BEYOND DISPARATE IMPACT:
Risk-based Pricing and Disparity in Consumer Credit History Scores

Working Paper: June 2009

By
Jonathan Spader

Center for Community Capital
The University of North Carolina at Chapel Hill

Prepared with financial support from The Ford Foundation

Center *for* Community Capital
Research and analysis on the transformative power of capital

UNC
COLLEGE OF
ARTS & SCIENCES

Abstract
Despite the increasing importance of credit scoring to an expanding range of activities, very little is known about the nature of the credit scoring process. This article examines the interaction of credit scoring with risk-based pricing, exploring the potential for credit scoring to contribute to the segmentation of low- and high-cost credit markets. Specifically, it uses a stylized example to illustrate the mechanism through which credit scores may capture disparities in surrounding credit markets, passing them into future periods and other credit markets. In the wake of mounting subprime foreclosures, this mechanism contains the potential to exacerbate the concentrated impacts of the subprime crisis in low-income and minority communities. The article examines this issue in the context of existing regulatory actions.

Introduction

The development of credit scoring is closely connected to the increased prevalence of risk-based pricing in mortgage and consumer credit markets. The aggregation of consumer credit data by the credit bureaus during the 1980s allowed creditors to more precisely observe consumers' past payment histories. By 2000, almost every mortgage origination referred to the borrower's credit history score. Today the influence of credit scores continues to grow, as scores are increasingly being used to evaluate applications for insurance, rental housing, employment, and other activities outside of lending and credit markets.[1] This expanded use of credit scores may create an efficiency gain to consumers, if the credit record provides valuable information to users. However, the expanded use of credit scores also raises the stakes involved in ensuring that both the calculation and use of credit scores do not disadvantage underserved groups.[2]

Existing studies document two concerns:

Lack of Data: The three major credit bureaus—Experian, Equifax, and TransUnion—all currently seek to collect comprehensive information on the credit histories of all individuals in the United States, collecting information from creditors, utility companies, legal records, and collection agency files. The credit bureaus lack legal rights to any information, instead soliciting voluntary reporting from these sources (Avery et.al. 2003).[3] The resulting data create substantial records for most borrowers, but do not reflect complete records of individuals' credit histories. The Information Policy Institute estimates that 35 to 54 million Americans do not have sufficient information to generate a score (PERC 2006). In particular, households that rely on small lenders and alternative financial service providers may not have the payments made on these accounts reflected in their credit records.

Factual Errors and Mistakes: The data used to construct individuals' credit records commonly contain mistakes and factual errors (Avery, Calem, and Canner 2004). The Fair Credit Reporting Act provides all individuals a right to review their credit record at no cost and mandates mechanisms for appeal and correction. However, no analysis has been performed of the ability of individuals to easily and successfully correct mistakes. Equally important, no analysis has been performed of which individuals review their records and pursue corrections. To the extent that

[1] Insurance providers regularly check borrower credit scores as a condition of approval and use the score in determining a premium. The Federal Trade Commission (2007) examined the potential for disparate impact of such use, finding small differences between existing scoring models and a corrected model. Citing evidence that credit-based insurance scores are strongly predictive of insurance claims and that no improved model could be found, the FTC concludes that the scores "appear to have little effect as a 'proxy' for membership in racial and ethnic groups in decisions related to insurance." Employers and landlords also have a legal right to access an applicant's credit record and to use the information in making employment and rental decisions.

[2] While individuals' incomes are not used directly in the calculation of generic credit history scores, these scores strongly correlate with household income. Imputing scores for households in the Survey of Consumer Finances, Bostic, Calem, and Wachter (2005) show that low-income and minority households were disproportionately concentrated in the lower score ranges in both 1989 and 2001. Furthermore, high-wealth and low-wealth households experienced divergent credit experiences between 1989 and 2001.

[3] The Fair Credit Reporting Act governs the bureaus' protection of individual information and the conditions under which a borrower's credit record or score can be disclosed.

underserved individuals lack knowledge of this right and/or lack access to the internet (see Lyons, Rachlis, and Scherpf 2007; Gates, Perry, and Zorn 2002), fewer individuals will be likely to pursue corrections.

This article presents a more fundamental issue related to the interaction of credit scores with the operation of consumer credit markets. Specifically, it outlines the potential for credit scoring to contribute to disparities in the segmentation of low- and high-cost credit markets, describing the consequences of defining credit scores with respect to a universal measure of default—serious delinquency on *any* open credit line. Definition of the outcome and payment history measures in this way fails to distinguish between default on a fixed-rate prime mortgage and default on a subprime ARM with a teaser rate and/or a balloon payment, credit products that carry very different incentives to prepay and default. The analysis shows that disparities in credit scores emerge when: (1) disparities exist in the processes through which borrowers are assigned to credit products and (2) alternative product characteristics carry varying risks of default.

In the wake of mounting subprime foreclosures, this mechanism contains the potential to exacerbate the concentrated impacts of the subprime crisis in low-income and minority communities. In particular, disparities in credit score outcomes may emerge as a result of the concentrated lending activities of brokers and subprime lenders, even in the absence of discrimination. Unfortunately, the proprietary nature of credit scoring has largely prevented scrutiny from regulators and academics.[4] The primary exception is that the Fair and Accurate Credit Transactions Act of 2003 (FACTA) directly instructed the Federal Reserve Board to test for disparate impact in the calculation of credit scores. The resulting report offers a seminal analysis of the credit scoring process. However, because the credit bureaus do not currently record the terms of an observed loan, the Federal Reserve analysis also relies on universal measures of payment history and default.

The first sections of this article present and discuss the calculation of generic credit history scores, focusing on the consequences of using universal measures. A stylized example is then developed to illustrate the mechanism through which credit scores may capture disparity in the surrounding credit markets, passing it into future periods and into other credit markets. While the use of simulated data in this model is partly motivated by the lack of public data on credit scores, this approach is also useful in that it circumvents the issues with omitted variables and unobserved risk that accompany analyses of default and credit risk. The resulting model of credit scoring shows that credit scores reflect the assignment of borrowers to credit products, in addition to borrowers' underlying risk of default.

The Definition and Development of Credit Scoring

[4] The lack of publicly available research is due at least in part to the proprietary nature of credit scores and the credit scoring formulas. The Fair Credit Reporting Act defines the permissible purposes under which a consumer's credit report can be obtained by outside entities—application for credit, insurance, employment, rental housing, or bank account (Avery et.al. 2003). Exceptions require the written consent of the consumer, a standard that is time consuming and expensive for large data collection efforts.

The first credit scoring models appeared in the 1950s and 1960s, designed for use by retail stores and finance companies. These lenders applied credit scoring models to their own internal data, as they lacked access to other credit providers' records of borrower payment (Hunt 2002). Only in the late 1980s did the technological requirements and institutional arrangements create widespread access to centralized data on consumers' payment histories. Lenders initially utilized this information (and the derived scores) to develop and market products, but also quickly applied this information to loan underwriting (Board of Governors 2007).

The application of these credit records to loan underwriting initially produced multiple credit scoring models that predicted lending risk with respect to individual credit products. Formally, a credit score reflects the likelihood that a borrower will default on a loan. This derivation differs for different products, as the predictors of mortgage default differ from those associated with small-business loans. In response, lenders developed a set of credit scoring models to underwrite individual products. For instance, a typical mortgage scoring model predicts the likelihood that a loan enters 90 day delinquency, assigning scores to different ranges along the distribution of predicted probabilities (Straka 2000; Avery et.al. 2000).

Alternatively, the three major credit bureaus—Experian, Equifax, and TransUnion—all produce generic credit history scores. These scores reflect the likelihood that a borrower will become seriously delinquent on any open credit account within 18-24 months. The Fair Isaac Corporation produces multiple variations of this generic credit history score, with different models (scorecards) used for predicting the scores of different subpopulations and for predicting the scores specific to each credit bureau. Because the scorecards are specific to the credit bureau's data, a borrower's score may differ slightly across bureaus. However, these differences tend to be small, and the generic credit history scores are generally grouped and referred to as the borrower's FICO score (Board of Governors 2007). For the remainder of this article, we use the term credit score to refer to the borrower's generic credit history FICO score, leaving alternative credit scoring and automated underwriting systems for future analyses.

While the precise models for computing borrowers' credit scores are not currently disclosed to the public, the bureaus do provide some information about the weight given to different sets of predictors: previous payment history (35%), outstanding debts (30%), length of credit history (15%), new accounts opened (10%), and types of credit used (10%).[5] In generating individual credit scores, the specific set of predictors differs depending upon the specific model or 'scorecard.' The scorecards used by the credit bureaus are also proprietary, so the Board of Governors (2007) approximates the credit scoring algorithm using three categories: the *clean file* scorecard includes credit records with no serious delinquency; the *thin file* scorecard includes credit records with a limited number of observed credit characteristics; and the *major derogatory* scorecard includes credit records with a serious delinquency on at least one open account.

[5] Fair Isaac Corporation. "Understanding Your FICO Score." 2005. [http://www.myfico.com/Downloads/Files/myFICO_UYFS_Booklet.pdf]

Recalling the definition of a generic credit history score, the credit scoring model for each scorecard identifies the association of the observed credit characteristics with the likelihood of default on *any* open credit line:

(1) $$Default_i = H_i\alpha + O_i\beta + L_i\gamma + A_i\delta + C_i\xi + \varepsilon_i$$

where H_i, O_i, L_i, A_i, and C_i reflect vectors of variables reflecting payment history, outstanding debts, length of credit history, origination of new accounts, and types of credit, respectively, and ε_i is the error term. For each scorecard, an individual's credit score is based on the predicted likelihood of default once weights have been assigned to each of the coefficient vectors in equation (1):

(2) $$Y_i = f(H_i\hat{\alpha} + O_i\hat{\beta} + L_i\hat{\gamma} + A_i\hat{\delta} + C_i\hat{\xi})$$

where Y_i is the individual's credit score and f() is a function that assigns credit scores to the distribution of predicted default probabilities.

In its seminal examination of credit scores, the Board of Governors (2007) disclosed a list of the 312 credit characteristics compiled by TransUnion in borrower credit records, offering greater detail about the individual measures that compile the above categories.[6] That report also displays the raw correlations of a subset of these variables with credit performance—defined as whether the record exhibited a serious delinquency on any account. However, neither the credit bureaus nor the Federal Reserve report discloses the estimated weight of each characteristic in predicting credit risk or the explanatory power of estimated credit scoring models.

What Risk is Credit Scoring Intended to Capture?

The credit scoring model defined in equation (1) predicts the likelihood of default on the basis of observed characteristics in consumers' credit records. In practice, scorecard developers begin with the hundreds of credit characteristics in consumers' credit records, but base the final scorecard on the eight to twelve characteristics that have the strongest predictive value. The resulting credit scores thus reflect the predicted likelihood of default on any open credit account on the basis of the characteristics in each scorecard, and the assignment of credit records to the clean file, thin file, and major derogatory scorecards. For instance, the Board of Governors (2007) developed a set of base scorecards intended to mirror the scorecards used by the credit bureaus. The 19 characteristics used in these scorecards are shown in Table 1.[7]

[INSERT TABLE 1 ROUGHLY HERE]

Examination of the credit characteristics in Table 1 gives rise to two insights with respect to the algorithms used to calculate credit scores. First, each of the credit characteristics is a universal measure, treating different types of credit products identically. Although several credit characteristics distinguish between revolving, installment, bankcard, and public accounts, none recognizes the interest rate and/or terms associated with the

[6] See Appendix B of the Federal Reserve report (Board of Governors 2007).

[7] The specific characteristics used by the Board of Governors (2007) for the clean file, thin file, and major derogatory scorecards are shown in Tables 12a, 12b, and 12c of that report.

underlying credit products. Similar to the universal measure of default, each of the credit characteristics groups a wide variety of credit products into a single measure.

The second observation regarding the credit characteristics in Table 1 is that the specification of each scorecard is detached from the theory of consumer default. Option theory suggests that the default decision can be defined as:

(3) $M(C(t), r(t), t) \geq C(t) + T(t)$

where M() is the outstanding value of the current credit product given collateral of value C(t), the current available interest rate r(t), and the number of monthly payments remaining t (Clapp et. al. 2001; Deng, Pavlov, and Yang 2005). The default decision is also influenced by the presence of transaction costs, T(t), which include both the direct costs associated with default and the increased cost of future credit. Thus, the inequality defined in equation (3) suggests that the default decision is determined by the individual's evaluation of the benefits of relinquishing the debt relative to the transactions costs associated with default and the loss of any collateral.

Definition of the default decision in this way directly informs the traditional default model, which defines default as the result of several factors:

Collateral: Borrowers considering default weigh the outstanding value of the loan against the collateral lost through the default process. Mortgage defaults, in particular, have been shown to be driven by the current loan-to-value ratio.

Transactions Costs: Borrowers considering default also consider the associated transactions costs in weighing the value of disposing their debt against the costs of default. For credit products such as credit cards that do not require collateral, the transactions costs of default are thought to be sufficiently large to induce payment.

Ability to Pay: Consumers responses to the incentives shown in equation (3) have been shown to be conditional on the borrowers' continued ability to pay. Interruptions such as job loss and divorce lead to default among some borrowers who otherwise would continue to make payments.

Consumer Attitudes: Consumers' responses to the financial incentive to default are moderated by internal attitudes toward default. In particular, the presence of social stigma attached to default, particularly bankruptcy and home foreclosure, may lead some borrowers to continue making payments when it is financial disadvantageous to do so.

While several of the characteristics in Table 1 may roughly reflect the theoretical determinants of default, they are also likely to reflect multiple other characteristics. As a result, the nature of the risk reflected by credit scores is ambiguously defined at best. Hollis Fishelson-Holstine of Fair Isaac asserts that the credit scoring process "analyzes all available relevant information to deliver a single score: a number that represents the risk (or odds of positive repayment) for a particular individual" (Fishelson-Holstine 2005). This general definition of default risk underscores the ambiguity of what is captured by generic credit history scores. The credit scoring algorithm is not specified to isolate the effect of a borrower's previous ability to pay or attitudes about default. Instead, the risk reflected by credit scores can only be defined generally as the likelihood of a serious delinquency on *any* open credit account, as predicted by the measures in Table 1.

This general definition of default risk increases the significance of credit scores' reliance on universal measures of default and credit characteristics. The credit scoring model defined by equation (1) predicts universal default on the basis of universal measures of both default and credit characteristics. Weighting strategies are used to adjust for differences between mortgages, auto loans, and consumer credit products. However, credit records do not allow scorecard developers to observe variation in the risk characteristics of the underlying credit products. For instance, Quercia, Stegman, and Davis (2007) show that prepayment penalties and balloon payments, characteristics common in many subprime mortgages, are associated with elevated default rates, even after controlling for the risk characteristics used to underwrite the mortgage. This observation suggests that estimation of equation (1) using universal measures of default and credit characteristics identifies both differences in individuals' risk of default on a given credit product and differences between the credit products used by different individuals. This result carries direct implications for consumers' access to credit, particularly in light of the growing diversity of higher-cost and non-traditional credit products.

Universal Risk Measures, Differential Effects, and Credit Score Outcomes

The mechanism through which the credit scoring algorithm captures disparities in borrower assignment to credit products is distinct from the traditional definition of a differential effect. Nevertheless, a brief discussion of differential effects is necessary to inform the analysis and its relevance to previous examinations of disparate outcomes in credit scoring.

The empirical methods for identifying differential effects were largely developed in response to the discussion of discrimination in loan denials (see Turner and Skidmore 1999; Ross and Yinger 2002), as researchers sought methods to test for the presence of disparate impact discrimination among lenders. The general form for a loan default model can be represented as follows:

(3) $$Y_i = X_i\beta + \varepsilon_i$$

where Y_i is a measure of loan performance such as a previous default, X_i is a vector of variables that predict default risk, and ε_i is the error term. A loan performance score can be derived following estimation of equation (3) by calculating:

(4) $$Score_i = f(\hat{Y})_i = f(X_i\hat{\beta})$$

where $\hat{\beta}$ is a vector of the coefficient estimates derived from estimation of equation (3), and f() is a function that assigns performance scores based on the predicted likelihood of default. Ross and Yinger (2002) argue that scores derived from models such as equation (3) may create differential effects if any of the risk characteristics in X_i is strongly associated with membership in a protected class.[8] Because the group membership variable is omitted, the variable X_i acts as a proxy for group membership to the extent that the omitted variable biases the associated coefficient β.

[8] The Equal Credit Opportunity Act and other federal anti-discrimination legislation defines multiple characteristics that cannot be used to distinguish between individuals. These protected classes include race, ethnicity, religion, country of origin, age, gender, family status, and disability.

In contrast to the above model, Ross and Yinger (2002) present a neutral scoring model that corrects for the omitted variables problem above:

(5) $Y_i = X_i\beta + M_i\delta + \varepsilon_i$

where Y_i is a measure of loan performance such as a previous default, X_i is a vector of variables that predict default risk, M_i is a set of indicator variables that reflect whether the individual is a member of each protected class, and ε_i is the error term. Neutral credit scores can be derived from equation (5) as:

(6) $Score_i = f(\hat{Y})_i = f(X_i\hat{\beta})$

Where $\hat{\beta}$ are the coefficient estimates derived from estimation of equation (5), and f() is a function that assigns credit scores based on the predicted likelihood of default. For simplicity, the model represented in equation (3) is termed the 'basic model' and the model represented in equation (5) is termed the 'neutral model.'

The credit bureaus and Fair Isaac calculate credit scores using the basic model defined in equation (1), as neither directly observes or records membership in several of the protected classes. The Fair and Accurate Credit Transactions Act of 2003 (FACTA) instructed the Federal Reserve Board to examine the potential existence of differential effects in credit scoring with respect to race/ethnicity, gender, age, immigrant status, and other protected classes. The resulting Board of Governors (2007) analysis includes a set of scorecards based on basic models of credit scoring that are asserted to mirror those used by the credit bureaus. The authors then contrast the scores derived from this model with scores derived from scorecards based on neutral models, using the comparison of coefficient weights to evaluate the presence of differential effects.[9] The difference between the basic and neutral models accords with the traditional definition of differential effect, whereby one or more of the independent variables acts as a proxy for group membership. To the extent that the coefficient on the proxy variable is biased in the basic model, the credit scores of all group members will be negatively affected.

The Board of Governors' (2007) analysis of differential effects in credit scoring provides the seminal analysis of the determinants of disparities in credit score outcomes. The resulting report documents the sizeable differences in credit scores across demographic groups, but attributes these gaps to differences in outstanding balance and payment history variables. In contrast, the researchers conclude that minimal evidence suggests that differential effects exist with respect to race/ethnicity, age, gender, or immigrant status, finding only that the length of credit history variables may act as a partial proxy for age and may understate the creditworthiness of recent immigrants.

The analysis in this article examines a related but distinct phenomenon. Beyond the presence of differential effects, disparities in credit score outcomes may result from the

[9] The Board of Governors (2007) analysis compares results from the base model to two neutral models. The first is comparable to the neutral model in equation (3), where an indicator variable is inserted in the scoring equation to adjust for membership in a protected class. The second neutral approach stratifies the sample by membership in the protected class and estimates the base model separately for the sample of members and for the sample of non-members, comparing coefficients across models.

interaction of credit scoring with the pathways through which borrowers select alternative credit products. Specifically, the analysis in this article examines the potential for the universal credit characteristic measures to proxy for the processes that assign borrowers to different types of credit products. In particular, measures of outstanding debt and/or previous payment history may proxy for the processes that determine borrower assignment to low- and high-cost credit products. To the extent that disparities in assignment are associated with both the universal credit characteristics and the universal measure of default, the resulting credit scores will reflect the processes of product assignment in addition to borrowers' underlying risk.

Put another way, disparities in credit score outcomes are shown to result when two conditions are present: (1) disparities exist in the processes through which borrowers are assigned to credit products and (2) alternative credit characteristics carry varying risks of default. It must be emphasized that the resulting disparities in credit score outcomes do not necessarily imply disparate impact discrimination, which carry specific legal standards regarding the identification of differential effects with respect to specific protected characteristics. Instead, this process is used more broadly to define a conceptual model for understanding the role of credit scoring in determining borrowers' access to low- and higher-cost credit.

To illustrate the consequences of the use of universal measures, the remainder of this article presents a stylized model that uses simulated data to clarify the processes through which disparate credit score outcomes arise. While the simulations are intended to reflect the actual operations of credit markets, the analysis seeks only to illustrate the mechanisms through which disparities in credit scores can be produced. The simulation results imply disparities in actual credit scores only to the extent that the simplified processes outlined below are replicated in the market. It should also be noted that the simulations define disparities to be between minorities and non-minorities, but that the mechanism illustrated is equally applicable to other characteristics. This article's focus on minority status is merely a reflection of the sizeable body of literature discussing discrimination in existing credit markets against minority individuals and neighborhoods (see Turner and Skidmore 1999).

Data

The datasets for analysis are created through two simulations that are designed to isolate the effect of disparate assignment to high- and low-cost credit products on credit score outcomes. Each simulation structures a simple model of the credit scoring process where disparity is imposed at the individual and neighborhood levels in the assignment of borrowers to credit products. The use of simulated data is useful in this pursuit, as it allows for the causal effect of the discrimination to be followed directly into the calculation of credit scores. The use of simulated data also clarifies the assumptions necessary for disparities in product assignment to be translated into disparate credit score outcomes.

The limitation of the resulting data is that inferences from the simulated data cannot be directly extrapolated to actual credit scores. As noted previously, the mechanism illustrated in this analysis implies disparities in actual credit scores only to the extent that the necessary assumptions are met in actual data. Specifically, the model imposes the two assumptions noted above:

1. Disparities, including but not limited to discrimination, must exist in the processes through which borrowers are assigned to low- and high-cost credit products.
2. Borrowers with similar underlying risk characteristics must be more likely to default on a high-cost credit product than on a low-cost credit product.

The latter assumption is consistent with the role of collateral and ability to pay in the theoretical model of default behavior. It is also supported by the importance of loan-to-value and debt-to-income ratios in empirical analyses of mortgage default.[10] The validity of the first assumption is discussed in greater detail in a later section.

<u>Simulation 1: Assignment to Product Types Differs by Race</u>[11]

The first dataset includes 100,000 observations with random assignment of 50,000 observations each to minority and non-minority status. The underlying risk of default for each observation is also randomly assigned with the distribution following a standard normal distribution.[12] Individuals are observed in two time periods.

Period 1: Lenders underwrite two credit products, a low-cost product and a high-cost product, on the basis of applicants' observed risks. Risks are observed with random error, and the observed risk variable is calculated as the sum of the individual's underlying risk and a random error term with a standard normal distribution.

Assignment to low- and high-cost products occurs with directly-imposed disparities, with the highest 30 percent of observed risk values among minority households and the highest 20 percent of observed risk values among non-minority households assigned to the high-cost product.[13]

The occurrence of default is random within products, with 10 percent of individuals with the low-cost product defaulting and 20 percent of individuals with the high-cost product defaulting.[14]

Period 2: Lenders again underwrite two credit products, a low-cost product and a high-cost product, on the basis of applicants' observed risks. The risks observed by lenders are again calculated as the sum of the individual's underlying risk and a

---

[10] A product with a higher interest rate by definition either reduces the speed of equity accumulation or increases the monthly payment obligation. With the exception of exploding ARMs, higher-cost products generally impose higher monthly payment obligations.

[11] The STATA code for each simulation is available from the author upon request.

[12] The distribution of the underlying risk variable is similar for minority and non-minority borrowers, as the underlying risk and minority variables are uncorrelated in the simulation dataset.

[13] While the disparity is labeled discrimination in this simulation, it could also be created if minority borrowers are disproportionately less likely to search for the best mortgage terms.

[14] A random default process within products is chosen for simplicity. Default could also be some function of underlying risk, the occurrence of unexpected events, and the product selected. The necessary assumption is that an individual's risk of default is higher with the bad product than the good product. In practice, higher-cost products often increase the monthly debt burden of the borrower and/or accrue equity more slowly than lower-cost products, either of which could increase the product's default rate.

random error term with a standard normal distribution.[15] Lenders also observe whether the individual defaulted in period 1.

Assignment to low- and high-cost products occurs in the same way, with the highest 30 percent of observed risk values for minority households and the highest 20 percent of observed risk values for non-minority households assigned to the high-cost product. Lenders also assign all households that defaulted in period 1 to the high-cost product.

The occurrence of default is again random within products, with 10 percent of individuals with the low-cost product defaulting and 20 percent of individuals with the high-cost product defaulting.

<u>Simulation 2: Assignment to Product Types Differs Across Neighborhoods</u>
The second dataset includes 100,000 observations with random underlying risk according to a standard normal distribution. Individuals are observed over two time periods in two distinct neighborhoods. The racial composition of Neighborhood 1 is equal, with a 50/50 share of minorities and non-minorities. In contrast, only 10 percent of the residents of Neighborhood 2 are minority, and 90 percent are non-minority. Minority status in both neighborhoods is again independent of the underlying risk variable.

Period 1: Lenders underwrite two credit products, a low-cost product and a high-cost product, on the basis of applicants' observed risks. Risks are observed with random error, and are calculated as the sum of the individual's underlying risk and a random error term with a standard normal distribution.

In this simulation, assignment to low- and high-cost products is influenced by neighborhood, with the highest 50 percent of observed risk values among Neighborhood 1 residents and the highest 20 percent of observed risk values among Neighborhood 2 residents assigned to the high-cost product.

The occurrence of default is random within products, with 10 percent of individuals with the low-cost product defaulting and 20 percent of individuals with the high-cost product defaulting.

Period2: Lenders again underwrite two credit products, a low-cost product and a high-cost product, on the basis of applicants' observed risks. The risks observed by lenders are again calculated as the sum of the individual's underlying risk and a random error term with a standard normal distribution. Lenders also observe whether the individual defaulted in period 1.

Assignment to low- and high-cost products occurs in the same way, with the highest 50 percent of observed risk values among Neighborhood 1 residents and the highest

[15] The random error used in this calculation in period 2 is independent of the random error variable used in period 1. Thus, a different but overlapping segment of the market receives the bad product in period 2.

20 percent of observed risk values among Neighborhood 2 residents assigned to the high-cost product. Lenders also assign all households that defaulted in period 1 to the high-cost product.

The occurrence of default is again random within products, with 10 percent of individuals with the low-cost product defaulting and 20 percent of individuals with the high-cost product defaulting.

Evaluation and Discussion

The results are evaluated with respect to the relative default rates and credit scores of minorities and non-minorities. First, the descriptive statistics presented in Table 2 offer additional insight into the simulations described in the previous section. The underlying risk variable in each simulation shows a standard normal distribution, and the minority status variables exhibit the proportions defined by the assignment rules. Similarly, the relative rates of default in each period directly reflect the simulated process. In particular, the higher rates of default in period 2 occur because all individuals who default in the first period are assigned to the high-cost product.

[INSERT TABLE 2 ROUGHLY HERE]

Table 3 presents the results of probit regressions of default on the underlying risk and minority status variables for both the basic and neutral models described in equation (1) and equation (3), respectively. The results are shown both for default in period 1 and default in period 2. For the period 2 regression, previous default is also included as a regressor to reflect the payment history variables used in credit scoring models. Credit scores are then calculated using the estimated coefficients of the underlying risk and previous default variables. For each model, the predicted likelihood of default for each individual is subtracted from one, generating credit scores that increase as the borrower's predicted default risk decreases.

The results for the period 1 models offer insight into the setup of the simulated process. Table 3 shows that underlying risk and minority status positively predict default following period 1. The coefficients on the underlying risk variable are also very similar in the basic and neutral models, as the risk and minority variables are uncorrelated. Table 4 presents t-tests comparing the mean credit scores of minority and non-minority individuals. The t-test results show that credit scores derived from both the basic and the neutral model following period 1 do not differ by minority status. This result stems from the definition of credit scores, which are calculated using only the underlying risk variable following period 1. Because the distribution of risk is identical for minority and non-minority borrowers, the resulting credit scores are also similar.

[INSERT TABLES 3 AND 4 ROUGHLY HERE]

The results for period 2 offer evidence of the consequences of the conditions imposed on the simulation. The credit scoring models estimated for period 2 include a previous default variable, which reflects whether the individual defaulted in period 1. The period

2 credit scores are then calculated on the basis of both the underlying risk and previous default variable.[16] The estimation models show that default in period 2 is predicted by previous default, underlying risk, and minority status. The t-test results presented in Table 4 for the period 2 model show that the credit scores of minorities are significantly lower than those of non-minorities. Because the distribution of the underlying risk variable is identical across minorities and non-minorities, the credit score differences are the result of the previous default variable capturing differential assignment to the low- and high-cost products. In this way, the credit scores reflect disparities in assignment through the payment history variables. Even in the neutral model, this effect persists because the values of the previous default variable are used in the calculation of credit scores to reflect previous payment history.

The results of Simulation 2 mirror those produced by Simulation 1. The period 2 models show that default in period 2 is predicted by underlying risk, minority status, and previous default (Table 5). Table 6 shows the results of t-tests of the mean credit scores of minorities and non-minorities. The credit scores for minorities and non-minorities again do not differ significantly after period 1, but show a significant difference after period 2 when the previous default variable is used to calculate scores. Each of the findings is robust to the use of the basic or neutral model.

[INSERT TABLES 5 AND 6 ROUGHLY HERE]

These results can also be evaluated with respect to the criteria used in the Board of Governors (2007) analysis. The Federal Reserve analysis uses multiple evaluative criteria, examining mean scores, performance residuals, model fit, and the correlation of membership in a protected class with the variables included in their credit scoring model. However, the definitive assessment comes from comparison of the coefficient estimates generated from a basic model with those generated from a neutral model.[17] A similar evaluation of the simulation results involves comparing the estimated coefficients of the default variables across the basic and neutral models.[18] For both Simulation 1 and Simulation 2, the coefficient on the default variable is slightly lower in the basic model than in the neutral model. This result accords with the results reported in the Board of Governors (2007) analysis, where estimation of the neutral model produced results that differ only minimally from those in the basic model.[19] The bivariate correlation of

[16] The credit scores are derived directly from the predicted likelihood of default using the estimated coefficients from the period 2 models. The predicted probability is then transformed, so that higher credit scores are assigned to lower default probabilities: $Score_i = 1 - \hat{Y}_i = 1 - X_i\hat{\beta}$.

[17] A base model is compared to two neutral models. The first is comparable to the neutral model estimated here, using a minority indicator. The second neutral approach stratifies the sample by membership in the protected class and estimates the base model separately for the sample of members and for the sample of non-members, comparing coefficients across models.

[18] Because the underlying risk and minority status variables are uncorrelated by design, the risk variable does not act as a proxy for minority status.

[19] The effects produced in the simulations in this comment result from the use of the default variable in the calculation of period 2 credit scores. The Board of Governors (2007) analysis concludes that the observed difference in the mean credit scores of white and black households results primarily from

minority status with period 1 default also remains well within the patterns documented in the Federal Reserve analysis (ρ=0.015).

Implications for Existing Credit Markets

The evidence provided in this paper is best described as an existence proof. Because no existing data source—including credit bureaus' proprietary records—allows investigation of this phenomenon, the analysis relies on simulated data. Nonetheless, the analysis implies that disparity in credit scores will emerge to the extent that the conditions imposed on the simulated dataset are replicated in actual credit markets. In practice, this requires that both of the assumptions presented in the data section must hold. First, discrimination and/or disparities must exist in the processes through which borrowers are assigned to low- and high-cost credit products. Second, borrowers with similar underlying risk characteristics must be more likely to default on a high-cost credit product than on a low-cost credit product.

The second assumption is directly supported by traditional underwriting practices, in which lending risk is measured by the loan-to-value ratio and the debt-to-income ratio. Higher-cost products by definition either increase the borrower's monthly payment obligations or slow the speed at which the borrower accumulates equity. Moreover, mounting evidence documents differences in the default risk associated with specific loan terms. For instance, Quercia, Stegman, and Davis (2007) show that the presence of prepayment penalties and balloon payments are associated with an elevated likelihood of default even after controlling for the risk characteristics used to underwrite the mortgage. Several additional studies further document the unique risk characteristics of hybrid ARMs and other subprime mortgage products (Ambrose, LaCour-Little, and Huszar 2005; Gerardi, Shapiro, and Willen 2007; Ding et.al. 2008a).

The veracity of the simulated model therefore depends on the first assumption, namely the extent to which disparities exist in product assignment. In the simulated models, disparity is directly imposed in a manner that is suggestive of discrimination. However, the only requirement is that minority and non-minority borrowers are assigned to the high-cost product at different rates. This could occur directly through discrimination.[20] However, the defined disparities may also arise through more subtle mechanisms related to the pathways through which borrowers learn about and select credit products. Given the growing information about the targeted marketing activities of mortgage brokers and subprime lenders, the concentration of subprime foreclosures in underserved neighborhoods raises concerns about increasing credit score disparities.[21]

---

the payment history variables. While such differences in empirical data likely reflect disparities in wealth and other socio-economic variables, they are not inconsistent with the type of effects described here.

[20] A large literature has discrimination in mortgage lending (see Turner and Skidmore 1999; Munnell et.al. 1996)

[21] Several recent studies conclusively document the disproportionate concentration of subprime mortgages in lower-income and minority communities (Calem, Gillen, and Wachter 2004; Calem, Hershaff, and Wachter 2004; Ding et.al. 2008b). Courchane, Surette, and Zorn (2004) also examine the extent to which borrowers get 'stuck in higher-cost markets.

Implicit in the simulated model is a feedback loop by which credit scores impact future credit options and thus also the individual's future credit characteristics and score. This feedback loop speaks directly to concerns over whether borrowers get 'stuck' in higher-cost credit markets (see Courchane, Surette, and Zorn 2004). It also raises the potential for credit scores to prolong the consequences of the current foreclosure crisis. As foreclosures mount, the impacts are being felt most in lower-income and minority communities. To the extent that credit scores reflect these outcomes, the concentrated effects of the foreclosure crisis will persist into the future.

The geographic nature of this process adds an additional dimension to this process. In addition to the households holding subprime ARMs, the spillover effects of foreclosures extend to the home prices and default rates of neighboring homeowners.[22] To the extent that accumulated equity and geographic location impact the future credit options of neighboring homeowners, the impact on consumer credit scores may include geographic effects that extend to individuals who diligently searched for low-risk prime credit options.

Future research is desperately needed in order to understand the extent to which the process described in this article contributes to observed gaps in consumer credit scores. While policymakers have taken initial steps to test for disparate impact in the calculation of credit scores, the widespread use of risk-based pricing implies that additional sources of disparity may be implicit in the collection of credit data and credit records. As a result, it is critical that researchers and policymakers understand the potential for each component of the credit scoring process to create and/or institutionalize disparities in credit outcomes.

Given the increasing centrality of credit scores to consumers' economic lives, this analysis also echoes the discussion over consumers' ability to advocate for themselves with respect to the credit scoring process. To date, the proprietary nature of credit data and the credit scoring algorithm has largely impeded consumers' abilities to understand and discuss the implications of credit scoring. As the influence of credit scoring increasingly influences consumer outcomes outside of access to financial services, policymakers must address both the credit scoring process itself and the role of consumer advocates and independent researchers in informing regulation and policy.

[22] An emerging literature examines the spillover effects of subprime foreclosures on neighboring homeowners (Immergluck and Smith 2006; Ding, Quercia, and Ratcliffe 2008).

References

Ambrose, Brent, Michael LaCour-Little, and Zsuzsa R. Huszar. "A Note on Hybrid Mortgages." *Real Estate Economics* 33 no.4 (2005): 765-782.

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

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 2003. "An Overview of Consumer Data and Credit Reporting." *Federal Reserve Bulletin* 89(2): 47-73.

Avery, Robert B., Paul S. Calem, and Glenn B. Canner. 2004. "Credit Report Accuracy and Access to Credit." *Federal Reserve Bulletin* 90(3): 297-322.

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." Report to Congress of the Federal Reserve Board, August, 2007.

Bostic, Raphael W., Paul S. Calem, and Susan M. Wachter. 2005. "Hitting the Wall: Credit as an Impediment to Homeownership." In Nicolas P. Retsinas and Eric S. Belsky. *Building Assets, Building Credit: Creating Wealth in Low-Income Communities*. Washington DC: Brookings Institution Press.

Calem, Paul S., Kevin Gillen, and Susan Wachter. "The Neighborhood Distribution of Subprime Mortgage Lending." *Journal of Real Estate Finance and Economics* 29 no.4 (2004): 393-410.

Calem, Paul S. Jonathan E. Herschaff, and Susan M. Wachter. "Neighborhood Patterns of Subprime Lending: Evidence from Disparate Cities." *Housing Policy Debate* 15 no.3 (2004): 603-622.

Clapp, J.M., Goldberg, G.M., Harding, J.P. & LaCour-Little, M. 2001. "Movers and Shuckers: Interdependent Prepayment Decisions." *Real Estate Economics* 29(3), 411-450.

Courchane, Marsha J., Brian J. Surette, and Peter M. Zorn. "Subprime Borrowers: Mortgage Transitions and Outcomes." *Journal of Real Estate Finance and Economics* 29 no.4 (2004): 365-392.

Deng, Y., Pavlov, A.D., & Yang, L. 2005. "Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale, and Default." *Real Estate Economics* 33(4), 739-764.

Ding, Lei, Roberto Quercia, Wei Li, and Janneke Ratcliffe. 2008a. "Risky Borrowers or Risky Mortgages: Disaggregating Effects Using Propensity Score Models." Center for Community Capital Working Paper, October.

Ding, Lei, Janneke Ratcliffe, Michael A. Stegman, and Roberto G. Quercia. 2008b. "Neighborhood Patterns of High-priced Lending: The Case of Atlanta." *Journal of Affordable Housing and Community Development Law* 17(3): 193-218.

Ding, Lei, Roberto G. Quercia, and Janneke Ratcliffe. 2008. "The Spillover Effect of Neighborhood Subprime Lending." Center for Community Capital Working Paper, May.

Federal Trade Comission. 2007. "Credit Based Insurance Scores: Impacts on Consumers of Automobile Insurance." Report to Congress by the Federal Trade Commission, July 2007.

Fichelson-Holstine, Hollis. 2005. "Credit Scoring's Role in Increasing Homeownership for Underserved Populations." In Nicolas P. Retsinas and Eric S. Belsky. *Building Assets, Building Credit: Creating Wealth in Low-Income Communities*. Washington DC: Brookings Institution Press.

Gates, Susan Wharton, Vanessa Gail Perry, and Peter M. Zorn. "Automated Underwriting in Mortgage Lending: Good News for the Underserved?" *Housing Policy Debate* 13 no.2 (2002): 369-391.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Federal Reserve Bank of Boston Working Paper #07-15, December 2007.

Hunt, Robert M. 2002. "The Development and Regulation of Consumer Credit Reporting in America." Federal Reserve Bank of Philadelphia. Working Paper no. 02-21. November, 2002.

Immergluck, Dan and Geoff Smith. 2006. "The External Costs of Foreclosure: The Impact of Single-Family Mortgage Foreclosures on Property Values." *Housing Policy Debate* 17(1): 57-79.

Lyons, Angela C., Mitchell Rachlis, and Erik Scherpf. "What's in a Score? Differences in Consumers' Credit Knowledge Using OLS and Quantile Regressions." *Journal of Consumer Affairs* 41 no.2 (2007): 223-249.

Munnell, Alicia H., Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86 no.1 (1996): 25-53.

Political and Economic Research Council. 2006. "Give Credit Where Credit is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data." Brookings Institution Urban Markets Initiative.

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

Quercia, Roberto G, Michael A. Stegman and Walter R. Davis. 2007. "The Impact of Predatory Loan Terms on Subprime Foreclosures: The Special Case of Prepayment Penalties and Balloon Payments." *Housing Policy Debate* 18(2): 311-346.

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

Turner, Margery Austin and Felicity Skidmore. 1999. *Mortgage Lending Discrimination: A Review of Existing Evidence.* Washington DC: The Urban Institute.

TABLES

Table 1: The 19 credit characteristics selected from the TransUnion database for use in the FRB base model scorecards (thin file, clean file, major derogatory)

	Credit Characteristic:
1	Total number of accounts in good standing, opened 18 or more months ago
2	Total maximum credit issued on open accounts reported in the past 12 months
3	Total number of months since the most recent account delinquency
4	Percentage of total remaining balance to total maximum credit for all open revolving accounts reported in the past 12 months
5	Percentage of total remaining balance to total maximum credit for all open installment accounts reported in the past 12 months
6	Percentage of total remaining balance to total maximum credit for all open bankcard accounts reported in the past 12 months
7	Percentage of accounts with no late payments reported
8	Number of accounts that have payments that are currently or previously 30 or more days past due within the past 24 months
9	Total number of accounts currently less than 120 days past due in the past 2 months
10	Greatest amount of time a payment was late ever on an account
11	Total number of months since the most recent occurrence of a derogatory public account
12	Total number of inquiries for credit
13	Total number of months since the most recent update on an account
14	Average age of accounts on credit report
15	Total number of open personal finance installment accounts reported in the past 12 months
16	Total number of open non-installment accounts with a remaining balance to maximum credit issued ratio greater than 50 percent reported din the past 12 months
17	Percentage of accounts that are open and active with a remaining balance greater than $0 reported in the past 12 months
18	Total number of different credit issuers
19	Total number of public records and derogatory accounts with an amount owed greater than $100

Table 2: Descriptive Statistics

	Simulation 1		Simulation 2	
	Mean	**Std Dev**	**Mean**	**Std Dev**
Underlying Risk	-.0001	.9997	.0015	1.0025
Minority Status	.5000	.5000	.3000	.4583
Default: Period 1	.1250	.3307	.1350	.3417
Default: Period 2	.1337	.3404	.1428	.3499

Table 3: Probit Regression of Basic and Neutral Models for Simulation 1

Dependent Variable:	**Period 1 Default**		**Period 1 Default**		**Period 2 Default**		**Period 2 Default**	
Model:	**Basic**		**Neutral**		**Basic**		**Neutral**	
	Coef.	**z**	**Coef.**	**z**	**Coef.**	**z**	**Coef.**	**z**
Underlying Risk	.109	21.46	.109	21.44	.084	16.81	.084	16.80
Minority Status			.049	4.78			.057	5.71
Previous Default					.299	21.48	.298	21.40

Table 4: T-tests of Mean Credit Score Differences for Simulation 1

	Mean		T-test
	Minority	**Non-Minority**	**p-value**
<u>*Period 1:*</u>			
Basic	1.000	1.000	.81
Neutral	1.000	1.000	.81
<u>*Period 2:*</u>			
Basic	.961	.964	<.001**
Neutral	.961	.964	<.001**

Table 5: Probit Regression of Basic and Neutral Models for Simulation 2

Dependent Variable:	**Period 1 Default**		**Period 1 Default**		**Period 2 Default**		**Period 2 Default**	
Model:	**Basic**		**Neutral**		**Basic**		**Neutral**	
	Coef.	**z**	**Coef.**	**Z**	**Coef.**	**z**	**Coef.**	**z**
Underlying Risk	.108	21.63	.108	21.62	.092	18.58	.092	18.58
Minority Status			.052	4.79			.049	4.63
Previous Default					.226	16.72	.225	16.66

Table 6: T-tests of Mean Credit Score Differences for Simulation 2

	Mean		**T-test**
	Minority	**Non-Minority**	**p-value**
Period 1:			
Basic	1.000	1.000	.83
Neutral	1.000	1.000	.83
Period 2:			
Basic	.967	.970	<.001**
Neutral	.968	.970	.002**