

Appendix

Safety and Soundness and Compliance Issues on Credit Scoring Models

SAFETY AND SOUNDNESS ISSUES

Credit scoring models, or score cards, can contribute to bank safety and soundness as long as bank management actively participates in the development, implementation, monitoring, and validation of the model. Operational factors, such as accurate implementation of the credit scoring model and adequate training of data input staff, also can have an impact on the effectiveness of models. Changes in the environment in which a bank operates can affect the predictive ability of a credit scoring model over time. For all these reasons, a credit scoring model's performance must be analyzed regularly. This section describes several issues on which examiners and bank management should focus to ensure effective use of scoring models.

Credit Scoring Model Development

Banks using credit scoring models purchased from third-party vendors and those using internally developed models must demonstrate that:

- Statistically developed credit scoring models are valid as defined in Regulation B.
- Credit scoring models accurately represent, or fit, the data used in development.
- The characteristics of the population used to develop the model are similar to those of a bank's current customers.

The last issue is particularly important if a bank chooses to use generic or pooled-data score cards or credit bureau score cards. Those score cards are developed using a population that includes applicants outside a bank's normal customer base. Information about population characteristics gathered in the initial validation process when the model is implemented will be necessary for future validation.

Validation

Score card vendors usually provide users with a list of recommended procedures for the periodic validation of the model. The OCC views those procedures as the minimum a bank should do to validate or revalidate its model(s) over time. Vendors usually recommend a minimum sample size for model validation; the sample size varies by vendor. For some products or customer groups (e.g., certain segments of the small business lending market), several quarters or years may be needed to generate a sufficient number of serious delinquencies (i.e., bad accounts) to evaluate the effectiveness of a model, even for banks that originate a large volume of loans. Users of credit scoring models have also benefited from independent model validations or an evaluation of the validation process by audit, credit review, or risk management staff.

Comparing the Model's Actual to Expected Performance

An integral part of the validation process involves comparing the model's actual performance to its expected performance. The expected outcomes are estimated at the time of model development. Without knowing the model's performance shortly after implementation, a bank cannot evaluate the magnitude and direction of the effect of changes in the market on the model.

At Implementation. The model's performance shortly after implementation should reflect the behavior of the underlying development population. When the model's results differ from the development data, management should analyze those differences to see if they support development of a new model.

During Solicitations. Solicitations must be tested or monitored early to ensure that the anticipated response by range is close to original projections. Often a bank will develop a marketing strategy meant to appeal to a certain set of borrowers within a range of scores. In developing the range, the bank generally arrives at an acceptable loss ratio for the target group as a whole and defines the range based on the expected acceptance rate of potential borrowers within each range. However, it is not unusual for a loss ratio to far exceed the projected ratio, although the performance of borrowers within each range performs remarkably close to model projections. The higher loss ratio is the result of much higher than anticipated acceptance of the solicitation by borrowers with weaker scores within the approved range and much lower response to the solicitation by those with higher scores. This pattern results in the solicitation portfolio being significantly skewed toward the weaker and more loss-prone borrowers and, in the longer term, results in higher than anticipated overall losses. The OCC is currently researching and reviewing the issue of adverse selection in credit solicitations.

Monitoring and MIS Reports on Performance of Models

Effective reporting and ongoing analysis is critical to controlling and managing risk in all bank activities, including the use of credit scoring models. Management should ensure that the bank has adequate MIS and staff trained to track each credit scoring model for accuracy and reliability.

OCC examiners will review the adequacy of the bank's monitoring reports and analyses during their examinations.

Monitoring reports. Banks that purchase models from a vendor or those that develop their own models must produce reports to monitor model accuracy and reliability. The reports are defined within two broad categories: model stability (front-end) analysis and model performance (back-end) analysis.

Front-end reports measure changes in the population at the time of customer application and

can serve as an early warning of deterioration in the model's performance. Examples of front-end reports include: population stability analysis or a comparison of the actual and expected score distributions; characteristic analysis or a comparison of applicants' score distributions by individual characteristics over time; and final score report or approve-and-deny score distribution analysis.

Back-end reports measure portfolio quality and the score card's effectiveness. Examples of back-end reports include: delinquency distribution analyses, including dynamic and vintage analyses; override tracking reports or reports that categorize the types and amounts of overrides; and the portfolio chronology log or a record of events relevant to the credit function. Many banks do not use a portfolio chronology log. The OCC recommends using such a log, because it can identify changes in the bank's strategy that may have affected a credit scoring model's performance. (Refer to Appendix A in the Credit Card Lending Handbook, distributed in October 1996, for a description of those reports.)

Report Frequency. Analysis of the model's effectiveness by competent staff members who understand credit scoring models should begin at implementation. Data for the front-end reports generally are available soon after implementation. The data to perform back-end reports will require more time to allow accounts to age. Formal, written reports should be prepared that summarize the results. The frequency of reporting depends upon the bank's loan volume. A bank's reporting cycle should reflect the length of time it takes to originate a sample of loans large enough to perform a statistically valid analysis. Smaller volume banks may take 6 to 12 months (or longer) to originate a sufficient loan volume to perform valid statistical analysis. A quarterly, if not monthly, reporting cycle should be set for large-volume banks.

Differences Between Expected and Actual Results

The back-end reports may reveal differences between expected and actual results. The reports may identify shifts in applicant behavior suggesting a model no longer effectively separates good from bad accounts and may identify deterioration in the model's ability to rank risk performance of the bank's current (or future) applicant pool. It may not be easy to determine whether this shift is due to a change in the lender's strategy or the performance of the model.

Bank management should consider the following factors in analyzing why a model is not performing as expected:

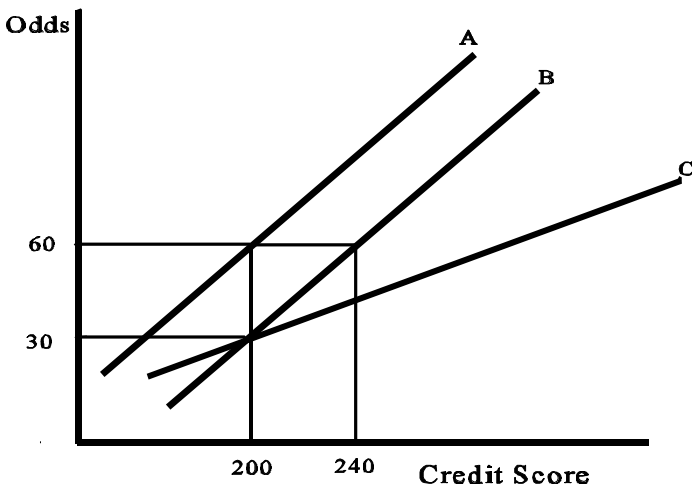
Model Development. Uncertainty about the model's performance may arise from the nature of the process used to develop a credit scoring model. Credit scoring models, in general, are derived solely from information specific to individual applicants. A model is used to predict the relationship between an applicant's profile and potential performance. This relationship is traced by analyzing the past performance of applicants from two to three years prior to model implementation. However, conditions in the market probably varied during this 24- to 36-

month period.

Market Shifts. Typical market shifts include changes in the economy, competitors' behavior, consumers' behavior, and the bank's marketing and lending strategies. As conditions change, the effectiveness of actuarially-based models may diminish. Banks may not observe shifts in the market directly. Rather, they may observe changes in the characteristics or types of applicants that result from market shifts. Small fluctuations in the distribution of applicant characteristics may also appear over time, because of random or seasonal fluctuations unrelated to market shifts. It is difficult to distinguish random shifts from market shifts, especially when the creditworthiness of the applicants accepted for credit normally does not become apparent for up to 18 to 24 months after origination.

Model Durability. Most credit scoring models are reasonably tolerant of changes in the market. As a result, market changes generally will not affect negatively the model's overall ability to predict credit performance. However, large and persistent changes in the population may affect future portfolio performance. A bank's response will depend on the model's ability to capture the impact of these changes. Model developers often assert that a model is valid if it continues to rank risk.

Example: a model ranks risk when it maintains the relative ranking of good-to-bad accounts (usually expressed as a good-to-bad-odds ratio) across score intervals. However, a model might continue to rank risk accurately, even when the absolute number of bad accounts exceeds expected levels of bad accounts within each score range. If line A represents the odds curve based on the development sample, the parallel downward shift in the odds curve to



line B represents an increase in the absolute bad account rate. At a score of 200, 60 good accounts exist for each bad account on line A, but only 30 good accounts for each bad account on line B. The relative risk remains the same. In models where increased scores represent better quality applicants, the number of good accounts increases on both lines as the scores rise from 200 to 240. Since the model still differentiates good accounts from bad, the model is considered a valid predictor of risk. Although relative risk may remain the same, the overall performance of

applicants evaluated by the scoring system may have declined.

A shift in the underlying behavior of the population might involve not only a downward

parallel shift of the odds curve, indicating increased absolute risk, but also a flattening of the slope of the curve, as shown by line C, representing a decrease in the ability of the model to rank applicants. Such a shift clearly shows that the performance of the current population differs from that of the development population. It is not clear whether the changes in behavior are significant enough to invalidate the model and require that a new credit score card be developed. This decision should be based on the bank's tolerance for risk; the desired approval rate or volume of business; and the tradeoff between the dollar losses from increased bad accounts and the cost of producing or purchasing a new and presumably better score card.

Economic Cycles. The OCC expects banks to closely monitor the performance of their credit scoring models throughout economic cycles, particularly when economic conditions change. For example, a number of banks use small business credit scoring models to extend small business loans. Those models have been effective in the prompt approval of credit at relatively small cost to the borrower. However, those loans have been extended in prosperous economic times. The predictability of small business credit scoring models developed in prosperous times should be retested in a declining business environment. Bank management must be particularly diligent in monitoring and tracking the performance of these and other credit scoring models if economic conditions change.

Use of Model with Inappropriate Population or Loan Products. Credit scoring models should be used only for the products and loan sizes for which they were developed (e.g., scoring applicants for \$1 million loans by using a small business scoring model developed for business loans less than \$250,000 may not be a safe and sound practice). The OCC recognizes that, as banks enter new markets, they often apply an existing credit scoring model based on similar attributes as a temporary decision making tool. This practice, by itself, may not be unreasonable as a second-best approach, until sufficient data become available to develop a more product-specific model. However, a decision to use this approach should be based on a preliminary analysis of the likely compatibility of the new market population with that used to validate the existing scoring model. An analysis similar in scope to the one used to validate the scoring model at implementation should be performed either before, or soon after (e.g., within three months), implementing the model in a new market. Once a sufficient sample of applicants and performance data on those applicants is generated in the new market, bank management should determine whether a model specific to that market should be constructed to replace one developed in a different market.

Although customized score cards based on the bank's own data may be more accurate in predicting applicants' performance within the specific market for which they were designed, they may be poor predictors when used in markets outside the original market. A customized score card reflects the specific relationship between the characteristic profile of the accounts used in the development sample and their behavior. If a bank uses the model in a market or geographic area in which the population is substantially different, the actual performance of the applicants may deviate from expectation. Therefore, a bank using a custom credit scoring model outside its intended market must be able to demonstrate the model's validity as soon as

a sufficient sample of applicants becomes available.

Use of two models together. Validated models can be successfully used independently, but when combined, their overall results may vary. Some banks use a two-dimensional strategy (using two scoring models to identify potential customers) which may distort the conclusions of their back-end analysis. A bank that chooses to modify the models must show the effect of those changes on performance and adjust the validation procedures. The actual data should be compared to development data to evaluate the validity of a two-dimensional strategy.

Example: a bank may pursue an account acquisition strategy that targets a group of quality applicants who are more likely to hold revolving balances. The bank would evaluate applicants using both a credit scoring model and a revolving balance scoring model. The latter model is designed specifically to identify those applicants most likely to hold large balances. Under that strategy, a bank would be likely to deny some good quality applicants with high credit scores, but low revolving-balance scores, because they are convenience users who pay accounts in full each month. In addition, banks would approve some applicants who have low credit scores, but high revolving-balance scores. To ensure proper monitoring, both the front-end and back-end analyses should be modified to reflect the joint distribution of the two score cards over credit and revolving-balance scores.

Use of Models for Decision Making

A bank may use the results of a credit scoring model as one of several factors on which credit decisions are based or as the sole (or primary) factor for credit decisions. If a bank bases its decisions on the model alone, it is particularly critical for bank management to regularly validate the model's performance. If a bank bases its credit decisions on human judgment as well as a credit scoring model, bank management should regularly confirm the validity of both the model and the judgmental bases for the credit decision.

Effect of Overrides

Bank management may choose to authorize overrides of the credit scoring model's results for the following reasons: bank policy, availability of additional information, and intuition. Policy overrides occur when management sets up special rules for certain kinds of applications. Informational overrides are based on information not available in the credit scoring model or only available in small quantities (e.g., applicant maintains large deposits at the bank or has sufficient collateral). Intuitive overrides are based on "gut feeling." Management should be able to identify the type and volume of overrides by category and measure the quality of their performances. OCC examiners will evaluate the reasonableness of the bank's override practices.

Level of Overrides. The manual that accompanies a credit scoring model bought from a vendor prescribes a reasonable level of overrides that minimize their effect on the model.

Banks that purchase a credit scoring model or develop their own should create a policy on the maximum override rates. Those levels should be monitored closely to determine if the models are missing factors identified as important determinants of behavior.

An override rate can be either excessive or reasonable from a business perspective. It depends on the product (auto loans, credit cards, mortgage, etc.); model purpose (the model is used as a tool in a judgmental decision process or as a decision making device within an automated underwriting system); or the stage in the transition from a judgmental to an automated underwriting system. The OCC is concerned that an excessive level of overrides negates the use of scoring models. If the scoring model properly reflects the bank's risk parameters, overrides should be used with considerable caution. A special tracking report should be generated for all overrides according to reason for override to determine if override decisions comply with policy and to monitor override performance.

OCC Definition of High-Side and Low-Side Overrides. Although vendors generally agree on the definitions of high-side and low-side override rates, the definitions used to evaluate the rates vary. Several definitions of overrides could be useful to management. However, for the determination of score card reliability, the OCC recommends that the following definitions be used to measure override rates: the high-side override rate is the number of declines with scores above the cut-off as a percentage of the number of applicants that score above cutoff; the low-side override rate is the number of approvals with scores below the cut-off as a percentage of the number of applicants that score below the cutoff.

Increase in Overrides. The OCC recognizes that business reasons may justify a temporary increase in overrides. For example, during the transition period to a new system (i.e., replacing a judgmental system with an automated system), override rates may exceed acceptable levels until a reasonable level of confidence in the new approach is achieved. Override analysis is an effective tool in demonstrating the strength of an automated system. However, overrides not demonstrated to be effective should be phased out as soon as possible.

Cutoff Bands. Some banks define a cutoff band instead of a cutoff score, under which an application is forwarded to an analyst for judgmental evaluation. On application scoring models, some banks define a cutoff band using scores from 190 to 210 for an additional judgmental review. Overrides are defined commonly as only those applicants that are denied with scores above the upper score of the cutoff band (i.e., high-side overrides), and applicants that are approved with scores below the lower score of the cutoff band (i.e., low-side overrides). Applicants in the gray area between the upper and lower cutoff scores are excluded from the override analysis. The OCC believes this practice is inappropriate and recommends that the upper cutoff score (in this example, the upper cutoff score is 210) should be used for override analysis. An evaluation of overrides in the gray area becomes important in documenting the effectiveness of the score card to assess the credit quality of marginal applicants.

Bank Actions to Limit Its Exposure to Losses

A bank may consider several approaches to limit losses from unanticipated changes in the performance of a credit scoring model, depending on the bank's ability to identify the most likely reason(s) for the shift in performance. A bank must have the staff capable of monitoring performance; analyzing the data to identify true shifts in behavior; and investigating shifts in performance from their expected levels. The easiest adjustment is to manage the cutoff score actively (by either raising or lowering the score) to maintain a targeted loss rate consistent with its profit objectives.

Second, a bank can adjust its underwriting policy to narrow the scope of its market to a targeted group believed to perform better than the population in general. This clearly assumes that the bank can identify such groups. Examples might include adjusting loan-to-value, debt-to-income or collateral requirements for select groups. This approach involves making specific, long-term changes in a bank's business strategy. For this reason, the approach, though useful for shifts in market trends, is limited as a short-term risk management tool.

Third, a bank could develop or purchase a credit scoring model based on more recent information about the current population. In this case, the bank must weigh the costs of developing or purchasing a model against that of carrying an increased number of bad accounts booked by an old model. The bank must also consider that a new model may take 12 to 24 months to develop and validate.

Fourth, a bank could do nothing. There may be sufficient evidence to show that the shift in behavior is small and likely to be of short duration. If that is the case, a shift in policy or changes to the model may not be warranted. However, if the changes are permanent, the bank will book more loans of poorer quality the longer it does not react. Since the strengths of a credit scoring model reside in its ability to better manage risk, the bank should be reasonably certain that a shift is temporary if it chooses to follow a "do nothing" policy.

COMPLIANCE ISSUES: ECOA (REGULATION B) AND THE FAIR HOUSING ACT

This section explains the principles the OCC will apply in addressing fair lending issues that arise in the use of credit scoring models. The principles are relevant whether the model is used to decide approval of an application for credit or to decide the terms (e.g., interest rate or credit limit) to be granted to a credit applicant, or both.

The Equal Credit Opportunity Act (ECOA) and its implementing regulation (Regulation B) are intended to promote the availability of credit to all creditworthy applicants without regard to race, color, religion, national origin, sex, marital status, age (provided the applicant has the capacity to contract), the receipt of public assistance, or the exercise in good faith of any right under the Consumer Credit Protection Act. The Fair Housing Act promotes the availability of residential real estate-related credit without regard to race, color, religion, sex, national origin,

handicap, or familial status. Accordingly, those laws prohibit practices that discriminate on any of those bases.

The Equal Credit Opportunity Act differentiates between credit decision systems of two categories: those that are "empirically derived, demonstrably and statistically sound," and those that are not. Credit decision systems that are not "empirically derived, demonstrably and statistically sound" are referred to as "judgmental systems." The regulatory definitions of "empirically derived, demonstrably and statistically sound" and "judgmental," as used in ECOA, may be found at 12 CFR 202.2 (p)(1), and 202.2 (t), respectively. To ensure that predictive ability is being maintained, creditors must review periodically the performance of the system. Several fair lending issues arise related to those models.

Overt Discrimination

Generally, in any system that evaluates creditworthiness, creditors may not consider explicitly any basis prohibited by the ECOA. Further, if the credit being sought is a residential real estate loan as defined by the Fair Housing Act, 42 USC 3605, creditors may not consider the additional bases of familial status and handicap. National banks can avoid overt discrimination by understanding the prohibited bases and ensuring that the credit scoring systems do not include them as predictive variables.

Regardless of whether a particular credit scoring system is validated according to the ECOA, and whether the validation is performed by a creditor or a vendor, a creditor cannot use a credit scoring system that assigns various points based on the applicant's race, national origin, or any other prohibited basis, subject to certain exceptions for age. Those exceptions are:

- In an empirically derived, demonstrably and statistically sound credit scoring system, a creditor may score an applicant's age as a predictive variable, provided that the age of an applicant who is 62 years of age or older is not assigned a less favorable factor or value. Therefore, an applicant who is 62 or older must be treated at least as favorably on the basis of age as anyone under the age of 62.
- In a judgmental credit scoring system, a creditor may consider the applicant's age only to determine a pertinent element of creditworthiness on an individualized case-by-case basis. For example, a creditor can consider an applicant's age to assess the significance of the applicant's length of employment.
- In any system used to evaluate creditworthiness, a creditor may consider the age of an applicant who is 62 years or older in order to favor the applicant in extending credit.

Credit Scoring Segmentation

A creditor may segment the population into score cards based on the age of an applicant. As noted in commentary issued by the Federal Reserve Board on Regulation B, such segmentation does not constitute scoring of age when one score card covers a narrow age range (for example, applicants in their twenties or younger) using attributes predictive for that age group, while a second score card covers all other applicants who are evaluated under the attributes predictive for that broader age group. As age is not scored, the system does not raise the issue of assigning a negative factor or value to the age of elderly applicants. (The system must be validated.)

The system scores age when it segments the population by age into multiple score cards and includes elderly applicants in a narrow age range. The OCC will scrutinize closely credit scoring systems that segment the population into more than two groups on the basis of age. National banks that use age in that fashion should ensure that the systems: have been validated according to Regulation B; and, either do not disadvantage applicants aged 62 or older, or disadvantage them only in a broad age range that is disadvantaged.

Segmenting the population by any other prohibited basis, regardless of whether the credit scoring system is validated, is illegal. Moreover, factors linked so closely to a prohibited basis that they may actually serve as proxies for that basis cannot be used to segment the population. For example, since those who speak Spanish as their primary language are likely to be of Hispanic national origin, a bank using a separate score card for persons who speak Spanish would be scrutinized by the OCC and may risk legal challenges by applicants.

Disparate Treatment

If credit scores are the sole basis for granting credit, the fact that two applicants have different scores means that they are not “similarly situated” in terms of their creditworthiness. There is no disparate treatment if applicants get different results. However, disparate treatment on a prohibited basis may occur for those applicants if: credit information is characterized subjectively before being entered for credit scoring; assistance is provided to improve one applicant's qualifications and not others; or overrides of the system are permitted.

Examiners will focus particularly on overrides. They may compare control group approvals involving overrides to denied prohibited-basis group applicants who scored approximately as well. They will evaluate whether the denied prohibited-basis applicants are as well qualified as those approved. Alternatively or additionally, they may compare prohibited-basis group denials involving overrides to approved control group applicants who scored approximately as well and evaluate whether those denied credit are as well qualified as those approved.

Illegal disparate treatment can be avoided by assuring that adequate controls exist during the pre-scoring, scoring, and post-scoring stages of the credit application process. Banks should

establish clear guidelines for:

- Providing similar assistance to all credit applicants.
- Overriding a credit score, including who may do so and under what circumstances.
- Training, monitoring, and supervising bank employees.

Disparate Impact

Disparate impact may occur in a credit scoring system when:

- A variable used in the credit scoring system is facially neutral (i.e., it does not discriminate on any prohibited basis overtly).
- That variable is applied evenly, without regard to any prohibited basis.
- That variable disproportionately adversely affects a segment of the population that shares a common characteristic that may not be considered legally.
- That variable cannot be justified by business necessity, or the business necessity can be achieved by substituting a comparably predictive variable that will allow the credit scoring system to continue to be validated, but also operate with a less discriminatory result.

All of those factors must be present to violate fair lending laws under disparate impact. For additional information on the disparate impact theory of proof under fair lending laws, see the joint agency Policy Statement on Discrimination in Lending, distributed as OCC Bulletin 94-30.

Bank regulatory agencies, law enforcement agencies, private organizations, and banks may evaluate the variables used in a validated credit scoring system to determine whether they have a disparate impact on any basis prohibited by the fair lending laws. However, the OCC will conclude a variable is justified by business necessity and does not warrant further scrutiny if the variable is statistically related to loan performance, and has an understandable relationship to an individual applicant's creditworthiness.

National banks should avoid including in their credit scoring systems variables that have little influence on the total credit score, yet disadvantage applicants on a prohibited basis to a statistically significant degree. The OCC will scrutinize closely a credit scoring system when a national bank substitutes a variable for one found to be predictive by a credit scoring vendor, or changes a system that a vendor has based on a demonstrably and statistically sound analysis of empirical data (by, for example, substituting a different variable or adjusting its weight).