DISCUSSION PAPER
PAYMENT CARDS CENTER

Fair Lending Analysis of Credit Cards

David Skanderson* and Dubravka Ritter**

August 2014

Summary:

This paper discusses some of the key fair lending risks that can arise in various stages of the marketing, acquisition, and management of credit card accounts, and the analysis that can be employed to manage such risks. The Equal Credit Opportunity Act (ECOA) and its implementing Regulation B prohibit discrimination in all aspects of credit transactions and include specific provisions relating to processes that employ credit scoring models. This paper discusses some of the areas of credit card operations that may be assessed in an effort to manage the risk of noncompliance with fair lending laws and regulations. Particular attention is focused on approaches to testing for the risk of disparate impact on a prohibited basis in credit scoring models and model-intensive prescreened marketing campaigns, as well as in judgmental credit card underwriting. The paper concludes by discussing how the fair lending risks associated with credit scoring models may be managed by synchronizing compliance oversight with an institution's model governance framework. The methods discussed in this paper are also applicable to other consumer credit products that utilize credit scoring models.

Keywords: ECOA, Regulation B, discrimination, fair lending, consumer lending, disparate treatment, disparate impact, credit card, scoring model, model governance

JEL Classification Numbers: G21, G28, K23

* Charles River Associates, 1201 F St. NW, Suite 700, Washington, D.C. 20004; e-mail: dskanderson@crai.com.
** Payment Cards Center, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106; e-mail: dubravka.ritter@phil.frb.org.

I. Introduction

The Equal Credit Opportunity Act (ECOA) and its implementing Regulation B prohibit lenders from discriminating on the basis of certain personal characteristics (or "prohibited bases") in any aspect of a credit transaction. We use the term "fair lending risk" in this paper to refer to the risk that "similarly situated" consumers who differ only in terms of a prohibited basis characteristic will receive different treatment or outcomes in some aspect of a credit transaction. Such differences can result in injury to consumers and in both liability and reputational damage for a financial institution.

Historically, mortgage lending has received the greatest attention in fair lending compliance testing and enforcement. This is due in part to the availability of data regarding the race, ethnicity, and sex of mortgage applicants.[1] However, the ECOA covers all types of consumer credit products and transactions. The Consumer Financial Protection Bureau (CFPB) has spurred a renewed emphasis on examining ECOA compliance in nonmortgage consumer credit products, using "proxies" or "surrogates" for demographic characteristics as the basis for analysis.[2] Among other things, the CFPB has clarified its expectation that consumer lenders will maintain effective "compliance management systems," including systems for detecting, monitoring, and controlling the fair lending risk of credit card operations.[3]

[1] Mortgage lenders report this information pursuant to the Home Mortgage Disclosure Act (12 U.S.C. §2801, et seq.) as implemented by Regulation C (12 C.F.R. §203).

[2] As discussed more fully in Section IV, "proxies" or "surrogates" refer to factors such as name or address information used to infer a consumer's likely race, ethnicity, or sex.

[3] See "CFPB Supervision and Examination Manual," version 2, October 2012, Part II, at http://files.consumerfinance.gov/f/201210_cfpb_supervision-and-examination-manual-v2.pdf (accessed March 16, 2014).

This paper discusses some of the key aspects of credit card operations in which fair lending risks can arise, as well as analysis approaches that may be employed to understand and manage these risks. There are many similarities between mortgage and credit card lending in terms of how fair lending compliance is evaluated. However, approaches for evaluating fair lending risk in credit card lending need to account for its more intensive reliance on automated scoring models and decision processes, the high-volume nature of the business, and its intensive use of direct marketing.

Scoring models and automated decision systems are employed in various ways and to various degrees across credit card issuers. Some issuers rely on generic credit scores obtained from credit bureaus (such as the FICO score) while others develop custom scoring models to better predict the risk of their specific customer base or products. Some issuers use entirely or almost entirely automated decision processes based on a score and various discrete decision rules. Others rely on automated decision systems only to approve the upper tail of the credit distribution and to decline the lower tail of the distribution, with judgmental manual underwriting playing a major role for marginal decisions.

Lenders can use automated decision systems to limit the potential for similarly situated credit applicants to be treated differently on a legally prohibited basis, whether deliberately or inadvertently. However, these automated systems can be a source of fair lending risk if they are not appropriately constructed, tested, and monitored. Assessing, quantifying, and weighing the fair lending risk of such systems is a complex technical endeavor.

In this paper, we describe a number of approaches for evaluating credit card lending processes for fair lending risk. The information in this paper may be helpful to financial institutions in connection with their fair lending programs. The paper may also be of interest to other parties concerned with equal access to credit or with the fair lending issues that can arise in credit card lending.

Section II discusses certain aspects of the ECOA and Regulation B that may apply to analyzing heavily model-oriented credit card processes.[4] Section III identifies some of the key areas of credit card operations that may be assessed for fair lending risk. Section IV discusses analytical methods that may be used to assess fair lending risk in key credit card operation areas, including prescreened marketing, underwriting, pricing, and credit line assignment. Section V discusses the evaluation of credit scoring models for fair lending risk. Section VI discusses how the management of fair lending compliance risks related to scoring models may be synchronized with an institution's model governance and model risk management framework in order to manage fair lending risk efficiently.

It is important that we remind the reader that the authors are not attorneys, and the views expressed in this paper do not constitute legal opinions or regulatory guidance.

[4] The term "model" is used here to refer generally to "a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates" (OCC Bulletin 2011-12). In this paper, we specifically refer to models meeting this definition that are used in the marketing, origination, or servicing of consumer credit card accounts, usually expressed in terms of some sort of score for rank-ordering consumers in terms of risk or some other behavioral propensity.

II. The ECOA and Regulation B

The ECOA makes it unlawful for "any creditor to discriminate against any applicant with respect to any aspect of a credit transaction" based on prohibited factors: race, color, religion, national origin (often referred to as "ethnicity"), sex, marital status, age (provided the applicant has the capacity to contract), the applicant's receipt of income from any public assistance program, or the applicant's exercise in good faith of rights under the Consumer Credit Protection Act.[5,6] The broad reach of the ECOA and Regulation B potentially extends to marketing, advertising, and solicitation practices; account origination; the setting of terms and conditions; and servicing practices.[7] In other words, nearly every aspect of credit card operations that directly affects actual or prospective customers has potential fair lending compliance implications.

A. Theories of Liability Under the ECOA

Three general theories of liability for discrimination may apply under the ECOA: overt discrimination, disparate treatment, and disparate impact.[8]

Overt discrimination refers to intentional discrimination on the basis of a legally prohibited factor. Such discrimination occurs when a creditor openly discriminates or uses a prohibited factor explicitly in a scoring model or decision rule.

---

[5] Equal Credit Opportunity Act: Title VII of the Consumer Credit Protection Act, as amended (15 U.S.C. §1601, et seq.). Board of Governors of the Federal Reserve System. *Equal Credit Opportunity Act, Regulation B* (12 C.F.R. §202), July 15, 2011, effective August 15, 2011, found at http://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&tpl=/ecfrbrowse/Title12/12cfr1002_main_02.tpl (accessed August 30, 2013).

[6] For a more detailed discussion of the ECOA and Regulation B, see Ritter (2012).

[7] See Regulation B, 12 C.F.R. §202, Official Staff Interpretations to §202.4(a).

[8] See CFPB Bulletin 2012-04 (Fair Lending), April 18, 2012.

Disparate treatment occurs when a creditor treats an applicant differently based on a prohibited factor. A conclusion of disparate treatment generally arises through comparing the treatment of similarly situated applicants. It is not necessary that the difference in treatment be motivated by prejudice or an intention to discriminate beyond the difference in treatment itself. Such a difference in treatment on a prohibited basis is considered to be intentional discrimination if no credible, nondiscriminatory reason explains the difference.[9] For example, disparate treatment risk can arise if similarly situated consumers of different races, ethnicities, or sexes receive different levels of consideration, service, or encouragement to apply for credit; or if the inconsistent application of judgmental underwriting criteria tends to disadvantage one group over another on a prohibited basis.

Disparate impact refers to situations in which a creditor practice (e.g., a policy, decision rule, or model) is neutral on its face but nevertheless has a disproportionately adverse impact on the basis of a prohibited factor *in effect*, even though the creditor has no intent to discriminate. A disparate impact may constitute illegal discrimination unless the practice in question meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact.[10]

[9] See Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18266, 18269 (joint statement by the Department of Housing and Urban Development, the Federal Trade Commission, and the Department of Justice). Also see Regulation B, 12 C.F.R. §202, Official Staff Interpretations to §202.4(a).

[10] The applicability of the disparate impact theory under the ECOA is not universally agreed upon. A footnote in Regulation B (12 C.F.R. §202.6, footnote 2) asserts that Congress intended for disparate impact liability to be available. The Department of Justice, the CFPB, the Federal Deposit Insurance Corporation, and other agencies have all pursued enforcement actions or have reached settlements with creditors in cases that included allegations of disparate impact under the ECOA. In addition, the CFPB announced its intention to continue employing a disparate impact "effects test" under the ECOA in its fair lending examinations and enforcement actions (CFPB Bulletin 2012-04, April 18, 2012). However, some have argued that the text of the statute does not permit disparate impact liability claims. See, for example, Cubita and Hartmann (2006).

Under the disparate impact legal theory of discrimination, the existence of a disproportionate adverse impact on a prohibited basis does not, by itself, mean that illegal discrimination has occurred. In *Wards Cove Packing Co. v. Atonio* (an employment discrimination case), the U.S. Supreme Court outlined a three-step process for assessing disparate impact and for assigning the burden of proof in such cases.[11] The first step requires the party bringing the claim to present evidence of a "substantial disparate impact" resulting from an identifiable policy or practice. If that burden is met, the second step shifts the burden of proof and requires the defendant (as it applies to fair lending, the creditor) to explain the legitimate business justification for the credit policy or practice that led to the disparate impact.[12] The third step shifts the burden back to the party, bringing the claim to show that there is an equally effective but less discriminatory option available to meet the credit issuer's legitimate business objective. If the third step is reached and an equally effective but less discriminatory alternative is identified, there could be a finding of illegal discrimination.[13]

Statistical analysis usually plays a central role in evaluating disparate impact claims, and analysis approaches used for this purpose can be constructed to account for

[11] 490 U.S. 642 (1989)

[12] Such a justification must be based on substantial evidence, and the burden remains with the party bringing the claim to refute the business justification that is proffered by the creditor. As stated in the *Wards Cove* decision, "A mere insubstantial justification in this regard will not suffice, because such a low standard of review would permit discrimination to be practiced through the use of spurious, seemingly neutral employment practices. At the same time, though, there is no requirement that the challenged practice be 'essential' or 'indispensable' to the employer's business for it to pass muster: this degree of scrutiny would be almost impossible for most employers to meet. … In this phase, the employer carries the burden of producing evidence of a business justification for his employment practice. The burden of persuasion, however, remains with the disparate impact plaintiff." — *Wards Cove*, 490 U.S. 642 at 659-660.

[13] "If respondents, having established a prima facie case, come forward with alternatives to petitioners' hiring practices that reduce the racially disparate impact of practices currently being used, and petitioners refuse to adopt these alternatives, such a refusal would belie a claim by petitioners that their incumbent practices are being employed for nondiscriminatory reasons." — *Wards Cove*, 490 U.S. 642 at 660-661.

the established legal standards for evaluating evidence of discrimination. For example, various economic characteristics that are correlated with creditworthiness (e.g., income, wealth, property ownership, education, and employment) may also be correlated with race, ethnicity, sex, or age. Therefore, the use of such factors in credit decisions may have an unequal impact on minorities compared with whites, women compared with men, or one age group compared with another. However, the legal test for disparate impact recognizes the possibility of such differences among demographic groups. Specifically, in order for illegal disparate impact to have occurred, it is not sufficient merely to establish that a credit policy or practice has a disproportionate adverse impact on members of a protected class *in general*. Rather, one must establish that the policy or practice in question has a disproportionate adverse impact on *qualified* members of a protected class and that the policy or practice lacks a business justification.[14] In Sections IV and V of this paper, we discuss the implications of this legal standard for the design of statistical methods to evaluate fair lending risk.

B. The "EDDSS" Standard

Regulation B groups credit scoring systems into two general types: "empirically derived, demonstrably and statistically sound" (EDDSS) credit scoring systems and "judgmental" systems. The distinction between an EDDSS scoring system and a judgmental scoring system is important. Creditors that use an EDDSS scoring system may take applicant age directly into account as a predictive variable (provided that

[14] The *Wards Cove* decision established that liability for disparate impact can only arise if *"otherwise-qualified"* members of a protected class of consumers receive less favorable credit outcomes than "*qualified"* members of another class. — *Wards Cove*, 490 U.S. 642, at 651-655.

elderly consumers are scored at least as favorably as younger consumers).[15] Judgmental systems may take age into account only to determine minimum legal requirements for a credit obligation or to treat elderly applicants more favorably than younger applicants. However, only an EDDSS scoring system may specifically score differences in credit risk that may be related to a consumer's age or use different sets of predictive variables for different age groups.

In addition, scoring systems that can be demonstrated to meet the EDDSS standard may result in a lower risk of fair lending compliance issues because rigorously derived scoring systems are more likely to have a demonstrable business justification.

A credit scoring system must satisfy all of the following criteria in order to be classified as EDDSS:

- *Empirical:* Based on a statistical analysis that is rigorous and derives empirical ways to distinguish between more and less creditworthy consumers, using data for applicants who applied for credit within a reasonable preceding period of time
- *Business justified:* Developed for the purpose of evaluating the creditworthiness of applicants with respect to a specific, legitimate business purpose of the creditor; and directly related to a legitimate business objective or necessity, such as (but not limited to) maximizing profit, limiting the risk of default, or limiting loss exposure in the event of a default
- *Statistically valid:* Developed and validated based on generally accepted statistical practices and methodologies

[15] "Elderly" consumers are defined under the ECOA as age 62 or older.

- *Periodically revalidated:* Reevaluated for statistical soundness from time to time and adjusted as necessary, using appropriate statistical methods and the creditor's own data, to maintain predictive ability

Forms of credit analysis that do not meet these standards are considered to be judgmental systems for the purposes of fair lending analysis. Even a statistically based credit scoring model might still be considered judgmental if it cannot be shown to meet all of the standards for an EDDSS scoring system. For example, models can deteriorate over time and lose predictive power as customer populations, credit policies, and economic conditions change. If a model drifts too far from its original predictive power, the foundation for its statistical validity and business justification may erode, and fair lending compliance risks may arise.

III. Focus Areas for Fair Lending Risk Assessment

Fair lending concerns can arise in several areas of credit card operations, and this section outlines the main areas of interest in this regard. A qualitative risk assessment and some amount of quantitative analysis (where possible) may be performed in each of the following areas to assess fair lending risk exposure.[16]

Marketing: A fair lending risk review may consider whether any facet of credit card marketing, advertisement, or solicitation tends to exclude, avoid, or discourage actual or potential applicants on a prohibited basis. Aspects of marketing to evaluate for fair lending risk may include:

[16] For in-depth guidance on conducting a comprehensive fair lending risk assessment in each of these areas, see the American Bankers Association's "Toolbox on Fair Lending" (2012) or the CFPB's "Interagency Fair Lending Examination Procedures" (2012).

- advertisements;
- marketing collateral or copy;
- media, languages, and images used;
- website presence;
- geographic focus;
- prescreened solicitations and credit offers; and
- statistical models used in marketing processes.

Underwriting: A fair lending risk review may consider whether the credit decision process results in different credit outcomes on a prohibited basis for similarly situated applicants, either because of specific evaluation criteria or due to inconsistency in the application of otherwise neutral criteria. Areas to evaluate for potential fair lending risk in underwriting may include:

- the content, specificity, and objectivity of underwriting policies, procedures, or guidelines (including whether any policies or guidelines may be viewed as potentially discriminatory);
- the construction and use of risk scoring models;
- automated underwriting systems or decision rules;
- judgmental or manual aspects of underwriting or processing (particularly the degree of consistency in applying policies and guidelines across applicants);
- the handling of escalations, reconsideration requests, overrides, and exceptions; and
- the adequacy of data and documentation to support the reasons for credit decisions.

Credit Line Assignment: A fair lending risk review may consider whether similarly situated applicants tend to receive different credit lines on a prohibited basis. Fair lending risk in credit line assignment that may arise at any of these stages:

- initial credit line assignments;
- increases or decreases in credit lines (whether based on account review or customer requests);
- models or decision rules used in the credit line assignment process; and
- judgmental credit line assignments or manual overrides to automated line assignments.

Pricing: A fair lending risk review may consider whether similarly situated applicants tend to receive different pricing, or other terms or conditions, on a prohibited basis. Areas to evaluate for potential fair lending risk in pricing may include:

- initial interest rates offered or assigned;
- the availability of promotional offers (e.g., introductory rates or balance transfer offers);
- postorigination changes in terms (including penalty/default interest rates);
- models or decision rules used in the pricing process; and
- judgmental overrides to automated pricing.

Account Servicing and Collection: A fair lending risk review may consider whether similarly situated applicants receive different levels of service or consideration on a prohibited basis. Areas to evaluate for potential fair lending risk in servicing or collection may include:

- whether there are any differences in the level or quality of customer service between English and foreign-language personnel or call centers or based on geography (which may correlate with a prohibited basis);
- treatment or level of service based on relationship status (membership level, profitability) or other criteria that may correlate with a prohibited basis; and
- collections processes or strategies, including payment plans, referrals for risk mitigation or collection, debt forgiveness, the application or waiver of fees or penalties, the availability of special customer assistance programs, and any models or automated systems used to determine collection or workout strategies.

Secured Cards: A fair lending risk review may consider whether consumers are targeted for or steered toward a secured card on a prohibited basis (including consideration of where and how secured cards are marketed compared with unsecured cards). It may also consider whether a "graduation" program exists that allows consumers who demonstrate creditworthiness to upgrade to a more favorable credit product appropriate for them.

Affinity Partners: A fair lending risk review may consider whether the selection of affinity partners has the effect of skewing overall account acquisitions in a way that correlates with a prohibited basis; it may also determine whether groups with a particular racial, ethnic, religious, age, or sex affinity tend to receive more or less favorable terms.

IV. Fair Lending Analysis Methods for Credit Cards

After performing a qualitative evaluation of the key areas of fair lending risk exposure in credit card operations, creditors may apply appropriate statistical methods for identifying and quantifying the fair lending risk. In this section, we discuss some quantitative analytical methods that may be used to examine fair lending compliance risk in prescreened marketing, underwriting, pricing, and credit line assignment.

A. Developing Proxies for Demographic Characteristics

Creditors typically capture the age or date of birth of credit applicants in their data because of the need to verify an applicant's identity and legal capacity to enter a contract or to meet other legal or regulatory requirements. However, credit card lenders usually do not have information about the race, ethnicity, or sex of credit card applicants. In order to perform statistical testing for fair lending risk based on race, ethnicity, or sex, regulators and creditors must resort to the use of statistical "proxies" or "surrogates" for these characteristics.

No formal regulatory guidance yet exists for developing or using proxies for demographic characteristics. Regulators have historically used either surname proxies (particularly for Hispanics) or geographic proxies (particularly for African Americans), or a combination of the two, to derive an estimated probability that a consumer belongs to a particular race or ethnicity group.[17] Regulators typically have used first names to proxy for sex.[18] The details of race, ethnicity, and sex proxy methods are beyond the scope of this paper, but we will assume that proxies are assigned through some means for purposes of our discussion of analysis methods.

Proxies are inherently subject to error in group assignment, including the risk of both false positives (e.g., incorrectly assigning nonminority consumers to a minority group) and false negatives (e.g., incorrectly assigning minority consumers to the nonminority group). Also, a nontrivial proportion of any given sample of consumer records might not be assigned to any specific race/ethnicity or sex group. For example, consumers living in areas that are racially and ethnically diverse or who have surnames that are not strongly associated with a specific race or ethnicity, typically are classified as "indeterminate" and are excluded from fair lending analysis. These limitations suggest that results derived from a proxy-based analysis should be treated with an appropriate degree of caution. Nevertheless, as long as the proxies are strongly correlated with consumers' actual characteristics, the analysis can be informative about potential compliance risk exposure.

[17] The CFPB has indicated that it uses a combination of surnames and geography in its proxy method. See "Preventing Illegal Discrimination in Auto Lending" (2013). Geographic proxies are derived from U.S. Bureau of the Census data regarding the race/ethnicity composition of detailed geographic areas (e.g., census tracts, block groups, blocks, or zip code areas). See the 2010 Census SF1 Summary File at http://www2.census.gov/census_2010/04-Summary_File_1/. Surname proxies are derived from Census data regarding the association between surnames and self-reported race and ethnicity. See "Genealogy Data: Frequently Occurring Surnames from Census 2000," http://www.census.gov/genealogy/www/data/2000surnames/index.html.

[18] Sex proxies have been based on Social Security Administration data or other data sources. See, for example, "Popular Baby Names," Social Security Administration, http://www.ssa.gov/oact/babynames/limits.html.

B. Prescreened Marketing[19]

A prescreened marketing campaign typically begins with the selection, from a broad credit bureau universe, of a group of consumers who will receive an offer of credit or an invitation to apply for credit. The creditor will generally utilize a set of screening criteria (often referred to as "suppression" criteria), decision rules, and/or statistical models that restrict the universe of consumers to a more targeted population of consumers who are most likely to (a) respond to an offer and (b) meet the issuer's credit risk standards and profitability targets. Prescreened mailing processes are usually highly automated — driven by objective criteria and models — and potentially affect a large volume of consumers in a repeated fashion. Though the standardization and automation of the process (if done well) may mitigate fair lending risk, the fact that so many consumers may be affected by the process implies that any fair lending issues that do arise can have a broad consumer impact.

Because of the automated nature of the prescreening process and because the decision rules utilized are typically facially neutral, the focus of fair lending analysis is typically on the risk of disparate impact. The overarching objective of the fair lending analysis is to determine whether the impact of the *process as a whole* tends to exclude

[19] The discussion in this section is based on the assumption that prescreened solicitation processes may give rise to fair lending compliance risk under the ECOA. However, there is some potential ambiguity in terms of whether and how the ECOA applies to prescreened marketing. According to the Official Staff Interpretation to §202.4(b) of Regulation B, "the regulation's protections apply only to persons who have requested or received an extension of credit." In addition, the Interagency Fair Lending Examination Procedures (Appendix Section VII.C.) states, "Pre-screened solicitation of potential applicants on a prohibited basis does not violate ECOA." However, the Official Staff Interpretation to §202.4(b) also states, "In keeping with the purpose of the Act — to promote the availability of credit on a nondiscriminatory basis — §202.4(b) covers acts or practices directed at prospective applicants that could discourage a reasonable person, on a prohibited basis, from applying for credit." We discuss potential fair lending risk associated with prescreened solicitations in this paper because the latter passage suggests the possibility that discriminatory solicitation factors could be viewed as evidence of unlawful discouragement under Regulation B.

certain groups from credit offers disproportionally, though it is also important to consider whether individual components of the process may add fair lending risk. For example, some individual suppression criteria or models used in the prescreening process may have disproportionate adverse impact on a prohibited basis, but other criteria or scoring models may have compensating favorable effects, resulting in little or no disparate impact of the overall process. We discuss this complexity in more detail in Section V.

Evaluating a prescreening process for disparate impact risk involves comparing the demographic distribution of the population selected to receive a prescreened mailing ("targeted" or "mailed-to" population), after all criteria and models have been applied, with a relevant "baseline" or "benchmark" population. If the demographic distribution of the consumers who are selected for an offer is materially different from that of the baseline population, disparate impact risk may be indicated.

One way to define the baseline population is as the set of applicants who meet at least the minimum eligibility criteria for the credit card product in question — that is, the initial credit bureau population less consumers who do not meet the creditor's most basic criteria for eligibility. The motivation for defining the baseline population in this way derives from the standards of evidence described in the *Wards Cove* decision, discussed previously, under which the relevant benchmark for establishing a disparate impact is the population of "otherwise-qualified" applicants.

For example, the population of all individuals with a credit bureau record typically would not form an appropriate baseline comparison group for the purposes of a disparate impact analysis of prescreening. Instead, the *Wards Cove* standard suggests that the relevant baseline population may comprise individuals who actually would be potential borrowers for the particular creditor and type of credit in question. The *Wards Cove* standard suggests that, for example, a creditor may not be held liable for discrimination solely because minority consumers were underrepresented among its credit card applicants, but the creditor could be held liable if minorities were underrepresented specifically because of a policy or practice that tended to exclude qualified minority applicants from access to credit or deterred such applicants from applying for credit.

Defining the appropriate baseline population typically requires analyzing the series of screening rules that are applied to an overall credit bureau population to arrive at a set of candidates for a credit solicitation. Some subsets of the credit bureau universe arguably do not belong in the baseline population, but some degree of judgment is required in deciding exactly where to draw the line.

Consumers who would have no chance of qualifying for a credit offer based on fundamental legal or credit policy criteria of the lender arguably do not belong in the baseline population. Criteria that filter out such consumers may include, for example, basic noncredit suppressions that exclude consumers who are not of legal age to enter a contract, are deceased, are outside the creditor's defined market area, do not have a valid mailing address, do not have a valid Social Security number, have no credit history or otherwise cannot be scored, have requested not to receive solicitations, or are already cardholders of the issuer. Such basic criteria do not appear to be subject to concern from a fair lending perspective.

Other suppressions that exclude consumers who fail to meet the lender's minimum credit qualifying criteria (e.g., by having recent derogatory credit history) arguably also could be excluded from the baseline population. The use of such minimum credit criteria in prescreening may not give rise to a heightened fair lending risk concern, provided the criteria in question are demonstrably related to business objectives and are consistently applied for the card product in question. However, if such criteria are potentially subject to override in a manual underwriting process, they arguably should not be applied in defining the baseline population. The potential impact of such criteria can then be included in evaluating the overall impact of the prescreening process.

Next, the prescreening process will typically define a series of credit and business criteria that further refine the target population. This may include criteria and models that attempt to predict the likelihood that a consumer will respond to an offer ("response models") or will default (a credit bureau score or custom credit score); criteria or models that attempt to predict the profitability of the customer relationship; and profit and cost objectives specific to a given marketing effort. Predictors of profitability may include criteria or scoring algorithms that attempt to predict whether a consumer is likely to carry a revolving credit balance (a "revolver") or to use the card mainly as a transaction medium and pay off the balance monthly (a "transactor"). Criteria or scoring algorithms may also attempt to predict whether the consumer is likely to be a "balance surfer," transferring balances among cards to take advantage of promotional rates. Such screening criteria can be thought of broadly as ways to prioritize the population of qualified or potentially qualified consumers to arrive at a set of targets for a given prescreened marketing campaign and are typically the main focus of disparate impact testing.

Overall, a well-developed program for assessing the disparate impact risk of prescreening processes may include some of the following steps:

- Qualitatively review the criteria used in the prescreening process and all factors used in scoring models or algorithms that enter into the prescreening process.
- Quantitatively evaluate the disparate impact risk of any models used in the process (e.g., credit scoring and response models), as discussed further in Section V.
- Define a baseline population that forms a valid basis for assessing the impact of the screening process on the demographic distribution of targeted consumers.
- Select a sample of targeted consumers for comparison (i.e., the group of consumers to which the creditor intends to mail, or has mailed, an offer in a representative marketing campaign).
- Apply proxy methods to impute demographic characteristics for the baseline and mailed-to populations.
- Compare the demographic distributions of the baseline and mailed-to populations.
- If significant differences in demographic distributions between the two populations exist, perform additional analysis to identify the source or sources of potential disparate impact within the prescreening process. This can be accomplished by evaluating the effects of individual screening criteria on the demographic distribution. As necessary, evaluate whether criteria giving rise to disparate impact risk have sufficient and demonstrable business justifications.

- Monitor and retest the screening process over time, focusing on changes to the process and any criteria that have been added or changed.

C. Underwriting

A fair lending assessment of underwriting may include qualitative review of underwriting criteria, quantitative disparate *impact* testing of scorecard model(s) and decision criteria, and analysis of potential disparate *treatment* in judgmental underwriting (as applicable).

In evaluating fair lending compliance risk in underwriting, it is important to distinguish between automated and manual decisions. Purely automated decision processes, if properly designed, involve little risk of disparate *treatment* because no human discretion or judgment is exercised in decision-making. Including both manual and automated decisions in a single underwriting analysis may lead to incorrect conclusions regarding the extent of fair lending disparate treatment risk, especially when automated decisions dominate the sample (as is often the case for credit card portfolios). Therefore, an analysis of potential disparate treatment risk in underwriting tends to focus on applications that were manually reviewed and to exclude applications that received a completely automated decision.[20]

[20] An analyst might also consider excluding from a disparate treatment analysis any applications that were declined based on nondiscretionary, noncredit criteria, such as the applicant not being of legal age to contract, confirmed fraud, an expired direct mail solicitation, or unverifiable information. Prior to making such exclusions, it is important to establish that applications of these sorts are consistently declined, without exception, and are not subject to underwriting discretion. The remaining records represent those applications where human judgment may have been used and, thus, where a potential for disparate treatment could arise.

Models and other automated decision criteria may be reviewed separately for disparate *impact* risk (as discussed in Section V). The remainder of this section describes two common methods for evaluating the potential disparate treatment risk arising from judgmental processes: statistical analysis and comparative file review.

1) Statistical Analysis of Differences in Application Denial Rates

Applications where judgment may have been used can be analyzed statistically by focusing on potential differences in denial rates on a prohibited basis. In other words, after accounting for legitimate underwriting factors, were members of a protected class denied at a higher rate than members of another class? Such a test typically involves a statistical (logistic or logit) model of approval/denial decisions that predicts the likelihood of an application's approval. The statistical model is structured to reflect the creditor's documented underwriting policies and guidelines. The exact logistic model specification will depend on the thresholds (or cutoffs) applicable to each credit criterion in the creditor's policies and guidelines. Ideally, it would take into account any interactions among underwriting criteria and any differences in criteria used for applications from different acquisition channels. In addition, separate models are typically applied to evaluate different population segments for which the lender applies significantly different underwriting policies or guidelines (e.g., based on product type or consumer segment).

Using a sample of approved and denied credit applications, a statistical model can be used to test for correlation between the underwriting outcome and prohibited factors, after accounting for objective credit factors used in the lender's underwriting process. This can be accomplished by including as an explanatory variable in the model an indicator (dummy variable) for demographic group membership (e.g., race/ethnicity or sex). The estimated coefficient on a demographic group indicator will measure the estimated marginal effect of group membership on the likelihood of being denied credit, after controlling for the effects of various objective underwriting criteria. If the estimated coefficient on the demographic group indicator is statistically significant (statistically different from zero), then members of that group were denied credit at a different rate than the comparison group, even after controlling for differences in credit qualifications. A positive sign on such a statistically significant coefficient would indicate that the statistical results are consistent with the presence of disparate treatment.[21]

However, a statistically significant difference in denial rates, in itself, is not sufficient for concluding that there actually is a disparate treatment issue because the model may not include all of the objective factors considered in underwriting and may not accurately represent the inherent "functional form" of the underwriting process. That is, there is potential for estimation bias due to omitted variables and/or model misspecification.

In particular, judgmental underwriting often involves evaluating the tradeoffs between indicators of elevated credit risk and "compensating" factors, and/or the consideration of "layered risk" factors. For example, if an applicant had a bankruptcy at some point in the past, the judgmental credit evaluation may consider whether he or she had reestablished an acceptable credit record since the bankruptcy occurred. Similarly, an

[21] This assumes that the dependent variable is defined as a binomial variable with a value of 1 representing an unfavorable outcome (e.g., denial) and a value of 0 representing a favorable outcome (e.g., approval).

applicant with a record of minor delinquencies may be viewed differently depending upon the level of their debt obligations in relation to their income. A logistic model cannot easily account for all these potential complexities. Thus, a statistically significant coefficient on an indicator for race, sex, or other prohibited basis in an underwriting logistic model likely signals the need for additional analysis to evaluate whether there truly is evidence of disparate treatment on a prohibited basis.

2) Comparative File Review

A standard practice in fair lending reviews is to investigate any statistically significant differences in denial rates on a prohibited basis by conducting a comparative file review. This review involves matching applicants who appear similarly qualified but experienced different outcomes (e.g., minority applicants who were denied compared with nonminority applicants who appear similar in terms of all measurable factors but were approved). Such a review attempts to reveal why a difference in outcomes occurred for apparently similar individuals and whether that difference is attributable to inconsistent treatment or, instead, to the consistent application of objective underwriting criteria.[22]

From the perspective of identifying potential fair lending issues, the available data and statistical evidence can be used to "risk-base" the file review sample more efficiently (i.e., target the sample to applicants for whom the actual credit decision appears to be inconsistent with the available data on credit qualifications). This can be done by using the predictions of a logistic model such as the one discussed in the

[22] For more information on comparative file reviews of this sort, see "Interagency Fair Lending Examination Procedures," (2012), Part III.C.

preceding section — the predicted probability of denial for each applicant conditional on their observed credit characteristics. For example, a comparative file review focusing on differences in outcomes related to race or ethnicity may examine the files of denied minority applicants (“target applicants”) for whom the model’s predicted probability of denial is relatively low. These files could be matched to files of approved nonminority applicants (“comparators”) who appear to be similar in terms of relevant credit characteristics. Conversely, the review may focus on approved nonminority applicants for whom the model’s predicted probability of denial was high. These files could be compared with the files of minority applicants with similar credit characteristics but whose applications were denied.[23]

The file examination may focus on identifying factors that led to the difference in underwriting decisions between apparently similar applicants for which the statistical model could not account. Such factors may include, for example, credit attributes not available in data form, data errors, or the judgmental evaluation of compensating factors and layered risk. Alternatively, the file examination may find that statistically significant differences in rates of denial are due to inconsistency in the exercise of judgment by underwriters or to discriminatory treatment.

If there are particular types of overrides or exceptions to automated underwriting decisions that also occur with differing frequency between the demographic groups of interest, then these may also be used as a basis for focusing the file

[23] This process of identifying applicants with a high predicted probability of denial who were approved, and applicants with a low predicted probability of denial who were denied, is a common method for identifying potentially “marginal” applicants, as that term is used in the “Interagency Fair Lending Examination Procedures,” Part III.C.

examination. For example, if the logistic model analysis found that a) Hispanic applicants were more likely to be denied than non-Hispanic white applicants, other things being equal, and b) analysis of override data found that overrides to the creditor's minimum credit score requirement occurred more frequently for non-Hispanic whites than for Hispanics, then this difference in overrides could be one of the reasons for the overall difference in denial rates. However, such a difference in overrides is not necessarily a fair lending issue because it may be the case that the non-Hispanic white applicants in the sample possessed compensating factors for low credit scores (e.g., stable employment history, low debt utilization, or long credit histories) more often than Hispanic applicants. To evaluate whether the difference in overrides may actually present a fair lending issue, a comparative file review may focus on non-Hispanic white applicants who received a credit score override and were approved compared with Hispanic applicants who had similar credit scores, and appeared to be similar in other relevant respects, but were nevertheless denied.

A review of loan files may include an evaluation of whether the file documentation (including underwriter notes regarding the reasons for their decisions) is sufficient to explain the reasons for each manual credit decision. A lack of sufficient documentation can increase fair lending compliance risk because, after the fact, the lack of documentation may make it difficult or impossible to explain the basis of the underwriting decisions or why apparently similar applicants were treated differently.

D. Pricing of New Credit

The assignment of purchase and cash advance interest rates to newly originated credit card accounts is typically automated, highly standardized, nondiscretionary, and based on measurable risk and/or profitability criteria. Any divergence from standard pricing is typically due to special promotional offers and direct marketing test offers, rather than discretion. If it can be confirmed that pricing is fully automated, with no scope for exceptions, then there is little disparate treatment risk. Nevertheless, automated pricing criteria may be reviewed for potential risk of disparate impact (see Section V). If exceptions or overrides to standard pricing are allowed, the frequency and size of those exceptions may be reviewed for potential differences on a prohibited basis. Manual postorigination changes in terms may be evaluated in a similar manner, including whether cardholder requests for interest rate reductions are approved and denied on a consistent basis.

E. Credit Line Assignment

Much like underwriting, credit line assignment is typically driven by automated decision rules based on specified risk factors, though scope for judgmental adjustment to system-assigned credit lines sometimes exists. A fair lending assessment of credit line assignment may include a qualitative review of decision criteria that enter the automated process, as well as quantitative testing of such criteria. For the portion of approvals where the credit line was assigned using judgment, card issuers may test for differences on a prohibited basis in the direction and size of divergence from lines recommended by the automated decision rules. Such an analysis may follow the general approach described previously for underwriting, but there are a few distinct differences.

Similar to the case of underwriting, a statistical assessment of differences on a prohibited basis in the incidence of judgmental credit line adjustments would be based on a statistical model that controls for the objective credit factors used in such evaluations, as specified in the creditor's policies and procedures. These factors might include whether and how much of a balance transfer was requested and whether the consumer had elevated credit risk not fully captured by the automated line assignment criteria. In this case, a multinomial logistic model may be appropriate because there are three potential outcomes rather than just two: line reduction, no change, or line increase. Potential differences in the magnitudes of credit line increases or decreases may be evaluated using ordinary least squares regression models or other estimation strategies (e.g., segmenting the sample between consumers who received a credit line increase and those who received a decrease, and evaluating whether there are prohibited basis differences in the average sizes of increases or decreases, after controlling for objective credit risk factors). As in the underwriting review process, any statistically significant differences on a prohibited basis may be investigated further through comparative file review.

V. Evaluating the Fair Lending Risk of Credit Scoring Models

Credit scoring models and other automated decision tools may help to control fair lending risk in credit decisions. By reducing or eliminating the amount of human judgment and discretion, credit scoring models may limit the potential for credit applicants to be treated differently on a prohibited basis, whether deliberately or inadvertently. However, to be effective in managing fair lending risk (as in managing credit risk), models must be developed and managed appropriately. Some of the key areas of fair lending risk related to credit scoring models are the use of variables closely related to prohibited factors, the misuse or mismanagement of the credit scoring system, and the potential disparate impact of the scoring system itself.[24]

A. Variables Closely Related to Prohibited Factors

Ostensibly neutral variables that predict credit risk may nevertheless present a risk of disparate impact on a prohibited basis if they are also highly correlated with a protected demographic characteristic. Two examples of variables sometimes considered for inclusion in custom scoring models and that may carry a risk of disparate impact are geographic location and income of the consumer, each of which may have some power to predict credit risk but also may be correlated with race or ethnicity. In these cases, the potential fair lending risk may come not simply from the variable's correlation with a prohibited basis but from that correlation together with the lack of a direct or obvious relationship between the variable and credit risk. In both of the mentioned examples, alternative predictors of credit risk may be available that capture the relevant aspect of credit risk but with less fair lending risk. For example, the level of a consumer's income might not be directly related to his capacity to repay debts. The ratio of his debts to his income might be used as a predictive factor instead of the level of income, and it may even be found to have better predictive power than income itself.

[24] Regulatory guidance regarding the examination of credit scoring systems for ECOA compliance can be found in "Considering Automated Underwriting and Credit Scoring," Part II of the Appendix to the CFPB's "Interagency Fair Lending Examination Procedures" (2012).

In the case of geographic predictive variables, it may be important to consider that location might simply be a convenient stand-in for some underlying economic factor(s), rather than inherently being a predictor of credit risk itself. For example, differences in economic performance or standard of living across states, metropolitan areas, or zip codes, as well as differences in state laws, may be associated with differences in the average risk of default. However, in some cases, there are distinct differences in minority population concentrations among states, metropolitan areas, or zip codes. As a result, geographic-based predictive variables may create a risk of disparate impact based on minority status. Such an impact might be reduced or avoided if the credit-related economic characteristics that differ among geographic areas can be measured and used directly in a scoring model, rather than using the "blunt instrument" of geographic location. Indeed, if the risk factor(s) underlying the predictive power of geographic indicators can be identified, they may have greater predictive power for credit risk than the geographic indicators, thus improving model performance while potentially reducing fair lending risk.

If it is not possible to identify the economic factors underlying geographic differences (e.g., if data are not available on such underlying factors), then it may be important to evaluate whether the contribution of such variables to the model's predictive power is sufficient to warrant their inclusion in light of their potential fair lending risk. The same thought process may apply to other variables that do not have a clear intuitive relationship to credit risk. If the decision is made to include a potentially suspect or controversial predictive variable in a model, then developing and documenting a rigorous empirical justification for its use may help to reduce potential fair lending compliance risk.

On a related point, bank card issuers sometimes give preferential treatment (via decision criteria or scoring models) to credit applicants based on their customer relationship status. Examples may include indicators of whether the applicant is an existing customer of the creditor or an affiliate, the number of the creditor's products or services the customer uses, or the amount of funds on deposit. To the extent that the demographic composition of the issuer's existing customer base differs significantly from that of the broader target market for its credit card products, such relationship-based factors might pose a risk of disparate impact because they may have the effect of disproportionately favoring one protected class over another. However, such criteria might nonetheless have a legitimate business justification, such as the value of relationship factors in predicting default risk and/or their contribution to the overall profitability of the customer relationship. Whatever the justification, again, fair lending risk may be reduced by documenting that justification rigorously.

B. Risks Arising from Mismanagement or Misuse of Models

The mismanagement or misuse of credit scoring models can be as much of a fair lending compliance issue as it is a credit risk issue. Mismanagement may include, for example, ad hoc adjustments to models, such as changes to score thresholds for underwriting approval, changes to weights on predictive variables, or the addition or removal of explanatory variables. Such changes, if performed on an ad hoc and judgmental basis, can undermine the demonstrable statistical validity of a model. If a model loses its previously demonstrated statistical validity, then it may be deemed to be a "judgmental scoring system," as discussed previously, with a resulting increase in fair lending compliance risk, if not an actual regulatory violation. For example, a model that uses age explicitly as a predictive variable might initially meet the EDDSS standard, but it may subsequently lose its EDDSS status as a result of ad hoc adjustments to the model over time. If that occurs, then the model's continued use of age as a predictive variable may no longer be permissible. Therefore, careful management of a model's fair lending compliance risk requires that changes to the model be based on rigorous and appropriately documented empirical analysis and that revised models be revalidated.

Fair lending risk may also arise when a model is applied to a consumer population for which it was not developed. For example, if a model was developed based on a sample of consumers with prime credit but then is applied to a nonprime product or consumer population, it might not be a statistically valid predictor of credit risk for this subpopulation. If the model development sample is not representative of the population to which the model is ultimately applied, the resulting model may be misspecified. In particular, variables that are predictive for prime consumers may be less important for predicting the risk of subprime consumers. Alternatively, a model derived from a prime population may omit variables that are important for predicting subprime credit risk. These issues may lead to incorrect rank-ordering of credit risk.[25] Beyond the implications for statistical validity, such misapplication or misspecification of a model may create

[25] See Avery et. al (2000).

disparate impact risk if the omitted predictive variables are more important for one ECOA-covered demographic group than for another.[26]

Fair lending risk may also arise when a creditor implements an otherwise valid model in an ad hoc way. For example, a CFPB enforcement action in 2012 included the settlement of an alleged ECOA violation resulting from a credit card issuer's failure to implement a credit scoring model consistently. In that case, the credit card issuer allegedly developed an age-split scoring model but only implemented it on a staged basis such that, for a period of eight months, the issuer had implemented the model for applicants age 35 and younger but not for applicants older than 35.[27] The CFPB contended that this violated the ECOA because the law requires credit scoring systems that take age into account to be designed and implemented properly.[28]

Fair lending risk may also arise from a failure to monitor appropriately a model's use and performance over time. First, model performance may degrade over time due to such factors as changes in the consumer population, changes in consumer behavior, or changes in credit policy. If a model's performance is not monitored regularly, as required by federal regulatory guidance, it may lose its statistical validity

---

[26] Statistical validity issues of the sort described in this paragraph may be inevitable when a creditor is launching a new product or entering a new market because the creditor would not have historical data on the population of interest for the purpose of developing a model. In such cases, potential fair lending risk may be managed by acquiring data from a data vendor, where possible, and by redeveloping the model as soon as sufficient performance data becomes available through the creditor's own portfolio.

[27] An age-split scoring model uses different scorecards or models based on the age of an applicant, with each scorecard containing variables that are predictive for a given age group. See Regulation B, 12 C.F.R. §202.6(b)(2).

[28] Federal Deposit Insurance Corporation and Consumer Financial Protection Bureau, In The Matter of American Express Centurion Bank, Joint Consent Order, Joint Order for Restitution, and Joint Order to Pay Civil Money Penalty, FDIC-12-315b, FDIC-12-316k, 2012 -CFPB-0002, October 1, 2012. See also the related CFPB press release, http://www.consumerfinance.gov/newsroom/cfpb-orders-american-express-to-pay-85-million-refund-to-consumers-harmed-by-illegal-credit-card-practices/.

over time.[29] To the extent that a model loses its claim to being EDDSS, fair lending risk may increase because the model may not have a sufficiently strong business justification to counter any potential disparate impact claims. And, as noted previously, a model using age as a predictive factor must meet the EDDSS standard.

Next, insufficient oversight and management of overrides or exceptions may create fair lending risk in two ways. First, an excessive number or frequency of overrides may undermine a model's claim to statistical validity. Second, overrides may result in different treatment of similarly qualified applicants who differ in terms of prohibited characteristics. Excessive overrides may be symptomatic of an implicit policy that could formally be written into guidelines or be built into the credit scoring model, thus reducing the need for discretion and the risk of fair lending compliance issues. Excessive overrides may also be the result of deficiencies in a credit scoring model's capabilities or performance. Again, the need for such overrides might be reduced by enhancing the scoring model. To the extent that score overrides are needed for legitimate business reasons, fair lending risk may be mitigated by (1) establishing clear guidelines regarding the allowable reasons for overrides, (2) requiring that underwriters document the reasons for granting an override, and (3) monitoring to ensure both that the guidelines are followed and that the volume or frequency of exceptions remains within an acceptable range.

Each of the potential fair lending risk issues discussed previously arises from basic problems with model risk management. Therefore, fair lending risk issues often may be avoided or mitigated simply by following sound model risk management

[29] OCC Bulletin 2011-12.

practices, including developing and maintaining documentation and evidence of each model's statistical validity, both at development and over time.

C. Testing a Model for Disparate Impact Risk

No guidance has been published by the federal regulators of financial institutions regarding how the disparate impact risk of credit scoring systems should be tested or how large a disparate impact needs to be before it becomes a regulatory compliance concern.[30] Absent official guidance, various reasonable approaches might be considered. It is beyond the scope of this paper to go into all of the technical details of such testing methods, but we provide an overview of some key considerations in performing such testing and note some statistical approaches that may be considered. The appropriate approach to testing in any given situation may depend upon the nature and functional form of the scoring model. Also, some degree of judgment may be required in evaluating whether the size of a disparate impact is large enough to warrant a regulatory compliance concern.

1) Some Considerations in the Testing of Models

Two key points should be noted at the outset. First, a thorough disparate impact analysis may consider both the impacts of predictive variables in a scorecard taken individually and the impact of the scoring system as a whole. Scoring models attempt to capture a multivariate relationship between a measure of credit performance and various

[30] In this section, we use the term "credit scoring model" broadly to include any model involved in the credit process, whether it is used to predict default risk, the likelihood of responding to an offer, profitability, or some other credit-related consumer behavior. For convenience, we carry out the discussion in terms of a model used to predict credit risk, but the concepts are transferrable to other credit model contexts.

predictive variables that may have varying degrees of correlation with demographic characteristics and with each other. Even if a particular variable in a model has a disparate impact on a prohibited basis when considered in isolation, the model as a whole may be free of disparate impact because negative effects related to one factor in a model may be offset by positive effects of other factors in the model. Therefore, multivariate analysis typically is required to assess fully the disparate impact risk of a scoring model or system. Nevertheless, it may be useful to evaluate the impacts of individual predictive variables, particularly if the overall model is found to have a risk of disparate impact.

Second, simple comparisons of average credit scores or score distributions of different demographic groups do not necessarily provide a conclusive analysis of disparate impact risk. Such comparisons may show that, for example, minority consumers have lower average scores than white consumers, which would tend to result in a higher rejection rate for minority credit applicants.[31] However, such a pattern, by itself, is not inherently an indication of an illegal disparate impact. It may simply be the case that minority consumers objectively tend to have higher average credit risk (i.e., lower unconditional good/bad odds)[32] than white consumers, as measured by the credit criteria included in the model, due to systematic differences in the populations in terms of income, wealth, employment, credit experience, or other economic factors. Such

[31] Avery, Brevoort, and Canner (2012) found that commercially available credit scores are lower on average for African American and Hispanic consumers compared with non-Hispanic white and Asian consumers. For example, the median scores of African American and Hispanic consumers are about 35% and 69%, respectively, of the median scores of non-Hispanic whites and Asians. About 62% of African American and 37% of Hispanic consumers are in the lowest quartile of the score distribution, compared with 20% of non-Hispanic whites and 16% of Asians.

[32] The "good/bad" odds are computed as the number of "good" accounts in a sample divided by the number of "bad" accounts in the sample; or, equivalently, the probability of being "good" divided by the probability of being "bad."

descriptive information about credit distributions does not indicate that applicants of the same risk (i.e., "similarly situated" applicants) were scored differently by a model.[33] Evaluating the disparate impact risk of a scoring model typically requires going beyond this level to determine whether the model tends to treat credit applicants *of the same risk* equally. Put differently, it typically requires evaluating whether a model has differential validity across demographic groups.

Quantitative disparate impact analysis may include some standard descriptive statistics that will be familiar to developers of credit scoring models and specific testing for disparate impact. Standard credit modeling statistics may be used to perform an initial analysis of whether a scoring model tends to assign similar scores to borrowers who have similar levels of risk. One simple visual approach to making such a comparison is to examine the relationships of good/bad odds to scores across demographic groups. A comparison of the relationships between good/bad odds and scores across the groups of interest, whether across the entire score distribution or at relevant approval thresholds, illustrates whether borrowers in different groups that have the same score tend to have the same risk (as measured by actual credit performance). Though useful as a high-level assessment of whether a model scores different populations differently, multivariate testing approaches may provide a more rigorous way to evaluate whether a significant disparate impact arises from the set of variables included in the model and which variables (if any) contribute to that impact.

---

[33] Some studies have found that, for at least some score ranges, minority or low- and moderate-income (LMI) borrowers at a given credit score actually have higher risk than white or higher-income borrowers, meaning that the scoring model actually treated minority or LMI borrowers more favorably than comparable whites or higher-income borrowers. See, for example, Martell et. al (1997), Board of Governors of the Federal Reserve System (2007), and Avery, Brevoort, and Canner (2012).

Multivariate approaches to testing models for disparate impact have been proposed by Fortowsky and LaCour-Little (2001), Ross and Yinger (2002), and Avery, Brevoort, and Canner (2012). These approaches are motivated by the intuition that disparate impact may arise when a predictive variable in the model acts as a statistical proxy for a demographic characteristic.[34] A predictive variable may act as a statistical proxy for a demographic characteristic if (1) both the variable and the demographic characteristic are correlated with credit performance and (2) all or part of the variable's ability to predict credit performance derives from its correlation with the demographic characteristic (i.e., it does not have predictive power independent of demographic status, or it has less predictive power once the predictive power attributable to demographic status has been factored out). Based on this intuition, the Fortowsky and LaCour-Little (2001), Ross and Yinger (2002), and Avery, Brevoort, and Canner (2012) approaches are designed to detect proxy effects for demographic characteristics by assessing whether controlling for the influence of demographic characteristics either significantly alters the estimated coefficients of the predictive variables in the model or would result in the selection of an alternative set of predictive variables for the model.[35]

Disparate impact testing typically accounts for model segmentation. Credit scoring systems often include criteria to segment the consumer population. Different

[34] Our use of the word "proxy" and "proxy effect" in this section is similar to, but distinct from, our use of the word "proxy" in inferring likely demographic group membership of consumers based on criteria such as names and addresses. Here we refer to proxies in the sense of variables used as predictors of credit risk that are so highly correlated with a prohibited demographic characteristic that they might be regarded as, in effect, a stand-in for the demographic characteristic such that consumers are treated differently because of that characteristic rather than because of an inherent difference in their credit characteristics. For a discussion of how such proxy effects may arise, see Chapter 9 of Ross and Yinger (2002).

[35] One may account for the influence of demographic factors on the predictive power of the credit variables either by including controls for demographic group membership as predictive variables in the model itself or by limiting the estimation sample to members of a single demographic group (e.g., a single race/ethnicity group).

models are developed and applied for different consumer segments, in order to account for differences in the set of credit variables that are predictive for different population segments and to account for differences across segments in the importance of each variable to the prediction. Each segment-specific model may be tested for disparate impact risk, and the possibility that the segmentation variables themselves could result in a disparate impact may be evaluated.

2) Evaluating Whether a Disparate Impact Represents a Fair Lending Issue

The finding of a statistical disparate impact in a scoring model may not be the end of the story with respect to the question of whether using the model might result in illegal discrimination. As discussed previously, a model or a model's predictive variable with a disproportionate adverse impact on a prohibited basis may still be legally permissible if it has a demonstrable business justification and there are no alternative variables that are equally predictive and have less of an adverse impact. Therefore, the next steps in the fair lending analysis may include (1) identifying the predictive variable or variables that are the source of the disparate impact, (2) evaluating whether the impact is large enough to be of potential concern, (3) scrutinizing the statistical evidence from the development of the model regarding the predictive power of the variable(s) in question, and (4) evaluating whether alternative predictive variables are available that have less of a disparate effect but do not significantly degrade the predictive power of the model.

Evaluation of the sizes of any disparate effects may include assessing whether they are both statistically and practically significant. Practical significance depends on whether and to what extent members of a given demographic group are harmed by the score. For example, taking into account the applicable score threshold or range for approval/rejection, does the model result in significantly higher rejection rates for credit for minority applicants than nonminority applicants with the same credit risk or profitability? Would the model cause minorities to be subjected, through a manual underwriting review, to rigorous scrutiny of their credit qualifications more likely than nonminorities, and does the resulting manual review tend to disadvantage minorities? Does the model result in minorities paying higher interest rates, on average, than nonminority applicants with the same credit risk? If the differential scoring of a model across demographic groups does not actually result in prohibited basis differences in credit outcomes, then it may not actually present a risk of disparate impact.

If there is evidence that a variable has a disparate impact, investigation of the business justification for the variable may include not just determining whether the variable has an empirical basis for inclusion in the model but also evaluating the contribution of the variable to the model's overall predictive power and to the business objectives of the model (such as achieving a target charge-off rate or level of profitability). If a variable contributes only marginally to predictive power or business objectives but produces a disparate impact on a prohibited basis, then it may be worth considering whether use of the variable presents an acceptable regulatory compliance risk.

The process of identifying and evaluating potential substitutes bears a resemblance to the variable selection process that a model developer follows to select variables for inclusion in the model from among a set of candidate variables that are correlated with credit performance. The evaluation of potential substitutes may involve examining variables that are highly correlated with credit performance, but were not originally selected for the model, and comparing each with the model variable that has a disparate impact in terms of their relative predictive power and degree of correlation with the demographic characteristic of interest. The evaluation may also include such practical considerations as model stability, parsimony, and implementation constraints.

If a viable substitute has been found that has less of a disparate impact but does not materially degrade the predictive power of the model, then the revised model may be retested to confirm that the introduction of the substitute variable has not created a source of potential fair lending concern with respect to another demographic group.

Finally, even if the analysis concludes that a variable with some level of disparate impact has a sufficient business justification and that there are no close substitutes available with less of a disparate impact, some professional and legal judgment normally comes into play in assessing whether to use the variable. The fact that a disparate impact exists may increase regulatory compliance risk (including the cost of defending the use of the variable in the face of regulatory scrutiny), and the tradeoffs between such risk and the benefits of using the variable should be weighed.

VI. Management of Fair Lending Risk Through Model Governance

Often, fair lending compliance specialists are not deeply involved in model governance or are not consulted at key stages of the model life cycle. In addition, fair lending compliance specialists do not always have sufficient statistical expertise to fully evaluate scoring models. As a result, compliance oversight may be reactive and detective, rather than proactive and preventive, often resulting in the costly retooling of models that are found to have fair lending risks only after they are already in production. At worst, the lack of involvement of fair lending compliance specialists in the model oversight process may result in serious regulatory enforcement actions. Such issues may be less likely to occur if fair lending risk management is properly embedded in a financial institution's overall model governance process.[36]

The risk management standards with which banks and other financial institutions are expected to comply already require rigorous documentation of the development, validation, use, and performance of predictive models. Those standards require a deliberate process of model risk management and oversight, which may include various checks and approvals throughout the model life cycle. Fair lending compliance risk may be managed efficiently, effectively, and proactively by integrating model compliance risk management with the broader model governance framework.

Creditors may design an effective system for model compliance risk management by ensuring that the model governance structure explicitly defines factors such as:

- responsibility and accountability for model compliance;
- the sets of models that require some level of fair lending compliance review;
- requirements for model documentation to meet fair lending compliance needs;
- requirements for validation, including documentation of the validation process and findings; and

[36] For an overview of regulatory expectations regarding risk model governance, see OCC Bulletin 2000-16 and OCC Bulletin 2011-12.

- the process for engagement of fair lending compliance staff by model owners in the organization.

Explicitly and clearly defining the types of models that require fair lending review is important to the process because it helps to avoid the chance that technical modeling specialists will make judgments affecting fair lending compliance risk that they may not be qualified to make. Generally speaking, the universe of models that may require some level of fair lending compliance review includes all “customer-facing” models — those relating to credit products or services that directly affect actual or potential customers. This may include models used for all of the following purposes:

- predicting risk or behavior for use in the extension of credit to consumers, including models used in marketing, solicitation, or underwriting;
- determining the amount of credit to offer or extend;
- setting terms and conditions (including pricing, rate-setting, or fees);
- determining customer retention offers or changes in terms;
- deciding collection strategies or other aspects of account servicing; and
- any other customer-facing purpose where the model contains demographic or geographic data.

It may also be helpful to explicitly define categories of models that do not require fair lending compliance review, in order to avoid unnecessary work and bureaucracy. Models declared “out-of-scope” for fair lending compliance review may include those that are not related to credit products or services and/or are not customer facing, including models used for financial analysis, financial management, or financial reporting; loss forecasting; capital adequacy; asset valuation; portfolio management or monitoring; market risk management; operational risk management (except perhaps customer-facing fraud models); reporting; and other back-office analysis or risk management purposes that do not directly affect decisions regarding actual or potential customer accounts.

The importance of model documentation and validation is clear when it comes to demonstrating that a model is EDDSS (as described in Section II). Its importance is also clear in demonstrating the business justification of a model, decision rule, or predictive variable. The statistics, logs, and other documentation created during model development may provide the requisite evidence justifying the use of each predictive variable in the model, as well as for the weight each variable receives in the decision process. The validation process typically ensures that such evidence has been independently verified. If a scoring system is, in fact, designed to use the most predictive combination of available credit factors, then it should be unlikely that someone could demonstrate that there is an equally effective alternative available, which the lender has failed to adopt.

The compliance engagement process may be organized to follow and work in tandem with the creditor's general model governance process. This may include defining the documentation and data requirements that the model owner must submit for review and scheduling the review, just as is typically done with respect to the model validation process. Indeed, the information required for model validation will likely include most or all of the information required for compliance review.

The schedule for model developer engagement with compliance staff ideally emphasizes early consultation. Fair lending compliance specialists may be consulted by model owners initially at the model proposal stage, when the general purpose and objectives of a proposed modeling effort have been defined. This allows for the early identification of potential fair lending risks so that they may be considered in the business decision to approve a model for development. Early consultation also may allow appropriate advance notice for planning and scheduling of the fair lending review, including accounting for the fair lending review in the model implementation schedule.

Next, engagement at the model development stage allows for the possibility of reviewing the list of candidate predictive and segmentation variables so that any off-limits or potentially problematic variables may be identified before there is extensive investment in the development process. Fair lending compliance specialists may also review the final model (in the event that any unanticipated variables with fair lending risk have made their way into the model) and may review the results of the model's validation process.

Creditor staff members with fair lending compliance oversight responsibilities have an obvious stake in the model validation process, and compliance approval of a model might be made conditional on successful (ideally independent) validation and, as necessary, disparate impact testing. For the sake of efficiency, any necessary disparate impact testing might take place concurrently with the validation process. Depending on the backgrounds of the model validation staff or consultants, it may also make sense to include disparate impact testing as part of the validation.[37] The fair lending compliance engagement process may also include involvement in model change control and model failure procedures so that any model revisions with fair lending implications may be identified and addressed and so that any fair lending impacts of model failures (such as implementation, programming, or production system errors) may be identified.

The chart that follows provides an overview of the linkages between the overall model governance process and the key fair lending risk management considerations discussed previously.

[37] For reasons of independence and objectivity, the staff responsible for model development typically would not have a direct role in disparate impact testing and ideally would be insulated from data on legally prohibited demographic characteristics, to the extent possible.

Key Components of Model Governance Process	**Linkages to Fair Lending Risk Management**
Model Governance Framework	
• Model Policies[38]	Define requirements for compliance involvement. Define types of models in-scope and out-of-scope for fair lending compliance review.
• Definition of Roles & Responsibilities	Assign ownership of fair lending risk. Define fair lending risk management and oversight responsibilities of model owner and compliance staff.
• Model Inventory	Document which models are subject to fair lending review and when review/sign-off occurred.
Model Life Cycle	
• Proposal/Concept Stage	Review model scope, objectives, target market, or consumer group.
• Development Stage	Review candidate predictive and segmentation variables; consult on potential disparate impact risks.
• Documentation/Developmental Evidence • Initial & Ongoing Validation/ Performance Monitoring • Change Control	Evidence to support initial and ongoing EDDSS and business justification. Identification of postimplementation fair lending risks.

By integrating compliance risk management with risk model governance, both sides of the business may become more efficient in managing the entire risk associated with scoring models — including compliance risk. When structured appropriately, fair lending compliance staff may have the information and modeler engagement necessary to help model owners in their financial institutions identify and mitigate fair lending compliance risk issues, while minimizing the amount of additional bureaucracy. The business line model owners, in turn, may avoid expenses associated with *ex post*

[38] Compliance aspects of models may alternatively be specified in compliance policies, but their inclusion in model governance policies may help to ensure that they are not overlooked by model owners and modeling staff.

detection of fair lending compliance issues that can result in the costly retooling of models already in production and may have a greater chance of avoiding the costs of regulatory actions.

VII. Concluding Comments

Rigorous statistical testing forms an important component of a fair lending compliance management system for credit card lending. However, statistical analysis approaches in this area — and particularly disparate impact testing approaches — are not well established, and there are no formal regulatory guidelines for conducting such analysis. In this paper, we have provided a broad outline of potential fair lending risk topics to which statistical testing methods may be applied, and we suggested various considerations in designing appropriate methods for such testing.

As we have described, proactively identifying, quantifying, and monitoring the fair lending risk of model- and rules-driven credit processes includes well-thought-out statistical approaches that take into account both the nature of the business processes and decision systems involved as well as the legal standards by which potential regulatory violations would be judged. However, statistical analysis and monitoring are only parts of a fair lending compliance management system for model-intensive credit card processes. The fair lending risk associated with scoring models may be best managed proactively by ensuring that fair lending compliance considerations are integrated with an institution's overall model governance process.

References

American Bankers Association, "Toolbox on Fair Lending," May 2012, http://www.aba.com/Tools/Toolboxes/Mem/Pages/FairLendingToolbox.aspx (available to ABA members only).

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner, "Credit Scoring: Statistical Issues and Evidence from Credit-Bureau Files," *Real Estate Economics*, 2000, 28:3, 523–547.

Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner, "Does Credit Scoring Produce a Disparate Impact?" *Real Estate Economics*, 2012, 40:1, S65–S114.

Board of Governors of the Federal Reserve System, "Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit," Submitted to the Congress pursuant to Section 215 of the Fair and Accurate Credit Transactions Act of 2003, August 2007.

*Collins, M. Cary, Keith D. Harvey, and Peter J. Nigro, "The Influence of Bureau Scores, Customized Scores and Judgmental Review on the Bank Underwriting Decision Making Process," *Journal of Real Estate Research*, 2002, 24:2, 129–152.

Consumer Financial Protection Bureau, "Interagency Fair Lending Examination Procedures," CFPB Examination Manual v.2 (October 2012), http://www.consumerfinance.gov/guidance/supervision/manual/.

Consumer Financial Protection Bureau, "Preventing Illegal Discrimination in Auto Lending," http://www.consumerfinance.gov/blog/ preventing-illegal-discrimination-in-auto-lending/, posted November 4, 2013.

Courchane, Marsha J., and David Skanderson. "Fair Lending Risk Management: Lessons from Recent Settlements," CRA Insights, November 2012, http://www.crai.com/uploadedFiles/Publications/FE-Insights-Fair-Lending-Risk-Management-1112.pdf.

Cubita, Peter N., and Michelle Hartmann, "The ECOA Discrimination Proscription and Disparate Impact — Interpreting the Meaning of the Words That Actually Are There," *The Business Lawyer*, 2006, 61:2.

Fortowsky, Elaine, and Michael LaCour-Little. "Credit Scoring and Disparate Impact," Working Paper, Wells Fargo Home Mortgage, December 13, 2001.

*Karlson, Kristian Bernt, Anders Holm, and Richard Breen, "Comparing Regression Coefficients Between Same-Sample Nested Models Using Logit and Probit: A New Method," *Sociological Methodology*, 2012, 42, 286–313.

*Maas, Ericca, “Credit Scoring and the Credit-Underserved Population,” *Community Dividend*, Federal Reserve Bank of Minneapolis, May 2008, http://www.minneapolisfed.org/pubs/cd/08-2/CommDiv_2008_2.pdf.

Martell, Javier, Paul Panichelli, Richard Strauch, and Sally Taylor-Shoff, “The Effectiveness of Scoring on Low-to-Moderate-Income and High-Minority Area Populations,” Fair Isaac Company, Inc., August 1997.

Office of the Comptroller of the Currency, “Risk Modeling,” OCC Bulletin 2000-16, May 30, 2000.

Office of the Comptroller of the Currency and Board of Governors of the Federal Reserve System, “Supervisory Guidance on Model Risk Management,” OCC Bulletin 2011-12, April 4, 2011.

Ritter, Dubravka. “Do We Still Need the Equal Credit Opportunity Act?” Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper 12-03, 2012.

Ross, Stephen L. and John Yinger, *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair–Lending Enforcement*, MIT Press, Cambridge, MA, 2002.

*Straka, John W. “A Shift in the Mortgage Landscape: The 1990s Move to Automated Credit Evaluations,” *Journal of Housing Research*, 2000, 11:2, 207–232.

*Turner, Michael A., Alyssa Stewart Lee, Ann Schnare, Robin Varghese, and Patrick D. Walker. “Give Credit Where Credit Is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data,” Political and Economic Research Council and the Brookings Institution Urban Markets Initiative, 2006, http://www.perc.net/wp-content/uploads/2013/09/alt_data.pdf.

Note: The sources marked with an asterisk are not specifically cited in the text but have been added as supplementary references.