NATIONAL COMMUNITY REINVESTMENT COALITION NCRC

August 10, 2004

Federal Trade Commission/Office of the Secretary
Room H-159 (Annex N)
600 Pennsylvania Avenue, NW
Washington DC 20580

RE: FACT Act Scores Study, Matter No. P044804

To Whom it May Concern:

The National Community Reinvestment Coalition (NCRC), the nation's economic justice trade association of 600 community organizations, urges the Federal Trade Commission (FTC) and the Federal Reserve Board (FRB) to conduct a comprehensive study on the impacts of credit scores on the availability and affordability of loans and insurance products for African Americans and Latino individuals and communities. An impartial and rigorous study should influence public policy regarding if and how to change the nation's credit scoring system in order to promote fairer access to credit and insurance.

NCRC recently conducted a study, the *Broken Credit System*, that assessed the relationship among creditworthiness, neighborhood racial and age composition, and housing stock characteristics (we have attached a copy of the study for your reference). NCRC obtained credit score information on a census tract level on a one-time basis from one of the three major credit bureaus. The study used proxies for race, income, and age of the borrower since NCRC is not aware of any database that merges Home Mortgage Disclosure Act (HMDA) data and credit scores. The proxies in the study were the race and age composition of census tracts. We believe that the proxies were reasonable and suggested a pervasive problem of price discrimination.

NCRC found that after controlling for creditworthiness and housing stock characteristics, the percentage of Subprime loans increased as the percentage of African-American and elderly residents increased in a census tract. These results occurred in ten large metropolitan areas and were particularly strong for African-Americans. We also found that Subprime lending increased as the percentage of neighborhood residents with supposedly higher credit risks increased.

Although the supposed higher risk translated into higher levels of Subprime lending, NCRC's study reveals a basic unfairness in the financial system in America. High cost loans surged in minority and elderly neighborhoods even after controlling for creditworthiness. This finding is consistent with a study conducted by a Federal Reserve economist using creditworthiness information similar to the data NCRC used.[1]

Nationwide testing conducted by NCRC of Subprime lenders revealed that African American loan applicants were steered to higher priced loans despite the fact that they indicated that they had higher credit scores then their White counterparts. The results of this testing and the above mentioned studies indicate that African Americans are both steered to, and targeted for higher priced loans than their credit scores would qualify them.

The FTC and the FRB study will make a positive contribution if it thoroughly investigates the factors in credit score systems and alternatives to credit scoring. Firstly, we urge the FTC and FRB to work with the credit bureaus to create a publicly available database with the major factors or variables used by credit scoring systems. These variables can then be included in multivariate equations similar to those in the NCRC and FRB studies. An analysis can be undertaken to determine which variables in credit score systems have the most influence on the availability of prime and sub-prime credit. Policy judgments should then be made regarding the variables with statistically significant impacts. For example, it has been revealed that credit score systems negatively consider consumers' use of finance companies. If the use of finance companies is a statistically significant variable in credit score systems, this would not be fair since finance companies are disproportionately located in minority neighborhoods, and represent the major choice for too many consumers. A second important example is that it has been revealed that the neighborhood of residence affects ones credit score. If true, we believe that this would have a negative and disparate impact based on race and national origin.

The FTC and FRB study should explore alternatives to credit scores. Some banks have expanded prime lending to traditionally underserved borrowers by using rental and utility payment histories, which are not variables included in credit scores. The FTC and FRB should survey banks using flexible underwriting and then assess if the replication of these approaches can increase fair access to credit.

Another major area of concern for us is that the Fair Isaac Corporation recently created a new Subprime credit score designed specifically for the "underserved market". Fair Isaac also created a new subsidiary to implement this new system in the underserved marketplace. The underserved market encompasses approximately 54 million Americans who have no credit records on file that will allow financial institutions to quickly evaluate the risk of lending to them. Usually, these populations lack credit history, and consequently are not eligible or qualified to obtain credit cards, mortgages and/or loans. Included in this group are recent immigrants, members of ethnic groups that traditionally have not used credit, the newly divorced or widowed, and the young.

[1] Paul S. Calem, Kevin Gillen, and Susan Wachter, *The Neighborhood Distribution of Subprime Mortgage Lending*, October 30, 2002. Available via pcalem@frb.gov.

This new “FICO Expansion score” for the underserved population’s credit score is based on FICO’s terminology “nontraditional data sources”. The criteria used ranges from how effective consumers handle payday loans and retail payment plans to whether they used overdraft protection on their checking accounts responsibly. NCRC believes that the use of these inherently negative, abusive, wealth-stripping products, (particularly payday loans and bounced check/loan protection services) to determine credit worthiness is problematic at best and more than likely will have a disparate impact on the basis of race and national origin and perhaps other classes protected by ECOA and the Fair Housing Act.

Let’s explore bounced check/loan protection services as an example. These loans disproportionately impact a small percentage of consumers who are usually minority, low-income and vulnerable. According to a survey conducted by the Consumer Federation of America (CFA) 28% of consumers overdraw their accounts. Of those who were most likely to overdraw their bank accounts, the results reflected the following: (1) African Americans (45%); 2) Moderate income consumers with household incomes of $25,000 to $50,000 (37%); 3) Those 25 to 44 years of age (36%) were most likely to have done so.

In reviewing this data alone, consumers most likely to be hurt by this new Subprime credit score product will be minorities, the young, and the poor and vulnerable. Lenders recognize that there is extensive growth in this underserved market (70%). Fair Isaac appears to have produced a Subprime product that bases its standards on negative products that strip equity from this population of people. The impetus driving this force appears to be increased revenues. Fair Isaac has come up with a quick fix solution to tap a potentially lucrative underserved population and make them the new high cost Subprime consumers. Fair Isaac has already marketed this design to auto and mortgage lenders who offer sub-prime products.

This arbitrary and capricious credit scoring design will adversely impact the minority, young, poor, and otherwise underserved populations. This population has little understanding of such a system, no voice to oppose it, and no input into a credit score system, which will ultimately control not only their access to credit, but also their access to insurance, employment, and housing as well.

The credit scoring industry argues that their systems are proprietary and the value and weight used in determining credit scores is proprietary as well. This argument for protecting the proprietary interests of the credit scoring industry rings hollow when we consider that every American’s future is determined by the use of some variation of a credit score. Clearly a consumers right to know how this important score is computed and weighted, outweighs any proprietary claims of the credit scoring industry.

In order for the FRB to have an effective study, full disclosure by the credit scoring industry as well as industry data must be included to assess the viability of their system. The Federal government has the right to demand full disclosure from all credit reporting agencies and the Fair Isaac Corporation regarding how an individual’s credit history is assessed and quantified, and the value and weight of the factors, which ultimately translate into a credit score, are calculated.

Finally, the FTC and FRB should expand their inquiries into credit-based insurance scores. A study by the state of Missouri has revealed a relationship between credit-based insurance scores and race of neighborhoods. Interestingly, some companies' credit scores were significantly less negative than others for residents in minority areas. Thus, credit scores of some insurance companies may not restrict choices of minorities as much as other companies' credit scores (the Missouri study is attached for your consideration). This suggests important differences in credit scoring systems that must be further investigated by the FTC and FRB study. Thirteen states, including Missouri, are about to embark upon an investigation of the credit-based insurance scores and the availability of insurance. NCRC urges cooperation between these states and the FTC and FRB.

In conclusion, NCRC shares the concerns of many other civil rights and consumer protection groups who fear that credit scores have become an increasingly dangerous tool which has been used by the lending industry to create a dual lending market. We remain concerned about the increasing incursion of a credit scoring system, which still remains secretive, into other industries including insurance, employment, and housing. We oppose the approach of creating a Subprime entry system through the use of a Subprime credit scoring system for those first entering the American credit system.

We strongly urge the FTC and the FRB to work for full public disclosure of the credit scoring models, and to thoroughly investigate the factors in the credit scoring system as well as alternatives to credit scoring.

Thank you for providing us with the opportunity to comment on this important matter. Please feel free to contact me on (202) 628-8866 if we can be of further assistance.

Sincerely,

John Taylor
President and CEO

credit score
discrimination
redlining
subprime lending
demographic variables
potential risk
advocate
counseling
unfair practices
community
enhance opportunity
data
African American
investment
regression analysis
elderly
neighborhood

[The Broken Credit System: Discrimination and Unequal Access to Affordable Loans by Race and Age]

Subprime Lending in Ten Large Metropolitan Areas

The National Community Reinvestment Coalition

The National Community Reinvestment Coalition (NCRC) is the nation's trade association for economic justice whose members consist of local community based organizations. Since its inception in 1990, NCRC has spearheaded the economic justice movement. NCRC's mission is to build wealth in traditionally underserved communities and bring low- and moderate-income populations across the country into the financial mainstream. NCRC members have constituents in every state in America, in both rural and urban areas.

The Board of Directors would like to express their appreciation to the NCRC professional staff who contributed to this publication and serve as a resource to all of us in the public and private sector who are committed to responsible lending. For more information, please contact:

John Taylor, President and CEO
David Berenbaum, Senior V.P. Policy and Director of Civil Rights
Joshua Silver, V.P. Policy and Research
Kelly Brinkley, Director of Legislative and Regulatory Affairs
Crystal Ford, Director of Legislative and Regulatory Affairs
Rachel Maleh, V.P. Communications

A special word of thanks to Mark Treskon, Milena Kornil, Josh Silver, and Dan Immergluck. As a former NCRC Research Analyst, Mark started this report and conducted the initial analysis that informed the methodology. Josh Silver and Milena Kornil teamed up to complete the data analysis and write the report narrative. Without their invaluable contributions, this report would not be as timely or comprehensive. Dr. Dan Immergluck, a professor at Grand Valley State University, provided expert peer review, consulting, and quick and thorough proofreading. His skilled assistance augmented the statistical rigor and meaning of the report.

NCRC

National Community Reinvestment Coalition
733 15th Street, NW
Suite 540
Washington, DC 20005
v: (202) 628-8866
f: (202) 628-9800
www.ncrc.org

Table of Contents

Executive Summary

The credit system is broken and discrimination is widespread in America. NCRC finds that African-American and predominantly elderly communities receive a considerably higher level of high cost subprime loans than is justified based on the credit risk of neighborhood residents. President Bush has declared an Administration's goal of 5.5 million new minority homeowners by the end of the decade. The widespread evidence of price discrimination, however, threatens the possibility of creating sustainable and affordable homeownership opportunities for residents of traditionally underserved neighborhoods.

The widespread evidence of price discrimination threatens the possibility of creating sustainable and affordable homeownership opportunities . . .

A subprime loan has an interest rate higher than prevailing and competitive rates in order to compensate for the added risk of lending to a borrower with impaired credit. NCRC defines a predatory loan as an unsuitable loan designed to exploit vulnerable and unsophisticated borrowers. Predatory loans are a subset of subprime loans. A predatory loan has one or more of the following features: 1) charges more in interest and fees than is required to cover the added risk of lending to borrowers with credit imperfections, 2) contains abusive terms and conditions that trap borrowers and lead to increased indebtedness, 3) does not take into account the borrower's ability to repay the loan, and 4) violates fair lending laws by targeting women, minorities and communities of color. Using the best available industry data on credit worthiness, NCRC uncovered a substantial amount of predatory lending involving rampant pricing discrimination and the targeting of minority and elderly communities.

Sadly, it is still the case in America that the lending marketplace is a dual marketplace, segmented by race and age. If a consumer lives in a predominantly minority community, he or she is much more likely to receive a high cost and discriminatory loan than a similarly qualified borrower in a white community. At the same time, the elderly, who have often built up substantial amounts of equity and wealth in their homes, are much more likely to receive a high cost refinance loan than a similarly qualified younger borrower. The disproportionate amount of subprime refinance lending in predominantly elderly neighborhoods imperils the stability of long-term wealth in communities and the possibilities of the elderly passing their wealth to the next generation.

Lending discrimination in the form of steering high cost loans to minorities and elderly borrowers qualified for market rate loans results in equity stripping and has contributed to inequalities in wealth. According to the Federal Reserve Survey of Consumer Finances, the median value of financial assets was $38,500 for whites, but only $7,200 for minorities in 2001. Whites have more than five times the dollar amount of financial assets than minorities. Likewise the median home value for whites was $130,000 and only $92,000 for minorities in 2001.[1]

This report confirms Americans' perceptions of bias in lending. In the winter of 2002, NCRC hired Republican pollster Frank Luntz and Democratic pollster Jennifer Laszlo Mizrahi to conduct a nationally representative poll of Americans' views of lending institutions. In the poll, fully 76 percent of Americans believed that steering creditworthy minorities and women to costly loan products was a significant problem. About 47

[1] Ana M. Aizcorbe, Arthur B. Kennickell, and Kevin B. Moore, *Recent Changes in U.S. Family Finances: Evidence from the 1998 and 2001 Survey of Consumer Finances*, Federal Reserve Bulletin, January 2003.

percent of the survey respondents believed that a white man would be more likely than an African-American man with the same credit history to be approved for a loan. Only 10 percent of the respondents believed that the African-American would be more likely to be approved for a loan. Among African-American survey respondents, 74 percent thought the white man would be approved, and only 3.6 percent thought that a similarly qualified African-American would be approved over the white man. Unfortunately, this report verifies that these perceptions of discriminatory treatment are reality in too many instances.[2]

The single most utilized defense of lenders and their trade associations concerning bias is that credit scoring systems allow lenders to be colorblind in their loan decisions. This study, the largest and among the first of its kind, debunks that argument and clearly makes the case that African-American and elderly neighborhoods, regardless of the creditworthiness of their residents, receive a disproportionate amount of high cost subprime loans.

NCRC selected ten large metropolitan areas for the analysis: Atlanta, Baltimore, Cleveland, Detroit, Houston, Los Angeles, Milwaukee, New York, St. Louis, and Washington, D.C. As expected, the amount of subprime loans increased as the amount of neighborhood residents in higher credit risk categories increased. After controlling for risk and housing market conditions, however, the race and age composition of the neighborhood had an independent and strong effect, increasing the amount of high cost subprime lending. In particular:

[2] A Laszlo/Luntz Poll, conducted January 21 to February 13, 2002. Overall poll of 1,258 adults, margin of error 3.3%. Available via NCRC.

- The level of refinance subprime lending increased as the portion of African-Americans in a neighborhood increased in nine of the ten metropolitan areas. In the case of home purchase subprime lending, the African-American composition of a neighborhood boosted lending in six metropolitan areas.

- The percent of African-Americans in a census tract had the strongest impact on subprime refinance lending in Houston, Milwaukee, and Detroit. Even after holding income, creditworthiness, and housing market factors constant, going from an all white to an all African-American neighborhood (100 percent of the census tract residents are African-American) increased the portion of subprime loans by 41 percentage points in Houston. For example, if 10 percent of the refinance loans in the white neighborhood were subprime, then 51 percent of the loans in an African-American neighborhood in Houston would be subprime. The portion of subprime refinance loans increased by 29, 26, and 20 percentage points in Milwaukee, Detroit, and Cleveland, respectively, from an all white to an all African-American neighborhood. Graph 1 provides details of this phenomenon across the metropolitan areas and shows a strong race factor in Atlanta, St. Louis, and Los Angeles as well.

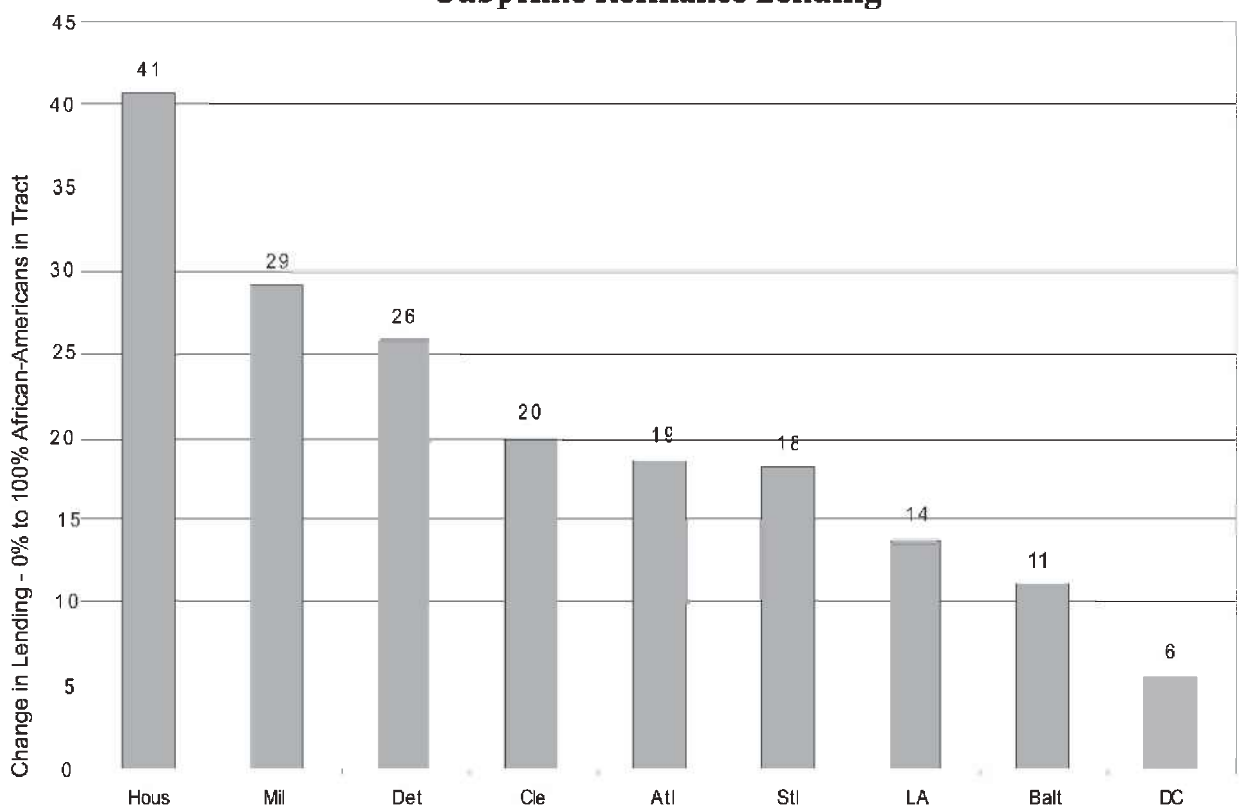
- Solely because the percentage of the African-American population increased, the amount of subprime home purchase lending surged in Cleveland, Milwaukee, and Detroit. From an all white to an all African-American neighborhood in Cleveland, the portion of subprime home purchase loans climbed 24 percentage points. Graph 2 reveals that the portion of subprime purchase loans similarly rose by 18 and 17 percentage points in Milwaukee and Detroit, respectively, in African-American neighborhoods compared to white neighborhoods.

- The impact of the age of borrowers was strong in refinance lending. In seven metropolitan areas, the portion of subprime refinance lending increased solely when the number of residents over 65 increased in a neighborhood.

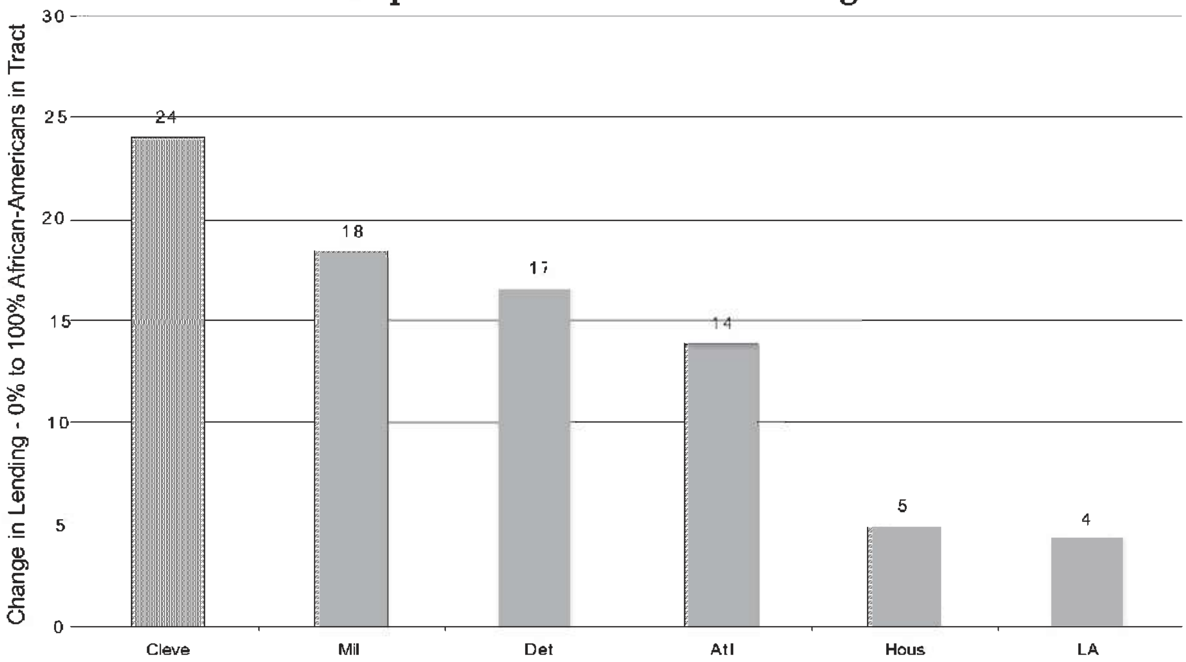
- Elderly neighborhoods experienced the greatest increases in subprime refinance lending in St. Louis, Atlanta, and Houston. Even after holding income, creditworthiness, and housing market factors constant, the portion of subprime refinance lending would surge 31 percentage points in St. Louis from a neighborhood with none of its residents over 65 to all of its residents over 65. Likewise, the increases were 27 and 25 percentage points in Atlanta and Houston, respectively. Although neighborhoods with such extreme age distributions (none or all residents over 65) are unusual, the regression analysis highlights and isolates the impacts of age on the level of subprime lending. Indeed, the level of subprime lending is likely to be considerably higher in neighborhoods with large concentrations of senior citizens.

- The level of subprime lending increased in a statistically significant fashion in the great majority of metropolitan areas as the percentage of neighborhood residents with no credit scores increased. Subprime refinance and home purchase lending climbed in nine and seven metropolitan areas, respectively, as the portion of neighborhood residents without credit scores increased. This is a significant issue for recent immigrants and other unbanked populations, many of whom are creditworthy for loans at prevailing interest rates, but receive high cost loans simply because they lack conventional credit histories.

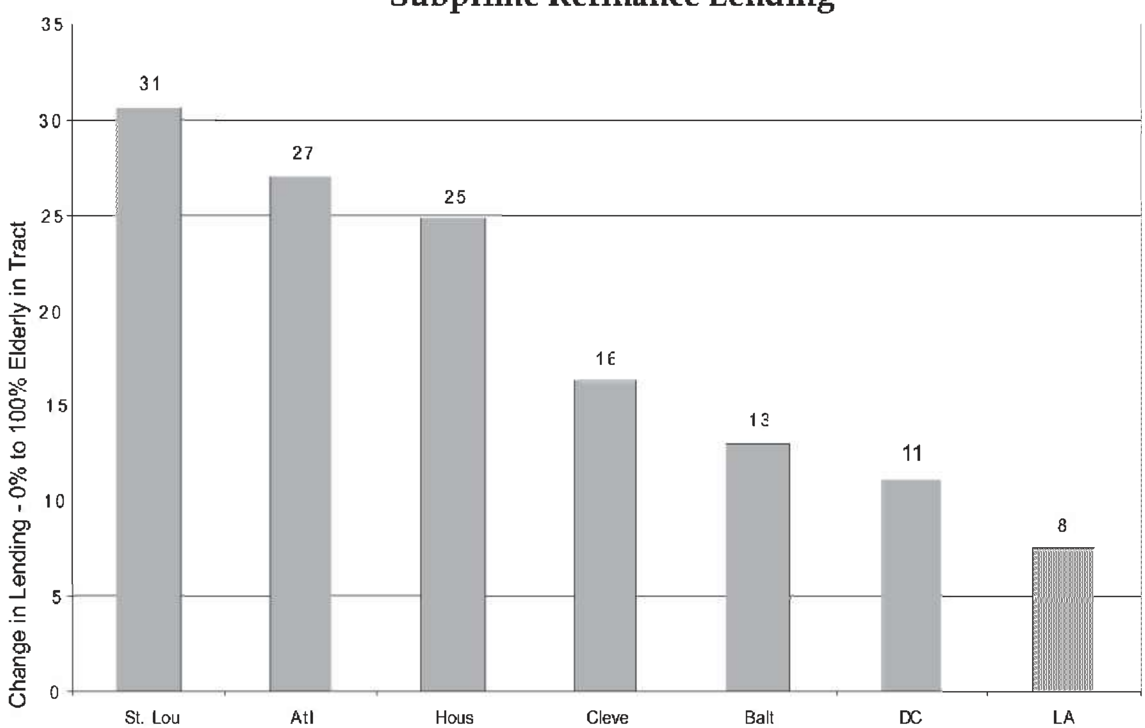
Graph 1: Index of Discrimination Against African-American Neighborhoods: Subprime Refinance Lending



Graph 2: Index of Discrimination Against African-American Neighborhoods: Subprime Home Purchase Lending



Graph 3: Index of Discrimination Against the Elderly: Subprime Refinance Lending



Recommendations

Legislative Recommendations

Reform FCRA to Mandate Complete and Accurate Credit Reports

As Congress renews the Fair Credit Reporting Act (FCRA), it must ensure that credit reports are complete and accurate. Anti-predatory lending bills introduced by members of Congress from both parties (Sarbanes and Ney) require creditors, once every three months, to provide a complete credit report and payment history to credit bureaus regarding all loans they made or serviced. A number of large subprime lenders currently withhold critical information regarding borrower on-time payments.[3] The practice of withholding information victimizes borrowers by trapping them in high cost loans and also victimizes lenders by reducing the overall reliability of the credit reporting system. A bipartisan consensus should be quickly achieved regarding this essential reform, yet the bipartisan House bill, HR 2622, does not contain this requirement. The FCRA bill proceeding in the Senate also does not require frequent reporting to the credit bureaus.

Our study also found that as the percent of neighborhood residents with no credit scores increases, so does the level of subprime lending. This is blatantly unfair since large numbers of consumers without traditional credit reports and credit scores are responsible and should qualify for loans at prevailing interest rates. One major reason why a large segment of consumers lack credit scores is that the credit reporting system does not capture non-traditional payment histories such as rental and utility

[3] Remarks by John D. Hawke, Jr., Comptroller of the Currency, Consumers Bankers Association Conference in San Francisco on June 7, 1999, available via http://www.occ.treas.gov.

payments. Congress must require the reporting of these two essential payment history items to the credit bureaus in order to reduce pricing discrimination and make the lending system fairer.

NCRC also recommends that an FCRA renewal bill requires additional studies on credit scoring and fund and promote nationwide financial education initiatives.

Comprehensive Anti-Predatory Legislation

Congress must enact comprehensive anti-predatory lending legislation along the lines of bills introduced by Senator Sarbanes and Representative Schakowsky. Comprehensive and strong anti-predatory lending legislation would eliminate the profitability of exploitative practices by making these practices illegal. It could also reduce the amount of price discrimination since fee packing and other abusive practices would be prohibited. A comprehensive anti-predatory law would also strengthen the Community Reinvestment Act (CRA) if regulatory agencies severely penalize lenders through failing CRA ratings when the lenders violate anti-predatory law.

Congress Must Pass a CRA Modernization Bill

In the 107th Congress, Representatives Luis Gutierrez and Thomas Barrett introduced HR 865, the CRA Modernization Act. This vital bill would increase the rigor of CRA exams by requiring the federal banking agencies to scrutinize the level of lending to minorities as well as low- and moderate-income borrowers. In addition, the CRA Modernization Act would expand CRA to cover independent mortgage companies and all non-depository affiliates of banks. Since price discrimination on the basis of race is prevalent, CRA must be used to prod lenders to offer more prime loans at prevailing interest rates to minorities. At the same time, expanding CRA to large numbers of lenders would also result in an influx of affordable loans to traditionally underserved communities.

Enhance the Quality of HMDA Data

NCRC believes that Congress and the Federal Reserve Board (which implements the HMDA regulations) must enhance HMDA data so that regular and comprehensive studies can scrutinize fairness in lending. Specifically, are minorities, the elderly, women, and low- and moderate-income borrowers and communities able to receive loans that are fairly priced? While NCRC is confident in the findings of our study, we believe that more information in HMDA data is critical to fully explore the intersection of price, race, gender, and income. HMDA data must contain credit score information similar to the data used in this report. For each HMDA reportable loan, a financial institution must indicate whether it used a credit score system and if the system was their own or one of the widely used systems such as FICO (a new data field in HMDA could contain 3 to 5 categories with the names of widely-used systems). The HMDA data also would contain one more field indicating which quintile of risk the credit score system placed the borrowers.

Using this data, regulators, researchers, the media, and the public could determine if any of the credit score systems were placing minorities and other protected classes in the higher risk categories a disproportionate amount of time. The data would facilitate more econometric analysis to assess whether the prices of loans are based on risk, race, gender, or age. In addition, other critical underwriting variables are needed in the HMDA data including information on debt-to-income ratios and loan-to-value ratios.

Financial Education Critical, Especially for Populations Lacking Credit Scores

In the metropolitan areas examined, about 15 percent of the population lacked credit scores. The percentage was even higher in minority census tracts. A significant finding of this report is that consumers are more likely to receive subprime loans when they lack credit scores. Increased financial education initiatives by Congress, government at all levels, the private sector, and the nonprofit sector are necessary to reach out to the segment of the population that lack credit scores and/or are "unbanked." The segment of the population without credit scores is unlikely to have a fair chance at receiving affordable loans as long as they lack credit histories and remain outside the financial mainstream. In order for financial education to be universal, NCRC recommends that the Department of Education require basic financial literacy to be part of the curriculum of all public schools.

A significant finding of this report is that consumers are more likely to receive subprime loans when they lack credit scores.

Regulatory Recommendations

Federal Agencies Must Step Up Enforcement of Existing Laws to Promote Full Product Choice and Prevent Product Steering

Periodically, the Federal agencies regulating financial institutions will make great fanfare announcing a settlement of a major discrimination lawsuit or the publication of new "interagency" fair lending guidelines. The sad fact, however, is that federal agency efforts to eliminate discrimination and steering creditworthy borrowers to expensive products are failing. The agencies must step up their enforcement of the Equal Credit Opportunity Act, the Fair Housing Act, the Community Reinvestment Act and other fair lending laws in order to ensure full product choice for all Americans.

Halt Preemption of State Anti-Predatory and Consumer Protection Law

The Office of the Comptroller of the Currency (OCC) has preempted Georgia's anti-predatory law for large national banks and has proposed to preempt anti-predatory and consumer protection laws in all states. The OCC's proposed regulations are much weaker in combating abusive practices than state law that would be preempted. At the same time, the Office of Thrift Supervision (OTS) has been preempting anti-predatory law, one state at a time, for federally chartered thrifts. Given the evidence of widespread pricing discrimination, anti-predatory and consumer protection law at all levels need to be strengthened, not weakened. For many decades, banking laws have co-existed on a Federal and state level in many areas such as privacy and disclosures of mortgage terms. This is precisely the wrong time to wipe out critical state anti-predatory and consumer protection law. The credit system is broken, and needs more oversight, not less.

Anti-predatory and consumer protection law at all levels need to be strengthened, not weakened

Federal Reserve Board Must Step Up Anti-Discrimination and Fair Lending Oversight

The General Accounting Office concluded that the Federal Reserve Board has the authority to conduct fair lending reviews of affiliates of bank holding companies. The Federal Reserve Board, however, continues to insist that it lacks this authority.[4] This issue must be resolved because comprehensive anti-discrimination exams of all parts of bank holding companies are critical. Most of the major banks have acquired large subprime lenders that are then considered affiliates and become off-limits to Federal Reserve examination. A pressing question is the extent to

[4] General Accounting Office, *Large Bank Mergers: Fair Lending Review Could be Enhanced with Better Coordination*, November 1999, GAO/GGD-00-16.

which the subprime affiliates refer creditworthy customers to the prime parts of the bank so that the customers receive loans at prevailing rates instead of higher subprime rates. Or does the subprime affiliate steer creditworthy borrowers to high cost loans? These questions remain largely unanswered. Consequently, we do not know the extent to which steering by subprime affiliates and/or their parent banks contributed to the discrimination documented by this report. Thus, it is past time for the Federal Reserve to examine affiliates as well as the parent bank.

Increase Fair Lending Enforcement of Non-Bank Lending

CRA and fair lending reviews cover depository institutions. Large non-bank lenders comprise a significant segment of subprime lenders but are not covered by regular CRA exams and fair lending reviews. As far as we know, neither the Department of Housing and Urban Development, the Department of Justice, nor the Federal Trade Commission has established a proactive program to conduct fair lending investigations of large non-bank lenders. The Department of Justice has settled lawsuits regarding price discrimination with the Long Beach Mortgage Company and other institutions.[5] These lawsuits, however, are usually reactive and in response to complaints or referrals from other regulatory agencies. In cooperation with state regulatory agencies, NCRC calls upon federal agencies to undertake a proactive and aggressive program to enforce the fair lending laws in the case of non-bank lenders.

CRA Exams Must Scrutinize Non-Prime Lending More Rigorously

Currently, CRA exams are not adequately assessing the CRA performance

[5] Department of Justice settlement with Long Beach Mortgage Company, September 5, 1996.

of subprime lenders. For example, the CRA exam of the subprime lender, Superior Bank, FSB, called its lending innovative and flexible before that thrift's spectacular collapse.[6] If CRA exams continue to mechanistically consider subprime lending, subprime lenders will earn good ratings since they usually offer a larger portion of their loans to low- and moderate-income borrowers and communities than prime lenders.

At this point, the regulatory agencies have stated in an "Interagency Question and Answer" document that banks will be downgraded if their lending violates federal anti-predatory law. NCRC has not seen rigorous action to implement this guidance. Fair lending reviews that accompany CRA exams do not usually scrutinize subprime lending for compliance with anti-predatory law, for possible pricing discrimination, or whether abusive loans are exceeding borrower ability to repay. NCRC recommends that all CRA exams of subprime lenders must be accompanied by a comprehensive fair lending and anti-predatory lending audit. In addition, CRA exams must ensure that prime lenders are not financing predatory lending through their secondary market activity or servicing abusive loans.

NCRC also recommends that any bank or thrift whose subprime lending exceeds a nominal amount such as 5 percent of its total loan amount must have a separate prime and subprime CRA lending exam. As NCRC stated in our comment letter during the Advance Notice of Proposed Rulemaking on the CRA during the fall of 2001, a bank or thrift must not pass its lending test if it does not score at least a satisfactory rating on the

[6] Office of Thrift Supervision Central Region's CRA Evaluation of Superior Bank, FSB, Docket #: 08566, September 1999. Available via http://www.ots.treas.gov, go to the CRA search engine and select "inactive" for the status of the institution being searched.

prime portion of its lending test. The lending test is currently the most important part of CRA exams for large banks and the only element of small bank exams. Prime lending must likewise be elevated as the most important part of the lending test. NCRC's study contributes to a significant amount of evidence that minority communities receive too much subprime lending due to discrimination. In order to correct for market failure and increase product choice in underserved communities, NCRC believes that prime lending must be emphasized on CRA exams.

Full Disclosure of Automated Underwriting Systems

This report focused on the impact of credit scores as well as race and age composition of neighborhoods in determining the level of subprime lending. Automated underwriting systems use credit scores and variables similar to the ones in this report in guiding financial institutions in their lending decisions. Since our report found a substantial amount of price discrimination, we believe that automated underwriting systems must be made more transparent in order to assess whether they are contributing to discrimination. Factors and the weights of factors used by the automated systems must be disclosed. The Department of Housing and Urban Development must release the results of its fair lending examination of Fannie Mae's and Freddie Mac's automated underwriting systems.

Recommendations for Lenders, Community Groups, and Consumers

Lenders Must Adopt Risk-Based, Not Race-Based or Age-Based Pricing: Best Practices Needed

This report finds that discrimination on the basis of race and age is widespread in America. Too many subprime lenders disregard risk, as measured by credit scores, in pricing their loans. NCRC calls upon the lending industry to adopt comprehensive best practices so that they can avoid pricing discrimination and other predatory practices. The best practices approach must also include rigorous compliance training for loan officers as well as mystery shopping and testing initiatives to identify and eliminate discriminatory practices. NCRC is in the process of completing a mystery shopper report that documents the need for additional industry compliance efforts because the report reveals disparate treatment regarding interest rate and loan terms for white and minority testers.

Community Groups Must Advocate and Offer Financial Education and Counseling Programs

NCRC's findings reinforce the need for community group advocacy as well as program delivery. Community groups must be active in the CRA process, offering comments during CRA exams and merger applications, particularly when they believe a lender is violating fair lending law and discriminating against minorities, women, and the elderly. Each time a community group and/or coalitions of community groups change the practices of a major lender (engaged in both prime and subprime lending), the impact on the industry as a whole is profound and cannot be underestimated. At the same time, community groups should continue pursuing programmatic opportunities, including mystery shopping, financial education, and counseling programs. Community groups should increase their skill and sophistication of using data compiled from their program delivery for their advocacy and policy positions.

Consumers Must Shop for Affordable Loans and Obtain Credit Reports, Credit Scores, and Pursue Inaccuracies

NCRC recommends that consumers consult with NCRC's *Best and Worst Lenders* at http://www.ncrc.org to find a list of lenders most likely to approve minorities, women, and low- and moderate-income consumers for affordable loans. *Best and Worst Lenders* provides detailed information on lenders in 25 major metropolitan areas. Consulting with *Best and Worst Lenders* increases the chances that consumers will be approved for loans. In addition, *Best and Worst Lenders* enables consumers to identify responsible banks that reinvest consumer deposits back into minority and low- and moderate-income communities instead of redlining local communities and investing their deposits elsewhere.

Once a year, consumers should also purchase their credit reports and scores from each major credit bureau (Experian at www.experian.com, Equifax at www.equifax.com; and Trans Union at www.transunion.com). If a consumer believes that his or her credit report contains an inaccuracy, he or she should ask the credit bureaus to investigate and correct any mistakes. If the consumer believes that the credit bureaus have not fairly resolved disputes over mistakes, he or she should contact the Federal Trade Commission at www.ftc.gov.

Background and Literature Review

NCRC benefited from industry data on creditworthiness in order to produce a comprehensive study on the relationship between loan pricing and the race and age of neighborhoods. NCRC used credit scoring data provided by one of the three large credit bureaus. A credit score is a numerical score estimating the chances a consumer will be delinquent in loan payments or default altogether. The credit score is derived from statistical analysis of information contained in credit reports regarding a consumer's past payment history and use of credit. On a census tract level, the credit scoring data indicated how many consumers were in various categories of risk. NCRC was then able to analyze the impact of credit scores on the level of subprime home lending by combining the credit scoring information with the Home Mortgage Disclosure Act (HMDA) data, and demographic and housing stock data from the Census Bureau.

NCRC employed regression analysis to predict the level of subprime lending on a census tract level in ten large metropolitan areas. The analysis allowed NCRC to determine whether increases in the African-American, Hispanic, or elderly population in a neighborhood led to increases in the amount of subprime loans after controlling for credit-worthiness (as revealed by the credit score data) and important housing stock characteristics. As stated above, the findings revealed that minority and elderly neighborhoods do, in fact, receive substantially higher levels of subprime lending than is justified based on the creditworthiness of their residents, housing values, and other measures of housing market conditions.

NCRC's findings are consistent with a body of research on subprime lending. A recent survey study conducted by Freddie Mac analysts finds that two-thirds of subprime borrowers were not satisfied with their loans, while three-quarters of prime borrowers believed they received fair rates and terms.[7] In previous years, Freddie Mac and Fannie Mae have often been quoted as stating that between a third to a half of

[7] Freddie Mac analysts Marsha J. Courchane, Brian J. Surette, Peter M. Zorn, *Subprime Borrowers: Mortgage Transitions and Outcomes*, September 2002, prepared for Credit Research Center, Subprime Lending Symposium in McLean, VA.

borrowers who qualify for low cost loans receive subprime loans.[8] Dan Immergluck, a professor at Grand Valley State University, was one of the first researchers to document the "hypersegmentation" of lending by race of neighborhood.[9] Like Immergluck's work, the Department of Housing and Urban Development found that after controlling for housing stock characteristics and the income level of the census tract, subprime lending increases as the minority level of the tract increases.[10] The Research Institute for Housing America, an offshoot of the Mortgage Bankers Association, released a controversial study in 2000 which concluded that minorities were more likely to receive loans from subprime institutions, even after controlling for the creditworthiness of the borrowers.[11]

A recent survey study conducted by Freddie Mac analysts finds that two-thirds of subprime borrowers were not satisfied with their loans,

NCRC's study is quite similar and builds upon important research conducted by a Federal Reserve economist and two researchers from the Wharton School at the University of Pennsylvania. Paul Calem of the Federal Reserve, and Kevin Gillen and Susan Wachter of the Wharton School also use credit scoring data to conduct econometric analysis scrutinizing the influence of credit scores, demographic characteristics, and economic conditions on the level of subprime lending. Their study found that after controlling for creditworthiness and housing market

[8] "Fannie Mae Vows More Minority Lending," in the Washington Post, March 16, 2000, page E01. Freddie Mac web page, http://www.freddiemac.com/corporate/reports/moseley/chap5.htm.

[9] Dan Immergluck, *Two Steps Back: The Dual Mortgage Market, Predatory Lending, and the Undoing of Community Development*, the Woodstock Institute, November 1999.

[10] Randall M. Scheessele, *Black and White Disparities in Subprime Mortgage Refinance Lending*, April 2002, published by the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development.

[11] Anthony Pennington-Cross, Anthony Yezer, and Joseph Nichols, *Credit Risk and Mortgage Lending: Who Uses Subprime and Why?* Working Paper No. 00-03, published by the Research Institute for Housing America, September 2000.

conditions, the level of subprime refinance and home purchase loans increased in a statistically significant fashion as the portion of African-Americans increased on a census tract level in Philadelphia and Chicago.[12]

Relatively few studies examine the relationship between the number of elderly residents of a neighborhood and the level of subprime lending although anecdotal evidence suggests that abusive lenders target the elderly. In one study, the South West office of Consumers Union found that every 1 percentage point increase in the portion of people over 65 in a neighborhood increased subprime refinance lending by 1.3 percentage points. The Consumers Union study examined neighborhoods in Dallas and Austin, and included demographic variables and a few underwriting variables such as loan amount to income ratios in its regression equations.[13] The AARP also conducted a national survey of elderly borrowers and found that older borrowers who were widowed, female, African-American, and less educated were more likely to receive subprime loans than their married, male, white, and more educated counterparts. The survey also found that seniors receiving subprime loans were more likely to have been approached by brokers, to have refinanced two or more times in the past three years, and to be dissatisfied with their loans.[14]

Another body of literature examines whether consumer credit reports are

[12] Paul S. Calem, Kevin Gillen, and Susan Wachter, *The Neighborhood Distribution of Subprime Mortgage Lending*, October 30, 2002. Available via pcalem@frb.gov.

[13] Consumers Union, *Elderly in the Subprime Market*, October 2002, www.consumersunion.org.

[14] Neal Walters and Sharon Hermanson, *Older Subprime Refinance Mortgage Borrowers*, AARP Public Policy Institute, Data Digest Number 74, July 2002, http://www.aarp.org/ppi.

accurate. If consumer credit reports are incomplete and inaccurate, then the credit scores used to assess risk could be seriously flawed. Troubling evidence suggests that substantial inaccuracies exist in credit reports and could be contributing to racial disparities in lending. In the summer of 2002, the Consumer Federation of America (CFA) shed more light on how credit report flaws can disproportionately impact borrowers on the edge between prime and subprime credit. CFA's analysis of credit scores in more than 500,000 merged credit files revealed that 29 percent of consumers had scores with a range of at least 50 points when using the credit reports from each of the three major bureaus. Focusing in more detail on 1,704 at-risk mortgage purchasers with marginal scores between prime and higher cost subprime credit, CFA found that at least one-fifth would be harmed, and one-fifth would benefit from score inaccuracy if they tried to purchase mortgage loans. The upshot of this finding is that at least 8 million Americans may be erroneously placed into subprime loans and thus pay tens of thousands of dollars each in unnecessarily high mortgage interest payments.[15]

In the winter of 2003, a Federal Reserve Bulletin article revealed that almost one third of sampled credit accounts lacked information on borrower credit limits, which is a key variable for credit scores. Furthermore, subprime specialists reported credit limits 77 percent of the time for their prime customers, but only 40 percent of the time for their subprime customers.[16] Not reporting the credit limit makes borrower credit appear

[15] Consumer Federation of America and National Credit Reporting Association, *Credit Score Accuracy and Implication for Consumers*, December 2002, http://www.consumerfed.org.

[16] Robert B. Avery, Paul S. Calem, Glenn B. Canner, Raphael Bostic, *An Overview of Consumer Data and Credit Reporting*, Federal Reserve Bulletin, February 2003, http://www.federalreserve.gov.

to be much worse than it actually is. The absence of this information results in borrowers appearing to be much closer to fully utilizing their credit cards and other open ended credit than they are in reality.

The findings of NCRC, the Calem, Gillen, and Wacther study, as well as other research, are disturbing but not surprising. Predatory lenders brazenly disregard credit scores and also do not engage in other conventional and prudent underwriting techniques. They discriminate by offering minority and elderly borrowers higher interest rate loans than is justified based on credit scores. At the same time, credit scores are not accurately predicting risk due to omitted variables that are key for traditionally underserved populations. In short, the credit system is broken and discrimination will only be eliminated if the recommendations outlined above are implemented.[17]

Methodology

As stated above, the key goal of the analysis is to determine the relationship between the portion of minority and elderly persons in a census tract and the percentage of home purchase and refinance loans that are made by subprime lenders. After controlling for economic and risk factors, does the portion of subprime loans increase as the minority and elderly population in a census tract increases? In other words, this study explores the likelihood of discrimination and reverse redlining in home

[17] Given the problems with credit reports, the credit scores used here are more likely to overstate risks for minority borrowers than for white borrowers. Accordingly, the scores are more likely to overstate the percent of borrowers in high risk groups in African-American rather than white census tracts. If such bias does occur in scores, then the use of these scores means that the true impact of race on subprime lending is higher than that indicated by the results found here. That is, our estimates of discrimination or redlining are biased low. The credit report and score data needs to be improved via renewal of Fair Credit Reporting Act.

lending. NCRC chose 10 metropolitan statistical areas (MSAs) from different parts of the United States and conducted a statistical analysis in each area. In particular, the MSAs selected are: Atlanta, Baltimore, Cleveland, Detroit, Houston, Los Angeles, Milwaukee, New York, St. Louis, and Washington DC. These areas have different demographic and economic characteristics, which will allow us to make credible and generalizable conclusions about the home lending patterns across large metropolitan areas. In the ten MSAs, the sample consists of about 7,000 census tracts (6,741 for home purchase and 7,097 for refinance). A multivariate regression approach controlled for demographic and risk factors.

NCRC conducted separate analyses for home purchase and refinance lending. We expected a higher degree of pricing disparities by race and age of neighborhood in refinance lending since subprime lenders specialize in refinance lending and make fewer home purchase loans. NCRC's previous work, including *Best and Worst Lenders,* also found more disparities in refinance lending than home purchase lending. Abusive subprime lenders are particularly active in refinance lending since their intention is to strip equity from homeowners through repeated refinancings or flipping.

Variables for the analysis belong to three categories: home lending, credit scoring, and demographics. NCRC used 2001 HMDA data for home lending, 1999 credit scoring data, and 1990 census tract demographic information. NCRC obtained the 1999 credit scoring data on a one-time basis from one of the three large credit bureaus. NCRC chose 2001 HMDA data, not 1999 data, as we believe that the distribution of credit scores on a census tract level does not vary significantly over a three year time period. NCRC ran regression equations using 1999 and 2000 home loan data to confirm the hypothesis. The results were similar over the years. Also, 2001 was a year of lower interest rates. NCRC wanted to see if minority neighborhoods were benefiting from lower interest rates as measured by a decrease in the statistical significance of race of neighborhood on the level of subprime lending. NCRC would have preferred to use 2000 census tract data, but the HMDA data will not use 2000 census data until the 2003 release in the summer of 2004. The 2001 HMDA data uses 1990 census tract boundaries. NCRC believes the results will be similar with HMDA data using 2000 census tract boundaries, but we intend to do follow-up research.[18]

HUD Subprime and Manufactured Home Lender List

In order to classify loans as subprime, NCRC used a list of subprime and manufactured home lenders developed by HUD. Since HMDA data does not have information on the Annual Percentage Rate (APR) or other loan terms and conditions, HUD developed its list by complementing data analysis with interviews of lending institutions and a literature search. As an additional step, HUD called the lenders on its list and asked them if they considered themselves subprime and manufactured home specialists. Generally speaking, a lender was included on the list if more than 50 percent of the loans in its portfolio was subprime or manufactured home.[19]

[18] Important characteristics of the HMDA data are discussed separately in an appendix.

[19] HUD itself admits that the list is not complete. A number of institutions considered to be prime specialists make a significant number of subprime loans, even if 50 percent or more of their loans are not subprime. Also, the list may not be complete due to name changes and omissions. HUD refines its lists on an annual basis and also corrects mistakes on previous years' lists. HUD's web page (http://www.huduser.org/datasets/manu.html) has more information about the lists and has copies of the lists.

Until more information on loan terms and conditions are available in HMDA data, HUD's list is a valuable resource for conducting subprime and manufactured home loan analysis. Although the list is incomplete, it still captures significant differences in lending behavior as revealed by this report and a substantial body of research.

Data and variables

Home lending data in the analysis represents only originations of home loans, not applications for the loans. We included all types of loans: conventional, and government insured (FHA, VA, and FSA/RHS) to owner-occupants only. NCRC also separated two types of home loans: home purchase loans and refinance loans. By doing so, we aimed to see for which loan type the race and age of neighborhood residents had a stronger influence. We excluded manufactured home lenders from the analysis as initial regressions revealed that the level of manufactured home lending did not vary in a statistically significant manner with the race of neighborhood residents.[20] Future research should explore this in more detail. The study excluded census tracts in which the number of originated loans was less than 20. This was done to ensure a sufficient number of loans for meaningful characterization of each tract's lending patterns.

[20] Manufactured home lenders specialize in making loans to borrowers purchasing manufactured homes. These lenders tend to make high interest rate loans; abusive lending has been widespread in the manufactured home sector as indicated by massive foreclosures and the failures of large national manufactured home lenders. According to HUD, "A manufactured home (formerly known as a mobile home) is built to the Manufactured Home Construction and Safety Standards (HUD Code) and displays a red certification label on the exterior of each transportable section. Manufactured homes are built in the controlled environment of a manufacturing plant and are transported in one or more sections on a permanent chassis." HUD has detailed information about manufactured housing on its web page of http://www.hud.gov.

The analysis chose the following variables that would hypothetically influence subprime lending in an area.

Home lending variables (dependent variables):

%subHP – percent of home purchase loans in a census tract that were subprime.

%subREF – percent of refinance loans in a census tract that were subprime.

Demographic variables included:

%black – percent of residents in a census tract who were African-American;

%hisp – percent of residents in a census tract who were Hispanic;

%65age – percent of residents in a census tract who were over 65 years old;

medage – dummy variable. The variable revealed the median age of houses in a census tract.

0 when the median age of housing was between 0-20 years old (built in 1970-1990);
1 when the median age of housing was between 21-50 years old (built in 1969-1940);
2 when the median age of housing was 51 years and older (built before 1940);

medhhinc – 1989 median household income in a census tract;

HT – housing turnover. This variable is a ratio of all home purchase loans made in 2001 divided by owner occupied units in 1990. The literature indicates that a higher amount of housing turnover (as revealed by larger values of this variable) suggests a more vibrant market and faster home value appreciation. This should make a census tract more attractive to prime lenders and thus decrease the portion of subprime lending.

capitaliz – The "capitalization" variable is a ratio of gross median rent divided by median housing value. The literature suggests that owner-occupied units appreciate slower in neighborhoods where the median rent is higher relative to the median housing value (higher ratio values for this variable). Therefore, prime lenders may find neighborhoods less attractive with higher values for the capitalization variable, meaning that the portion of subprime loans will be higher in these neighborhoods.

Credit scoring variables included:

%vhigh – is a credit score variable that indicated the percent of people in a census tract in the very high credit risk category;

%NC – is the percent of neighborhood residents lacking credit scores;

vh+h+m – the cumulative percent of neighborhood residents in very high, high, and moderate credit risk categories added together.

The credit risk scores used in this report measure the likelihood of future delinquencies and foreclosures. The database had a credit score range from 0 to 1,000 with lower scores indicating lower risk or chance of borrower delinquency. The scores were divided into five equal categories or quintiles of risk; the specific categories are Very Low, Low, Moderate, High and Very High risk. The credit score range was separated into quintiles, not the population totals within the quintiles. In other words, each score quintile did not have equal numbers of people, but each score range was of equal length (about 200 units for each quintile since the total range is from 0 to 1,000).

For each census tract, the database contains the number and percent of neighborhood residents in each of the five risk categories, and the number and percent of neighborhood residents with no credit scores.

NCRC's analysis focuses on the "vh+h+m" credit score variable. Our regression analysis was iterative. One equation (Column 1 on Tables 1 through 10) included the combined risk variable of "vh+h+m" and the NC or no credit score variable. Column 2 is another regression in which the very high risk and no credit score variables are included as separate variables (see the tables below).

Columns 3 through 4 repeat the iterative approach for the risk variables in the same order as Columns 1 through 2. The difference between Columns 1 and 2 and Columns 3 and 4 is that the race and age variables are omitted in Columns 3 and 4. This is done in order to understand better the added explanatory power obtained by including the race and age variables (see discussion below in the Functional Form section).

The "vh+h+m" variable was statistically significant across all ten MSAs for home purchase lending and nine MSAs for refinance lending. The impact of the variable was as expected; that is, subprime lending was more prevalent as the percentage of people in a census tract with very high, high, and moderate risk increased. The regression equations including only the very high risk and no credit score variables had very similar outcomes to the equations with the "vh+h+m" combined risk and no credit score variables. Although the very high risk equations (Column 2) were similar to the "vh+h+m" equations (Column 1), we focused on the "vh+h+m" equations since subprime lenders would likely make loans to consumers with high and moderate risk as well as very high risk. The coefficients and R squares in the "vh+h+m" equations were consistent with these expectations.

In contrast to our report, the Calem, Gillen, and Wacther study focuses on the equations with the very high risk and no credit score variables. The fact that two different series of equations (those with very high risk and no credit score variables and those with the combined risk and no credit score variables) produced similar results adds to the robustness of the overall findings.

Impact of Demographic Versus Economic Factors

As stated above, we conducted multivariate regression analysis with the dependent variable represented by the percentage of subprime loans in a census tract and independent variables that control for demographic, economic and risk factors. Our variables of interest were the minority and elderly populations in a census tract. NCRC hypothesized that the percent of minorities and elderly people in a census tract was positively related to the percent of subprime loans originated in a census tract.

. . .the percent of minorities and elderly people in a census tract was positively related to the percent of subprime loans originated in a census tract.

Table 11 shows the statistical significance of variables at the 10%, 5%, and 1% precision level, sign of estimated coefficients, and adjusted R square for every regression. The adjusted R square was rather high for most MSAs and loan types (the higher the R square, the better the equation accounts for and explains patterns of subprime lending on a neighborhood level). The R square was higher for refinance than home purchase, suggesting that our model was better at predicting patterns in refinance lending. For refinance lending, the R square ranged from 0.5252 in Los Angeles to 0.8993 in Detroit. For home purchase lending, the R square fell between 0.0843 in Baltimore and 0.6865 in Cleveland. The R square was above 0.3 in five out of ten MSAs in home purchase lending. In contrast, the R square was above 0.3 in all MSAs in refinance lending. Overall, we believe our model is robust and a good predictor of lending patterns. The model's results were consistent with the Calem, Gillen, and Wachter study.

The African-American population in a census tract was statistically significant in six MSAs for home purchase lending and in nine MSAs for refinance lending. As expected, after controlling for risk and housing stock characteristics, the effect of the percentage of African-American population on the portion of subprime loans in a census tract was positive in all MSAs. Lenders still associated high risk with race and thus, compensated by making a substantially higher level of subprime loans in African-American than white tracts.

The percent of Hispanic population in a census tract was significant in only one MSA for home purchase and in five MSAs for refinance lending. The sign of the coefficients was not consistent for each MSA.[21] The sign was negative in one MSA for home purchase lending and in two MSAs for refinance lending. In contrast, the sign was positive in three MSAs for refinance lending, meaning that the level of subprime refinance lending increased as the portion of Hispanics increased in a census tract. Our study results suggest no consistent relationship between the level of subprime lending and the portion of Hispanics in a neighborhood. However, the portion of Hispanics in a neighborhood was associated with an increase in subprime lending, all else equal, in a subset of the MSAs.

The portion of people over 65 was a strong factor for three out of ten MSAs for home purchase lending. For refinance lending, the age of the census tract population was significant in eight MSAs. For refinance and

[21] A coefficient expresses the effect of an independent variable on the dependent variable. In this report, the portion of subprime loans is the dependent variable. The level of subprime lending changes because of the racial composition of the neighborhood and other "independent" variables. For the racial composition of the neighborhood, the coefficient measures the impact in percentage point terms. For every percentage point increase in African-American or Hispanic residents in a census tract, the portion of subprime loans increases or decreases by a certain number of percentage points as revealed by the value and sign of the coefficient. The coefficient only has an impact if it is statistically significant (as revealed by legends in the charts capturing the regression results).

home purchase lending, the sign of the coefficients was positive in all MSAs except in two of the eleven cases. This supports the contention that abusive lenders target the elderly to take advantage of the fact that the elderly have substantial amounts of equity but are often short on cash. These results contradict those obtained by Calem, Gillen, and Wachter. They mentioned that this variable "yielded no additional insights," but their study looked at only two MSAs.

Median household income of a census tract was statistically significant in four out of ten MSAs in home purchase lending and in refinance lending. Except in one case, the sign of the coefficients was positive, which is counterintuitive. The literature, however, discusses that a segment of high income borrowers do not report income level to lenders nor do they want to undergo a lengthy application process. Hence, they receive subprime loans. It must be added that the coefficient values were very small, meaning that the income variable had a small impact on the level of subprime lending in census tracts.

Except for Detroit refinance lending, the combined risk variable in all MSAs for both loan types was statistically significant. Coefficients were positive, meaning that a larger percentage of people with higher risk factors was associated with a higher percent of subprime loans in a census tract. These findings are quite consistent with those discussed in the Calem, Gillen, and Wachter report. Also, the level of subprime home purchase and refinance lending increased in a statistically significant fashion in the great majority of MSAs as the percentage of neighborhood residents with no credit scores increased.

The other variables including housing turnover and capitalization behaved in the expected manner. Housing turnover was significant in most MSAs and the coefficients' signs were negative, which supported our expectations. Higher housing turnover indicates more vibrancy in the market of the neighborhood, which in turn leads to less subprime lending. The capitalization variable was significant in six MSAs for home purchase and in ten MSAs for refinance lending. Except in one case, it also had the expected effect on subprime lending. Specifically, it was positively related to the percent of subprime loans, proving that faster appreciation of the owner-occupied units (smaller capitalization ratios) leads to less subprime lending in a neighborhood.

Subprime lending increased significantly as the portion of African-Americans and elderly people increased in a neighborhood.

In summary, after controlling for risk and housing stock characteristics, subprime lending increased significantly as the portion of African-Americans and elderly people increased in a neighborhood. Pricing discrimination is widespread in the dual lending marketplace in America.

Metropolitan Areas Compared

Tables 12 through 14 sort MSAs by the effect of race and age factors on the level of subprime home purchase and refinance lending in a census tract. As Table 12 reveals, the percentage of African-Americans in a census tract imposed the strongest effect on subprime home purchase lending in Cleveland, Milwaukee, Detroit, and Atlanta. The African-American variable had the largest effect in Houston, Milwaukee, Detroit, and Cleveland for refinance lending. For example, in Houston a ten percentage point increase of African-Americans in a census tract, holding all other variables constant, would lead to an increase in the portion of subprime refinance loans of 4.058 percentage points. In contrast, in Baltimore a 10 percentage point increase in the portion of African-Americans would lead to only a 1.107 percentage point increase in the portion of subprime refinance loans.

In Tables 12 through 14, the coefficients with one, two, or three asterisks are coefficients estimated at the 10%, 5%, and 1% level of statistical significance, respectively. In other words, these coefficients are valid in predicting the portion of subprime loans. In contrast, when the coefficients do not have asterisks, they cannot be used to predict the level of subprime loans.

The coefficient values for the African-American variables in this report are consistent with those in Calem, Gillen, and Wachter. The ordinary least squares regressions in the Calem, Gillen, and Wachter study estimated the African-American coefficient at about 0.2, which was approximately the median coefficient in our equations as reported in Table 12.

The portion of Hispanics in a census tract had the strongest impact in the Detroit and Houston MSAs for refinance lending, according to Table 13. In Detroit for example, a 10 percentage point increase in the Hispanic population would lead to 1.282 percentage point increase in the portion of subprime refinance lending.

The portion of people over 65 was a relatively strong variable in Detroit and Houston for home purchase lending and in St. Louis, Atlanta, and Houston for refinance lending. In particular, in the St. Louis MSA, a 10 percentage point increase of people over 65 would lead to a 3.065 percentage point increase in the portion of subprime refinance loans in a neighborhood.

In refinance and home purchase lending, the African-American portion of people in a census tract increased subprime lending regardless of the level of segregation in a MSA (see Table 12 which shows segregation levels as well as estimated coefficients for the African-American variable). For African-Americans, discrimination poses great difficulties across a wide swath of MSAs of different economic and demographic conditions. Regardless of the level of segregation, the African-American variable increased subprime refinance lending. No trends appeared regarding the level of segregation and the impact of the Hispanic variable on the amount of subprime lending.

Functional Form

Another dimension that should be discussed in this analysis is functional form: how it affects the results and what conclusions it informs. As stated above, NCRC used two forms when running the regressions: including and excluding race and age factors. The outputs are presented in the Tables 1 through 10. In most cases, the R square was lower when the race and age variables were excluded (this is observed clearly when comparing Columns 1 and 3 with the vh+h+m combined risk variable). This suggests that the equations explained a greater amount of the variation in the dependent variable when the race and age variables were included.

Calem, Gillen, and Wachter took a different iterative approach, but their findings were similar to our study. They ran some regressions with only demographic characteristics while we ran some regressions with only non-race variables. The end result of both approaches was that the R square was higher when the race variables were included.

Conclusion

After controlling for risk and housing market conditions, the race and age composition of the neighborhood had an independent and strong effect, increasing the amount of high cost subprime lending. The level of refinance subprime lending increased as the portion of African-Americans in a neighborhood increased in nine of the ten metropolitan areas. In the case of home purchase subprime lending, the African-American composition of a neighborhood boosted lending in six metropolitan areas. The impact of the age of borrowers was strong in refinance lending. In seven metropolitan areas, the portion of subprime refinance lending increased solely when the number of residents over 65 increased in a neighborhood. In America today, lenders engage in widespread price discrimination, making high cost loans based on the race and age of neighborhoods, not solely based on risk.

Appendix

HMDA Data: Its Strengths and Weaknesses

Enacted by Congress in 1975, the Home Mortgage Disclosure Act (HMDA) requires banks, savings and loan associations, credit unions, and other financial institutions to publicly report detailed data on their home lending activity. Under HMDA, lenders are required to disclose annually the number of loan applications by census tract, and by the income, race, and gender of the borrower. The law also requires institutions to indicate the number and dollar amount of the loans made.

Prior to 1990, lenders were required to report the census tract containing the property for which the applicant succeeded or failed in obtaining a home loan. The Financial Institutions Reform, Recovery, and Enforcement Act (FIRREA) required lenders to report the race, gender, and income of loan applicants and borrowers starting in 1990. Thus, HMDA data before 1990 reveals information only on the census tract location of the application or loan, whereas HMDA data after 1990 includes information on borrower characteristics. Also, starting in 1993, independent mortgage companies were required to report HMDA data.

HMDA requires lenders to report on a number of possible actions or "dispositions" on loan applications. Each year, the lender must report the number of loan applications it approved and denied. The lender must also indicate how many of its loan approvals were unaccepted (the bank approved the application but the applicant did not want the loan). Finally, the lender must specify how many applications were withdrawn (the applicant withdrew his application before the bank made a credit decision), and how many applications were incomplete (the application was not considered because the applicant did not provide all the necessary information).

Housing loans covered by HMDA include home purchase, home improvement, and refinance loans for single family dwellings (1 to 4 units) and loans for multi-family units. Lenders must disclose whether the loan was a conventional loan or a loan insured by a government agency such as the Federal Housing Administration (FHA), the Veterans Administration (VA), the Farm Service Agency (FSA), and the Rural Housing Service (RHS). Additional information reported includes the occupancy status of the property (owner occupied or non-owner occupied). The lender must also indicate if the loan was purchased on the secondary market and the type of institution that bought the loan (for example, another bank or Fannie Mae or Freddie Mac).

Who is Covered by HMDA

A depository institution (bank, thrift, and credit union) must report HMDA data if it has a home office or branch in a metropolitan statistical area (MSA) and has assets above a threshold level that is adjusted upward every year by the rate of inflation. Before 1997, small depository institutions were exempt if they had assets less than $10 million. The Economic Growth and Regulatory Paperwork Reduction Act of 1996 amended HMDA to adjust the exemption level to take into account annual inflation as measured by the Consumer Price Index for Urban Wage Earners and Clerical Workers. For the 1997 data, the asset level for exemption was increased from $10 million to $28 million (to take into account inflation occurring between 1975, the first year of HMDA data, through 1996). For 1998 and 1999 data collection, the Federal Reserve increased the asset level for exemption to $29 million. For the year 2000 and 2001, the Federal Reserve set the asset level for exemption to be $30 million and $31 million, respectively.

In addition, a depository institution is not required to report HMDA data if it did not make a home purchase loan on a 1-to-4 unit dwelling (or if it did not refinance a home purchase loan) during the previous calendar year.

Many non-depository institutions must also report HMDA data. An example of a non-depository institution is a mortgage company that does not accept deposits but raises funds for lending by borrowing from investors. A non-depository institution must report HMDA data if it has more than $10 million in assets and it originated 100 or more home purchase loans (including refinances of home purchase loans) during the previous calendar year. A non-depository institution is exempt from HMDA reporting requirements if its home purchase loans (including refinances of home purchase loans) were less than 10 percent of all of its loan originations, measured in dollars, during the previous calendar year.

Gaps in HMDA Data

Small lenders and lenders with offices only in non-metropolitan areas (as noted above) are exempt from HMDA data reporting requirements. Data for rural areas is also incomplete, particularly information on the census tract location of loans. If banks and thrifts have assets under $250 million dollars (or are part of holding companies under $1 billion dollars), they do not have to report the census tract location for loans in MSAs (metropolitan statistical areas) in which they do not have any branch offices. They also do not have to report the census tract location for loans outside of MSAs.

Non-depository institutions do not have to report the census tract location of loans made in non-metropolitan areas. They have to report the census tract location of loans in those MSAs in which they received applications for, originated, or purchased five or more home purchase or home improvement loans during the preceding calendar year.

Another area of incompleteness concerns race and gender data of applications taken via the telephone. When applications are made in person, the loan officer is required to ask the applicant about his/her race. If the applicant refuses, the loan officer is required to record race on the basis of visual observation or applicant surname. The loan officer is required to inform the applicant that federal law designed to combat discrimination requires this information. In contrast, when applications are received over the phone, the loan officer is not required to ask for the race and gender of the applicant (but this is about to change, see immediately below). When applications are received through the mail, the lending institution is required to ask for the race and gender of the applicant.

In the case of the electronic media, the official staff commentary of the Federal Reserve Board regarding the HMDA regulation states that lenders are required to ask for race and gender when applications are received over the Internet. When lenders are using electronic media with a video component, lenders are to use the same procedures as if the application is made in person.

Finally, a lender is not required to report the race, gender, and income data for loans that they purchase from another institution.

Improvements in HMDA Data

In the summer of 2002, the Federal Reserve Board made some significant changes to HMDA (the Federal Reserve Board has statutory responsibility to promulgate HMDA regulations). Lending institutions will be required to ask borrowers applying over the phone for their race and gender, starting in 2003.

In 2004, non-depository institutions making at least $25 million in home purchase loans will be required to report HMDA data. This will capture more non-depository institutions as HMDA reporters than the thresholds described above. Lending institutions will be required to indicate in the HMDA data if the loans were for manufactured homes or traditional single family residences. The Federal Reserve Board will also require lenders to report price information if the APR on their loans exceeds the rate on Treasury securities by three percentage points for first-lien loans and five percentage points for second-lien loans.

Other changes to HMDA data beginning in 2004 include improving the definition of home improvement and refinance loans, requiring an indication if a loan is covered by the Home Ownership and Equity Protection Act, and requiring pre-approvals to be reported for home purchase loans. Finally, but importantly, lenders will be required to indicate the identity of their parent companies in the HMDA data.

Table 1: Detailed Regressions for Atlanta

Atlanta - Home Purchase

Variable	Column 1	Column 2	Column 3	Column 4	Variable
Intercept	-0.0736	0.0001	-0.2301	-0.0743	Intercept
	-1.6899	0.0057	-6.9928	-3.4637	
%black [est. coeff.]	0.1393	0.1327			%black
[t-Score]	***8.4146***	***7.4253***			
%hisp [est. coeff.]	-0.2080	-0.2475			%hisp
[t-Score]	-1.3761	-1.6392			
%65age [est. coeff.]	0.0845	0.0404			%65age
[t-Score]	1.2000	0.6217			
medage [est. coeff.]	-0.0060	-0.0052	0.0114	0.0104	medage
[t-Score]	-0.9145	-0.7775	*1.7122*	1.6101	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	**2.0566**	1.6146	***3.8901***	***3.1293***	
HT [est. coeff.]	-0.0007	0.0000	-0.0042	-0.0034	HT
[t-Score]	-0.3130	-0.0374	**-1.9974**	*-1.6600*	
capitaliz [est. coeff.]	2.2945	2.3405	0.3412	0.0582	capitaliz
[t-Score]	1.3955	1.4269	0.1905	0.0336	
%vhigh [est. coeff.]		0.1635		0.4289	%vhigh
[t-Score]		***2.8298***		***8.9836***	
% NC [est. coeff.]	0.0756	-0.0036	0.5576	0.2826	%NC
[t-Score]	0.8172	-0.0403	***7.3417***	***3.4278***	
vh+h+m [est. coeff.]	0.1621		0.3740		vh+h+m
[t-Score]	***2.8550***		***7.7943***		
Adj R-square	0.4566	0.4564	0.3429	0.3684	**Adj R-square**

Atlanta - Refinance

Variable					Variable
Intercept	-0.2316	-0.0823	-0.4070	-0.1572	Intercept
	-4.9917	-3.1144	-10.8020	-6.5746	
%black [est. coeff.]	0.1886	0.1682			%black
[t-Score]	***11.1936***	***9.2579***			
%hisp [est. coeff.]	-0.2456	-0.3350			%hisp
[t-Score]	-1.5388	**-2.1166**			
%65age [est. coeff.]	0.2701	0.1899			%65age
[t-Score]	***3.6791***	***2.8195***			
medage [est. coeff.]	0.0016	0.0043	0.0325	0.0310	medage
[t-Score]	0.2257	0.6160	***4.2526***	***4.3506***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***2.7783***	**1.9990**	***4.0840***	***3.1652***	
HT [est. coeff.]	-0.0021	-0.0008	-0.0065	-0.0052	HT
[t-Score]	-0.8715	-0.3277	***-2.7204***	**-2.3121**	
capitaliz [est. coeff.]	7.9826	7.7769	5.7983	4.8837	capitaliz
[t-Score]	***4.7224***	***4.6556***	***2.9185***	***2.6230***	
%vhigh [est. coeff.]		0.3827		0.7148	%vhigh
[t-Score]		***6.2345***		***13.6511***	
%NC [est. coeff.]	0.1760	0.0061	0.8036	0.3462	%NC
[t-Score]	*1.8166*	0.0654	***9.1324***	***3.7494***	
vh+h+m [est. coeff.]	0.3458		0.6046		vh+h+m
[t-Score]	***5.6966***		***11.0804***		
Adj R-square	0.6903	0.6944	0.5654	0.6091	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 2: Detailed Regressions for Baltimore

Baltimore - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept	-0.0274	0.0012	-0.0174	0.0128	Intercept
	-0.9384	0.0629	-0.9437	0.8683	
%black [est. coeff.]	0.0063	-0.0096			%black
[t-Score]	0.5582	-0.7825			
%hisp [est. coeff.]	-0.0890	-0.1080			%hisp
[t-Score]	-0.5333	-0.6547			
%65age [est. coeff.]	0.0367	0.0270			%65age
[t-Score]	0.9263	0.7600			
medage [est. coeff.]	0.0014	0.0017	0.0027	0.0026	medage
[t-Score]	0.3706	0.4567	0.7710	0.7620	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.6878	1.1145	0.4214	0.7548	
HT [est. coeff.]	-0.0209	-0.0133	-0.0267	-0.0164	HT
[t-Score]	-1.0024	-0.6474	-1.3083	-0.8145	
capitaliz [est. coeff.]	-1.5117	-2.3430	-1.4297	-2.1868	capitaliz
[t-Score]	-1.2807	*-1.9550*	-1.2171	*-1.8440*	
%vhigh [est. coeff.]		0.1912		0.1605	%vhigh
[t-Score]		***4.1024***		***5.0770***	
%NC [est. coeff.]	0.1625	0.1064	0.1432	0.0865	%NC
[t-Score]	**2.4925**	*1.6110*	**2.3639**	1.3829	
vh+h+m [est. coeff.]	0.1096		*0.1076*		vh+h+m
[t-Score]	***2.7570***		***3.9710***		
Adj R-square	0.0843	0.1028	0.0864	0.1059	**Adj R-square**

Baltimore - Refinance

Variable					**Variable**
Intercept	-0.1032	-0.0535	-0.1591	-0.0692	Intercept
	-2.7780	-2.0886	-6.0809	-3.2914	
%black [est. coeff.]	0.1107	0.1016			%black
[t-Score]	***8.0671***	***6.7403***			
%hisp [est. coeff.]	-0.4806	-0.5125			%hisp
[t-Score]	**-2.2312**	**-2.3859**			
%65age [est. coeff.]	0.1307	0.1012			%65age
[t-Score]	**2.5661**	**2.2017**			
medage [est. coeff.]	0.0041	0.0044	0.0104	0.0096	medage
[t-Score]	0.8486	0.9049	**2.0732**	**1.9929**	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.2127	0.1780	0.3565	0.8598	
HT [est. coeff.]	-0.1173	-0.1081	-0.1724	-0.1429	HT
[t-Score]	***-4.3461***	***-4.0315***	***-5.9525***	***-5.1085***	
capitaliz [est. coeff.]	11.4350	11.0128	12.1084	10.2778	capitaliz
[t-Score]	***7.4773***	***7.0691***	***7.2380***	***6.2013***	
%vhigh [est. coeff.]		0.1915		0.4338	%vhigh
[t-Score]		***3.2109***		***9.8300***	
%NC [est. coeff.]	0.3391	0.2854	0.3476	0.2013	%NC
[t-Score]	***3.9410***	***3.2582***	***3.9729***	**2.2663**	
vh+h+m [est. coeff.]	0.1471		0.3089		vh+h+m
[t-Score]	***2.9374***		***8.0034***		
Adj R-square	0.6306	0.6320	0.5539	0.5801	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 3: Detailed Regressions for Cleveland

Cleveland - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept	-0.0968	-0.0667	-0.2787	-0.1445	Intercept
	-2.4616	-2.6279	-9.6417	-6.9277	
%black [est. coeff.]	0.2400	0.2159			%black
[t-Score]	***15.6258***	***11.9307***			
%hisp [est. coeff.]	-0.0317	-0.0693			%hisp
[t-Score]	-0.5279	-1.1269			
%65age [est. coeff.]	0.0698	0.0496			%65age
[t-Score]	1.2876	1.0664			
medage [est. coeff.]	0.0114	0.0104	0.0029	0.0008	medage
[t-Score]	**2.1543**	**1.9885**	0.4430	0.1363	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.0055	0.5456	**2.3867**	***4.2976***	
HT [est. coeff.]	-0.0425	-0.0405	-0.2003	-0.1330	HT
[t-Score]	-0.8212	-0.7884	***-3.1160***	**-2.2735**	
capitaliz [est. coeff.]	8.3768	7.5255	10.5030	6.1981	capitaliz
[t-Score]	***5.2034***	***4.5995***	***5.1443***	***3.2482***	
%vhigh [est. coeff.]		0.2395		0.8201	%vhigh
[t-Score]		***3.3621***		***15.3546***	
%NC [est. coeff.]	0.1226	0.0691	0.2533	0.0019	%NC
[t-Score]	**2.2792**	1.2988	***4.0533***	0.0307	
vh+h+m [est. coeff.]	0.1274		0.5215		vh+h+m
[t-Score]	**2.2510**		***10.6801***		
Adj R-square	0.6865	0.6904	0.4906	0.5747	**Adj R-square**

Cleveland – Refinance

Variable					**Variable**
Intercept	-0.2596	-0.1557	-0.3936	-0.1729	Intercept
	-6.1378	-5.8013	-13.4316	-8.6214	
%black [est. coeff.]	0.1988	0.1238			%black
[t-Score]	***12.4492***	***6.7255***			
%hisp [est. coeff.]	0.0693	-0.0251			%hisp
[t-Score]	1.1136	-0.4123			
%65age [est. coeff.]	0.1635	0.1104			%65age
[t-Score]	***2.8461***	**2.2404**			
medage [est. coeff.]	0.0134	0.0094	0.0028	0.0019	medage
[t-Score]	**2.1879**	1.6132	0.3966	0.3124	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	-0.5386	1.0357	0.8153	***2.8402***	
HT [est. coeff.]	0.0142	0.0298	-0.2029	-0.0665	HT
[t-Score]	0.2246	0.4945	***-2.8433***	-1.0777	
capitaliz [est. coeff.]	16.4428	14.1417	16.9059	12.1840	capitaliz
[t-Score]	***9.4880***	***8.3802***	***8.4575***	***6.9456***	
%vhigh [est. coeff.]		0.7923		1.1672	%vhigh
[t-Score]		***10.3537***		***24.0454***	
%NC [est. coeff.]	0.3718	0.1896	0.4998	0.1288	%NC
[t-Score]	***5.9831***	***3.1951***	***7.5462***	**2.1248**	
vh+h+m [est. coeff.]	0.4403			0.8241	vh+h+m
[t-Score]	***7.0236***			***16.8755***	
Adj R-square	0.8108	0.8268	0.7400	0.8060	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 4: Detailed Regressions for Detroit

Detroit - Home Purchase

Variable	Column 1	Column 2	Column 3	Column 4	Variable
Variable					**Variable**
Intercept	-0.1612	-0.0673	-0.2883	-0.1217	Intercept
	-6.5514	-4.5959	-15.3291	-10.5391	
%black [est. coeff.]	0.1661	0.1414			%black
[t-Score]	***17.3528***	***12.6615***			
%hisp [est. coeff.]	0.0645	0.0671			%hisp
[t-Score]	0.8549	0.8940			
%65age [est. coeff.]	0.1606	0.1108			%65age
[t-Score]	***4.5974***	***3.5032***			
medage [est. coeff.]	-0.0009	-0.0006	0.0073	0.0064	medage
[t-Score]	-0.2483	-0.1527	*1.6466*	1.5942	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***7.0185***	***7.2346***	***9.5542***	***11.2168***	
HT [est. coeff.]	-0.0487	-0.0422	-0.0668	-0.0487	HT
[t-Score]	***-2.7491***	***-2.3909***	***-3.1544***	**-2.5180**	
capitaliz [est. coeff.]	0.9817	0.2664	2.6210	-0.0667	capitaliz
[t-Score]	1.5908	0.4177	***3.6241***	-0.0964	
%vhigh [est. coeff.]		0.2817		0.5624	%vhigh
[t-Score]		***7.9450***		***21.2638***	
%NC [est. coeff.]	0.2134	0.0892	0.3806	0.0654	%NC
[t-Score]	***4.3575***	*1.7369*	***7.1284***	1.2392	
vh+h+m [est. coeff.]	0.2435		0.4483		vh+h+m
[t-Score]	***7.3623***		***15.2271***		
Adj R-square	0.6267	0.6302	0.4622	0.5494	**Adj R-square**

Detroit - Refinance

Variable					**Variable**
Intercept	0.0163	0.0239	0.0160	0.0166	Intercept
	1.2207	2.3102	0.7742	1.0967	
%black [est. coeff.]	0.2577	0.2578			%black
[t-Score]	***40.0263***	***40.0004***			
%hisp [est. coeff.]	0.1282	0.1295			%hisp
[t-Score]	***2.6175***	***2.6440***			
%65age [est. coeff.]	-0.0634	-0.0633			%65age
[t-Score]	**-2.2064**	**-2.2031**			
medage [est. coeff.]	0.0059	0.0059	0.0071	0.0070	medage
[t-Score]	1.6232	1.6277	1.2371	1.2299	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***-5.1794***	***-5.1494***	***-5.6100***	***-5.5512***	
HT [est. coeff.]	-0.0940	-0.0940	-0.1672	-0.1674	HT
[t-Score]	***-4.2685***	***-4.2686***	***-4.6023***	***-4.6095***	
capitaliz [est. coeff.]	12.4840	12.4769	21.6557	21.6289	capitaliz
[t-Score]	***25.9571***	***25.9340***	***32.1928***	***32.1477***	
%vhigh [est. coeff.]		0.0088		-0.0266	%vhigh
[t-Score]		0.4675		-0.8586	
%NC [est. coeff.]	-0.0270	-0.0244	-0.0912	-0.0518	%NC
[t-Score]	-0.9466	-0.6699	*-1.9387*	-0.8615	
vh+h+m [est. coeff.]	0.0190		-0.0006		vh+h+m
[t-Score]	0.9414		-0.0181		
Adj R-square	0.8993	0.8992	0.7224	0.7226	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 5: Detailed Regressions for Houston

Houston - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept	-0.0716	-0.0121	-0.0638	0.0024	Intercept
	-2.3607	-0.6369	-2.4380	0.1439	
%black [est. coeff.]	0.0492	0.0061			%black
[t-Score]	***3.5117***	0.3776			
%hisp [est. coeff.]	-0.0260	-0.0244			%hisp
[t-Score]	-1.4890	-1.4337			
%65age [est. coeff.]	0.1597	0.1507			%65age
[t-Score]	***2.5969***	***2.5793***			
medage [est. coeff.]	-0.0021	-0.0009	0.0026	0.0037	medage
[t-Score]	-0.3409	-0.1577	0.5345	0.8384	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.9668	*1.6872*	1.0104	*1.9404*	
HT [est. coeff.]	-0.0030	0.0002	-0.0025	-0.0003	HT
[t-Score]	-1.0546	0.0876	-0.8813	-0.0933	
capitaliz [est. coeff.]	-0.3612	-1.4909	-1.0640	-2.2156	capitaliz
[t-Score]	-0.3971	-1.6291	-1.1510	**-2.5192**	
%vhigh [est. coeff.]		0.3416		0.3347	%vhigh
[t-Score]		***7.2297***		***9.3429***	
%NC [est. coeff.]	0.0590	-0.0969	0.0596	-0.1120	%NC
[t-Score]	1.0204	*-1.6705*	1.0468	**-1.9726**	
vh+h+m [est. coeff.]	0.2145		0.2307		vh+h+m
[t-Score]	***5.3134***		***6.4863***		
Adj R-square	0.1762	0.2121	0.1302	0.1969	**Adj R-square**

Houston - Refinance

Variable					**Variable**
Intercept	-0.2230	-0.1553	-0.4695	-0.2285	Intercept
	-4.2211	-4.7643	-8.2199	-7.2035	
%black [est. coeff.]	0.4058	0.3194			%black
[t-Score]	***17.8827***	***11.8561***			
%hisp [est. coeff.]	0.0694	0.0660			%hisp
[t-Score]	**2.2102**	**2.1770**			
%65age [est. coeff.]	0.2483	0.2632			%65age
[t-Score]	**2.2765**	**2.5762**			
medage [est. coeff.]	0.0397	0.0446	0.0859	0.0888	medage
[t-Score]	***3.7532***	***4.3637***	***8.0243***	***10.2813***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.2985	1.3561	0.9242	***2.9685***	
HT [est. coeff.]	-0.0296	-0.0227	-0.0206	-0.0101	HT
[t-Score]	***-6.1039***	***-4.6654***	***-3.2921***	*-1.8924*	
capitaliz [est. coeff.]	14.4833	11.5724	10.9087	4.9465	capitaliz
[t-Score]	***9.0106***	***7.1455***	***5.1527***	***2.8008***	
%vhigh [est. coeff.]		0.6078		1.2788	%vhigh
[t-Score]		***6.9964***		***18.2973***	
%NC [est. coeff.]	0.2893	-0.0187	0.5737	-0.2016	%NC
[t-Score]	***2.6597***	-0.1652	***4.0848***	-1.5846	
vh+h+m [est. coeff.]	0.3045		0.8178		vh+h+m
[t-Score]	***4.1601***		***10.1633***		
Adj R-square	0.7364	0.7529	0.5333	0.6690	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 6: Detailed Regressions for Los Angeles

Los Angeles - Home Purchase

Variable	Column 1	Column 2	Column 3	Column 4	Variable
Intercept	-0.0148	0.0871	-0.0453	0.0472	Intercept
	-0.5055	4.7543	-2.0613	3.4345	
%black [est. coeff.]	0.0434	0.0278			%black
[t-Score]	***3.7431***	**2.2361**			
%hisp [est. coeff.]	-0.0738	-0.0662			%hisp
[t-Score]	***-6.5858***	***-6.0490***			
%65age [est. coeff.]	-0.0702	-0.1048			%65age
[t-Score]	*-1.6689*	**-2.5966**			
medage [est. coeff.]	0.0094	0.0088	0.0066	0.0050	medage
[t-Score]	**2.1647**	**2.0267**	1.5305	1.1809	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.4378	0.8086	*1.7249*	***3.0392***	
HT [est. coeff.]	-0.0514	-0.0332	-0.0211	-0.0031	HT
[t-Score]	*-1.9595*	-1.2885	-0.8087	-0.1218	
capitaliz [est. coeff.]	-7.2678	-8.6568	-7.7193	-11.1339	capitaliz
[t-Score]	***-3.8854***	***-4.5039***	***-4.0284***	***-5.8148***	
%vhigh [est. coeff.]		0.3435		0.4428	%vhigh
[t-Score]		***7.7136***		***11.8946***	
%NC [est. coeff.]	0.1144	-0.0043	0.0208	-0.1125	%NC
[t-Score]	**2.4322**	-0.0945	0.5577	***-2.9010***	
vh+h+m [est. coeff.]	0.2952		0.3193		vh+h+m
[t-Score]	***7.3164***		***9.0717***		
Adj R-square	0.1407	0.1441	0.0644	0.0997	**Adj R-square**

Los Angeles - Refinance

Variable					Variable
Intercept [est. coeff.]	-0.0906	-0.0129	-0.1650	-0.0638	Intercept
	-4.3821	-1.0019	-9.8654	-6.2372	
%black [est. coeff.]	0.1378	0.1286			%black
[t-Score]	***16.9109***	***14.6106***			
%hisp [est. coeff.]	0.0280	0.0342			%hisp
[t-Score]	***3.5810***	***4.4814***			
%65age [est. coeff.]	0.0756	0.0452			%65age
[t-Score]	**2.5679**	1.6024			
medage [est. coeff.]	0.0091	0.0087	0.0194	0.0177	medage
[t-Score]	***2.9504***	***2.8080***	***5.8533***	***5.5704***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***3.0705***	***3.1206***	***3.3433***	***5.2530***	
HT [est. coeff.]	-0.0318	-0.0192	-0.0829	-0.0660	HT
[t-Score]	*-1.7193*	-1.0509	***-4.2070***	***-3.5052***	
capitaliz [est. coeff.]	5.5637	4.8410	7.4860	3.8030	capitaliz
[t-Score]	***4.2604***	***3.6001***	***5.1977***	***2.7021***	
%vhigh [est. coeff.]		0.2280		0.4768	%vhigh
[t-Score]		***7.3062***		***17.5866***	
%NC [est. coeff.]	0.1631	0.0799	0.2772	0.1393	%NC
[t-Score]	***4.9454***	**2.5321**	***9.9885***	***4.9591***	
vh+h+m [est. coeff.]	0.2113		0.3472		vh+h+m
[t-Score]	***7.4171***		***13.0532***		
Adj R-square	0.5252	0.5247	0.4009	0.4467	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 7: Detailed Regressions for Milwaukee

Milwaukee - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept [est. coeff.]	-0.0561	0.0130	-0.1595	-0.0106	Intercept
	-1.3438	0.3896	-5.7474	-0.4008	
%black [est. coeff.]	0.1844	0.1457			%black
[t-Score]	***6.8455***	***4.3336***			
%hisp [est. coeff.]	-0.0610	-0.0752			%hisp
[t-Score]	-0.6171	-0.7587			
%65age [est. coeff.]	0.0231	-0.0225			%65age
[t-Score]	0.4227	-0.4502			
medage [est. coeff.]	-0.0010	-0.0006	-0.0124	-0.0095	medage
[t-Score]	-0.1977	-0.1161	***-2.4492***	**-2.0155**	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	-0.3238	-0.6619	0.9549	0.5907	
HT [est. coeff.]	-0.1624	-0.1526	-0.1719	-0.1504	HT
[t-Score]	***-3.8946***	***-3.6747***	***-3.8059***	***-3.6134***	
capitaliz [est. coeff.]	3.8248	2.5950	7.2203	1.5137	capitaliz
[t-Score]	*1.6469*	1.0752	***2.9384***	0.6136	
%vhigh [est. coeff.]		0.2419		0.5094	%vhigh
[t-Score]		**3.3803**		***10.5301***	
%NC [est. coeff.]	0.0356	-0.0717	0.0597	-0.2022	% NC
[t-Score]	0.3727	-0.7106	0.6883	-2.2449	
vh+h+m [est. coeff.]	0.1751		0.3760		vh+h+m
[t-Score]	***3.1259***		***7.8538***		
Adj R-square	0.5929	0.5953	0.4931	0.5567	**Adj R-square**

Milwaukee - Refinance

Variable					**Variable**
Intercept [est. coeff.]	-0.1289	-0.0553	-0.3075	-0.0990	Intercept
	-3.3313	-1.9004	-9.9169	-4.1451	
%black [est. coeff.]	0.2913	0.2290			%black
[t-Score]	***13.4897***	***8.8845***			
%hisp [est. coeff.]	0.0253	-0.0129			%hisp
[t-Score]	0.3411	-0.1760			
%65age [est. coeff.]	0.0682	0.0207			%65age
[t-Score]	1.2791	0.4296			
medage [est. coeff.]	-0.0010	-0.0014	-0.0226	-0.0161	medage
[t-Score]	-0.2040	-0.2998	***-3.7912***	***-3.2240***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.9831	1.0871	**2.4469**	***3.0354***	
HT [est. coeff.]	-0.2229	-0.2103	-0.2733	-0.2261	HT
[t-Score]	***-5.4905***	***-5.3254***	***-5.1182***	***-5.0763***	
capitaliz [est. coeff.]	7.0170	5.3346	13.0116	5.1581	capitaliz
[t-Score]	***3.6779***	***2.7993***	***5.4563***	**2.4298**	
%vhigh [est. coeff.]		0.3505		0.7782	%vhigh
[t-Score]		***6.0860***		***18.1084***	
%NC [est. coeff.]	0.2398	0.1268	0.3423	0.0121	%NC
[t-Score]	***2.8523***	1.5293	***4.1184***	0.1611	
vh+h+m [est. coeff.]	0.2216		0.5925		vh+h+m
[t-Score]	***4.4829***		***11.8902***		
Adj R-square	0.8391	0.8470	0.7107	0.7952	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 8: Detailed Regressions for New York

New York - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					Variable
Intercept	-0.0831	-0.0156	-0.0693	-0.0026	Intercept
	-3.7671	-1.1341	-5.2760	-0.2874	
%black [est. coeff.]	-0.0028	-0.0333			%black
[t-Score]	-0.2905	***-2.9956***			
%hisp [est. coeff.]	-0.0176	-0.0175			%hisp
[t-Score]	-1.1753	-1.1991			
%65age [est. coeff.]	0.0245	-0.0133			%65age
[t-Score]	0.8318	-0.4858			
medage [est. coeff.]	0.0063	-0.0049	-0.0066	-0.0052	medage
[t-Score]	**-2.2128**	*-1.7481*	**-2.3241**	*-1.8580*	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.8508	1.2882	0.8606	0.9210	
HT [est. coeff.]	-0.0671	-0.0652	-0.0698	-0.0650	HT
[t-Score]	***-5.1135***	***-5.0214***	***-5.3603***	***-5.0273***	
capitaliz [est. coeff.]	4.5458	4.0967	4.5306	4.1659	capitaliz
[t-Score]	***4.6141***	***4.1908***	***4.6271***	***4.2846***	
%vhigh [est. coeff.]		0.3385		0.2506	%vhigh
[t-Score]		***8.6606***		***10.5744***	
%NC [est. coeff.]	0.1373	0.0628	0.1113	0.0342	%NC
[t-Score]	***3.1419***	1.4733	***3.0438***	0.8812	
vh+h+m [est. coeff.]	0.2211		0.2046		vh+h+m
[t-Score]	***7.0687***		***9.6398***		
Adj R-square	0.2235	0.2412	0.2237	0.2366	**Adj R-square**

New York - Refinance

Variable					Variable
Intercept	-0.3449	-0.0956	-0.3494	-0.1038	Intercept
	-15.0857	-5.5738	-16.6523	-7.0802	
%black [est. coeff.]	-0.0045	-0.0048			%black
[t-Score]	-0.5259	-0.5912			
%hisp [est. coeff.]	-0.0181	-0.0238			%hisp
[t-Score]	-1.3867	*-1.9461*			
%65age [est. coeff.]	-0.0054	-0.0127			%65age
[t-Score]	-0.1350	-0.3377			
medage [est. coeff.]	0.0244	0.0173	0.0246	0.0175	medage
[t-Score]	***4.8576***	***3.6681***	***5.0704***	***3.8485***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	0.9236	1.1906	0.9846	1.2698	
HT [est. coeff.]	-0.2578	-0.2235	-0.2623	-0.2303	HT
[t-Score]	***-5.0285***	***-4.6395***	***-5.1396***	***-4.7978***	
capitaliz [est. coeff.]	8.2697	5.9878	8.3394	6.0702	capitaliz
[t-Score]	***3.7790***	***2.9259***	***3.8197***	***2.9704***	
%vhigh [est. coeff.]		0.8740		0.8669	%vhigh
[t-Score]		***25.6367***		***25.5495***	
%NC [est. coeff.]	0.6245	0.3339	0.6313	0.3443	%NC
[t-Score]	***9.7477***	***5.2304***	***9.8874***	***5.4100***	
vh+h+m [est. coeff.]	0.7021		0.6974		vh+h+m
[t-Score]	***21.3501***		***21.3121***		
Adj R-square	0.5878	0.6363	0.5881	0.6358	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 9: Detailed Regressions for St. Louis

St. Louis - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept	-0.3851	-0.2098	-0.3840	-0.2093	Intercept
	-10.3472	-8.2588	-10.7522	-8.4073	
%black [est. coeff.]	0.0060	0.0068			%black
[t-Score]	0.5060	0.6852			
%hisp [est. coeff.]	0.2666	0.3189			%hisp
[t-Score]	1.2764	*1.6922*			
%65age [est. coeff.]	-0.0294	-0.0279			%65age
[t-Score]	-0.4692	-0.4977			
medage [est. coeff.]	0.0287	0.0140	0.0290	0.0148	medage
[t-Score]	***3.2903***	*1.7411*	***3.9000***	**2.1538**	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***5.2746***	***6.0803***	***5.2586***	***6.0563***	
HT [est. coeff.]	-0.2985	-0.2102	-0.3006	-0.2131	HT
[t-Score]	***-3.9183***	***-3.1781***	***-3.9678***	***-3.2254***	
capitaliz [est. coeff.]	10.5586	4.7064	10.6740	4.9026	capitaliz
[t-Score]	***4.6207***	**2.1910**	***4.7203***	**2.2988**	
%vhigh [est. coeff.]		0.8341		0.8276	%vhigh
[t-Score]		***12.1652***		***12.2001***	
%NC [est. coeff.]	0.5673	0.1533	0.5672	0.1557	%NC
[t-Score]	***6.4062***	*1.7063*	***6.4251***	*1.7330*	
vh+h+m	0.4893		0.4862		vh+h+m
[t-Score]	***7.3599***		***7.4763***		
Adj R-square	0.5441	0.6289	0.5453	0.6284	**Adj R-square**

St. Louis - Refinance

Variable					**Variable**
Intercept	-0.4462	-0.2706	-0.5173	-0.2867	Intercept
	-8.9409	-8.9943	-12.3150	-10.8358	
%black [est. coeff.]	0.1822	0.1405			%black
[t-Score]	***10.4092***	***8.0440***			
%hisp [est. coeff.]	0.2816	0.2517			%hisp
[t-Score]	0.7563	0.7189			
%65age [est. coeff.]	0.3065	0.2401			%65age
[t-Score]	***4.2338***	***3.7708***			
medage [est. coeff.]	0.0189	0.0192	0.0347	0.0322	medage
[t-Score]	***2.8394***	***3.0790***	***4.9275***	***5.2674***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***4.7326***	***5.0831***	***5.3023***	***5.8190***	
HT [est. coeff.]	-0.1380	-0.1004	-0.3125	-0.2252	HT
[t-Score]	*-1.8453*	-1.4468	***-3.7234***	***-3.0865***	
capitaliz [est. coeff.]	15.1680	12.6709	15.6756	11.5736	capitaliz
[t-Score]	***8.7029***	***7.5884***	***7.7473***	***6.3298***	
%vhigh [est. coeff.]		0.7636		1.0054	%vhigh
[t-Score]		***10.3399***		***14.6164***	
%NC [est. coeff.]	0.5985	0.2600	0.9368	0.3687	%NC
[t-Score]	***6.8804***	***2.9608***	***10.9743***	***4.0613***	
vh+h+m [est. coeff.]	0.5096		0.6599		vh+h+m
[t-Score]	***7.0111***		***9.2071***		
Adj R-square	0.8156	0.8368	0.7509	0.8032	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 10: Detailed Regressions for Washington, D.C.

Washington - Home Purchase

	Column 1	Column 2	Column 3	Column 4	
Variable					**Variable**
Intercept	-0.0921	-0.0403	-0.0839	-0.0303	Intercept
	-4.7182	-3.9111	-6.9137	-3.8307	
%black [est. coeff.]	0.0007	-0.0162			%black
[t-Score]	0.0815	*-1.9010*			
%hisp [est. coeff.]	-0.0230	-0.0117			%hisp
[t-Score]	-1.0384	-0.5382			
%65age [est. coeff.]	0.0415	0.0265			%65age
[t-Score]	1.6110	1.1546			
medage [est. coeff.]	0.0035	0.0043	0.0050	0.0034	medage
[t-Score]	1.4144	*1.7684*	**2.3703**	*1.6626*	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	***6.7120***	***7.7899***	***7.4649***	***7.9575***	
HT [est. coeff.]	-0.0152	-0.0083	-0.0159	-0.0082	HT
[t-Score]	**-2.5370**	-1.4396	***-2.6972***	-1.4197	
capitaliz [est. coeff.]	2.7519	1.2741	2.8480	1.7619	capitaliz
[t-Score]	***3.2323***	1.4574	***3.4670***	**2.1306**	
%vhigh [est. coeff.]		0.2455		0.1992	%vhigh
[t-Score]		***8.2219***		***11.1844***	
%NC [est. coeff.]	0.1122	0.0371	0.1043	0.0239	%NC
[t-Score]	***4.0712***	1.5746	***4.7132***	1.0587	
vh+h+m [est. coeff.]	0.1611		0.1530		vh+h+m
[t-Score]	***5.8323***		***9.3834***		
Adj R-square	0.1876	0.2180	0.1853	0.2168	**Adj R-square**

Washington - Refinance

Variable					**Variable**
Intercept	-0.0885	-0.0067	-0.1401	-0.0285	Intercept
	-4.4291	-0.6134	-10.6061	-3.3379	
%black [est. coeff.]	0.0557	0.0522			%black
[t-Score]	***6.6773***	***6.0619***			
%hisp [est. coeff.]	-0.1044	-0.0916			%hisp
[t-Score]	***-4.7428***	***-4.1683***			
%65age [est. coeff.]	0.1105	0.0694			%65age
[t-Score]	***3.9719***	***2.7602***			
medage [est. coeff.]	0.0015	0.0014	0.0126	0.0094	medage
[t-Score]	0.6225	0.5641	***5.4239***	***4.2054***	
medhhinc [est. coeff.]	0.0000	0.0000	0.0000	0.0000	medhhinc
[t-Score]	1.5437	0.4820	***3.1343***	***2.5557***	
HT [est. coeff.]	-0.0326	-0.0234	-0.0469	-0.0296	HT
[t-Score]	***-4.9534***	***-3.6294***	***-6.7679***	***-4.4176***	
capitaliz [est. coeff.]	5.3927	4.4650	4.8013	2.8950	capitaliz
[t-Score]	***6.2500***	***4.8876***	***5.3119***	***3.2051***	
%vhigh [est. coeff.]		0.2274		0.3725	%vhigh
[t-Score]		***7.3702***		***19.4870***	
%NC [est. coeff.]	0.0900	-0.0049	0.1492	0.0014	%NC
[t-Score]	***3.1698***	-0.2003	***6.0717***	0.0573	
vh+h+m [est. coeff.]	0.2006		0.3043		vh+h+m
[t-Score]	***7.2331***		***17.2681***		
Adj R-square	0.5908	0.5917	0.5151	0.5473	**Adj R-square**

Italic - 10% level of significance
Bolded - 5% level of significance
Bolded and Italicized - 1% level of significance

Table 11: Summary of Regression Results

Home Purchase Lending

Variable	Atl.	Balt.	Cleve.	Det.	Hous.	LA	Milw.	NYC	St. L.	D.C.
%black	+++		+++	+++	+++	+++	+++			
%hisp						- - -				
%65age				+++	+++	-				
medage			++			++		- -	+++	
medhhinc	++			+++					+++	+++
HT				- - -		-	- - -	- - -	- - -	- -
capitaliz			+++			- - -	+	+++	+++	+++
NC		++	++	+++		++		+++	+++	+++
vh+h+m	+++	+++	++	+++	+++	+++	+++	+++	+++	+++
Adj R-square	0.4566	0.0843	0.6865	0.6267	0.1762	0.1407	0.5929	0.2235	0.5441	0.1876

Refinance Lending

Variable	Atl.	Balt.	Cleve.	Det.	Hous.	LA	Milw.	NYC	St. L.	D.C.
%black	+++	+++	+++	+++	+++	+++	+++		+++	+++
%hisp		- -		+++	++	+++				- - -
%65age	+++	++	+++	- -	++	++			+++	+++
medage			++		+++	+++		+++	+++	
medhhinc	+++			- - -		+++			+++	
HT		- - -		- - -	- - -	-	- - -	- - -	-	- - -
capitaliz	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
NC	+	+++	+++		+++	+++	+++	+++	+++	+++
vh+h+m	+++	+++	+++		+++	+++	+++	+++	+++	+++
Adj R-square	0.6903	0.6306	0.8108	0.8993	0.7364	0.5252	0.8391	0.5878	0.8156	0.5908

\+ positive relationship
\- negative relationship
\+ or - 10% significance level
++ or - - 5% significance level
+++ or - - - 1% significance level

Table 12: Impact of Number of African-Americans in a Neighborhood

Percent African-Americans in a census tract

Home Purchase

	Estimated coefficient	Level of Significance	White/African-American Segregation Index
Cleveland	0.2400	***	79.7
Milwaukee	0.1844	***	84.4
Detroit	0.1661	***	86.7
Atlanta	0.1393	***	68.8
Houston	0.0492	***	71.8
Los Angeles	0.0434	***	70.5
Baltimore	0.0063		71.8
St. Louis	0.0060		78.0
Washington	0.0007		66.2
New York	-0.0028		84.3

Refinance

	Estimated coefficient	Level of Significance	White/African-American Segregation Index
Houston	0.4058	***	71.8
Milwaukee	0.2913	***	84.4
Detroit	0.2577	***	86.7
Cleveland	0.1988	***	79.7
Atlanta	0.1866	***	68.8
St. Louis	0.1822	***	78.0
Los Angeles	0.1378	***	70.5
Baltimore	0.1107	***	71.8
Washington	0.0557	***	66.2
New York	-0.0045		84.3

* - 10% level of significance
** - 5% level of significance
*** - 1% level of significance

The dissimilarity index varies between 0 and 100, and measures the percentage of one group that would have to move across neighborhoods to be distributed the same way as the second group. A dissimilarity index of 0 indicates conditions of total integration. A dissimilarity index of 100 indicates conditions of total segregation. For more information see www.CensusScope.org of the Social Science Data Analysis Network at the University of Michigan.

Table 13: Impact of Number of Hispanics in a Neighborhood

Percent Hispanics in a census tract

Home Purchase

	Estimated coefficient	Level of Significance	White/Hispanic Segregation Index
St. Louis	0.2666		36.7
Detroit	0.0645		48.3
New York	-0.0176		69.3
Washington	-0.0230		52.5
Houston	-0.0260		59.2
Cleveland	-0.0317		59.0
Milwaukee	-0.0610		60.6
Los Angeles	-0.0738	***	64.4
Baltimore	-0.0890		40.3
Atlanta	-0.2080		56.8

Refinance

	Estimated coefficient	Level of Significance	White/Hispanic Segregation Index
St. Louis	0.2816		36.7
Detroit	0.1282	***	48.3
Houston	0.0694	**	59.2
Cleveland	0.0693		59.0
Los Angeles	0.0280	***	64.4
Milwaukee	0.0253		60.6
New York	-0.0181		69.3
Washington	-0.1044	***	52.5
Atlanta	-0.2456		56.8
Baltimore	-0.4806	**	40.3

* - 10% level of significance
** - 5% level of significance
*** - 1% level of significance

The dissimilarity index varies between 0 and 100, and measures the percentage of one group that would have to move across neighborhoods to be distributed the same way as the second group. A dissimilarity index of 0 indicates conditions of total integration. A dissimilarity index of 100 indicates conditions of total segregation. For more information see www.CensusScope.org of the Social Science Data Analysis Network at the University of Michigan.

Table 14: Impact of Number of Elderly Residents in a Neighborhood

Percent People over 65

Home Purchase

	Estimated coefficient	Level of Significance
Detroit	0.1606	***
Houston	0.1597	***
Atlanta	0.0845	
Cleveland	0.0698	
Washington	0.0415	
Baltimore	0.0367	
New York	0.0245	
Milwaukee	0.0231	
St. Louis	-0.0294	
Los Angeles	-0.0702	*

Refinance

	Estimated coefficient	Level of Significance
St. Louis	0.3065	***
Atlanta	0.2701	***
Houston	0.2483	**
Cleveland	0.1635	***
Baltimore	0.1307	**
Washington	0.1105	***
Los Angeles	0.0756	**
Milwaukee	0.0682	
New York	-0.0054	
Detroit	-0.0634	**

* - 10% level of significance
** - 5% level of significance
*** - 1% level of significance

NCRC Board Members

Marva Smith Battle-Bay
Vermont Slauson Economic Development Corporation

Lee Beaulac
Rural Opportunities, Inc.

Gail Burks
Nevada Fair Housing Center

Malcolm Bush
The Woodstock Institute

Alan Fisher
California Reinvestment Comm.

Devorah Fong
Spring Creek Community Corporation

Pete Garcia
Chicanos Por La Causa, Inc.

Edward J. Gorman, III
Vermont Slauson Economic Development Corporation

Charles Harris
Housing Education and Economic Development

Irvin Henderson
Community Reinvestment Association of North Carolina

Jean Ishmon
Northwest Indiana Reinvestment Alliance

Alan Jennings
Community Action Committee of the Lehigh Valley

Elbert Jones, Jr.
Community Equity Investments, Inc.

Matthew Lee
Inner City Press/Community on the Move

Maryellen Lewis
Community and Economic Development
Michigan State University

Dean Lovelace
Dayton Community Reinvestment Institute

Eugene Lowe
U.S. Conference of Mayors

Moises Loza
Housing Assistance Council

Gene Ortega
Home Education Livelihood Program

Odalis Reyes

Shelley Sheehy
John Lewis Coffee Shop

Hubert Van Tol
Fairness in Rural Lending

Morris Williams
Coalitition of Neighborhoods

Veronica Williams

Ted Wysocki
LEED Council

Insurance-Based Credit Scores: Impact on Minority and Low Income Populations in Missouri

Brent Kabler, Ph.D.
Research Supervisor
Statistics Section

January 2004

Table of Contents

Charts and Figures

Abstract and Overview

The widespread use of credit scores to underwrite and price automobile and homeowners insurance has generated considerable concern that the practice may significantly restrict the availability of affordable insurance products to minority and low-income consumers. However, no existing studies have effectively examined whether credit scores have a disproportionate negative impact on minorities or other demographic groups, primarily because of the lack of public access to appropriate data.

This study examines credit score data aggregated at the ZIP Code level collected from the highest volume automobile and homeowners insurance writers in Missouri. Findings—consistent across all companies and every statistical test—indicate that credit scores are significantly correlated with minority status and income, as well as a host of other socio-economic characteristics, the most prominent of which are age, marital status and educational attainment.

While the magnitude of differences in credit scores was very substantial, the impact of credit scores on pricing and availability varies among companies and is not directly examined in this study. The impact of scores on premium levels will be directly addressed in studies expected to be completed by late 2004.

Missouri statue prohibits sole reliance on credit scoring to determine whether to issue a policy. However, there are no limits on price increases that can be imposed due to credit scores, so long as such increases can be actuarially justified.

This study finds that:

1. **The insurance credit-scoring system produces significantly worse scores for residents of high-minority ZIP Codes.** The average credit score rank[1] in "all minority" areas stood at 18.4 (of a possible 100) compared to 57.3 in "no minority" neighborhoods – a gap of 38.9 points. This study also examined the percentage of minority and white policyholders in the lower three quintiles of credit score ranges; minorities were overrepresented in this worst credit score group by 26.2 percentage points. Estimates of credit scores at minority concentration levels other than 0 and 100 percent are found on page 8.

2. **The insurance credit-scoring systems produces significantly worse scores for residents of low-income ZIP Code.** The gap in average credit scores between communities with $10,953 and $25,924 in *per capita* income (representing the poorest and

[1] Results are presented here as ranks, or more accurately, *percentiles*. Because of significant differences in the scoring methods of insurers, many of the results in this report are presented as *percentiles* rather than as *percentage differences* in the raw credit scores. Anyone who has taken a standardized test should be familiar with the term. Scores for each company in the sample are ranked, and each raw score is then translated according to its relative position within the overall distribution. For example, a score ranked at the 75th percentile means that the score is among the top one-fourth of scores, and that 75 percent of recorded scores are worse. If the average for non-minorities was at the 30th percentile, and the minority average at the 70th percentile, the *percentile difference* is 40 percentiles. The *percentile difference*, calculated from the statistical models, is used herein as a convenient way to summarize results for the non-technical reader.

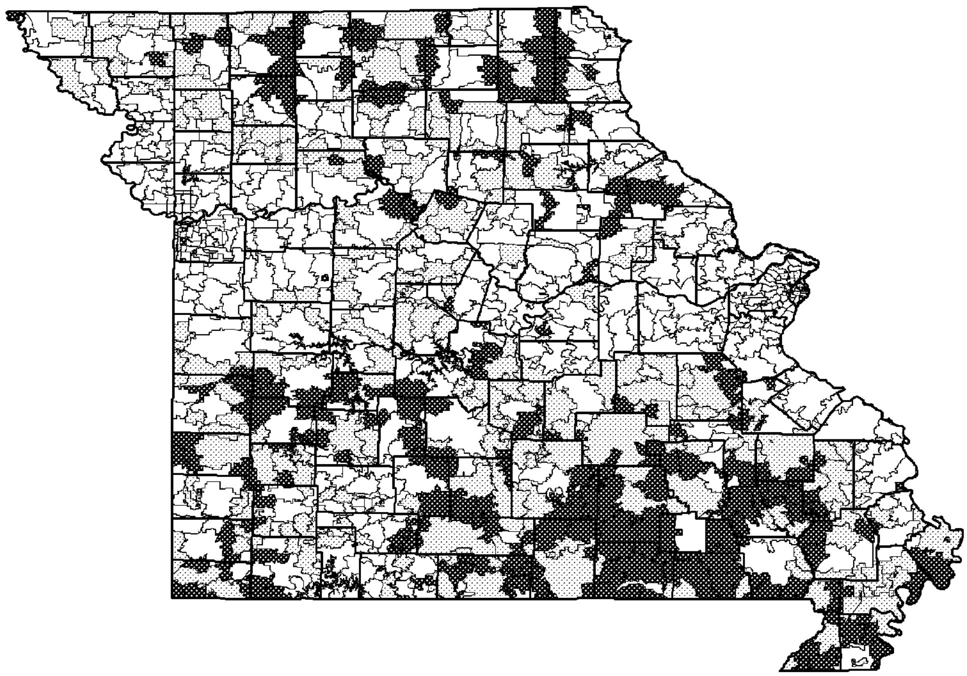
wealthiest 5 percent of communities) was 12.8 percentiles. Policyholders in low-income communities were overrepresented in the worst credit score group by 7.4 percentage points compared to higher income neighborhoods. Estimates of credit scores at additional levels of *per capita* income are found on page 9.

3. **The relationship between minority concentration in a ZIP Code and credit scores remained after eliminating a broad array of socioeconomic variables, such as income, educational attainment, marital status and unemployment rates, as possible causes.** Indeed, minority concentration proved to be the single most reliable predictor of credit scores.

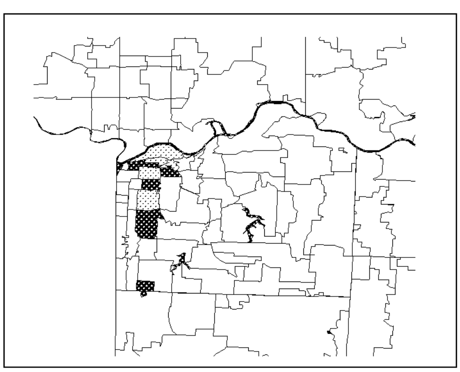
4. **Minority and low-income *individuals* were significantly more likely to have worse credit scores than wealthier individuals and non-minorities.** The average gap between minorities and non-minorities with poor scores was 28.9 percentage points. The gap between individuals whose family income was below the statewide median versus those with family incomes above the median was 29.2 percentage points.

The following maps indicate the areas in Missouri that are most negatively affected by the use of credit scores.

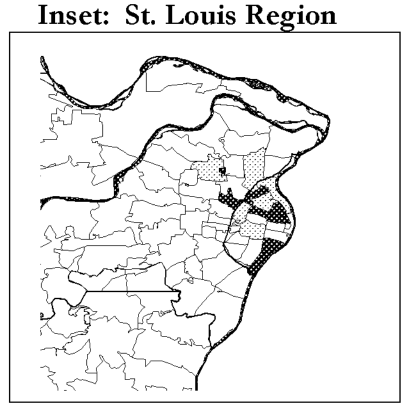
Lower Income Areas of Missouri Most Affected by Credit Scoring



Inset: Kansas City Region



Inset: St. Louis Region

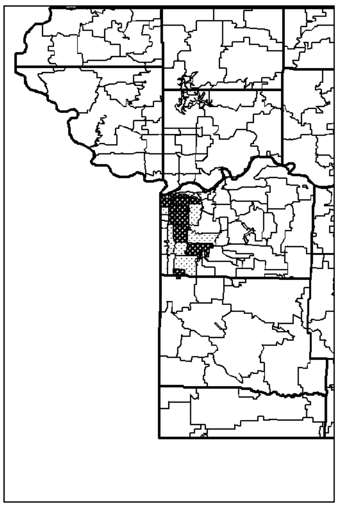


Bottom Quartile = 253 Zip Codes (out of 1,015), with 562,453 persons,
($6,153 - $13,335) or 10% of 5.6 million Missourians

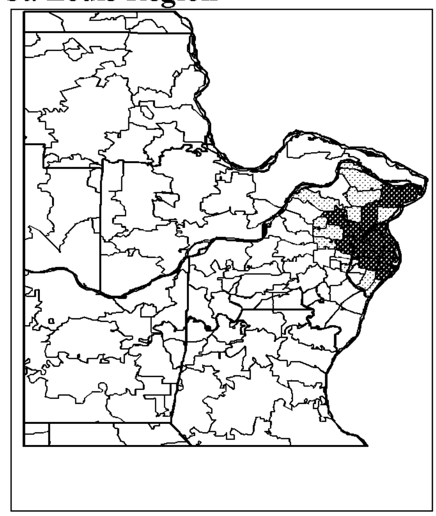
Second Quartile = 254 ZIP Codes with 839,281 persons, or 15% of 5.6
($13,336-$15,326) million Missourians

Areas of Missouri With High Minority Concentration Most Affected by Credit Scoring

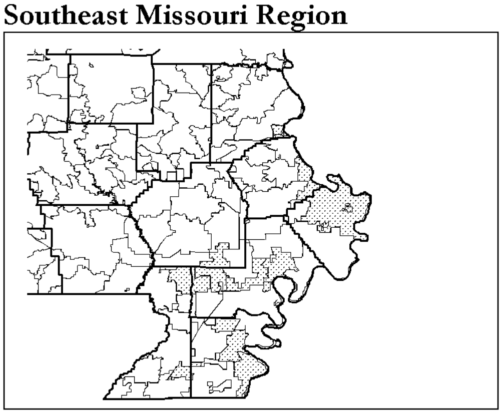
Kansas City Region

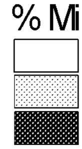


St. Louis Region



Southeast Missouri Region





Missourians in High-Minority ZIP codes

% Minority	White, Non-Hispanic	African-Americans and Hispanics	Other	Total
20% to 50%	337,631	165,441	11,953	515,025
Over 50%	134,541	397,430	10,817	542,788
Total Missouri Population	**4,687,837**	**815,325**	**92,049**	**5,595,211**

Executive Summary

The use of individuals' credit histories to predict the risk of future loss has become a common practice among automobile and homeowners insurers. The practice has proven to be controversial not only because of concerns about how reliably credit scores may predict risk. Many industry professionals, policymakers, and consumer groups have expressed concern that the practice may pose a significant barrier to economically vulnerable segments of the population in obtaining affordable automobile and homeowners coverage.

This study finds evidence that justifies such concerns.

Four questions are addressed in the study:

1. Is there a correlation between place of residence and insurance-based credit scores (called "credit scores" or "scores" throughout the remainder of this report)? Specifically, do residents of areas with high minority concentrations have worse average scores?

2. Do residents of poorer communities have worse average scores?

3. If credit scoring has a disproportionate impact on residents of communities with high minority concentrations, what other socioeconomic factors might account for this fact?

4. Do minorities and poorer individuals tend to have worse scores than others, irrespective of place of residence?

For this report, the category 'minority' includes all Missourians who identified themselves as African-American or Hispanic in the 2000 census. A separate analysis of African-Americans resulted in no substantive difference from the results presented here.

Data

Credit score data was solicited from the 20 largest automobile and homeowners writers in Missouri for the period 1999-2001. Of these, 12—individually or combined with sister companies—had used a single credit scoring product for a sufficient period of time to generate a credible sample. In some instances, a single company is displayed as two separate "companies" representing separate analyses of automobile and homeowners coverage. In other instances, sister companies were combined to yield a more statistically credible sample. The net result of these combinations is the 12 "companies" presented in the report.

Companies That Submitted Data for this Report

NAIC Code	Name
16322	Progressive Halcyon Insurance Co.
17230	Allstate Property & Casualty Insurance Co.
19240	Allstate Indemnity Co.
21628	Farmers Insurance Co., Inc.
21660	Fire Insurance Exchange
21687	Mid-Century Insurance Co.
22063	Government Employees Insurance Co.
25143	State Farm Fire And Casualty Co.
25178	State Farm Mutual Automobile Insurance Co.
27235	Auto Club Family Insurance Co.
35582	Government General Insurance Co.
42994	Progressive Classic Insurance Co.

Additional information about how the Missouri's largest insurers use credit scores can be found at the MDI web site, www.insurance.mo.gov.

The companies provided average credit scores by ZIP Code, as well as the distribution of exposures (automobiles and homes) across five credit score intervals representing equal numeric ranges. Both the average score and the percent of exposures in the worst three intervals are used to assess to the degree to which race and ethnicity and socioeconomic status are correlated with credit scores.

Because of the nature of the data, results are presented from two categorically distinct levels of analysis:

1. *Aggregate level*—Inferences about **residents in areas with high minority concentrations or areas with lower incomes**. This level of analysis does not purport to make inferences about minority or lower-income individuals *per se*.

2. *Individual level*—Assessments of the likely impact of credit scores on minority **individuals**, without reference to place of residence. These results make use of statistical models that are widely employed in the social sciences, but findings are somewhat more speculative than are the aggregate level results.

Findings

1. On average, residents of areas with high minority concentrations tend to have significantly worse credit scores than individuals who reside elsewhere.

2. On average, residents of poor communities tend to have significantly worse credit scores than those who reside elsewhere.

Given the variation in credit scoring methodologies, raw credit scores possess no intrinsic meaning, and comparing raw scores across companies is of limited value. Normalized or "standardized" results afford more meaningful comparisons. Averaged across all companies, the spread in standardized scores between "no minority" and "all minority"[2] ZIP Codes was 38.9 percentiles—a very considerable gap.[3] For more than half of the companies, the average scores of individuals residing in minority ZIP Codes fell into the bottom one-tenth of scores (that is, at or lower than the 10th percentile). The average score of individuals residing in non-minority ZIP Codes fell into the upper one-half of scores for every company.

The last three columns of the table display percentile differences by income group. On average, ZIP Codes with a *per capita* income of $25,924 (the top 5 percent of ZIP Codes) had scores that were 12.8 percentiles higher than ZIP Codes with a *per capita* income of $10,953 (the bottom 5 percent of ZIP Codes).

---

[2] The statistical models incorporate data from all ZIP Codes to determine the overall relationship between minority concentration and credit scores. Estimates derived from the models are presented here at the extremes of 0 percent and 100 percent minority concentration for expository reasons (the meaning of values at the extremes is usually more intuitive). For example, if the regression model indicated that every percentage point increase in minority concentration is associated with a decrease in credit scores of 1.68 points, the impact of increasing minority concentration to 100 percent would be a decline of 168 points. In reality, there are no ZIP Codes whose residents are all minorities, though several ZIP Codes have more than 95 percent minority concentration.

[3] Percentile differences are based on normalized scores ranging from 0 to 100, and represent the rank of a score relative to all other scores in the sample. Such percentiles are exactly analogous to those used for reporting standardized test results. For example, a score falling in the 75th percentile means the score is among the top one-fourth of scores. The numbers reported in the table below represent the percentile difference between high and low minority ZIPs. For example, if the average score of high minority ZIP Codes was at the 20th percentile, and those for low minorities at the 80th percentile, the difference is 60 percentiles.

Standardized Credit Scores (Percentiles) by Minority Concentration and *Per Capita* Income in ZIP Code

Results of Weighted OLS Regression of Average Credit Score

Scores Coded So that a *Lower* Score is *Worse*

	Average Score Percentile by Minority Concentration (on a scale of 100)			Average Score Percentile by *Per Capita* Income (on a scale of 100)		
Company[4]	**100% Minority**	**0% Minority**	**Percentile Difference**	**$10,953 (Poorest 5% of ZIP Codes)**	**$25,924 (Wealthiest 5% of ZIP Codes)**	**Difference**
A	24.2	54.0	29.8	35.9	51.6	15.7
B	2.1	59.5	57.4	37.8	52.4	14.6
C	5.8	59.1	53.4	30.5	52.4	21.9
D	11.9	56.4	44.5	44.4	52.8	8.4
E	12.3	57.9	45.6	46.8	54.8	8.0
F	30.5	59.5	29.0	46.0	57.9	11.9
G	29.1	59.1	30.0	42.9	56.8	13.9
H*	22.4	56.0	33.6	45.2	52.8	7.6
I*	33.0	50.8	17.8	41.3	48.0	6.7
J	14.2	59.9	45.6	40.5	55.2	14.7
K	25.1	55.6	30.4	44.0	53.6	9.6
L	9.7	59.5	49.8	34.8	55.2	20.3
Average (Unweighted)	**18.4**	**57.3**	**38.9**	**40.9**	**53.6**	**12.8**

**These two companies were unable to provide MDI with raw credit scores. Data thus consists of scores that have been furthered modified based on non-credit related information prior to being used for rating / underwriting.*

In addition to average credit scores by ZIP Code, the number of exposures[5] in five equal credit score intervals was also collected; each interval represents the range of scores divided by five.[6] The proportion of exposures in the worst three intervals was used, as a parallel measure to average scores, to assess the association between race and income and credit scores. On average, a 26.2 percentage point difference existed in the proportion of exposures in the worst credit score group between "all minority" and non-minority ZIP Codes. The corresponding gap between the wealthiest and poorest income groups was 7.4 percentage points.

Estimates for additional levels of minority concentration and *per capita* income are displayed in the following four tables.

[4] This report represents an analysis of credit scoring in general, and not the compliance of a specific company with any laws, nor the degree to which a company deviated from the norm. Thus, no individual companies are identified when displaying results.

[5] One "exposure" is equal to one year of coverage for one automobile or home.

[6] For clarification, credit score intervals are not quintiles where each interval represents an equal number of exposures. Rather, each interval is an equal numeric range in credit scores, and exposures are not distributed equally between intervals.

Percent of Exposures in Worst 3 Credit Score Intervals by % Minority and *Per Capita* Income in a ZIP Code

Results of Weighted OLS Regression

	Scores in Worst Group by Percent Minority			Scores in Worst Group by *Per Capita* Income		
Company	**0% Minority**	**100% Minority**	**Difference**	**$10,953 (Poorest 5% of ZIP Codes)**	**$25,924 (Wealthiest 5% of ZIP Codes)**	**Difference**
A	41.4%	64.8%	23.4%	52.4%	44.4%	8.0%
B	8.9%	53.7%	44.9%	19.4%	12.5%	6.9%
C	20.5%	61.7%	41.2%	35.8%	25.1%	10.7%
D	26.7%	57.2%	30.6%	34.4%	28.2%	6.2%
E	33.7%	73.2%	39.5%	42.6%	35.9%	6.7%
F	38.9%	62.3%	23.5%	50.9%	39.5%	11.3%
G	14.5%	31.9%	17.4%	22.9%	16.2%	6.7%
H	21.7%	37.1%	15.5%	26.7%	22.9%	3.8%
I	68.3%	79.7%	11.4%	75.0%	68.0%	7.0%
J	12.1%	30.4%	18.3%	19.0%	13.8%	5.2%
K	13.2%	28.4%	15.2%	18.6%	14.2%	4.4%
L	21.8%	55.5%	33.7%	35.9%	24.1%	11.8%
Average (Unweighted)	**26.8%**	**53.0%**	**26.2%**	**36.1%**	**28.7%**	**7.4%**

Standardized Credit Scores (Percentiles) by % Minority in a ZIP Code

Results of Weighted OLS Regression of Average Credit Score

Scores Coded So that a *Lower* Score is *Worse*

Company	**0% Minority**	**25% Minority**	**50% Minority**	**75% Minority**	**90% Minority**	**100% Minority**
A	54.0	46.0	38.2	30.9	26.8	24.2
B	59.5	37.1	18.4	7.2	3.6	2.1
C	59.2	41.3	24.2	13.1	8.2	5.8
D	56.4	42.9	30.5	20.1	14.9	11.9
E	57.9	44.4	31.6	20.6	15.2	12.3
F	59.5	48.0	44.8	37.5	33.0	30.5
G	59.1	48.4	43.6	36.3	31.9	29.1
H	56.0	46.8	37.8	29.8	25.1	22.4
I	50.8	46.0	41.7	37.1	34.5	33.0
J	59.9	46.8	34.1	23.0	17.4	14.2
K	55.6	47.6	39.4	31.9	27.8	25.1
L	59.5	44.0	29.8	17.9	12.5	9.7
Average	57.3	44.9	34.5	25.4	20.9	18.4

Percent of Exposures in Worst 3 Credit Score Intervals
by % Minority in a ZIP Code
Results of Weighted OLS Regression

Company	0% Minority	25% Minority	50% Minority	75% Minority	90% Minority	95% Minority	100% Minority
A	41.4	47.2	53.1	58.9	62.4	63.6	64.8
B	8.9	20.1	31.3	42.5	49.2	51.5	53.7
C	20.5	30.8	41.1	51.4	57.6	59.6	61.7
D	26.7	34.3	42.0	49.6	54.2	55.7	57.2
E	33.7	43.6	53.5	63.3	69.2	71.2	73.2
F	38.9	44.7	50.6	56.5	60.0	61.2	62.3
G	14.5	18.9	23.2	27.6	30.2	31.0	31.9
H	21.7	25.5	29.4	33.3	35.6	36.4	37.1
I	68.3	71.2	74.0	76.9	78.6	79.2	79.7
J	12.1	16.7	21.2	25.8	28.5	29.5	30.4
K	13.2	17.0	20.8	24.6	26.9	27.6	28.4
L	21.8	30.2	38.6	47.1	52.1	53.8	55.5
Average	26.8	33.4	39.9	46.4	50.4	51.7	53.0

Standardized Credit Scores (Percentiles) by *Per Capita* Income in ZIP Code
Results of Weighted OLS Regression of Average Credit Score
Scores Coded So that a *Lower* Score is *Worse*

Company	Bottom 1% ($8,642)	Quartile 1 ($13,335)	Quartile 2 ($15,326)	Quartile 3 ($18,092)	Top 1% ($50,536)
A	33.4	38.2	40.5	43.3	76.1
B	35.9	40.1	42.1	44.8	74.5
C	27.4	33.7	36.7	40.5	84.1
D	43.3	45.6	47.2	48.4	65.9
E	45.2	48.0	49.2	50.4	67.7
F	44.0	48.0	49.6	51.6	75.5
G	40.9	45.2	46.8	49.6	76.7
H	44.0	46.4	47.6	48.8	64.4
I	40.1	42.5	43.3	44.4	59.1
J	38.2	42.9	44.8	47.6	77.0
K	42.5	45.6	46.8	48.4	68.4
L	31.9	37.8	40.5	48.8	83.7
Average (Unweighted)	**38.9**	**42.8**	**44.6**	**47.2**	**72.8**

Percent of Exposures in Worst Three Credit Score Intervals
by *Per Capita* Income a ZIP Code
Results of Weighted OLS Regression

Company	Bottom 1% ($8,642)	Quartile 1 (13,335)	Quartile 2 (15,326)	Quartile 3 (18,092)	Top 1% (50,536)
A	53.6	51.1	50.1	48.6	31.6
B	20.5	18.3	17.4	16.1	1.4
C	37.4	34.1	32.6	30.7	7.9
D	35.3	33.4	32.6	31.4	18.3
E	43.6	41.5	40.6	39.4	25.1
F	52.6	49.1	47.6	45.5	21.3
G	23.9	21.8	20.9	19.7	5.4
H	27.3	26.1	25.6	24.8	16.7
I	76.1	73.9	73.0	71.7	56.8
J	19.8	18.2	17.5	16.5	5.5
K	19.3	17.9	17.3	16.5	7.2
L	37.7	34.0	32.4	30.2	5.1
Average (Unweighted)	**37.3**	**34.9**	**34.0**	**32.6**	**16.9**

3. Credit scores are significantly correlated with minority concentration in a ZIP Code, even after controlling for income, educational attainment, marital status, urban residence, the unemployment rate and other socioeconomic factors.

Statistical models were used to control for—i.e., remove—the impact of socioeconomic factors that might account for the correlation between race/ethnicity and credit scores. The inclusion of such controls slightly weakened, but by no means eliminated (or accounted for) the association between minority status and credit scores. Among all such control variables, race/ethnicity proved to be the most robust single predictor of credit scores; in most instances it had a significantly greater impact than education, marital status, income and housing values. It was also the only variable for which a consistent correlation was found across all companies.

Other variables found to be significantly correlated with credit scores across the majority of companies were educational attainment, age, marital status, and urban residence.

Why scores should be correlated with minority status, even after controlling for such broad measures of socioeconomic status, is not immediately clear. Such a result indicates that the variable "minority concentration" contains unique characteristics not contained in the "control" variables. For example, credit scores may reflect factors uniquely associated with racial status (such as limited access to credit, for example). The results clearly call for further study.

4. The minority status and income levels of *individuals* are correlated with credit scores, regardless of place of residence.

Three different statistical models were used to assess differences in scores between minority and low-income **individuals,** as opposed to **residents of high minority or low-income areas** (not all of whom, of course, are minorities or poor). **Based on the most credible of the three models, African-American and Hispanic insureds had scores in the worst credit score group at a rate of about 30 percentage points higher than did other individuals (for example, where 30 percent of one group may have poor scores, compared to 60 percent of another group). A gap of 30 percentage points also existed between individuals earning below and above the median family income for Missouri.** Across companies, the gap for minority status ranged from 14 percent to 48 percent; and for income the gap ranged from 17 to 46 percent.

Difference in % of individuals in the worst 3 (of 5) credit score intervals

Estimates of Gary King's Ecological Inference (EI) Model[7]

Company	**Minority Status** (% of minorities with low scores minus % of non-minorities with low scores)	**Income** (% of lower-income individuals with low scores minus % of higher-income individuals with low scores)
A	19.1%	27.7%
B	39.5%	16.8%
C	42.1%	46.1%
D	30.6%	22.5%
E	47.9%	28.5%
F	25.8%	35.6%
G	14.5%	21.0%
H	29.1%	32.8%
J	15.0%	26.7%
K	15.3%	26.4%
L	38.5%	37.2%
Unweighted Average	**28.9%**	**29.2%**

[7] The EI model is one of three employed in this report to make individual-level inferences. The other two are Goodman's Regression and the "Neighborhood" model, each of which is explained in the body of the report.

While considerable variation exists among the three models with respect to the magnitude of estimates, all three consistently estimated a disproportionate impact based on the minority status of individuals and an individual's family income.

Because the data is composed of ZIP Code level aggregates, inferences about individual-level characteristics are somewhat more speculative than are inferences about the demographic characteristics of place of residence. Individual-level estimates in this report result from three of the most widely-used statistical models for such purposes. *While the model results are not "proof" of an* ***individual-level*** *disproportionate impact, the evidence appears to be substantial, credible and compelling*.

I. Introduction

Use of credit scores by insurers has come into prominence within the last ten years. A recent study found that more than 90 percent of personal lines insurers use credit scores for rating or underwriting private automobile insurance (Conning & Co., 2001), and many insurers also use credit scoring for homeowners coverage. Such scores are distinguished from credit scores used in financial underwriting. While both lending and insurance scores have many elements in common, insurance-based credit scores purport to predict the risk of insurance loss rather than the risk of financial default.

The insurance industry has produced studies indicating that credit scores are predictive of both loss frequency and severity for a wide variety of coverages. For example, for private passenger automobile insurance, one study found credit scores highly predictive of liability (both BI and PD), collision, comprehensive, uninsured motorist and medical payment losses (Miller and Smith, 2003. See also Tillinghast-Towers Perrin, 1996; Monaghan, 2000; and Kellison, Brockett, Shin, and Li, 2003).

This study does not examine the relationship between credit scores and the likelihood of insurance losses. Regulators and consumer groups have expressed growing concern that use of credit scores may restrict the availability of insurance products in predominantly minority and low income communities, markets that already show signs of significant affordability and access problems (Kabler, 2004).

Components common to most scoring models have been made public: high debt to limit ratios, derogatory items such as collection actions, liens, and foreclosures, the number of loan and credit card applications, and the number of credit accounts. Many of these items are known to be correlated with both income and minority status. The largest study of its kind, the Freddie Mac Consumer Credit Survey, concluded that both African-Americans and Hispanics were significantly more likely to have derogatory items on their credit history than were their white counterparts. Similar gaps were observed between income groups (Freddie Mac, 1999).

Many analysts also contend that credit scores, which weigh items that signify financial distress or limited availability of credit, are correlated with minority status. Significant debate has continued about lending practices that restrict access to credit in minority communities—a factor that could have a significant impact on insurance-based credit scores. Minority communities in core urban areas also are more typically vulnerable to economic dislocations, such as significantly elevated un- and under-employment rates, that produce the kind of financial distress likely to be measured by credit scoring models.

Unfortunately, no rigorous studies have directly examined what, if any, impact the growing prevalence of insurance credit scores has had on the availability of insurance coverage in poor and minority communities.

The studies that have entered the public domain have been largely inconclusive or suffer from serious methodological deficiencies. A study funded by the American Insurance Association (AIA), an industry trade association, found no correlation between income and credit scores (AIA, 1998). However, the AIA study appears to suffer from methodological flaws so serious that no conclusions are warranted.[8]

The Virginia Bureau of Insurance sponsored a study based on ZIP Code aggregates. Unfortunately, the numeric results of the analysis were never publicly released. Rather, the Bureau's report stated that "Nothing in this analysis leads the Bureau to the conclusion that income or race alone is a reliable predictor of credit scores, thus making the use of credit scoring an ineffective tool for redlining"—a statement that could reasonably be made even with a finding of a very significant disproportionate impact (Commonwealth of Virginia, 1999).[9]

More recently, the Washington Department of Insurance sponsored a consumer survey that matched demographic information obtained from telephone interviews with credit scores (Pavelchek and Brown, 2003). While the study found a statistically significant association between credit scores and income, the findings regarding the racial impact of scoring were inconclusive, primarily because of the small number of minorities included in the survey sampled from the relatively homogonous population of the state of Washington .

A literature review by the American Academy of Actuaries (2002) has also concluded that existing studies were inconclusive with respect to the disproportionate impact issue. This study begins filling that void.

Caveats and Limitations of Study

This study is based on ZIP Code-level credit score averages and is subject to certain limitations. Unlike a survey of individuals, in which demographic data such as race and income are obtained directly, this analysis makes inferences based on patterns observed in aggregate relationships (such as average credit score in a ZIP Code). The reader is therefore

---

[8] The study suffers from two serious flaws. First, based on conversations with the data provider, the data used in the study is not a random sample of the population about which inferences are made. Rather, it is a marketing sample that systematically excludes poorer individuals, renters, and individuals who had recently relocated. Secondly, the dependent variable, income, is not directly measured but rather estimated via a procedure that is not explained.

[9] Based on conversations with Virginia analysts, the study does not appear to have been designed to measure disproportionate impact. The study's conclusion is relevant only to acts of intentional discrimination, where in the Bureau's opinion credit scores are ineffective for such purposes due to the fact that many non-minorities also have poor scores, and that credit scores may be related to other socioeconomic characteristics such that the *sole* use of scores is "ineffective." In technical terms, this conclusion is based on the R-squared value of the regression models used (which measure how "precise" scores are at targeting only minorities). Unfortunately, the R-Squared values were not reported, and there is clearly an element of subjective judgment about what level of R-Squared renders credit scoring an effective tool for "intentional" discrimination, let alone what might constitute a significant disproportionate impact. For example, one could conclude that, while 60 percent of minorities have poor scores, because 30 percent of non-minorities have poor scores that scores are not precise enough to be used as a "redlining" tool. However, such results would indicate a substantial disproportionate racial impact.

alerted to the dangers of conflating two categorically distinct levels-of-analysis contained in the report:

1. **Macro or Aggregate Level-of-Analysis**

Inferences made about the correlation between average credit scores and demographic characteristics of ZIP codes.

2. **Micro or Individual Level-of-Analysis**

Inferences made about the correlation between **individual traits** and credit scores, irrespective of place of residence

The macro-level analysis (# 1) based on ZIP Code characteristics can produce valid inferences about "individuals that reside in poorer ZIP Codes,' or "individuals that reside in areas with large minority concentrations," but **not** about **minority individuals** or **poor individuals** *per se*; data limitations prevent any **direct** inferences about the relationship between credit scores and individual characteristics such as race/ethnicity or socioeconomic status (see methodological appendix).

However, the ecological or aggregate relationship is meaningful on its own terms, and possesses broad implications for important public policy issues. Federal courts, as well as statutes in many states, restrict or prohibit the use of geographic area as a rating or underwriting factor in personal lines. Such "redlining" issues are most directly relevant to the racial mix of an area, and not necessarily the race or ethnicity of *individuals* residing in such areas who might be harmed. In fact, non-minorities have been recognized in both lending and insurance litigation as possessing an actionable claim if they are harmed by business practices with negative consequences associated with the racial composition of areas in which they reside (Cf. United Farm Bureau Mutual Insurance Co v. Metropolitan Human Relations Commission, 24F.3d 1008 (7th Circuit, 1994).

The individual-level analysis (# 2) is based on statistical procedures that model underlying individual-level distributions that could account for the observed ZIP Code level distributions. Thus, the results are somewhat more speculative than are the direct ZIP Code level observations. The results of three different models for each company/ insurance line combination are presented. These results, *taken together*, provide credible and compelling, if not irrefutable, evidence for conclusions.

An additional limitation of this study is that some sparsely populated ZIP Codes were not included in the analysis due to a lack of data. This problem was acute in some cases where companies used scores for new business only, or did not use scores over the entire study period (1999-2001). For the aggregate-level analysis, this problem was minimized by the use of "weights" based on ZIP Code exposures. For the individual-level analysis, ZIP Codes lacking credible data were deleted. In all instances, the number of ZIP Codes included in the analysis, as well as the percent of Missouri's population residing in those ZIP Codes, is reported for each table.

Among the findings of the report are:

Aggregate analysis

1. Mean credit scores are significantly correlated with the minority concentration in a ZIP Code.

2. Mean credit scores are correlated with socioeconomic characteristics, particularly income, educational attainment, marital status, and age.

3. The correlation between minority concentration and credit scores remains even after controlling for numerous other socioeconomic characteristics that might be expected to account for any disproportionate impact of credit scores on minorities. Indeed, minority concentration proved to be a much more robust predictor of credit scores than any of the socioeconomic variables included in the analysis.

Individual-Level Analysis

1. Credit scores appear to be significantly correlated with race/ethnicity and with family income.

Data and Methodology

Credit score data aggregated at the ZIP Code level was solicited from the 20 largest home and automobile insurance writers in the state. A total of 12 insurers had credible data for at least one line of insurance for the study period of 1999 to 2001. The data contained the following elements for each Missouri ZIP Code:

1. Mean credit score
2. The number of exposures for each of five equal credit score intervals

These two data elements constitute our dependent variables, with the second measured by the percent of exposures (insured automobiles or homes) falling into the worst three of five credit score intervals. Demographic data for each Zip Code was obtained from the 2000 decennial census.

The aggregate analysis was performed using weighted regression, where each observation weight was based on number of exposures. The individual-level inferences are the product of three different models: Goodman's Regression, the Neighborhood Model, and Gary King's EI method. Each model entails different requisite assumptions. Conclusions are presented only in those instances in which the results of each model are concordant. In addition, the maximum possible bounds for individual-level estimates are presented. These models are more fully described in the methodological appendix.

The Dependent Variable: Disproportionate Impact

The primary purpose of this study is to measure the level of disproportionate impact between credit scores and race/ethnicity, and credit scores and socioeconomic status. Disproportionate impact is defined as the ***bivariate*** relationship between credit scores and the independent variable of interest, such as race/ethnicity or income. That is, for purposes of *measuring the level of disproportionate impact*, no attempt is made to control for possible confounding variables, or factors that might **explain** a disproportionate impact should one be identified.

A secondary purpose of this study—for which the data is less well suited—is to tentatively identify *causal* explanations for any disparities that might be observed. This causal analysis does employ statistical controls for possible confounding variables related to socioeconomic status. However, the reader should bear in mind the differing purposes of the **bivariate** and **multivariate** analyses: the first is the **measure** of disproportionate impact; and the second a rudimentary **causal** analysis of disproportionate impact. Multivariate regression is employed for the aggregate analysis only. Due to both data and methodological limitations, the individual-level analysis is not amenable to a multivariate analysis of any complexity.[10]

This interpretation of disproportionate impact conforms to various judicial interpretations. A clear judicial statement regarding the statistical issues was issued by the Supreme Court in **Thornburg v. Gingles, 478 U.S. 30 (1986).** While there were separate concurring opinions, there was no disagreement regarding the statistical problem associated with the case. At issue was alleged gerrymandering that diluted the voting strength of minorities across several districts. Given the relevancy of the court's opinion to issues discussed above, the decision is worth quoting at some length:

"Appellants argued that the term 'racially polarized voting' must, as a matter of law, refer to voting patterns for which the principal cause is race. Courts erred by relying only on bi-variate analysis which merely demonstrated a correlation between the race of the voter and the level of voter support for certain candidates, but which did not prove that race was the primary determinant of voters' choices. The court must also consider party affiliation, age, religion, income, educational levels, media exposure…"

....

"Appellant's argument [was] that the proper test was not voting patterns that are "merely correlated with the voter's race, but to voting patterns that are determined primarily by the voter's race, rather than by the voter's other socioeconomic characteristics."

---

[10] One can postulate a variety of causal paths: race (or racial discrimination) *causes* lower incomes relative to majority groups. Lower incomes in turn might *cause* lower credit scores. Such causal chains are not well identified in models that implicitly assume that all causal variables operate **simultaneously and independently** upon credit scores. Multivariate analyses such as multiple regression asks the question "if African-Americans were identical to whites with respect to income, education, occupation, etc, would racial status still be correlated with credit scores?" This is not necessarily the most important question for our purposes. However, our (aggregate) data **do not** permit a full path analysis whereby complex causal relationships can be more appropriately modeled. Our analysis is limited to identifying whether any residual correlation between race / ethnicity remains that cannot be accounted for by socioeconomic variables. We recognize that such an analysis may raise more questions than it answers.

The Court refused the appellants' argument that a demonstration that minorities vote in recognizable patterns that differ from majority voting must use multivariate analysis to determine the **causes** of differences in voting; and that voting differences must persist after **removing** or **controlling** for such causes (i.e. income, etc.).

Justices Brennan, Marshall, Blackman, and Stevens wrote:

"The reasons black and white voters vote differently have no relevance to the central inquiry….[regarding the legal test]…It is the difference between the choices made by blacks and whites-not the reasons for that difference-that results in blacks having less opportunity than whites to elect their preferred representative…only the correlation between race of voter and selection of certain candidates, not the causes of the correlation, matters."

"A definition of racially polarized voting which holds that black bloc voting does not exist when black voters' choice of certain candidates is most strongly influenced by the fact that the voters have low incomes and menial jobs- when the reason most of those voters have menial jobs and low incomes is attributable to past or present racial discrimination…"

Justice O'Connor, joined by Justices Powell and Rehnquist, issued a concurring opinion:

"Insofar as statistical evidence of divergent racial voting patterns is admitted solely to establish that the minority group is politically cohesive and to assess its prospects for electoral success, such a showing cannot be rebutted by evidence that the divergent voting patterns may be explained by causes other than race.

Results

Regression results for each company are displayed for each of the following relationships:

Aggregate-Level (Macro) Analysis:

1. The bivariate relationship between credit scores and % minority in a ZIP Code
2. The bivariate relationship between credit scores and per capita income in a ZIP Code
3. A multivariate analysis incorporating race /ethnicity, income, and additional socioeconomic variables.

For each of the three general types of relationships, two different measures of credit scores is used: mean credit score, and the percent of individuals that fall into the worst three of five credit score intervals (as defined above). Since the nominal value of credit scores possesses no intrinsic meaning, regression results are presented as standard deviations from the sample mean, with mean=0 and standard deviation=1.

Individual-Level (Micro) Analysis

1. The bivariate relationship between minority status and the percent of exposures in the worst three credit score intervals
2. The bivariate relationship between family income and the percent of exposures in the worst three credit score intervals

This report contains no information that would identify specific companies.

The Relationship Between Demographic Characteristics of an Area and Credit Scores

Regression coefficient estimates for each company/line of business combination (called "companies" in the following tables) are displayed in the Tables 1-5. The racial/ethnic composition of ZIP Codes is strongly correlated with the average credit score of a ZIP Code for all companies. Table 1 indicates that, averaged across companies, a one percent increase in minority concentration is associated with a change in credit score of -.012 standard deviations. That is, as the minority concentration in a ZIP Code approaches 100 percent, the average credit score is 1.2 standard deviations below (i.e. worse than) ZIP Codes with no minority residents. In a few instances, average credit scores decreased by over two standard deviations. In no instance was a credit score not significantly correlated with racial composition.

The R-Squared values, representing the proportion of the variation in credit scores "explained" by the model, are displayed in the final column. R-Square values range from .0419 to .5261, so that in at least some instances, the single variable (minority concentration) accounts for a majority of the variability in credit scores across ZIP Codes. In other instances, minority concentration accounts for little of such variability.

Table 1: Mean Credit Score (Standard Deviation) = $B_1 + B_2$ (% Minority) + e
Weighted OLS Regression
(Coded so that lower score results in less favorable terms of insurance)

Company	B_1 (Intercept)	Parameter Estimate for B_2 (% Minority)	Significance Level (P – Value)	R-Squared
A	.096311	-.007964	.0003 / .0001	.1882
B	.236896	-.022663	.0001 / .0001	.4677
C	.234784	-.018088	.0001 / .0001	.5261
D	.156336	-.013346	.0001 / .0001	.2578
E	.204466	-.013667	.0001 / .0001	.1355
F	.242645	-.007525	.0001 / .0001	.1957
G	.234755	-.007851	.0001 / .0001	.1294
H	.149917	-.009123	.0001 / .0001	.1005
I	.020339	-.004620	.4828 / .0001	.0419
J	.247975	-.013219	.0001 / .0001	.2841
K	.140280	-.008133	.0001 / .0001	.1204
L	.235147	-.015372	.0001 / .0001	.3433
Unweighted Average	**.18332**	**-.011798**		

Table 2 provides a parallel measure of the relationship between minority composition and credit scores. Data included the distribution of exposures along five equal numeric intervals. The following table displays the results of a regression of percent minority on the percent of exposures in the three intervals containing the worst scores. For each percentage point increase in minority density, the percent of exposures in the worst credit score intervals ranged from .11 to .44.[11] The average estimate across all companies was .26.

[11] Again, the reader can think of these estimates in terms of comparing ZIP Codes with 0 percent and 100 percent minority population. For example, the parameter estimate for Company A indicates that high minority concentration in a ZIP Code is associated with a 23.4 percentage point increase of the number of exposures in the worst credit score intervals.

Table 2: % of Exposures in Worst Credit Score Interval(s) = B_1+B_2(% Minority) + e

Company	B_1 (Intercept)	B_2 (% Minority)	Significance Level (P – Value)	R-Squared
A	41.390861	.233971	.0001 / .0001	.1349
B	8.867530	.448665	.0001 / .0001	.4810
C	20.459163	.412182	.0001 / .0001	.5062
D	26.689941	.305530	.0001 / .0001	.2433
E	33.732080	.394545	.0001 / .0001	.1176
F	38.8656692	.234620	.0001 / .0001	.1590
G	14.545614	.173579	.0001 / .0001	.1263
H	21.660166	.154712	.0001 / .0001	.0394
I	68.32027	.114139	.0001 / .0001	.0300
J	12.112518	.182560	.0001 / .0001	.2303
K	13.218579	.151518	.0001 / .0001	.1130
L	21.813759	.336678	.0001 / .0001	.2655
Unweighted Average	**26.80635**	**.261892**		

The relationship between per capita income and credit scores is also positive in all cases. Tables 3 and 4 measure the impact on credit scores of each $10,000 increment in per capita income in ZIP Code. Across all companies, a $10,000 increase in per capita income is associated with an increase in average credit scores of .22 standard deviations (Table 3), and a 4.93 percentage point increase in the number of exposures in the worst three credit score intervals (out of five). As with tables 1 and 2, there is considerable variability in the estimates across different companies.

Table 3: Mean Credit Score (Standard Deviation) = $B_1 + B_2$ * Per Capita Income (Per 10k Increments) + e
(Coded so that lower scores results in less favorable terms of insurance)

Company	Intercept	Parameter Estimate for B1 (*Per Capita Income*)	Significance Level (P – Value)	R-Squared
A	-.659632	.270907	.0001 / .0001	.1480
B	-.569438	.242403	.0001 / .0001	.0561
C	-.928092	.382609	.0001 / .0001	.2247
D	-.291691	.138827	.0001 / .0001	.0557
E	-.232981	.136252	.0001 / .0001	.0394
F	-.319388	.199621	.0001 / .0001	.1221
G	-.425798	.228680	.0001 / .0001	.2111
H	-.252602	.124069	.0001 / .0001	.0378
I	-.345479	.113245	.0001 / .0011	.0177
J	-.510392	.247263	.0001 / .0001	.2025
K	-.323383	.158699	.0001 / .0001	.0731
L	-.770462	.345873	.0001 / .0001	.2049
Unweighted Average	**-.469112**	**.2157**		

Table 4: % of Exposures in Worst Credit Score Interval(s) =$B_1 + B_2$ * Per Capita Income (Per 10k Increments) + e

Company	B_1 (Intercept)	B_2 (*Per Capita Income*)	Significance Level (P – Value)	R-Squared
A	58.205403	-5.315069	.0001 / .0001	.0473
B	24.465080	-4.615034	.0001 / .0001	.0533
C	43.569153	-7.125176	.0001 / .0001	.2056
D	38.893367	-4.116010	.0001 / .0001	.0881
E	47.491322	-4.468555	.0001 / .0001	.0441
F	59.143437	-7.562138	.0001 / .0001	.1463
G	27.753627	-4.469898	.0001 / .0001	.1611
H	29.455088	-2.546238	.0001 / .0002	.0217
I	80.165443	-4.681817	.0001 / .0001	.0357
J	22.795670	-3.462954	.0001 / .0011	.1468
K	21.814874	-2.927337	.0001 / .0001	.0616
L	44.491601	-7.874	.0001 / .0001	.1713
Unweighted Average	**41.520339**	**-4.9304**		

For each company (i.e. company/line of business combination), multiple regression was used to determine whether any residual relationship between minority concentration and credit scores remained after controlling for additional socioeconomic variables. Included are numerous variables that provide a broad measure of socio-economic status: per capita income, average age, unemployment rate, percent of renters, percent of population residing in an urban area, percent of adults without post-secondary education, the divorce rate, and the median value of owner occupied homes. Stepwise regression was used to delete variables from the analysis that were not correlated with credit scores with at least a .05 significance level. Variables that were deleted are indicated by the absence of a corresponding parameter estimate.

Somewhat surprisingly, controlling for such factors did little to diminish the correlation between racial /ethnic concentration and average credit score below the level of correlation found in the bivariate models. Controlling for socioeconomic status, minority concentration was significantly correlated with both measures of credit scores for all companies without exception. Indeed, race/ethnicity proved to be among the strongest and most robust single correlate of credit scores, in many instance having a significantly greater impact than education, marital status, income, and housing values. **It was also the only variable for which a consistent correlation was found across all companies (A – L).** Other variables highly correlated to credit scores across many companies were the percent the adult population without college education, percent divorced, average age, and percent urban. Per capita income and the median value of homes were not consistently correlated with credit scores, after controlling for the additional socioeconomic variables.

Why scores should be correlated with minority status, even after controlling for such broach measures of socioeconomic status, is not immediately clear. Such a residual correlation indicates that the variable "minority status" includes information not contained in the socioeconomic "control" variables. Either a relevant variable(s) has been omitted from the model (perhaps additional socioeconomic characteristics), or credit scores capture factors uniquely associated with racial status (such as impediments on access to credit, for example). The results would indicate that further study is necessary.

Table 5: Credit score, race / ethnicity, and socio-economic status

Multivariate Weighted OLS Regression
All scores coded so that a lower score results in less favorable terms of insurance

Company A

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**Est.**	**P-Value**
Intercept	-1.08165870	.0020	81.10301598	.0001
% Minority	-.00602571	.0001	.24208715	.0001
Per Capita Income (10k Increments)				
Average Age	.03922638	.0001	-.97675761	.0003
% Unemployed				
% Rent	.00467218	.0055	-.16692035	.0065
% Urban	-.00243239	.0035		
% Without College Ed	-.01086974	.0001	.1652206	.0009
% Divorced				
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.28624571**		**.17123689**	

Company B

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**Est.**	**P-Value**
Intercept	-.54258067	.0445	13.30431564	.0124
% Minority	-.02145699	.0001	.43192738	.0001
Per Capita Income (10k Increments)				
Average Age	.03538828	.0001	-.42958138	.0001
% Unemployed	-.2379533	.0106	.48889572	.0077
% Rent	.01853674	.0001	-.34449232	.0001
% Urban	-.00354218	.0001	.06114996	.0001
% Without College Ed	-.01239611	.0001	.21138434	.0001
% Divorced	-.02786944	.0003	.59142332	.0003
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.56774300**		**.56021731**	

Company C

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**Est.**	**P-Value**
Intercept	.45641238	.0048	14.25448656	.0182
% Minority	-.01563090	.0001	.39531608	.0001
Per Capita Income (10k Increments)			1.93345161	.0444
Average Age	.02008501	.0001	-.47502897	.0005
% Unemployed				
% Rent	.00803030	.0001	-.21809311	.0001
% Urban	-.00268132	.0001	.05365846	.0002
% Without College Ed	-.01387117	.0001	.32258164	.0001
% Divorced	-.04404118	.0001	.85056141	.0001
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.67065158**		**.59802404**	

Company D

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**Est.**	**P-Value**
Intercept	-.39050190	.0705	33.39785282	.0001
% Minority	-.01304273	.0001	.27985290	.0001
Per Capita Income (10k Increments)				
Average Age	.02859810	.0001	-.47916453	.0001
% Unemployed	-.02673679	.0001	.65611396	.0001
% Rent	.00809207	.0001	-.19735467	.0001
% Urban	-.00120566	.0078	.03690904	.0005
% Without College Ed	-.01005798	.0001	.22315803	.0001
% Divorced	-.01154343	.0460	.32579527	.0118
Median Value, Owner Occupied Homes (10k Increments)	-.01228151	.0084		
R-Squared	**.36885902**		**.37683128**	

Company E

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	.52177336	.0001	18.95408275	.0001
% Minority	-.01170901	.0001	.34730453	.0001
Per Capita Income (10k Increments)				
Average Age				
% Unemployed	-.04011977	.0001	1.15508251	.0007
% Rent			-.15690245	.0315
% Urban				
% Without College Ed	-.00400652	.0004	.12953732	.0004
% Divorced			.78091287	.0036
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.18830144**		**.17753363**	

Company F

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	-.15067768	.4624	38.61297213	.0001
% Minority	-.00740184	.0001	.22781643	.0001
Per Capita Income (10k Increments)				
Average Age	.00899694	.0455		
% Unemployed				
% Rent	.00319185	.0022	-.09109792	.0043
% Urban				
% Without College Ed	-.00471283	.0007	.17478909	.0001
% Divorced				
Median Value, Owner Occupied Homes (10k Increments)	.01354553	.0049	-.56861050	.0004
R-Squared	**.28435611**		**.27976738**	

Company G

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	-1.97713972	.0001	45.19496618	.0001
% Minority	-.01131468	.0001	.23095202	.0001
Per Capita Income (10k Increments)			-3.24019300	.0001
Average Age	.05511056	.0001	-.55194374	.0001
% Unemployed	.04034641	.0001	-.62670129	.0012
% Rent	.00961211	.0001	-.27087221	.0001
% Urban	.00175568	.0202		
% Without College Ed	-.00694914	.0001		
% Divorced	-.03830223	.0001	.84837163	.0001
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.45424970**		**.35721908**	

Company H

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	-1.31393291	.0001	28.30623389	.0001
% Minority	-.00937620	.0001	.15167450	.0001
Per Capita Income (10k Increments)	.09985755	.0001	-6.48418501	.0011
Average Age	.02471241	.0002		
% Unemployed				
% Rent	.00558516	.0049		
% Urban				
% Without College Ed				
% Divorced				
Median Value, Owner Occupied Homes (10k Increments)			.73808454	.0162
R-Squared	**.14546558**		**.06154477**	

Company I

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	-.157612	.6390	75.245498	.0001
% Minority	-.00258036	.0168	.07657484	.0059
Per Capita Income (10k Increments)				
Average Age	.01395931	.0456		
% Unemployed				
% Rent				
% Urban	-.00235209	.0115		
% Without College Ed	-.00693470	.0004		
% Divorced				
Median Value, Owner Occupied Homes (10k Increments)			-.6716167	.0001
R-Squared	**.06621077**		**.05615157**	

Company J

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	1.05804537	.0001	.49764027	.0001
% Minority	-.01098292	.0001	.15120341	.0001
Per Capita Income (10k Increments)				
Average Age				
% Unemployed				
% Rent				
% Urban				
% Without College Ed	-.00834227	.0001	.13548650	.0001
% Divorced	-.04362875	.0001	.54580532	.0068
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	**.44924324**		**.36923936**	

Company K

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	.20562127	.0146	8.0153226	.0047
% Minority	-.00589409	.0001	.13753958	.0001
Per Capita Income (10k Increments)				
Average Age				
% Unemployed				
% Rent			.06166797	.0070
% Urban			.02508670	.0189
% Without College Ed			.12533573	.0001
% Divorced	-.02553375	.0001		
Median Value, Owner Occupied Homes (10k Increments)	.01756473	.0001	-.1878982	.0413
R-Squared	**.19969154**		**.19227795**	

Company L

Variable	Mean Credit Score (Standard Deviation) Est.	P-Value	% in Worst Credit Score Interval(s) Est.	P-Value
Intercept	.58930427	.0535	-3.59560078	.2084
% Minority	-.01538083	.0001	.3142610	.0001
Per Capita Income (10k Increments)				
Average Age	.01260286	.0417		
% Unemployed				
% Rent	.01508428	.0001	-.31634580	.0001
% Urban	-.00170738	.0235	.07104571	.0005
% Without College Ed	-.01569382	.0001	.40441733	.0001
% Divorced	-.03655970	.0004	.78705329	.0054
Median Value, Owner Occupied Homes (10k Increments)				
R-Squared	.52526256		.42966710	

Individual-Level Analysis

Three widely used models were employed to estimate the individual-level differences in credit scores based on patterns observed in the aggregate data: the *neighborhood model*, *Goodman's Regression*, and King's *EI model*. Each model requires different requisite assumptions about the underlying distribution of credit scores across demographic groups that might account for the observed aggregate patterns discussed in the previous section. Goodman's Regression and the neighborhood model make polar opposite assumptions. Goodman's regression assumes that all variation in credit scores between groups is associated with variation **within** each ZIP Code, such that no differences exist between minorities residing in different ZIP Codes with respect to credit scores. The neighborhood model assumes that all variation is attributable to differences **between** ZIP Codes, such that no differences exist between minorities and non-minorities residing in the same ZIP Code. The much newer EI model, published by Gary King in 1997, assumes that average credit scores follow a truncated bivariate normal distribution across ZIP Codes, and are thus permitted to vary both **between and within** ZIP Codes.

It is our opinion that the EI model is the most plausible of the three. However, for the purposes of this study, conclusions are made only to the degree to which all three models produce concordant results (that is, they all either show or fail to show a disproportionate impact). Such concordance is interpreted as strong and credible evidence for the conclusions indicated, particularly given the results of the multivariate models presented above. In addition to the estimates produced by the three models, total bounds are also calculated, indicating the maximum and minimum possible percentage of minorities and non-minorities that fall within the worst credit score intervals.

Ecological inference models are not well suited for "controlling" for additional variables. For this reason, only the bivariate relationships between credit score and income, and credit score and race/ethnicity, are estimated. As argued above, the bivariate relationship is the defining measure of disproportionate impact.

The individual-level relationships between race / ethnicity and credit score proved to be as consistent and robust as the aggregate relationship measured by ZIP Code averages. In all instances, both minority status and income is strongly related to whether an individual's score falls into the worst three credit score interval. The percentage point differences in the EI model estimates are displayed in Table 6. An average of 28.9 percentage points was associated with race/ethnicity, and 29.2 percentage points divided individuals earning above and below the median family income of Missouri.

Table 6: Percentage Point Difference
% of minorities in worst interval - % of non-minorities in worst interval
% of high income in worst intervals - % low income in worst intervals
Estimates Based on EI Model (King, 1998)

Company	Minority Status	Income
A	19.0%	27.7%
B	39.5%	16.8%
C	42.1%	46.1%
D	30.6%	22.5%
E	47.9%	28.5%
F	25.8%	35.6%
G	14.5%	21.0%
H +I Combined	29.1%	32.8%
J	15.0%	26.7%
K	15.3%	26.4%
L	38.5%	37.2%
Unweighted Average	**28.9%**	**29.2%**

The EI estimates are very close to those produced via Goodman's Regression. The Neighborhood Model, however, consistently produced much smaller differences between racial /ethnic groups as well as between income groups. In some instances, the estimated percentage point difference was negligible. Nevertheless, all three models estimated a disproportionate impact in every case. In no case did the models produce discordant results.

Absolute bounds, within which the true (and unknown) values must fall, are also presented in the following tables. In every case, the bounds are far too broad to permit one to make inferences about disproportionate impact. For example, while the EI model estimates that 61.6 percent of minorities have scores within the worst credit score interval(s), the bounds indicate that the true value must[12] lie somewhere between 24.1 percent and 85.3 percent. The bounds for non-minorities are 33.2 percent and 57.5 percent. Different assumptions about the underlying distribution giving rise to the observed aggregate relationship can produce results not consistent with our conclusion about the level of disproportionate impact. For example, one might assume that the aggregate relationship between minority concentration and poorer average credit scores is produced by lower credit scores among **non-minorities** that reside in high minority ZIP Codes. At the extreme, such an assumption would produce a reverse disproportionate impact whereby non-minorities tend to have poorer credit scores. For Company A, for example, an estimate that 24 percent of minorities have credit scores in the worst interval(s), compared to 57.5 percent of non-minorities, is mathematically possible given the bounds. However, we believe that such assumptions are far less plausible than those of the three models presented. Our belief is reinforced by the robustness of the correlation between minority concentration and credit scores, even controlling for a fairly comprehensive set of area socioeconomic characteristics.

[12] Mathematically, the true (and unknown) value **must** lie within the interval.

Nevertheless, the bounds are presented for those that might wish to entertain alternative assumptions.

Table 7
% of Demographic Groups With Credit Scores in Worst Credit Score Interval(s)

Company A

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	61.6 (.0158)	42.5 (.0063)	19.1%
Goodman	61.10 (.0346)	42.8% (.0157)	17.6%
Neighborhood	52.6%	45.0%	7.6%
Bounds	24.1% to 85.3%	33.2% to 57.5%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	65.4% (.0339)	38.7% (.0177)	26.7%
Goodman	64.4 (.0492)	38.7% (.0267)	25.7%
Neighborhood	47.9%	45.4%	2.5%
Bounds	5.3% to 90.1%	32.0% to 76.7%	

N=143
Population: 3,353,615

Company B

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	49.9% (.0188)	10.4 (.0033)	39.5%
Goodman	53.0% (.0211)	10.0 (.0060)	43.0%
Neighborhood	31.0%	15.8%	15.2%
Bounds	7.6% to 74.2%	6.0% to 17.9%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	27.6% (.0200)	10.8% (.0099)	16.8%
Goodman	27.9% (.0291)	9.7% (.0175)	18.27%
Neighborhood	20.3%	17.1%	3.2%
Bounds	0.1% to 47.4%	0.1% to 24.1%	

N=265
Pop=4,319,018

Company C

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	62.6% (.0153)	20.5% (.0042)	42.1%
Goodman	60.9% (.0244)	21.0 (.0100)	39.9%
Neighborhood	41.1%	25.5%	15.6%
Bounds	18.0% to 82.6%	15.0% to 32.7%	

By Income

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	61.2% (.0231)	15.1% (.0105)	46.1%
Goodman	58.9% (.0402)	15.2% (.0215)	43.7%
Neighborhood	31.9%	26.9%	5.0%
Bounds	4.0% to 81.3%	6.0% to 41.2%	

N=176
Population: 3,748,671

Company D

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	57.3% (.0149)	26.7% (.0021)	30.6%
Goodman	58.3% (.0229)	27.5% (.0051)	30.8%
Neighborhood	41.0%	30.5%	10.5%
Bounds	15.1% to 83.4%	21.7% to 35.8%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	45.6% (.0187)	23.1% (.0088)	22.5%
Goodman	44.8% (.0197)	21.1% (.0141)	23.7%
Neighborhood	33.8%	31.1%	2.7%
Bounds	3.0% to79.2%	7.5% to 47.7%	

N=500
Population: 5,108,469

Company E

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	81.1% (.0279)	33.2% (.0044)	47.9%
Goodman	82.0% (.0439)	32.4% (.0125)	49.6%
Neighborhood	47.8%	38.5%	9.3%
Bounds	10.8% to 98.8%	30.4% to 44.3%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	60.1% (.0320)	31.6% (.0127)	28.5%
Goodman	60.1% (.0427)	28.7% (.0224)	31.4%
Neighborhood	41.3%	38.4%	2.9%
Bounds	2.5% to 93.7%	18.2% to 54.5%	

N=131
Population: 3,067,775

Company F

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	62.8% (.0103)	37.0% (.0031)	25.8%
Goodman	62.5% (.0234)	37.6% (.0089)	24.9%
Neighborhood	50.5%	40.7%	9.8%
Bounds	21.9% to 86.8%	31.6% to 47.9%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	66.6 (.0177)	31.1% (.0088)	35.5%
Goodman	66.8 (.0298)	29.6% (.0169)	37.2%
Neighborhood	45.2%	41.3%	3.9%
Bounds	1.7% to 66.7%	0.8% to 31.0%	

N=202
Population: 4,034,991

Company G

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	29.6% (.0165)	15.1% (.0033)	14.5%
Goodman	31.2% (.0216)	17.5% (.0070)	13.7%
Neighborhood	24.2%	18.4%	5.8%
Bounds	6.7% to 62.0%	9.6% to 22.5%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	33.1 (.0248)	12.1% (.0086)	21.0%
Goodman	32.8 (.0254)	13.2% (.0136)	19.6%
Neighborhood	20.9%	18.7%	2.2%
Bounds	0.0% to 57.8%	1.6% to 28.9%	

N=254
Population=4,318,544

Company H & I Combined

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	69.4% (.0205)	40.2% (.0049)	29.2%
Goodman	65.4% (.0335)	40.5% (.0117)	24.9%
Neighborhood	51.9%	44.2%	7.7%
Bounds	20.4% to 89.6%	35.5% to 51.6%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	69.2% (.0320)	36.4% (.0130)	32.8%
Goodman	70.7% (.0469)	34.2% (.0220)	36.5%
Neighborhood	47.5%	44.7%	2.8%
Bounds	4.8% to 97.3%	23.6% to 63.3%	

N=126
Population= 3,242,541

Company J

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	27.5% (.0180)	12.5% (.0035)	15.0%
Goodman	30.7 (.0270)	7.36 (.0157)	23.3%
Neighborhood	20.9	14.1	6.8%
Bounds	6.7% to 54.6%	6.0% to 17.9%	

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	33.9% (.0199)	7.2% (.0081)	26.7%
Goodman	30.68 (.0270)	7.4 (.0157)	23.3%
Neighborhood	17.8	14.5	3.3%
Bounds	0.0% to 49.6%	0.5% to 22.4%	

N=146
Population: 2,345,518

Company K
By % Minority

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	27.7% (.0169)	12.4% (.0033)	15.3%
Goodman	28.8% (.0245)	13.0% (.0082)	15.8%
Neighborhood	20.0%	15.0%	5%
Bounds	5.0% to 57.3%	6.8% to 18.3%	

By Income

Method	Individuals Earning Less than Median Income	Individuals Earning More Than Median Income	Percentage Point Difference
EI	33.7% (.0199)	7.3% (.0080)	26.4%
Goodman	30.7% (.0270)	7.4% (.0157)	23.3%
Neighborhood	17.0	15.4	1.6%
Bounds	0.0% to 46.9%	4.8% to 23.8%	

N=316
Population: 4,684,292

Company L

Method	Minorities	Non-Minorities	Percentage Point Difference
EI	63.4% (.0123)	24.9% (.0032)	38.5%
Goodman	62.9% (.0237)	25.0% (.0087)	37.9%
Neighborhood	44.2%	29.4%	14.8%
Bounds	20.6% to 85.6%	17.2% to 42.0%	

Method	Below Median Income	Above Median Income	Percentage Point Difference
EI	64.6% (.0211)	27.4% (.0204)	37.2%
Goodman	60.5% (.0311)	25.6% (.0178)	34.9%
Neighborhood	40.9%	36.8%	4.1%
Bounds	5.4% to 89.6%	13.4% to 54.6%	

N=209
Pop=3,951,569

Conclusion

Based on the aggregate-level analysis, it can confidently be stated that individuals that reside in areas with large minority concentrations tend to have significantly worse credit scores than those that reside elsewhere. The aggregate regression models were robust, and in every case without exception indicated a substantial correlation between minority concentration and credit score, even controlling for a wide variety of other socioeconomic characteristics.

This analysis also indicated substantial differences in the level of disproportionate impact across companies. While all scoring products examined negatively impacted individuals residing in high minority areas, some did so to a much greater extent than others. This suggests that there may be ways to design credit scores with far less potential to restrict the availability of affordable insurance products in high minority areas.

The evidence regarding the individual-level relationships presented herein should be interpreted in light of well-known caveats associated with making individual-level inferences from aggregate data. However, *interpreted in totality*, the evidence appears to be credible, substantial, and compelling that credit scores have a significant disproportionate impact on minorities and on the poor. Additional study is necessary to determine how the practice of credit scoring impacts premium levels and declinations among minorities.

Methodological Appendix

This study is based on credit score and demographic data aggregated at the ZIP Code level. As a result, different levels of analysis were presented, each of which involves categorically distinct interpretations. Differences between individual-level and aggregate-level analyses can be illustrated by the types of questions each method can answer:

Individual-Level

"*Do members of minority groups tend to have lower (or higher) credit scores on average than do members of non-minority groups?*"

"*If such differences exist, is there a correlation between the minority status of individuals and credit scores, after controlling for individual characteristics such as income, employment status, and marital status?*"

Aggregate Analysis

"*Do individuals who reside in areas with high minority concentrations tend to have lower (or higher) credit scores on average than do individuals residing in areas with few minorities?*"

"*If such differences exist, is there a correlation between the minority concentration of an area and credit score, after controlling for the median income, unemployment rate, and divorce rates (etc) of such areas?*"

Note that the existence of an ecological or aggregate—level correlation does not necessarily imply that minorities *per se* have higher or lower credit scores, since the ecological inference problem prohibits ***direct*** individual-level inferences. Nothing in the statistical methods rules out the possibility that non-minorities residing in high minority areas lower the overall average credit score in an area. *However,* as argued above, the ecological or aggregate correlation is meaningful in its own terms where public policy concerns are directed precisely at business practices with negative consequences for residents of areas with high minority concentrations, including non-minority residents of such areas.

Ecological Fallacy

While inferences about aggregate relationships based on aggregate data are non-problematic, considerable controversy surrounds methods that make inferences about individuals based on aggregate data. William S. Robinson's (1950) well-known article is generally considered a seminal statement of potential perils associated with ecological inferences. The problem can be stated quite simply: it is a mistake to assume that relationships observed in aggregate data **necessarily** obtain for individual-level relationships. Robinson's example illustrates the problem. Data was obtained for each of the 48 contiguous states for aggregate (English language) literacy rates and the percent of each state's population that was of foreign birth. The correlation between these two variables, aggregated at the state level, was .53 (with 0 representing no correlation, and 1 representing a perfect correlation), suggesting the counterintuitive result that non-native speakers were more English literate than native speakers. However, the individual-level correlation between foreign--birth and literacy was -.11. The aggregate positive correlation was obtained simply because individuals of foreign--birth were more likely to reside in more affluent coastal states where the *native-born* had higher literacy rates than the national average.

However, there are often questions in the social sciences that cannot be addressed via survey methods, and researchers across many fields often rely on aggregate data. In many instances, survey data does not exist (as with historical voting patterns), is prohibitively costly to collect, or is known to be unreliable (as is the case with some elections). For this reason, methodologists have developed statistical techniques for making individual inferences based on aggregate data. Such methods are valid, so long as certain assumptions are met. Various methods have been recognized as valid in federal courts in instances when survey data is unavailable.

Rather than relying solely on a single model, a more methodological conservative approach is adopted here. The following three strategies were pursued:

1. Perform an aggregate analysis without attempting to make inferences about individuals. Assess the level of correlation between protected classes and credit scores as defined by the demographic characteristics of an area. Both univariate and multivariate analysis are performed.

2. Produce estimates of individual-level correlations from the aggregate data, using a variety of existing methods. Each method requires certain statistical assumptions. If all methods produce the concordant results (i.e. all either show or fail to show a correlation between protected classes and credit score), the results can reasonably be considered reliable and strong, if not irrefutable, evidence of whether a disparate impact exists based on individual-level characteristics, irrespective of place of residence.

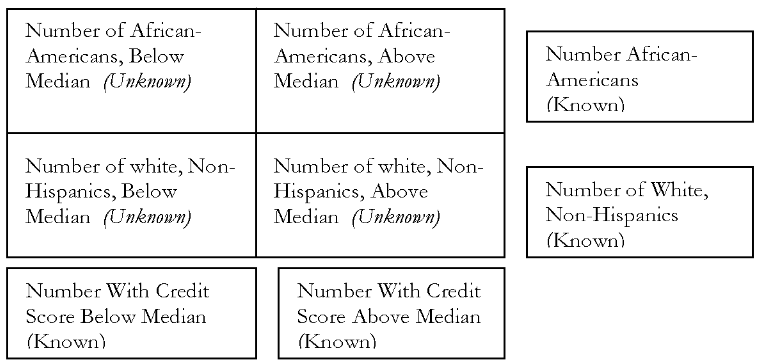
3. If the three methods produce contradictory results, then the evidence should be considered inconclusive. However, even in this event, reasonable **tentative** conclusions can be made as to which set of assumptions are more likely to have been met.

Methods of Ecological Inference

Ecological inference methods provide estimates of unknown quantities of interest based on patterns observed in aggregate data. Each method can produce valid estimates, ***so long as necessary assumptions are satisfied.***

The quantities to be estimated are illustrated in the following diagram, using ethnicity and credit score as an example. The ZIP Code aggregates (called *marginals* and represented by the sum of the cells across column and rows*)* are known from aggregate data. For example, the number of African-Americans residing in a ZIP Code can be obtained from census data, while numbers above or below an average or median score could be obtained from insurers. The unknown quantities of interest are represented by the individual cells: the number of African-Americans above and below the mean credit score, and the corresponding figures for white, non-Hispanics. Since insurers do not possess all of the required demographic information, the cell--quantities are unknown and have to be estimated. Once estimated, they can then be summed over all areas (over all ZIP Codes or census tracts in a state) to provide estimates for each demographic group within the state population.

Illustration of Ecological Inference Problem



Unfortunately, the range of possible cell values is in many instances so wide that little useful information about the relationship between minority status and credit score could be gleaned from the marginals. The hypothetical distributions below illustrate the point. Assume that in a given ZIP Code, we know the following:

1. From census data, we know that of the 2,400 residents, 800 are non-minorities, and 1,600 are minorities.

2. From credit score data, we know that 1,200 individuals have bad credit scores, and 1,200 have good credit scores (however defined).

Therefore, we know the following (marginal) values:

Known ZIP Code Totals

	Credit Score		
Minority Population	**Number in Worst Credit Score Group**	**Number in Best Credit Score Group**	**Totals**
Non-Minorities	*Unknown*	*Unknown*	**800**
Minorities	*Unknown*	*Unknown*	**1,600**
Total	**1,200**	**1,200**	**2,400**

From the known data, what can be inferred about the relationship between minority status and credit score? The examples below indicate that in this instance, no valid inferences can be made. All possible relationships between minority status and credit score would be consistent with the known marginal values. Example 1 illustrates the zero correlation case, where an equal percent of minority and non-minorities have poor credit scores. Example 2 shows a negative relationship between credit score and minority status, and Example 3 illustrates a positive relationship. All such relationships are consistent with the given known ZIP Code totals.

Hypothetical Distributions Illustrate How Different Relationships Are Consistent with the Same Marginal Values

Example 1: No Relationship between Minority Status and Credit Score

	Credit Score		
Minority Population	**Number in Worst Credit Score Group**	**Number in Best Credit Score Group**	**Totals**
Non-Minorities	400	400	**800**
Minorities	800	800	**1,600**
Total	**1,200**	**1,200**	**2,400**

Example 2: Non-Minorities Tend to Have Lower Scores

	Credit Score		
Minority Population	**Number in Worst Credit Score Group**	**Number in Best Credit Score Group**	**Totals**
Non-Minorities	700	100	**800**
Minorities	500	1,100	**1,600**
Total	**1,200**	**1,200**	**2,400**

Example 3: Minorities Tend to Have Lower Scores

	Credit Score		
Minority Population	**Number in Worst Credit Score Group**	**Number in Best Credit Score Group**	**Totals**
Non-Minorities	100	700	800
Minorities	1,100	500	1,600
Total	**1,200**	**1,200**	**2,400**

However, incorporating data from all ZIP Codes can significantly narrow the range of reasonable estimates for cell values. Nevertheless, *all methods of producing cell estimates entail simplifying* ***assumptions,*** *though such assumptions may be subject to at least limited verification.* The approach adopted here was to produce estimates for different sets of assumptions under differing conditions. While the term *assumption* may sound immediately suspect to some readers, it should be noted that virtually **all** statistical techniques require specific assumptions. Preferably, such assumptions can be verified or tested. Where they cannot, then the analyst should produce estimates under all plausible assumptions. For example, this would be akin to an economic forecast producing estimates of economic growth under differing possible interest rate levels. **If the same result is obtained under the differing sets of assumptions, then such results should be interpreted as strong (if not irrefutable) evidence that the indicated relationship is the correct relationship.**

Variations of three methods have been widely employed to provide estimates of the missing cell quantities: the *neighborhood model*, *Goodman's Regression*, and more recently, Gary King's "*EI Model*." The methods differ primarily in terms of the assumptions about how specific group characteristics might vary across ZIP Codes.

Using the percent of the population in a ZIP Code with credit scores below the state-wide median and minority status as an example:

Goodman's Regression assumes that there is no variation **across** ZIP codes in the percent of minorities and non-minorities with low credit scores. The model constrains estimates to equalize across ZIP Codes. In other words, the model assumes that there are no contextual effects, as would be the case if the percent of minorities with low credit scores were correlated with other ZIP Code characteristics.[13]

The Neighborhood Model makes the diametrically opposite assumption that there is no variation **within** each ZIP Code between minorities and non-minorities with respect to low credit scores. The model assumes that any differences of credit scores based on ethnicity are entirely a function of geographic effects, whereby differences in credit scores result from socio-economic differences across ZIP Codes. Hypothetical examples of distributions that would conform to each set of assumptions is displayed in the following table.

[13] In many applications, the minority population characteristic of interest is correlated with the concentration of minorities. One example is a well-known observation that the minority vote tends to be more cohesive in areas with high concentrations of minorities.

ZIP Codes of Equal Populations	% Minority	Hypothetical Distribution under Goodman Assumptions (% Minority with low credit scores / % Non-Minority With Low Credit Scores)	Hypothetical Distribution under the assumptions of the "Neighborhood" Model
ZIP Code A	25%	50% / 20%	20% / 20%
ZIP Code B	58%	50% / 20%	50% / 50%
ZIP Code C	92%	50% / 20%	80% / 80%
Total	58%	50% / 20%	62% / 34%

The requisite assumptions for each model would likely be strictly satisfied only in rare instances. However, estimates produced by the models may be useful if *both produced similar results, indicating that results are relatively robust under wildly differing assumptions.*

Gary King's "EI" model offers a more recent alternative to both Goodman's Regression and the Neighborhood Model. King's model combines elements of the Goodman and neighborhood approaches, so that the percent of minorities and non-minorities with low credit scores is allowed to vary **both within and across** ZIP Codes, though according to probabilities associated with a truncated bivariate-normal distribution, and within additional known constraints.
According to King (1997), the EI method has the following advantages over other ecological inference methods:

1. Necessary assumptions can be tested by observable features of the data. An analyst can be alerted to possible departures from assumptions via various diagnostic tests.

2. The model is robust to departures from assumptions.

2. Remedial measures can be taken in those instances when assumptions are violated.

3. The model is robust against aggregation bias[14]

4. The model takes advantage of all information in the data, considerably narrowing the bounds of allowable estimates. Estimates must fall within known constraints.

5. Estimates can be assigned levels of uncertainty, such as confidence intervals or p-values (significance levels), and are thus comparable to any inferential statistic (such as correlation or regression coefficients, etc).

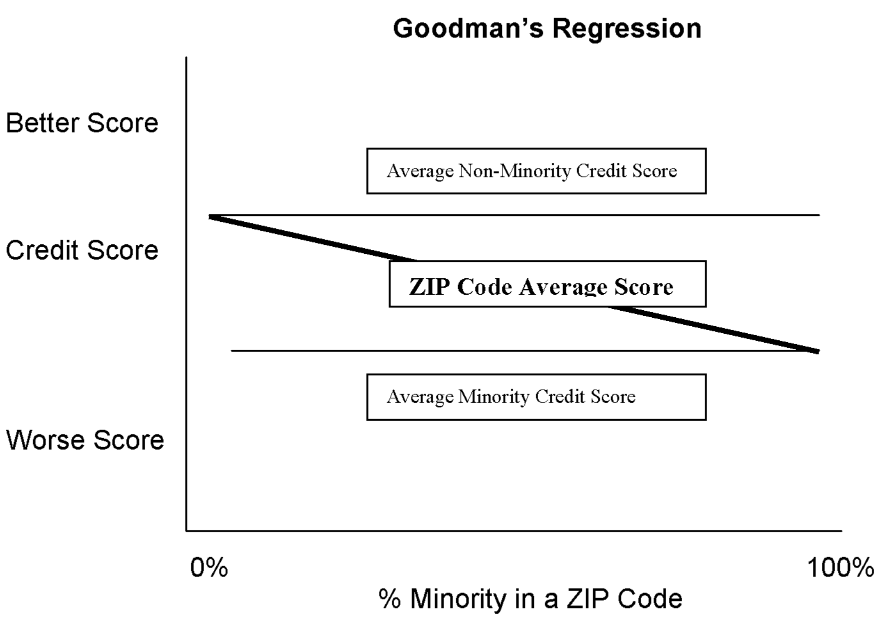
The EI model has generated much comment in the scholarly literature since its publication in 1997, not all of it necessarily favorable. In addition, pieces that have employed

[14] Aggregation bias occurs when differing results are obtained for different levels of aggregation. For example, using ZIP Codes versus census tracts.

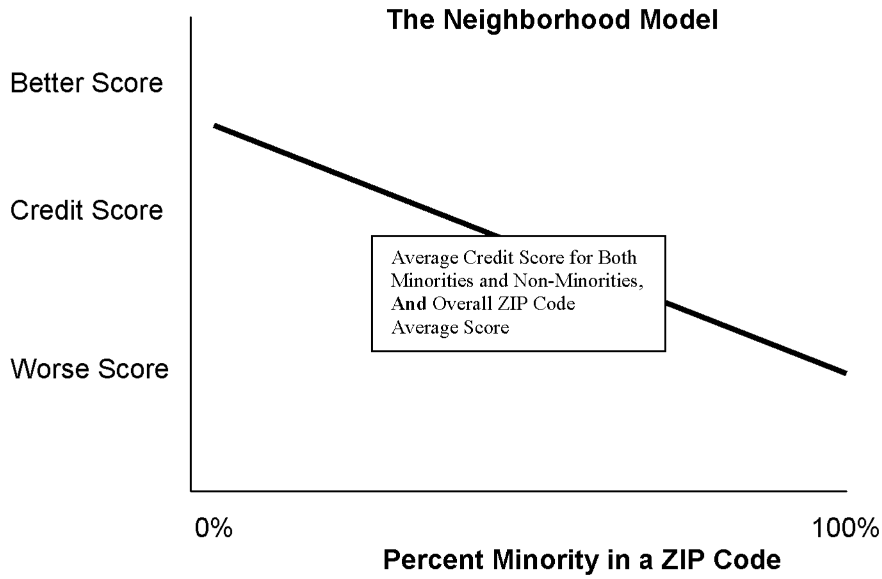
the method have begun appearing in peer reviewed scholarly publications, indicating that the method is enjoying broadening acceptance. See bibliography for citations.

More information about King's model can be found on his internet site at http://Gking.Harvard.Edu Gary King has also made software freely available that implements the EI model.

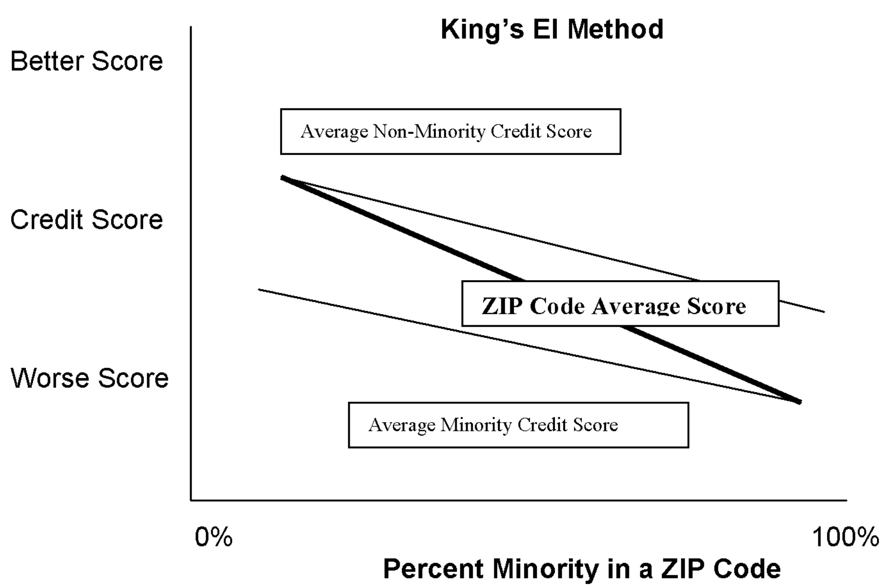
The assumptions of the three methods of ecological inference are displayed graphically below.



Goodman's Regression assumes no variation in credit scores across ZIP Codes; all variation between minorities and non-minorities is produced by within-ZIP Code differences. The bold line represents the overall ZIP Code average score, which approaches the average score for minorities as minority concentration approaches 100%. The bold line representing the overall ZIP Code average is a pattern that is observed in the aggregate data. The two lines representing minority and non-minority average scores are ***unobserved*** *and* ***unknown.*** *Assumptions about the relationship between the unobserved underlying trends, and how they might account for the observed overall ZIP Code average, distinguish the three models.*

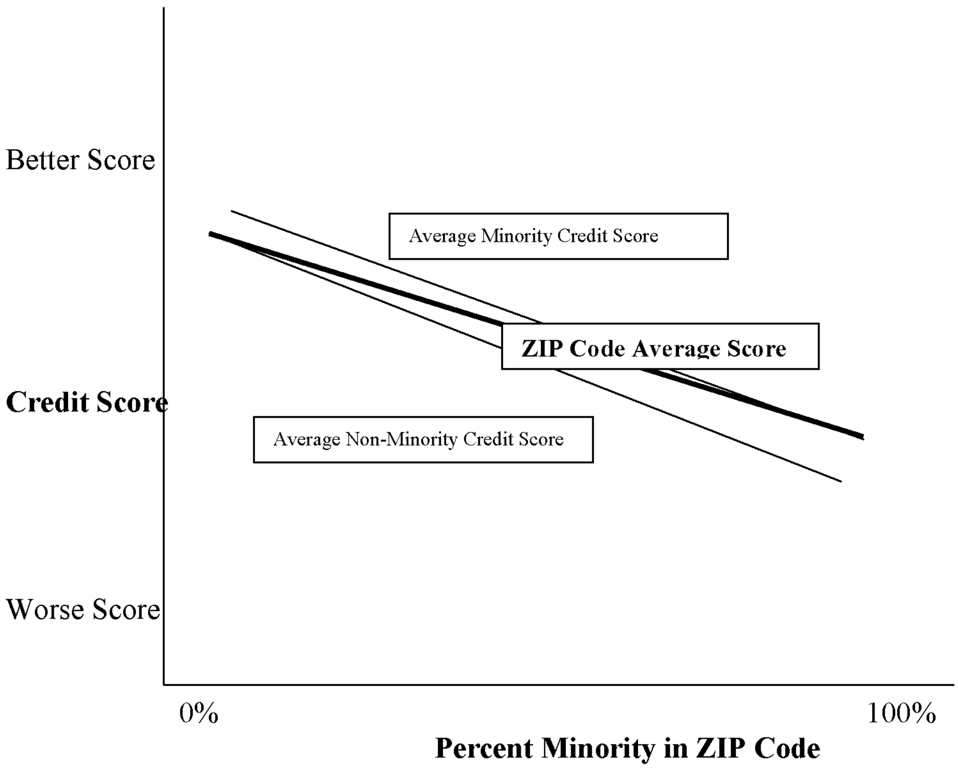


*The Neighborhood Model assumes no variation in credit scores **within** ZIP Codes; all variation between minorities and non-minorities is produced by between ZIP Code differences*



The EI method permits variation both within and between ZIP Codes, subject to a truncated bivariate normal distribution, as well as additional known constraints.

Alternative Assumptions



The three models do not exhaust the range of ***possible*** *assumptions, though we believe they exhaust all* ***plausible*** *assumptions. Above is a hypothetical distribution consistent with an observed correlation between minority concentration and average score, but in which* ***non-minorities*** *have lower average scores than minorities. King (1997), however, does present voluminous evidence, based both on statistical simulations and tests where the true values are known, that support the credibility and reliability of EI estimates. While others have demonstrated that the EI method can fail, such results appear to be based on datasets contrived to seriously violate the assumptions of EI, and are not likely to represent distributions encountered in practical applications (see Freedman, et. al, 1998, and King, 1999).*

Nevertheless, readers should keep such alternatives in mind when interpreting results. Ultimately, interpretation should be based on which set of assumptions readers believe are reasonable.

Sources

AIA. 1998. Statement of the American Insurance Association On the Impact on Insurance Availability, Affordability and Accessibility Of Insurance Underwriting Use of Credit History and Credit Scoring. Washing, DC: AIA.

American Academy of Actuaries. 2002. The Use of Credit History for Personal Lines of Insurance: Report to the National Association of Insurance Commissioners. Washington, DC: American Academy of Actuaries.

Commonwealth of Virginia. 1999. Report of the State Corporation Commission's Bureau of Insurance on the Use of Credit Reports in Underwriting to the State Commerce and Labor Committee of the General Assembly of Virginia.

Conning and Co. 2001. Insurance Scoring in Personal Automobile Insurance: Breaking the Silence. Conning Insurance Research and Publications: Hartford, CT.

Freddie Mac. 1999. Consumer Credit Survey. Freddie Mac Special Reports and Publications.

Kabler, Brent. 2004. Affordability and Availability of Personal Lines Insurance in Underserved Markets. Jefferson City, MO: Missouri Department of Insurance. Forthcoming.

Kellison, Bruce, Patrick Brockett, Seon-Hi Shin, and Shihong Li. 2003. A Statistical Analysis of the Relationship Between Credit History and Insurance Losses. Bureau of Business Research, University of Texas at Austin.

King, Gary. 1997. **A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.** Princeton, NJ: Princeton University Press.

Miller, Michael J., and Richard A. Smith. 2003. The Relationship Between Credit- Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity. Bloomington, IL.: Epic Actuaries, LLC.

Monaghan, James E. 2000. The Impact of Personal Credit History on Loss Performance in Personal Lines. Casualty Actuarial Society, 2000 Winter Forum.

Pavelchek, Dave, and Bruce Brown. 2003. A Report to the Legislature: Effect of Credit Scoring on Auto Insurance Underwriting and Pricing. State of Washington, Office of the Insurance Commissioner.

Robinson, William S. 1950. Ecological Correlation and the Behavior of Individuals. **merican Sociological Review.** 15: 351-357.

Tillinghast-Towers Perrin. 1996. Insurance Bureau Scores vs Loss Ratio Relativities. Prepared on behalf of Fair, Isaac. Tillinghast-Towers Perrin: St. Louis, Mo.

Published Works Employing King's EI Method

Burden, Barry C, and David C. Kimball. 1998. A new approach to the study of ticket splitting. ***American Political Science Review.*** 92 (September): 533-44.

Gay, Claudine. 2001. The effect of black congressional representation on political participation. ***American Political Science Review.*** 95: 589.

Gimpel, James G., and Jason E. Schuknecht. 2002. Political and demographic foundations for sectionalism in state politics: The Connecticut case. ***American Politics Research.*** 30(March): 193-213.

Liu, Baodong "Paul." 2001. The positive effect of black density on white crossover voting: Reconsidering social interaction theory. ***Social Science Quarterly.*** 82(September): 600-613.

Lublin, David, and D. Stephen Voss. 2002. Context and Francophone support for sovereignty: An ecological analysis. ***Canadian Journal of Political Science.*** 35 (March): 75-101.

Voss, Stephen D., and Penny Miller. 2001. Following a false trail: The hunt for white backlash in Kentuchy's 1996 desegregation vote. ***State Politics and Policy Quarterly.*** 1(March): 63-82.

Methodological Issues Associated With Ecological Inference

Adolph, Christopher, and Gary King, with Michael C. Herron and Kenneth W. Shotts. 2002. A Consensus on Second Stage analyses in Ecological inference models. Working paper, September 22, 2002.

Anselin, Luc, and Wendy K. Tam Cho. 2002. Spatial effects and ecological inference. ***Political Analysis.*** 10(3): 276-297.

Cho, Wendy K. Tam. 1998. Iff the assumption fits: A comment on the King ecological inference solution. ***Political Analysis***. 7: 143-163.

Freedman, D.A., S.P. Klein, M. Ostland, and M.r. Robers. 1998. Review of "A Solution to the Ecological Inference Problem." ***Journal of the American Statistical Association.***

Herron, Michael C., and Kenneth W. Shotts. *Using ecological inference point estimates as dependent variables in second state linear regressions.* Paper presented at the 2000 Political Methodology Summer Meetings, and available on the internet.

King, Gary. 2002. Isolating spatial autocorrelation, aggregation bias, and distributional violations in ecological inference. ***Political Analysis.***

King, Gary. 1999. The future of ecological inference research: A reply to Freedman et. Al. ***Journal of the American Statistical Association.***