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

State of Missouri
Department of Insurance

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.

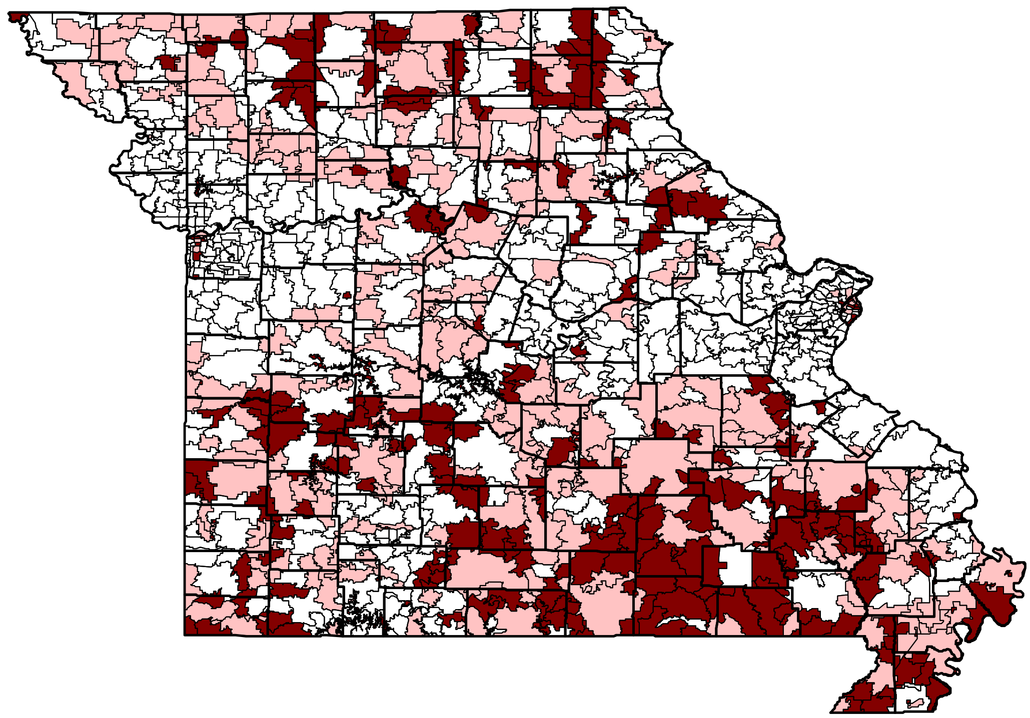
wealthiest 5 percent of communities) was 12.8 percentiles. Policyholders in low-income communities were overrepresented in the worst credit score group by 7.4 percentage points compared to higher income neighborhoods. Estimates of credit scores at additional levels of *per capita* income are found on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

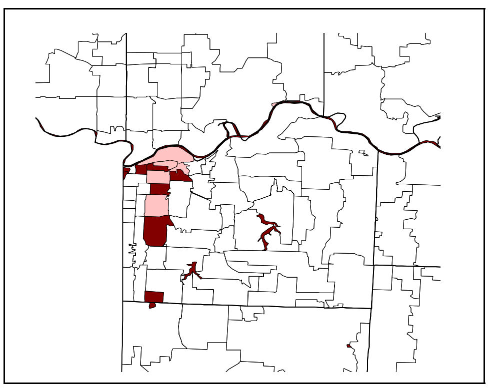
4. Minority and low-income *individuals* were significantly more likely to have worse credit scores than wealthier individuals and non-minorities. The average gap between minorities and non-minorities with poor scores was 28.9 percentage points. The gap between individuals whose family income was below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

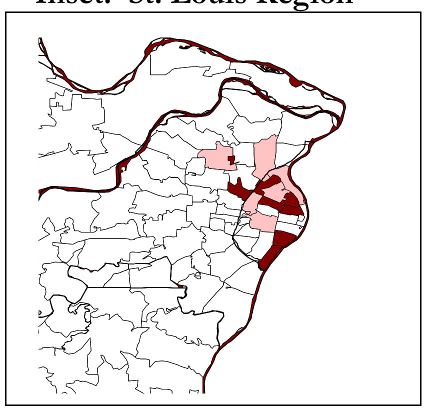
Lower Income Areas of Missouri Most Affected by Credit Scoring



Inset: Kansas City Region



Inset: St. Louis Region

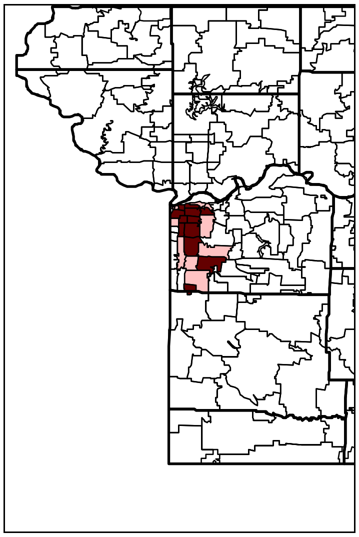


Bottom Quartile ($6,153 - $13,335) = 253 Zip Codes (out of 1,015), with 562,453 persons, or 10% of 5.6 million Missourians

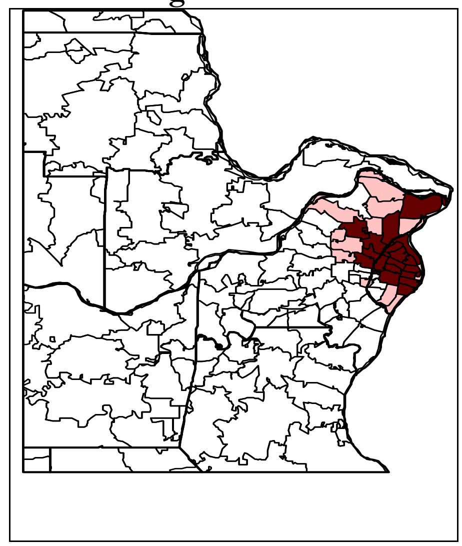
Second Quartile ($13,336-$15,326) = 254 ZIP Codes with 839,281 persons, or 15% of 5.6 million Missourians

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

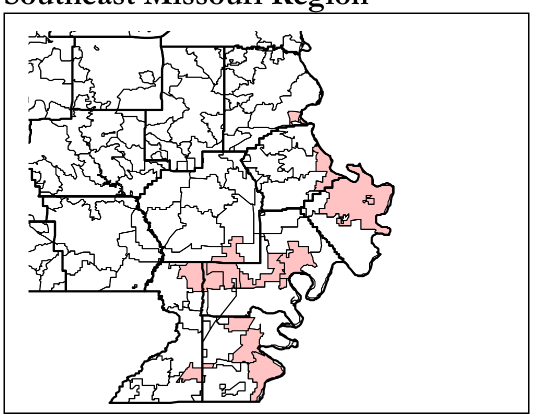
Kansas City Region

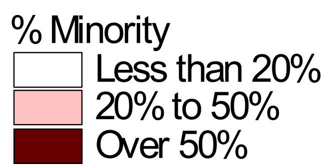


St. Louis Region



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 (7^{th} 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.	Mean Credit Score (Standard Deviation) P-Value	% in Worst Credit Score Interval(s) Est.	% in Worst Credit Score Interval(s) 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.	Mean Credit Score (Standard Deviation) P-Value	% in Worst Credit Score Interval(s) Est.	% in Worst Credit Score Interval(s) 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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

	Mean Credit Score (Standard Deviation)		% in Worst Credit Score Interval(s)	
Variable	**Est.**	**P-Value**	**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

Number of African-Americans, Below Median *(Unknown)*	Number of African-Americans, Above Median *(Unknown)*	Number African-Americans (Known)
Number of white, Non-Hispanics, Below Median *(Unknown)*	Number of white, Non-Hispanics, Above Median *(Unknown)*	Number of White, Non-Hispanics (Known)
Number With Credit Score Below Median (Known)	Number With Credit Score Above Median (Known)	

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).

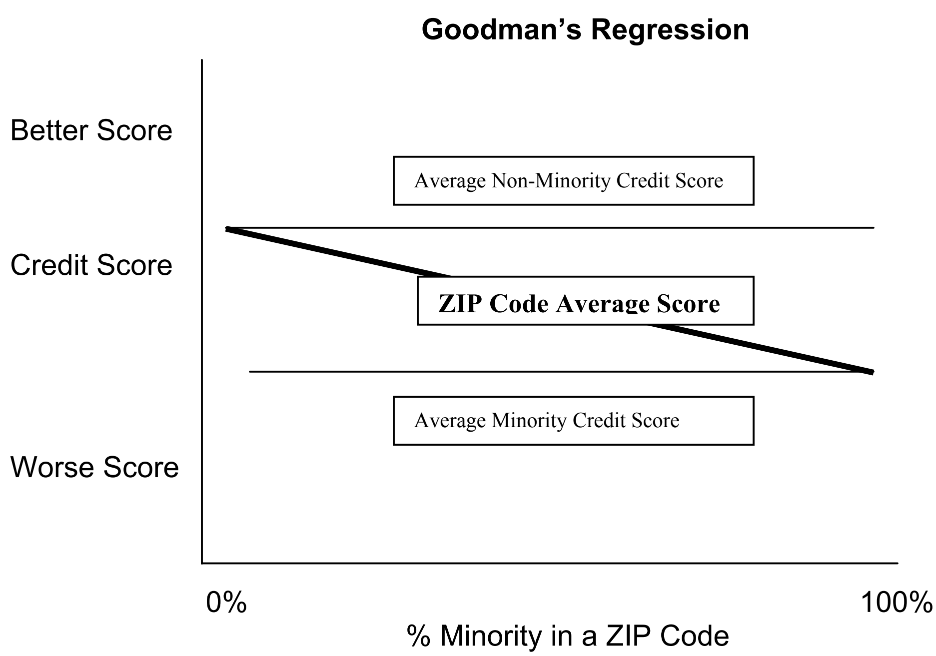
The EI model has generated much comment in the scholarly literature since its publication in 1997, not all of it necessarily favorable. In addition, pieces that have employed

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

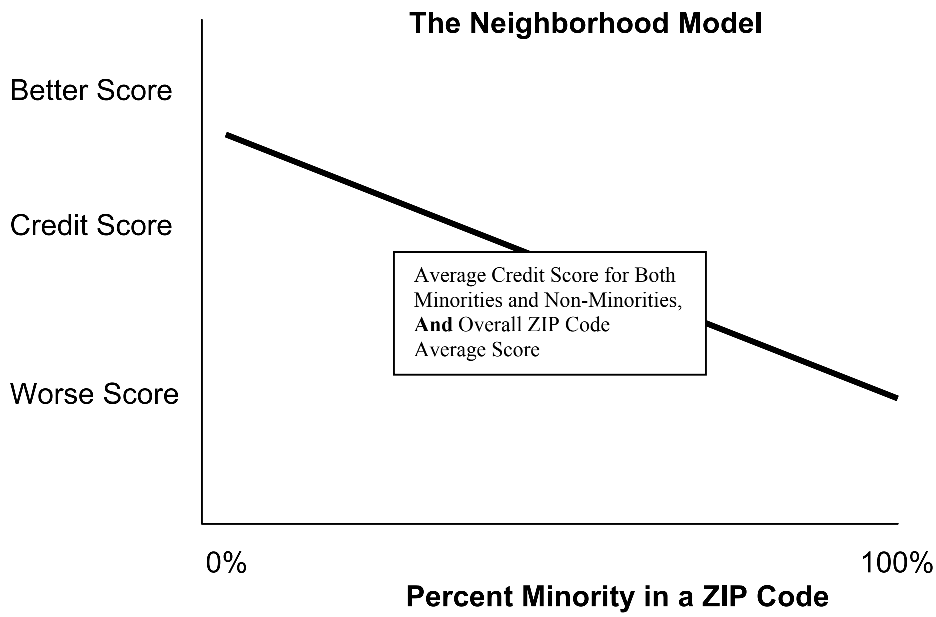
the method have begun appearing in peer reviewed scholarly publications, indicating that the method is enjoying broadening acceptance. See bibliography for citations.

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.

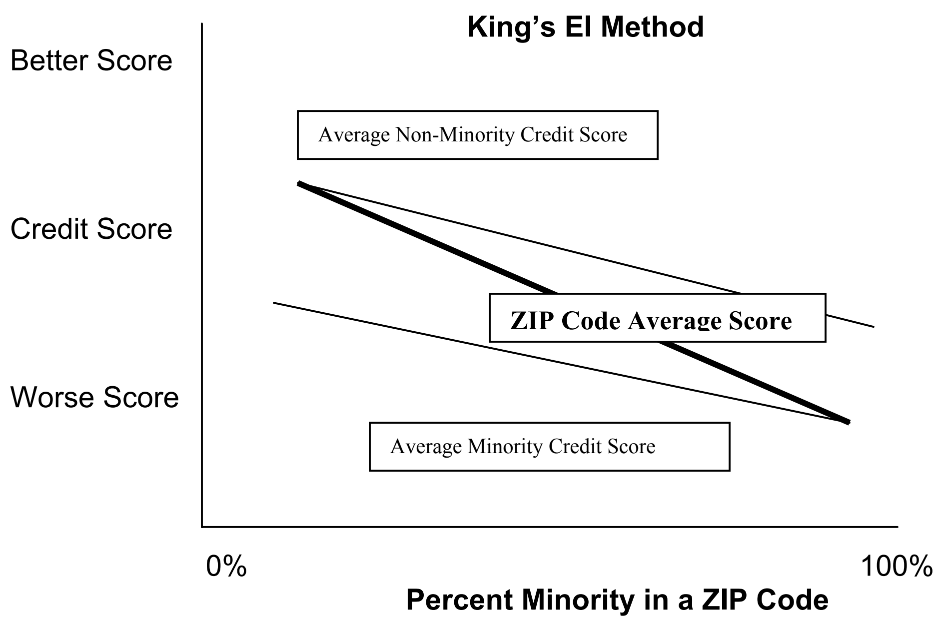
The assumptions of the three methods of ecological inference are displayed graphically below.



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.*

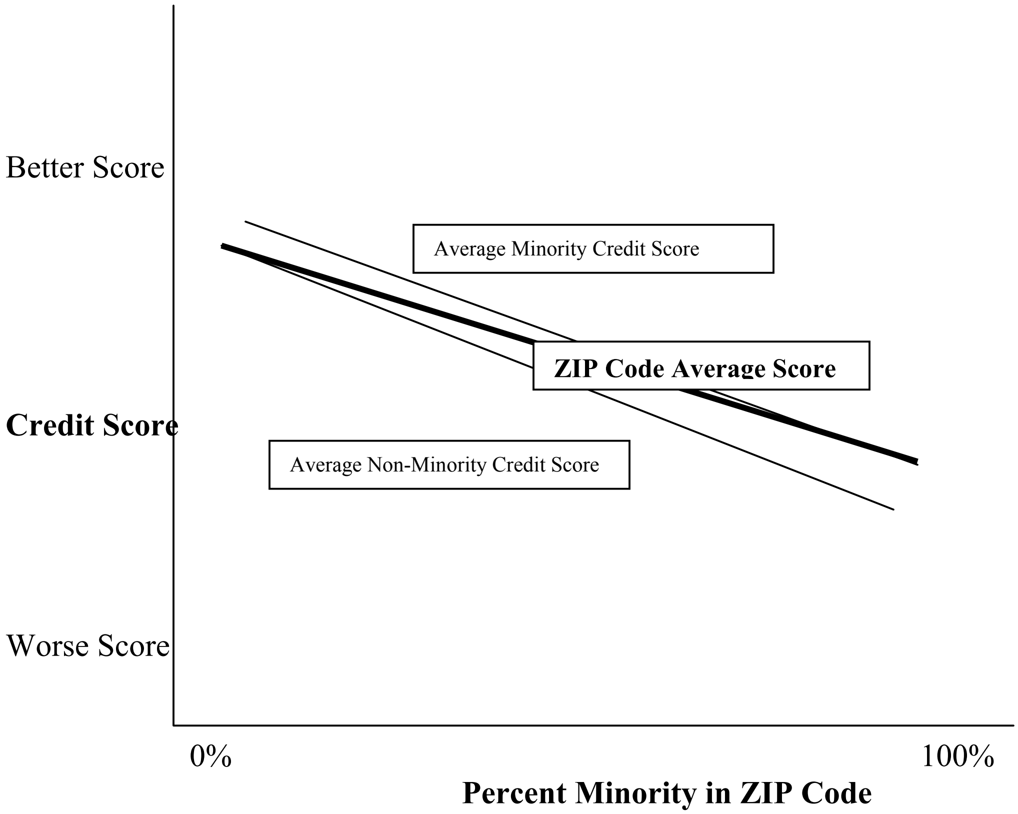


The Neighborhood Model assumes no variation in credit scores ***within*** *ZIP Codes; all variation between minorities and non-minorities is produced by between ZIP Code differences*



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.***