F.2d 1290, 1293 (8th Cir. 1975). Indeed, some opinions view statistical evidence as essential to establishing the prima facie case. Senter v. General Motors Corp., 532 F.2d 511, 527 (6th Cir. 1976); Sager v. Yellow Freight System, Inc., 529 F.2d 721, 729 (5th Cir. 1976).


\textsuperscript{165} This may no longer be true, given the increasing use of androgynous and hyphenated names.


\textsuperscript{167} 401 U.S. 424, 426 (1971).
Past customers fall readily into satisfactory and unsatisfactory groups. The predictions of the scoring system are confirmed or belied by the subsequent credit history of accepted applicants, and the scoring system’s rate of error is precisely quantified. This increased ability to measure individual credit performance and overall credit predictions should make determinations of manifest relationship and the acceptability of alternative scoring systems obvious.

Third, an employer has a limited number of job openings. It will hire only enough applicants to fill its openings. In contrast, the quantity of credit available depends mainly upon its cost. At the proper price, unlimited supplies of credit probably exist. At another price, unlimited demand for credit occurs. Instead of seeking the one applicant best qualified to receive the credit, the credit analyst can accept any number of creditworthy applicants. If the creditor disburses all of its funds and it still has qualified applicants willing to pay for credit, it can simply buy more money.

Fourth, the creditor lacks the employer’s continuing control over its accepted applicants. If an employee proves unsatisfactory, the employer can always dismiss the worker. If done rapidly, the employer can minimize the unsatisfactory employee’s damage to the organization. Credit usually lacks such safeguards. Once the creditor disburse the principal, it largely loses control over the borrower. It has few decisive steps, analogous to dismissal, that it can take to minimize the damage done by an unsatisfactory borrower. Acceleration of the obligation, repossession of any collateral, and aggressive collection efforts may recover some of the credit’s principal, but these measures do not always produce complete repayment. This inability to effectively limit losses after disbursement greatly increases the importance of accurate credit analysis.

Fifth, credit transactions typically occur with greater frequency than the process of hiring. Presumably, most Americans only hold one or two jobs. However, the same worker may simultaneously engage in multiple credit transactions. The large number of “events” processed by a scoring system permits far more reliable statistical con-
clusions than do the smaller number of "events" processed by a given hiring procedure.

Sixth, employers draw their job applicants from a nonunique local market. The same people who carry coal from one end of the yard to the other can also load sawdust into a papermaking machine. Except for positions requiring special qualifications, essentially all the inhabitants around the employer’s place of business comprise potential job applicants. Courts measure disproportionate effect by comparing the demographic composition of defendant’s employees to that of the surrounding community. In contrast, the applicants to a particular creditor do constitute a unique subpopulation.\textsuperscript{173} Adjacent creditors may appeal to entirely different segments of the local populace. This specialized identification may derive from the community’s perception of the creditor’s policies, the nature of its business, the direction of its advertising, or other factors. Because of this segmentation in credit markets, comparison of community demographics with the makeup of the creditor’s borrowers does not necessarily show disparate impact upon its applicant population. Proper methodology would contrast only that creditor’s applicants with its borrowers.

\textit{Allocation of proofs}

Given the preceding considerations, it is necessary to consider how the courts might apply the foregoing statement of the effects test to credit scoring. To begin with, the middle step should largely disappear. Any moderately accurate scoring system automatically has a manifest relationship with creditworthiness.\textsuperscript{174} The user, for purely business reasons, rapidly discards any system which fails to perform effectively.\textsuperscript{175} A cursory showing that default rates remain at acceptable levels should suffice to prove the system’s continuing predictive power. As with system validation, probably anything which classifies applicants better than a completely random process would qualify as a business necessity.

The burden of proof in a scoring-effects-test case therefore falls

\footnotesize{search Program of the National Bureau of Economic Research, Studies in Consumer Instalment Financing No. 6, 1940).}

\textsuperscript{172} See note 163 \textit{supra}; see text accompanying note 206 \textit{infra}.

\textsuperscript{173} Hsia, \textit{supra} note 159, at 794 n.87; Main, \textit{supra} note 1; see authorities cited note 206 \textit{infra}.

\textsuperscript{174} \textit{ECO, supra} note 14, at 107.

\textsuperscript{175} Scoring’s primary economic justification is the profitability it confers through more accurate credit analysis and reduced operating costs. See note 30 & accompanying text \textit{supra}.}
almost solely upon the plaintiff, who must show both disparate impact and the availability of a less discriminatory alternative. The incompleteness of demographic information about the applicant population increases the plaintiff's evidentiary burden in connection with the first step of the effects test. In addition, presentation of a less discriminatory credit system and testing of its effects requires a full scale research and development effort entailing all of the steps enumerated above. The creditor can assume the five and six figure cost of system development because the business advantage of highly accurate credit analysis justifies the expense. The credit applicant turned disgruntled litigant, however, lacks this economic justification for system development efforts. Finally, because the rejected applicant lacks quantifiable damages, it seems unlikely that even a successful class action would yield sufficient proceeds to warrant creating a substitute method of credit analysis.

The would-be plaintiff faces one further obstacle to proffering an alternative scoring system. Its development requires a representative sample of the defendant-creditor's applicants. The system user controls access to this data and might justifiably resist surrendering it. First, disclosure of information about credit applicants to a third party might contravene the protections of the Fair Credit Reporting Act. Second, it might intrude upon the applicants' personal privacy. Third, since the composition of applicant populations varies considerably from creditor to creditor, it might well constitute proprietary material not subject to unlimited discovery.

176. See text accompanying note 163 supra.
177. See text accompanying note 30 supra. See generally sources cited by S. MILLER, A BIBLIOGRAPHY OF CONSUMER FINANCIAL SERVICES AND REGULATIONS 29 (Purdue Credit Research Center Monograph No. 7, 1977).
179. See text accompanying note 67 supra.


Tradeoffs between protected classes

Many important and more difficult effects-test issues remain unresolved under Title VII despite the availability of a large volume of employment case law. However, application of this analogous body of law to credit scoring often obfuscates rather than illuminates these issues.

Foremost among the unresolved issues pertaining to the credit effects test is the question of whether an alternative system proposed by the plaintiff under step three of the effects test can reduce discrimination against one protected class while increasing its disparate effect on another protected class. Varying or replacing the existing system could give substantially more credit to one protected class while maintaining the same overall default rate and volume of accepted applicants. The alternative therefore has less discriminatory impact on that class while still predicting creditworthiness equally well. It may, however, simultaneously have an adverse effect on another protected class.

Suppose, for example, that the plaintiff proposes eliminating homeownership as a characteristic. Because of housing discrimination, relatively few minority applicants own homes. Therefore, eliminating homeownership might substantially increase the volume of creditworthy minority applicants accepted. Assume that the addition of more creditworthy minority borrowers to the creditor’s portfolio would not affect its size or profitability because of their small numbers relative to the balance of the portfolio. Further suppose, however, that no longer considering homeownership would also have the effect of rejecting increased numbers of creditworthy female applicants. Moderate numbers of females own homes and this attribute predicts their creditworthiness very accurately. If women homeowners make excellent credit risks, then eliminating consideration of this “pertinent element of creditworthiness” reduces the creditor’s predictive accuracy with respect to female applicants and, thereby, the quantity of credit they receive.183

Recent reverse discrimination cases may help resolve this dilemma. The Supreme Court has held that Title VII prohibits “[d]iscriminatory preference for any [racial] group, minority or majority.”184 Elsewhere, the same opinion refers to special disadvan-

---

183. G. CHANDLER & D. EWERT, DISCRIMINATION ON THE BASIS OF SEX UNDER THE EQUAL CREDIT OPPORTUNITY ACT (Purdue Credit Research Center Working Paper No. 8, 1976) [hereinafter cited as CHANDLER & EWERT].
tages imposed upon any group. Altering an otherwise demonstrably and statistically sound, empirically derived credit system solely to favor one protected class at the expense of another protected class would therefore appear to contravene this judicial principle.

**Degree of discrimination reduction**

Another unresolved ECOA effects-test issue concerns the degree by which the plaintiff's proposed alternative must reduce discriminatory effect. Must it result in substantially less adverse impact on the protected class in order to establish the effects-test case or will very slight improvement in the status of the class suffice?

Employment-discrimination decisions have not yet resolved this question under Title VII. Although employment effects-test litigation has generated a large volume of case law incorporating much sweeping language, no opinions have found it necessary to reach the third step and apply the pertinent facts to proposed alternative-employment practices. Future litigants may draw upon at least two different lines of analysis, deriving from the earlier steps of disparate-impact analysis.¹⁸⁵

One approach comes from decisions analyzing the prima facie case showing disparate impact.¹⁸⁶ These decisions rely heavily on statistical tests to establish discrimination. The percentage of job applicants hired from the protected class is compared to the defendant's overall hiring rate. Mathematically derived tests calculate whether differences in the hiring rates occur randomly or as a result of systematic discrimination.¹⁸⁷ The Supreme Court has indicated that if the tests were otherwise valid that statistically significant differences at the ninety-five percent level of reliability would constitute sufficient evidence for a prima facie, effects-test discrimination case.¹⁸⁸

Under this approach to the third step of the ECOA effects test, the plaintiff would have to establish that the alternative proposal would result in statistically significant reductions in disparate impact. Establishing such reductions would involve a comparison of the existing

---


scoring system's approval rate for creditworthy applicants from the protected class and for the complete applicant population, with the projected approval rate for each population by the proposed alternative system. Properly selected contingency tests would determine whether differences are statistically significant at the required level.\textsuperscript{189} Under this standard, the plaintiff's alternative would have to produce a relatively large reduction in discrimination in order to establish the effects test's final step.

A second possible approach to divining the degree of improvement in nondiscrimination required of the alternative proposal derives from dicta with respect to the second step of the effects test. When commenting on manifest relationship and business necessity, the courts have repeatedly said that an employer must go "as far as possible" to eliminate discrimination.\textsuperscript{190} It must use the practice having the least possible disparate impact to effectively argue that the practice amounts to a business necessity.

An alternative scoring system, under this standard, would merely have to marginally reduce discrimination. Even a slight lessening of the adverse impact on the protected class would suffice to establish the final step of the ECOA effects test. Although employment cases have not yet progressed to the point of choosing between these alternative standards for measuring the third step of the effects test, credit scoring imposes certain additional considerations. These special distinctions between scoring and employment practices support the judicial adoption of the former, less rigorous standard.

In the first place, use of the latter, very strict standard will impose more development costs on an already very expensive technology. Its adoption would place system developers in the position of having to discover all available combinations of characteristics, attributes, and scores which maximize the separation between creditworthy and non-creditworthy applicants.\textsuperscript{191} The system developer must then select from among them the system that has the least unequal impact on the classes of applicants protected by ECOA. Such redundant develop-

\textsuperscript{189} Id.; Blalock, supra note 46, at 275-76.


\textsuperscript{191} See text accompanying note 56 supra.
ment of multiple systems would proportionately multiply the creditor's cost of installation. Such additional costs will have the secondary result of discouraging creditors from using scoring systems, and encouraging use of the relatively more discriminatory judgmental systems. In addition, the resources diverted by both creditor and applicant into development of multiple systems to attain marginal improvements in nondiscrimination could be far more cost-effective if used to combat discrimination elsewhere in the credit process.

Requiring pre-use selection among multiple alternative systems may require that the creditor make tradeoffs between different protected classes, as described above. One system may select creditworthy elderly applicants with unusual accuracy while performing poorly with married women. Another system may accurately select creditworthy minorities while operating to the disadvantage of certain religious groups. How can the system user select among them? Although advocates of linear programming may feel that this tradeoff will readily yield to the dispassionate calculation of the greatest good for the greatest number, such detachment will hardly mollify the specific applicant, rejected by the selected system, who would have received credit from an alternative system.

Employment practices do not present quite such graphic choices. Employment cases frequently use the effects test to detect intentional discrimination disguised by facially neutral practices. The court frequently lacks sufficient data on the effect of a proposed alternative employment practice on other protected classes. In addition, the challenged practice may affect only one particular group. An unreasonable height and weight requirement or a policy of terminating pregnant employees clearly has disparate impact on females. Altering

192. See text accompanying note 129 supra.
193. See text accompanying note 183 supra.
such practices will only marginally affect racial minorities or particular religious groups.

Credit scoring methodology presents the parties with more information about the applicant population and the exact effect of alternative systems upon the population. Potential tradeoffs between cost and marginal changes in adverse impact between different protected classes, therefore, become overt. The advent of more accurate measurements does not necessarily mean that courts should automatically embrace the most rigorous standard susceptible of measurement. Rather, they should balance all the known considerations and avail themselves of the unresolved nature of Title VII interpretation to select the standard which will in the long run minimize discrimination against all credit applicants.

Degree of increased losses

Some consumer advocates suggest that reducing discrimination by using the effects test may entail less accurate credit analysis and increased loan losses. This argument capitalizes on the fact that credit-scoring technology can quantitatively compare the performance of alternative prediction systems and their effect on the various protected classes of applicants. A court, therefore, no longer needs to wallow in vague conclusions about “unavailable alternatives,” “manifest relationship,” or “business needs.” Confronted with statistical evidence that 0.01% loss of predictive accuracy will result in fifty percent reduction in discriminatory impact, it could rule that the effects test requires the system user to accept this minor increase in losses in order to sizeably reduce discrimination.

Such a rule, however, might appear potentially undesirable. First, mere availability of more accurate credit data does not necessarily mandate its immediate use. Second, creditors use scoring systems to make profits. For them, maximizing profit constitutes a business necessity. A court may properly require changes that reduce discriminatory impact without affecting profitability. Any diversion of a business’

196. CHANDLER & EWERT, supra note 183; G. Chandler, Special Purpose Credit Scoring Models for Protected Classes (Oct. 7, 1977) (paper presented before the Citicorp Conference on Special Purpose Credit Programs); Control Systems, supra note 92; Roy & Lewis, supra note 7, at 17; Roy & Sanderson, supra note 17.
profits to serve purely social ends, however, distorts the economy and constrains the free flow of funds to those areas best able to use them.

Finally, any line-drawing process presents a certain potential for unanticipated excess.\textsuperscript{199} A 0.01\% decrease in the repayment rate may seem like a good tradeoff for a fifty percent decrease in disparate impact, but what about 2\% for 20\%, 15\% for 1\%, or 100\% for 0.01\%? Selection among relative repayment and approval rates compares apples and oranges. No common basis for undertaking a cost-benefit analysis exists. Accordingly, judicial forums uniformly should avoid selection of mandatory decreases in predictive accuracy as an acceptable price for increased acceptance of creditworthy applicants from the protected class.

\textit{Inclusion of judgmental elements}

The earlier description of how credit scoring systems operate assumed that the system has the final decision with regard to which applicants get credit. Everyone above the cutoff score receives credit. Everyone below the cutoff receives an adverse action notice. In reality, the credit process tends to lack this mathematical purity.\textsuperscript{200} Certain judgmental elements usually intrude.

This judgmental intrusion occurs because creditor managements do not fully understand scoring systems or trust scoring techniques. The scoring sheet may not confirm their intuitive beliefs about what characteristics associate well with creditworthiness. Given the amount of money at stake, the creditor refuses to blindly trust the system's predictions and insists upon second guessing it judgmentally. This override typically takes the form of a reexamination by a credit officer of the scoring system's credit approvals or denials, depending upon whether the creditor's present business strategy calls for expansion of approvals or reduction of default rates.


Addition of judgmental overrides probably does not increase the predictive accuracy of a creditor’s final decisions. The perfect system of credit analysis, which accepts all creditworthy applicants and rejects all noncreditworthy applicants, has not yet been developed. An empirically based credit system using all available information, makes the most accurate prediction statistically possible and predicts the rate at which it will make classification errors. Judgmental second guessing based on the same information would, in the aggregate, misclassify more applicants correctly classified by the scoring system than it would correctly reevaluate applications misclassified by scoring. In addition, judgmental overrides increase the potential for discrimination in credit analysis. Reintroduction of the human element necessarily includes human biases and prejudices. The reviewing credit officer may use information about the applicant’s appearance or demeanor that the system does not find predictive or that the ECOA brands illegal.

In contrast to a judgmental system of credit analysis, an empirically derived credit system which is demonstrably and statistically sound has the legal advantages of the statutory authorization to directly consider the applicant’s age and a potentially lighter burden of proof than in an effects-test lawsuit. Accordingly, users of so-called “hybrid” systems want to bring their credit analysis within the Regulation B definition of a demonstrably and statistically sound, empirically derived credit system, notwithstanding their use of judgmental overrides. In response to such requests, an official staff interpretation, promulgated shortly after Regulation B’s effective date, authorizes inclusion of nonscoring components in empirically derived credit systems. The nonscoring component must, however, comply with all requirements imposed on judgmental systems and may not consider the applicant’s age directly. The entire system, including both the scoring and nonscoring components, remains subject to the full impact of the effects test.

The considerations that may have influenced the staff to approve the use of hybrid systems parallel those affecting the Board’s decision regarding how rigorous to make standards for system validation.

201. Rosenberger, Measuring the Impact of Overrides, Part 1, Viewpoints, Fall/Winter 1977, at 1; Rosenberger, Measuring the Impact of Overrides, Part 3, Viewpoints, Winter/Spring 1978, at 1. Often the officer considers information not available to the system development process, such as applicant demeanor.


204. See text accompanying note 129 supra.
Admittedly, judgmental overrides probably decrease the overall accuracy of the system's predictions and may offer greater potential for credit discrimination. However, purely judgmental systems tend to discriminate illegally to an even greater extent. Nondiscrimination policies should therefore encourage use of scoring, even with judgmental overrides, as the lesser of two unavoidable evils.

This conclusion necessitates the classification of hybrid systems as empirical. Calling them judgmental systems would for several reasons tend to discourage the continued development of scoring. In particular, defining hybrid systems as empirically derived permits judgmental creditors, undecided about the virtues of scoring, to gradually adopt and experiment with such systems while retaining their familiar judgmental techniques as overrides. In addition, this designation permits hybrid systems to continue direct consideration of age. To the extent that age makes an excellent predictive characteristic, its consideration improves the accuracy of the system's credit predictions and benefits the elderly.

**Borrowed systems**

Regulation B's final pronouncement on credit scoring technology concerns use of borrowed systems.\(^{205}\) The Act itself neither authorizes nor prohibits the transfer of scoring systems between creditors. However, because of the high development costs associated with even a single scoring system the question of whether creditors can share their systems with each other naturally arises. Other creditors need to be able to borrow systems because their applicant populations do not have sufficient size for statistical analysis in the development process.

Again, as with the standards for validation, competing scientific and practical considerations apply. Researchers believe that the composition of applicants varies from creditor to creditor.\(^{206}\) Accordingly, the scoring system that predicts well for one creditor may predict poorly for another. Similarly, it may have a discriminatory impact on one applicant population and not another. These factors militate against permitting the unrestricted transfer of systems among different users. On the other hand, scoring probably discriminates less than

---

judgmental analysis. The government should therefore avoid imposing unduly rigorous performance and nondiscrimination standards on scoring systems for fear of discouraging their adoption by additional judgmental creditors.

Regulation B successfully accommodates these countervailing interests. It permits borrowing of scoring systems but requires that the creditor undertake validation tests as soon as it accumulates sufficient information to do so. This approach enjoys the dual advantage of permitting a creditor to avoid development costs by borrowing an existing system while still affording its credit applicants considerable assurance that the system processing their request will make accurate decisions.

Discrimination and Scoring

Notwithstanding the detailed proscriptions of Regulation B, its limitations in no way foreclose all possible discrimination by credit scoring systems. The Supreme Court describes discriminatory practices as "invidious" and acknowledges that new ones will continually appear. Obviously, no single source can enumerate all the possible scoring practices that might constitute illegal discrimination. Imaginative users and applicants will eventually utilize, and challenge as discriminatory, practices not presently known.

This section nevertheless seeks to highlight certain areas of scoring that present potential ECOA problems. It will also attempt to point out the pertinent considerations that will guide courts when they confront potentially illegal scoring practices in the future. Deciding what practices constitute credit discrimination depends upon both the immutable technical and economic aspects of the credit process and the case law that interprets employment and housing antidiscrimination statutes.

Per se discrimination

As described in the preceding section, Regulation B imposes certain basic restrictions on scoring procedures in an attempt to reduce the most egregious forms of credit discrimination. A system may not score prohibited bases such as race, color, religion, national origin, sex, marital status, receipt of public assistance, and exercise of consumer
Age may not receive a negative factor or value. Neither may the system score additional attributes that Regulation B deems to constitute indirect discrimination, such as alimony income, parenthood plans, telephone listings, pension income, and availability of credit insurance.\textsuperscript{209} Notwithstanding the express nature of these prohibitions, one occasionally encounters pre-ECOA systems that overtly violate even these basic proscriptions. The author’s experience includes systems that score religion and systems that score telephone listings in the applicant’s own name, an obvious handicap for married women. These nonconforming systems typically represent the creditor officers’ intuitive impressions rather than empirical conclusions. As public recognition of ECOA spreads, these few per se violations will presumably disappear. The consumer compliance examinations of regulated creditors, such as banks, conducted by federal authorities, as well as the Act’s provision for substantial civil penalties, will undoubtedly speed this process.\textsuperscript{211}

Other discriminatory characteristics

More frequently, scoring systems employ characteristics that do not constitute per se violations but that may discriminate indirectly. The probability of discrimination varies from situation to situation. Consider a system that scores “housewife” as an occupation category. This attribute consists primarily of married women. If “housewife” receives few points, does this constitute sex or marital status discrimination? Certainly it has an adverse impact upon a protected class. However, empirically this treatment may be justified. Conversely, prohibiting scoring of “housewife” could result in reclassification of such applicants as “unemployed,” presumably a category earning even fewer points. Indeed, use of “unemployed” can itself present a problem of discrimination on the basis of age (retired) or receipt of public assistance.

Even a practice as ordinary as scoring homeownership may present potential discrimination. Certain protected classes of applicants rent more frequently than either society at large or the creditor’s applicants taken as a whole. Even if homeownership proves an excellent predictor of creditworthiness, this seemingly innocuous practice may

\textsuperscript{209} Regulation B, 12 C.F.R. § 202.6(b)(1) (1978).
\textsuperscript{210} Regulation B, 12 C.F.R. § 202.6(b)(3)-(5) (1978).
violate the effects test or some other indirect measure of discrimination. The preceding section suggests that future judicial action will articulate new standards for measuring credit discrimination but that their eventual perimeters remain obscure. Almost all applicant characteristics correlate negatively with some protected class.

Currently, much controversy surrounds the use of geographic characteristics by scoring systems. Some systems processing applicants on a nationwide basis assign different numbers of points to residents of different states or multistate regions. Other systems allocate points based on the applicant’s ZIP code or even smaller geographic zones. Still other systems score urban, suburban, and rural addresses differently. In a properly developed system, this geographic disparity presumably derives from the varying credit performance of local residents.

Critics charge that consideration of geographic characteristics may result in pretextual or even inadvertent discrimination against minorities. Others criticize this practice as illegal “redlining” contributing to the unnecessary deterioration of marginal neighborhoods. Altogether too much controversy and too little nonpartisan research surrounds these charges because of their highly political nature.

This criticism ignores several countervailing considerations. First, the degree of disparate impact on minorities will vary from system to


213. See authorities cited note 3 supra.

system. Controlled research might well reveal that certain systems scoring geographic characteristics, particularly those whose attributes embrace large areas, have no discriminatory effect. Second, while redlining may seem socially undesirable, this does not necessarily make it illegal. Mere stupidity should not automatically inspire governmental intervention. Third, banning use of geographic characteristics may not have the effect desired by its advocates. If the “location of the applicant’s residence” predicts creditworthiness well and no substitute for it exists, then its exclusion will mean less accurate credit analysis. This increased risk will make creditors more reluctant to lend to applicants who are economically marginal. Economically marginal applicants comprise the group most in need of credit, the group most likely to live in redlined neighborhoods, and the group critics of geographic scoring are most eager to protect. Outright prohibition of redlining could have unanticipated, counterproductive effects.

Conversely, if “applicant location” predicts well, and a good substitute for it does exist, then the redlining problem merely is displaced one step further. System developers will replace location with a substitute characteristic, such as age and condition of dwelling. The end result will not change, although the substitute may not appear as “discriminatory” and may not be the object of political criticism. It may, however, have the same aggregate effect on the same group of applicants. Advocates of direct substitution argue that the replacement characteristic relates to applicant creditworthiness, whereas the applicant’s ZIP code does not. This argument ignores the fact that discriminant analysis only selects the characteristics that predict creditworthiness well. It does not investigate why they predict well. Scoring has not yet advanced to the point of including causal analysis.

Authorities on equal credit opportunity foresee considerable litigation before resolving the geographic scoring issue.

Multiple systems

A related problem concerns the use of multiple systems by a single

---


CREDIT SCORING

creditors. Creditors currently use different systems for different applicants based upon geography, demographics, type of credit, or other preliminary screening considerations. Thus, some creditors process applications from all over the country using a single system, with or without scoring geographic characteristics. Other national creditors employ a different system for each of half a dozen regions. Still others build a different system for each branch or store.

Other users of scoring systems route applicants under thirty years of age to one system and those over thirty to another system. Researchers have experimented with processing male and female applicants into different systems. Full service banks offering several credit cards, overdraft accounts, and various installment programs frequently apply scoring to some lines of business and not to others. A few creditors direct old customers to one means of analysis and new applicants to another, or have applicants rejected by one system reanalyzed for possible reconsideration by a second system.

To date, the number of systems used by a creditor, and the applicant subpopulations they cover, has remained a business decision. The creditor deploys multiple systems because of the increased accuracy they confer upon the various subpopulations affected. It continues to subdivide the population into smaller and smaller units until the increased profit from improved accuracy no longer offsets the costs of additional systems development. From a technical standpoint, only the need for a large applicant base for each system’s construction prevents the use of multiple systems for applicants with different attributes, rather than simply incorporating and scoring those differences.

Multiple systems now present the question of whether their use or nonuse can discriminate. Consider creditor $A$ that puts applications from both males and females into a single system. If eighty percent of its applications come from males, then the development sample from which the system derives would also include about eighty percent males. Frequently, characteristics that predict well for male applicants do not work as well for female applicants. In the single system, however, the better predictors for females are submerged in the largely male development sample.

Contrast the single system user with creditor $B$ that develops two systems based on separate samples of male and female applicants. Al-

---

218. See text accompanying note 70 supra.
219. CHANDLER & EWERT, supra note 183.
though its future applicants may consist primarily of males, its few fe-
male applicants are processed by a separate system that incorporates 
those characteristics that best predict female creditworthiness. As a re-
result of this increased accuracy, creditor B extends much more credit to 
females than creditor A at the same level of risk. It should also confer 
slightly more credit on male applicants.\textsuperscript{220}

The business preferability of one system as opposed to multiple 
systems will vary depending upon the individual creditor’s own situa-
tion. Whether ECOA compels one approach or another remains uncer-
tain. ECOA prohibits credit analysis from considering the applicant’s 
membership in a protected class such as sex.\textsuperscript{221} Certainly a system 
could not legally make the applicant’s sex a scored characteristic. The 
question is whether analyzing the creditworthiness of applicants using 
different criteria because of their sex, constitutes taking “a prohibited 
基础 into account.” It would certainly appear so, especially if discrimi-
nation means treating one applicant \textit{differently} from another because of 
a prohibited bias.\textsuperscript{222}

On the other hand, Regulation B may support such dual systems 
because it defines discrimination as treating “an applicant \textit{less 
favorably} than other applicants.”\textsuperscript{223} Increasing the information avail-
able for credit analysis makes the system more accurate and allows the 
creditor to approve more applications while maintaining the same level 
of risk. If it can take the applicant’s sex into account by using two 
systems, as suggested above, usually more applicants will receive credit. 
The multisystem creditor, that sets its cutoff scores so that its systems 
have equal repayment rates or comparable rates of profitability, could 
persuasively argue that it does not treat any applicants less favorably 
because of their sex. It does handle their applications differently be-
cause of their sex, but not less favorably.

A similar conclusion follows with respect to similar divisions of 
applicant population for development of multiple systems. The credi-
tor that has a different system for each store or branch can make more 
accurate decisions. Presumably, the credit applicants in suburban 
Scarsdale have different attributes from those in rural Mississippi. 
However, the extra accuracy means developing more systems and in-
curring proportionately higher costs. At least indirectly, it appears to

\textsuperscript{220} \textit{Id.}
\textsuperscript{222} \textit{Compare} Baker v. California Land Title Co., 349 F. Supp. 235, 238 (C.D. Cal. 
\textsuperscript{223} Regulation B, § 202.2(n) (1978).
take into account the location of the applicant's residence. On the other hand, this type of geographic consideration gives considerably more credit to the applicants from marginal neighborhoods, who would be rejected by a single system because of their lower economic status.

Clearly, consideration of location by multiple systems warrants social encouragement, but how does it differ from merely scoring ZIP code? Possibly, the way the competing alternatives use geography serves to distinguish them. A single system that scores the location of an applicant's residence, of the applicant's employer, or even of the branch of the creditor at which the application was submitted can use the attribute to automatically reject almost all applicants from that area. A system that scores only applications from the same area automatically grants credit to a predictable fraction of the requests. Because of its better predictive ability, fewer local applicants would be rejected by the multiple system.

The scoring of age and the prohibition against assigning it a negative factor or value may provide a partial analogy. By dividing the applicant population along some basis, prohibited or not, and analyzing the subpopulations separately, everybody gains. Each subpopulation receives more credit and the creditor achieves greater accuracy. Unfortunately, the ECOA contains no express authorization for favorable consideration of prohibited basis other than age. This oversight may stem from the fact that Congress did not realize that multiple systems can convert a zero-sum situation, one under which one subpopulation must lose in order that another may gain, into a nonzero-sum situation where everybody can gain.

Conversely, neither does ECOA expressly prohibit multiple systems. Indeed, under certain circumstances, the general principle that creditors should aggressively eliminate discrimination arguably compels development of multiple systems.224 Accordingly, the eventual legal resolution of the multiple system remains uncertain. Only the technical requirement of a large enough applicant base for systems development and the economic requirement of sufficient increased profitability to offset extra development costs constrain the widespread use of multiple systems and the eventual judicial disposition of the issues they raise.

Overrides

As suggested in the preceding section, many creditors elect to over-

224. See text accompanying note 190 supra.
ride the system's decision in selected cases.\textsuperscript{225} The addition of a nonmechanical "decision screen" reintroduces considerable opportunity for discriminatory credit decisions notwithstanding the use of a properly developed, nondiscriminatory scoring system. The potential for discrimination through overrides will vary from case to case, depending upon the pertinent facts and policies. ECOA litigation challenging overrides will therefore closely parallel litigation over purely judgmental systems of credit analysis. For example, after scoring at least sixty-two points, the applicant must also have an affirmative credit report, certain collateral, or a present cash-flow ratio. Whether this override discriminates will vary from creditor to creditor.

\textbf{Prescreening}

A related practice concerns prescreening of applicants before they can be scored. Even using scoring, credit analysis has a cost. Therefore, many creditors limit in arbitrary ways the applications they will consider for scoring. For example, banks frequently will make loans only to depositors who have used the bank for a certain period of time or who maintain certain balances. Many creditors make a practice of only extending credit to "local" residents. Another common prescreening device requires that applicants meet an income requirement or have certain credit references before the creditor will consider processing their credit requests.\textsuperscript{226} A more controversial practice consists of refusing loans to finance the purchase of homes in certain neighborhoods.

Prescreening has two undesirable effects on otherwise neutral scoring systems. First, the screen may itself discriminate, at least under the effects test. Prescreening that eliminates applicants who have not maintained bank accounts for at least the two preceding years might have a disproportionate impact on the young or on recipients of public assistance.\textsuperscript{227} Screening out loans to residents of selected neighborhoods may discriminate against minority applicants. Most prescreening appears vulnerable to similar challenges, although the degree of disparate

\textsuperscript{225} See text accompanying note 200 supra.

\textsuperscript{226} \textit{E.g.}, Letter from Edward S. Cogan to the Director of the Division of Consumer Affairs, Federal Reserve Board (Sept. 30, 1977) (on file under Docket No. EC-513); Letter from Lucien A. Dancause to the Director of the Division of Consumer Affairs, Federal Reserve Board (Feb. 3, 1977) (on file under Docket No. EC-37).

\textsuperscript{227} \textit{National Comm'N on Electronic Fund Transfers, EFT in the United States 71 (Final Report, 1977); accord, I. Friend, Individuals' Savings: Volume and Compassion (1954); R. Goldsmith, A Study of Savings in the United States (1955); U.S. Savings & Loan League, Consumer Survey on Saving Habits (1954).}
impact will vary depending upon the circumstances surrounding that particular creditor or device.

A second objection to prescreening arises on technical grounds. Both scientific principles and Regulation B require that the system predict the creditworthiness for the population consisting of all the creditor’s applicants. Only a system developed from a sample of this “true applicant population” can accurately predict credit performance.

Prescreening on any basis prevents creditor’s files from fully representing the true applicant population. Because the files reflect only a portion of the true applicant population, any system developed using a sample drawn from those files will reflect the statistical biases of the subsample. Not only will the resultant system predict poorly, but it may also have unnecessary adverse effects on protected classes of applicants. The creditor exacerbates these problems if it applies the screening device inconsistently or intentionally changes it. It then puts into the scoring system a totally different subpopulation from the one it used to construct the system. Inaccuracy necessarily ensues with possibly unnecessary discriminatory impact.

Notwithstanding the legal and technical undesirability of prescreening, its suppression will entail considerable difficulty. Creditors will naturally evince considerable reluctance to abandon traditional practices that seem to have a business purpose and to work effectively. Plaintiffs will encounter evidentiary problems when proving the existence of the screening device and measuring its effect.

A corollary to the prescreening problem recurs in the form of self-screening and steering. The preceding discussion assumes that the true applicant population consists of all those who want credit from this creditor and that screening occurs only after they formulate this desire. Therefore, by obtaining written applications from everyone

228. Regulation B, § 202.2(p)(2)(i) (1978); see text accompanying note 78 supra.
229. BLALOCK, supra note 46, at 527-28; HUFF, HOW TO LIE WITH STATISTICS 11 (1954).
230. Suppose the creditor’s official “policy” requires $12,000 minimum income. Examination of the institution’s files reveals, however, several loans to persons not meeting this standard. These persons turn out to be the credit manager’s son-in-law, the local minister, creditor employees, and other “special” cases. The creditor has applied its screen inconsistently. Any sample drawn from the creditor’s files suffers from the double statistical bias of underrepresenting applicants with less than $12,000 income, while concomitantly reflecting an unrepresentative group of those applicants.
231. This true population does not consist of all local residents or all local credit customers. Cf. Green v. Missouri-Pacific R.R., 523 F.2d 1290 (8th Cir. 1975) (alternative definitions for the applicant population). A person who does not want credit clearly should not appear in the applicant population. Whether the population includes a person who wants credit but applies elsewhere remains unclear.
who walks through the creditor's door (often referred to as the "through door population" by system developers), the adverse impact of screening is excised, or at least mitigated. Unfortunately, the real story becomes somewhat more complex because screening can also occur before the applicant enters the creditor's premises, indeed before the applicant can decide to apply to any particular creditor. The third party that actively "steers" the would-be applicants to certain creditors and away from others, before they file applications, changes the composition of the "through door population."

Similarly, a successful advertising campaign that reaches applicants not previously served by that creditor will also alter the demographic and economic characteristics of the true applicant population. Because of changes in the public's perception of the institution's credit standards and of the nature of its clientele, new types of applicants "self-steer" themselves towards and away from the creditor. Indeed, even without special advertising, a creditor projects a certain image to the public that encourages or discourages applications from certain segments of the populace.

This behavior pattern raises a quasi-ontological question: What persons comprise a creditor's true applicant population? Clearly the "through the door population" is included. What about persons who decided to apply with that creditor but who somehow were steered away? Indeed, how about those who needed credit, but did not think of applying with the creditor because of a fear of rejection or for some other reason? Presumably the consumer who never heard of the credi-

232. Such third parties might include a real estate broker, a seller of consumer goods, or other persons who advise the consumer about sources of financing incident to a sale. Certain types of steering have provoked considerable controversy in housing-discrimination cases, United States v. Henshaw Bros., 401 F. Supp. 399 (E.D. Va. 1975); Topic v. Circle Realty, 377 F. Supp. 111 (C.D. Cal. 1974); Zuch v. Hussey, 366 F. Supp. 553 (E.D. Mich. 1973). Credit scoring methodology objects to all such steering and screening practices, but for different reasons. Steering skews the composition of the "through the door" population and makes it statistically unrepresentative of the true applicant population, quite apart from whether it discriminates illegally.

tor does not fall into the true applicant population, regardless of the creditor's advertising practices.

Some observers may conclude that only the applicants actually walking through the door and requesting credit should constitute the true applicant population. One becomes an "applicant" only by actually asking for credit. Others would argue, however, that such a view takes a too limited perspective of the effect of creditor advertising policies and of creditor responsibilities to the community it serves. A creditor that indirectly discourages a protected class of applicants from ever applying discriminates just as much as one that scores a prohibited characteristic, even though the steering may result in no tangible evidence of disparate impact. Thus, a valid argument can be made that the true applicant base includes people who wanted to apply for credit, but did not do so. Only by comparing this more inclusive population to the "through the door population" can the observer detect discriminatory steering. Isolating this true applicant base also gives the system developer more complete data from which to construct a more accurate scoring system that deteriorates less when the "through the door population" changes.

Unfortunately, determining the practices that cause the various types of steering and measuring their effects on the "through the door population" present considerable scientific and evidentiary uncertainties. How does one determine who did not apply? One possible solution uses consumer-behavior theory. Researchers hypothesize that consumer actions occur through a four-step process, problem recognition, problem comprehension, attitude formation, and overt behavior. Thus, the consumer first recognizes a need for certain goods or money. Analysis of this need results in recognition that credit would assuage this need and that several possible sources of credit exist. The consumer then analyzes the sources and develops a favorable attitude towards one of them, possibly because of price, convenience, or some similar consideration. Finally, the consumer takes action on the basis of the attitude, and actually applies for credit.

The question then becomes, why did the consumer not choose to apply to this particular creditor? The choice might depend upon "ra-
tional" factors such as interest rate, or on potentially discriminatory factors such as third party steering or slanted advertising. At present, survey research has only just commenced investigating this hypothetical behavioral process. However, no superior approaches to the steering problem have yet appeared. In conclusion, prescreening in all forms seems most undesirable, but the point at which it becomes illegal remains uncertain.

Cutoff scores

Another potential discriminatory practice occurs at the other end of the credit-analysis process, when the creditor sets its cutoff score. As described in Section III, the system developer generates a repayment-probability table that shows, for a range of scores, what repayment rate and acceptance rate the user can expect. The creditor selects its cutoff score based upon its business strategy and the degree of risk it wishes to support. It applies this cutoff to all future applications and may adjust the cutoff in response to its changing business outlook.

This apparently mechanical process may discriminate because of the complete creditor discretion in establishing the cutoff. By selecting a particularly high cutoff, the creditor can intentionally exclude a very high percentage of minorities, young people, or working women, for example. Where the creditor selects a high cutoff because it desires to do business only with white, male executives, it reflects improper intent.

Even where the creditor selects the cutoff for purely economic reasons, a violation of the effects test may occur. The creditor merely intends to reduce credit losses to an absolute minimum during a time of high costs for money. However, if this decision excludes disproportionate numbers of applicants from a protected class, then at least a prima facie case of discrimination exists.

Creditors also use scoring cutoffs to set credit limits on open-end credit such as credit cards and check-overdraft plans. Applicants scoring more points receive higher credit limits. Setting the required score unreasonably high or the credit limits unreasonably low, could

237. Orgler, supra note 27.
239. See text accompanying notes 151-60 supra.
240. See text accompanying notes 18-19 supra.
under circumstances similar to those posited above, violate the ECOA as discriminatory.

Assuming consistency

Section II notes that all credit analysis rests upon the assumption that the future will resemble the past. Future satisfactory applicants will have the same attributes as past satisfactory applicants. Strong statistical and historical evidence supports the validity of this assumption.

Such an assumption would, however, provoke any civil-rights advocate familiar with other types of discrimination litigation. The relevant case law repeatedly attacks practices based on an assumption of consistent behavior over time. The fact that prison inmates of different races have proven disruptive when sharing cells does not justify a policy of permanently segregating prisoners by race. A high dropout rate among married high school students does not imply that a particular student should not participate in extracurricular activities for fear of interfering with marital responsibilities. A job which requires lifting heavy weights or working long hours does not necessarily require a male to fill it. Civil-rights laws are purposely designed to challenge traditional assumptions and stereotypes about the behavior of particular groups and their proper roles in society.

Proponents of this theory argue that the scoring system does not know how rejected applicants would have performed. It denies credit to many applicants who would have repaid satisfactorily if given the opportunity. Furthermore, even if a particular group did prove non-creditworthy in the past, this does not mean that it will necessarily retain that attribute in the future.

These arguments may have validity for other types of discrimination and possibly for judgmental systems of credit analysis. The empirical foundation of scoring, however, rebuts each of the assertions. First, in the aggregate, the system development process takes into account how rejects would have performed. The repayment probability table and the augmented frequency count of the accepted applicants

241. See text accompanying note 42 supra.
serve as a basis for calculation of both the number of applicants rejected at any cutoff score and the percentage of creditworthy applicants among the rejects.\textsuperscript{246} By comparing the total number of creditworthy applicants below the cutoff with the total of noncreditworthy applicants, one arrives at the projected repayment rate for the reject population, seventy-two percent for example. This repayment rate among the reject population must necessarily amount to less than the ninety-two percent repayment rate among the accepted population.

Notwithstanding the preceding conclusion that most rejected applicants would repay if given credit, it does not necessarily follow that the creditor should approve their requests. The system cannot fully distinguish between creditworthy rejects and noncreditworthy rejects over most of a range of scores. Accepting more of the reject population means incurring a marginal repayment rate of only seventy-two percent. This compromise may seem to exclude a considerable number of creditworthy applicants, but their acceptance could entail high losses because of the lopsided relationship between profit on a satisfactory account and loss on an unsatisfactory account.\textsuperscript{247}

Finally, scoring accepts the principle that a particular group's credit performance may change over time.\textsuperscript{248} As change occurs, the system's performance decays. The developer rebuilding the system will investigate the causes of the deterioration and allow for changes in each group's creditworthiness. This process has the advantage of adjusting to changes as they actually occur, instead of requiring, as judgmental systems do, the creditor to either guess at the nature of the changes or increase risks by actually extending credit to measure whether change has occurred.

**Potential Technical Objections to Scoring**

The foregoing concludes the substantive part of this Article. There remain certain minor, technical issues that pertain to the relationship between scoring technology and Regulation B's specific rules for non-discriminatory credit administration. Most of these flow from over-literal interpretation of the regulation's text or the scientific constraints on the methodology, rather than from any potential for real discrimination.

\textsuperscript{246} See Figure 1, \textit{supra}.
\textsuperscript{247} See note 4 & accompanying text \textit{supra}.
Mandatory income consideration

Section 202.6(b)(5) states that "Where an applicant relies on alimony, child support, or separate maintenance payments in applying for credit, a creditor shall consider such payments as income to the extent they are likely to be consistently made." Prior to ECOA, creditors commonly refused to consider alimony income as a source of payment for credit. They apparently reasoned that ex-spouses rarely pay alimony as agreed or awarded, making this type of income unreliable. This practice has highly adverse effects on divorced and separated women trying to obtain credit.

Applied literally to scoring, this provision may mean that if an applicant discloses alimony income the system must score it. Income, however, does not appear as a characteristic at all on most scoring sheets. It rarely predicts creditworthiness well enough to warrant inclusion. It would seem most anomalous to mandate scoring of one type of income while essentially barring use of another type of income, especially in cases where neither contributes to the system's accuracy.

This anomaly stems in part from overly broad drafting. In the sentence preceding the quoted passage, the drafters apparently seek to prohibit similar, arbitrary exclusion of retirement income for consideration by creditors. There they provide that "[A] creditor shall not discount or exclude from consideration the income of an applicant or the spouse of the applicant because of a prohibited basis or because the income is derived from part-time employment, or from an annuity, pension, or other retirement benefit . . . ." This formulation does not mandate scoring of retirement income unless the creditor also scores other types of income, thereby avoiding the whole problem.

Indeed, this alternative statement may point to a way of defining away the whole problem. Regulation B authorizes a creditor to consider any information obtained, so long as it does not use it to discriminate on a prohibited basis. In the absence of discrimination, the creditor may evaluate the information any way it sees fit. This system does not consider income at all. It therefore accords alimony income exactly the same weight as all other types of income. Where no discrimination occurs, Regulation B authorizes the usage of such information, regardless of the type of system involved. This reasoning may

seem sophisticated, but would rebut this purely technical ECOA violation quite nicely.

Joint applications

The preceding discussion assumes that the applicant requests credit alone. Frequently, however, several persons apply for a loan jointly. The mechanism by which scoring should process joint applications has not yet received much discussion.

Regulation B hardly mentions joint applications and their evaluation.253 A creditor does not have to offer joint credit. However, the prohibition against marital status discrimination clearly indicates that if the creditor does offer joint credit, then it cannot limit its availability to married couples.254 This paucity of regulatory guidance presently leaves the question of scoring joint applications largely unresolved. Suppose the system scored home ownership. One joint applicant owns a home, but the coapplicant rents. How should this be scored? On the other hand, suppose the system scores income, and one applicant has a large income and the other a small income. Should their joint income be scored? Would the same rationale apply to scoring the age of a young applicant and an old applicant?

Several reasonable solutions exist. One approach would construct separate scoring systems for joint applications and individual applications.255 The characteristics analyzed for possible inclusion during the development of the joint application system would include: first applicant's characteristic (auto, for example), second applicant's characteristic (auto), and their joint characteristic (two autos). Conceivably, the car owned by the second applicant might predict creditworthiness, while the car of the first applicant would not. Alternately, the joint debt-to-income ratio might perform well, when the individual ratios perform poorly alone.

A second approach would develop a single system for all applications, but include analysis of joint applicants' characteristics and of combined characteristics. Under certain circumstances, the final system would score certain attributes of a second applicant or of both applicants, with varying impacts dependent on the data used.

255. See text accompanying note 217 supra.
A final, although possibly less satisfactory, approach would incorporate the number of applicants or cosigners as a second characteristic, if its predictive power so warrants.\textsuperscript{256} Thus, an individual application might get fourteen points and a joint application only three. To the extent a characteristic predicts, the system should use it. Traditional credit analysis assumes credit to be more secure if the transaction has a cosigner. However, scoring may not always confirm this presumption. If joint applications default at a higher rate than individual applications, then the former characteristic could cause the application to actually lose points.

Assumption of normality

For highly technical reasons beyond the scope of this Article the mathematics of discriminant analysis require that the two subpopulations used in system development form normal curves of equal variance.\textsuperscript{257} Thus, a plot of frequency counts over scores received should describe two curves of approximately the same size and shape.\textsuperscript{258} The curves can center on different parts of the graph, but should have the same shape. The difference between the centers of the curves depicts the scoring system's success at separating creditworthy and non-creditworthy applicants.

![Figure 9—Normal subpopulations of equal variance](image)

Applicant populations rarely satisfy this "normality" requirement. More frequently, the developer finds the creditworthy subpopulation clustered together forming a curve of narrow width and great height, while the noncreditworthy subpopulation covers a great width with little height.\textsuperscript{259}

\textsuperscript{256} Interview with Howard Schneider, Vice President Bankers Trust, in New York City (May 2, 1977).
\textsuperscript{257} See authorities cited note 56 supra.
\textsuperscript{258} DURAND, supra note 6, at 109.
\textsuperscript{259} Concepts of Scoring, supra note 17, at 10.
At present, no good solution exists to the problem of applicant populations that violate the assumption of normality. Developers often elect to ignore it, reasoning that use of nonnormal subpopulations will not downgrade overall system performance significantly. They also note that existing methods make correct predictions at a rate deemed acceptable for business purposes. At the same time, theoreticians are investigating techniques for adjusting the applicant population to statistically conform to the assumption of normality.\footnote{260}{Eisenbeis, supra note 65, at 16; Pitfalls, supra note 65. A similar technical problem may arise from artificial constraints on the development process as described at note 102 supra.}

**Conclusion**

Almost everything about the interaction of credit scoring and ECOA remains unsettled. Regulation B defines a demonstrably and statistically sound, empirically derived credit system, but users of scoring do not know whether their systems conform to its requirements. The law also defines credit discrimination, but lawyers cannot decide whether scoring systems infringe this prohibition. This universal uncertainty stems from the underdeveloped condition of the law, insufficient dissemination of knowledge about the operation of credit scoring, and the evidentiary difficulty of establishing whether a particular system discriminates.

This Article seeks to dispel some of the obscurity surrounding the first two causes of uncertainty. It suggests that future judicial decisions should give full attention to the empirical aspects of scoring when applying principles borrowed from other areas of antidiscrimination litigation not having comparable levels of statistical certainty. The third cause of uncertainty, however, must await judicial clarification on a case by case basis.