Mortgage Pricing and Race: Evidence from the Northeast*

Kevin A. Clarke[†]
Lawrence Rothenberg[§]

May 9, 2016

*We thank Michael Peress and Jake Bowers for comments and suggestions. Bradley Smith provided research assistance. Errors are our own.

[†]Corresponding author. Associate Professor, Department of Political Science, University of Rochester, Rochester, NY 14627-0146. Email: kevin.clarke@rochester.edu.

[§]Corrigan-Minehan Professor of Political Science, Department of Political Science, University of Rochester, Rochester, NY 14627-0146. Email: lawrence.rothenberg@rochester.edu.

Abstract

The putative existence of race-based discrimination in mortgage pricing is both a scholarly and societal concern. Efforts to assess discrimination empirically, however, are typically plagued by omitted variables, which leave any evidence of discrimination open to interpretation. We take a two-pronged approach to mitigating the problem. First, we analyze a data set comprising mortgage overages assessed by loan officers working for a large mortgage company. The data set is unique in that it includes three credit scores for each borrower. By focusing on mortgage pricing (overages), as opposed to the approval decision, we avoid the need to measure default risk. Second, we perform a formal sensitivity analysis that quantifies the impact of potentially omitted variables. Our results suggest that minority borrowers pay more on average for mortgages than non-minorities, and that this effect persists even in the presence of unmeasured confounders.

1 Introduction

In July of 2012, Wells Fargo Bank agreed to a $175 million fine on the grounds that it had discriminated against African-American and Hispanic borrowers between 2004 and 2009 in violation of fair-lending laws. The bank, while acquiescing to the penalty, nonetheless denied the government's charges, claiming that it settled due to the prohibitive litigation costs involved. The bank later asserted that a corporate decision to cease providing mortgages through independent brokers was a completely independent choice. The Wells Fargo payment closely followed a $335 million fine agreed to by Bank of America in late 2011. The Department of Justice complaint in this case alleged that more than 200,000 black and Hispanic borrowers paid more for loans than white borrowers because of race, not borrower risk.[1]

Despite these troubling charges, statistical evidence for racial discrimination in the mortgage industry is both scarce and of questionable quality.[2] Determining the existence of discrimination in mortgage lending would be straightforward if the data comprised identically credit-worthy and sophisticated individuals applying for comparable loans to the same, or randomly assigned, loan officers. Instead, scholars must rely on observational data and statistical controls, and the threat from omitted variable bias is ever present. Banks rarely make sensitive borrower information available to re-

[1] There have also been a number of far smaller recent settlements, such as a $700,000 fine agreed to by the Texas Champion Bank of Alice, TX for discriminating against Latinos in February of 2013 and a $33 million settlement by New Jersey's Hudson City Savings Bank for racial discrimination in September of 2015.

[2] Fines do not provide definitive proof of bias as lenders have an incentive to settle even if they are non-discriminators.

searchers, and scholars must make due with data that are far from ideal.[3] Kau et al. (2012), for example, rely on neighborhood racial and ethnic compositions because they lack actual data on the borrowers themselves. Even the well-known data set collected by the Federal Reserve of Boston lacks clean measures of applicant creditworthiness (for the best known analysis of these data, see Munnell et al. 1996; for recent reanalyses, see Goenner 2010 and Han 2011).

We attack the omitted confounders problem by combining a unique data set with a formal sensitivity analysis. The data set is unusual in two ways. First, it includes three credit scores, one from each of the three major bureaus, for each borrower. We are fortunate in that credit scores are rarely, if ever, made available directly to researchers. Second, the data set concerns mortgage pricing as opposed to the approval decision. The dependent variable is the overage, which is a fee assessed after the decision to grant a mortgage has been made. Many of the variables that are routinely unobserved by data analysts are factored in at the approval stage of the mortgage negotiations and thus do not affect overages. We can show, for example, that credit scores are unrelated to overage size.

Despite the unique nature of our data set, there remain variables we would like to have, but to which we were not granted access. Our response

[3] The U.S. government requires the collection of race data under law; for a discussion, see Taylor 2012. The data nearly always fall short of the information available to lenders. Researchers studying peer-to-peer lending where some individuals request and others provide loans (notably from the website Prosper.com) have access to far better data. Both Pope and Syndor (2011) and Ravina (2012) find similar evidence of black borrowers facing some additional costs but also defaulting more. Regardless, this type of situation is qualitatively different, and arguably less important from a public policy perspective, from that with respect to the choices made by for-profit lenders.

is to employ a sophisticated sensitivity analysis that allows us to assess whether any additional unobserved variables exist that could change our results significantly. The results of the sensitivity analysis suggest that our findings are not particularly sensitive to omitted confounders.

Our empirical strategy unfolds in three stages. First, we use regression (overage is a continuous variable) to show that being a minority is associated on average with larger overages, while credit scores (default risk) are unrelated to overages. Second, we report a number of robustness checks and demonstrate that our findings are consistent across many different specifications. These alternative specifications include general additive models to check for nonlinearities and matching estimators to assess model dependence. Third, we introduce and describe the results from the sensitivity analysis. Our conclusion is that a non-trivial amount of discrimination exists in these data.

The paper proceeds as follows. In section 2, we describe the problem of unobserved heterogeneity in discrimination research and discuss some recent attempts at solutions. Section 3 introduces the idea of an overage, and section 4 describes our data set. The empirical analysis is in section 5, and the sensitivity analysis is in section 6. Section 7 concludes with a brief discussion of discrimination law, and how the law relates to our results.

2 Discrimination and unobserved heterogeneity

Omitted variables bedevil the study of discrimination in mortgage lending. As Ross and Yinger (2002, 108) note, the Boston Federal Reserve Study has more control variables than any previous study (an additional 38), but was still widely attacked for leaving out important explanatory variables. Researchers studying discrimination in other areas experience similar problems and have developed two ways of dealing with missing variables. The first is known as audit or matched pair testing where two volunteers (or confederates) of different races attempt to get help at a retail outlet, obtain a loan, or be hired for a job. The key to this design is that the two volunteers must be alike in every way except race (or gender) to the store clerk, mortgage officer, or employer. Any difference in service or outcome may then be attributed to discrimination.

Heckman (1998) criticizes audit tests for fragility in the face of assumptions regarding unobservable variables. No two people are alike in every discernible way except for race (or gender). Researchers have attempted to overcome this flaw by using correspondence tests, where the employer, for example, does not observe the matched applicants (Neumark, 2012). Heckman and Siegelman (1993), however, argue that discrimination is unidentified even in correspondence tests when the variance of unobserved productivity differs across groups. Even assuming that this problem could be addressed (Neumark (2012) proposes a test), the experimental approach is necessarily limited in scope.

A second method widely used by researchers is the outcome test first suggested by Becker (1993). Outcome tests consider the difference in average outcomes between two groups that have had a common experience (Glaser, 2014). If the difference is significant, we can infer that the two groups have been treated unequally. In Becker's original formulation, if minorities default on loans at a lower rate than non-minorities, there is evidence that minorities faced more stringent requirements than non-minorities to secure a loan (Glaser, 2014). Outcome tests are not susceptible to omitted variables bias because the variables that are unobserved by the researcher are observed by the loan or police officer, and under the hypothesis of no discrimination, are taken into account.

Ayres (2001) criticizes outcome tests for what he terms the inframarginality problem. Discrimination is most likely to occur in marginal cases; well-qualifed borrowers receive loans, and non-qualifed borrowers do not (Glaser, 2014). Marginal rates, however, are unobservable, and comparing averages confuses these marginal cases with the nonmarginal (inframarginal) cases. An exception exists when we expect an equilibrium to hold under the null hypothesis of no discrimination. In that case, no difference exists between marginal and non-marginal cases.[4]

Neither approach has been widely used to study mortgage lending discrimination. Correspondence studies require online transactions, and nearly 50% of first-time homebuyers meet with a loan representative in person (Garrison, 2014). To our knowledge, the only correspondence study relating to mortgages is Hanson et al. (2016), who look at mortgage loan originators (MLO).[5] They find an effect of being African-American on MLO response roughly equivalent to a credit score that is 50 points lower. MLOs also offer

[4]Knowles et al. (2001) argue, for example, that motorists must in equilibrium carry illegal substances with equal probability regardless of race (Anwar and Fang, 2006). The reasoning is that if one subgroup carries contraband more often than another, the police would focus their searches on those motorists. The subgroup would respond by carrying contraband less often.

[5]Licensed mortgage sales workers who assist customs with loan applications. They have discretion over how they respond to customer inquires (Hanson et al., 2016, 2).

more details to whites and use friendlier language in email correspondence.

An outcome test applied to mortgage decisions would suffer from the inframarginality problem as there is no reason to assume equilibrium behavior. We do not, for example, expect to observe mortgage lenders targeting minorities to the exclusion of non-minorities with the expectation of higher profits.

Our approach is to focus on mortgage pricing, namely overages, which are monies that borrowers may pay after the decision to grant a mortgage has been made. The reason is simple: many of the variables, observed or not, that determine whether a borrower receives a loan do not determine whether that borrower pays an overage. The determinants of loan approval are singularly concentrated around the ability to repay the loan. Common variables related to the approval decision include purchase price, personal assets, personal liabilities, income, employment history, and credit history. Not all the determinants of risk can be observed, however, most often due to privacy concerns. The unobserved measures then raise the specter of omitted variable bias. Once the decision to make a loan has been made, risk of default is no longer a central concern, and the loan officer's attention turns to the cost of the loan.

That being said, looking at mortgage pricing instead of the approval decision does not completely free us of omitted variables. As we argue in the following sections, we have proxies for some of these variables, and the sensitivity analysis will address the others.

3 The Nature of Overages

An overage is the extra amount that a borrower may pay beyond the posted price of a mortgage (the combination of interest rate and points). Thus, we can think of an overage as the total loan points minus any fees (such as origination fee) minus the posted rate. Loan officers generally share in the monies from overages, suggesting that in an efficient market we will see no discrimination based on race as officers have an incentive to extract the maximum from all concerned. However, loan officers typically have a great deal of flexibility with respect to setting overages. Furthermore, most borrowers are unaware of the existence of overages (Black et al., 2003, 1141). Mortgage pricing negotiations are therefore ripe for potential discrimination.

To understand how overages work, consider a situation where the posted price (the price on the rate sheet given by the mortgage company to the lending officer) on a particular loan is 6% and 0 points. Now imagine that the lending officer induces the borrower (who is not shown the rate sheet) to pay 6% and 1 point. That point is an overage. On a loan of $300,000, a 1 point overage is worth $3000. Typically, loan officers pocketed half of the money generated by the overage.[6] Thus, the loan officer could increase his or her commission by $1500.

Not all borrowers pay positive overages. Some pay no overage, and others pay negative overages (called underages). Underages are usually the result of competition between lending institutions over a particular type of loan or a particularly desirable customer. Financially sophisticated borrowers

[6] Bank of America banned the practice in 2010 (Guttentag, 2010).

can rate shop among competing lenders and drive down the price of their mortgage (Harney, 1993). Underages are rare, and our dependent variable is correspondingly skewed to the right.

Only two major studies of overages exist.[7] Courchane and Nickerson (1997) analyze data from three banks following an investigation by the Office of the Comptroller of the Currency. The loans were made between 1992 and 1994. They provide regression results for two of the banks, bank A and bank B, and although small differences are found, they conclude that the banks were profit-maximizing, but not discriminatory. Black et al. (2003) study 1996 data from a major bank. They find no evidence of discrimination once other differences in the borrower pool, such as bargaining ability, were controlled.

Our data and analysis differ significantly from these previous studies. Courchane and Nickerson (1997) and Black et al. (2003), for example, consider data obtained from banks in the early to mid 1990s, whereas our data comes from a mortgage brokerage company in 2000. These differences are meaningful because they relate to both the number of borrowers with overages and the size of those overages. Banks making mortgage loans in the mid 1990s were under heightened scrutiny following the 1992 Federal Reserve mortgage lending study. Black et al. (2003) have a sample size of 2002, but only 17.9% of the sample have positive overages. The bank they study also limited overages to 2%. In contrast, 54.5% of the borrowers in our sample have positive overages, and nearly 8% of the overages in the sample are

[7] We were unable to obtain the data from either study. The Black et al. (2003) data are proprietary, and Courchane and Nickerson (1997) were unable to locate their data.

larger than 2 (the largest overage is 6; see Table 1).

Courchane and Nickerson's (1997) sample from bank A contains over 33,000 loans, but only 38.7% have positive overages. In their regression, however, the authors appear to have selected on the dependent variable and included only those borrowers with positive overages. The sample from bank B has a larger percentage of borrowers with overages, but the regression results are questionable as the authors include interest rate as a control variable. The resulting endogeneity bias likely explains their strikingly counterintuitive finding that "minorities are likely to be charged smaller amounts of overage points" (Courchane and Nickerson, 1997, 147).

Finally, Black et al. (2003) persuasively argue that credit risk should not explain much variation in mortgage pricing because risk is included in the base interest rate of the loan. Neither of the two previous studies, however, have direct access to credit scores with which to assess the claim. We can show that average overage size is orthogonal to credit score.

4 The data

The data comprise 2129 observations on mortgage overages (measured in points) assessed by loan officers for a large national mortgage company during the second quarter of 2000 (the loans were made between August 1, 2000 and October 31, 2000). The overages in the data set are from loans actually taken, not just negotiated. Although the firm had branch offices and brokered loans in all quadrants of the country, the majority of loans in the data were made in and around the Northeast. Providing adequate information to assess the representativeness of this mortgage broker is difficult as the data were provided to the authors by a principal of the firm on the condition of anonymity for the firm and the loan officers. The fact that our data are from a single firm (as are the data in Black et al. (2003)) is concomitant with its uniqueness. In addition, the firm, which no longer exists an an independent entity, bore an instantly recognizable household name. The differences between our data and the data from Black et al. (2003) and Courchane and Nickerson (1997) are due to institution type (banks v. mortgage broker) and temporal proximity to federal scrutiny (1992).

Any mortgage data from the 2000s raises questions concerning subprime lending. The major increases in subprime mortgages, however, occurred between 2003 and 2006 (JCHS, 2008, 4). Furthermore, through 2001, the ratio of home prices to income had remained steady between 2.9 and 3.1 for two decades. The ratio rose to 4.0 only in 2004, and 4.6 in 2006 (Steverman and Bogoslaw, 2008). There is little reason to assume, then, that our data are either wildly unrepresentative or the product of the subprime crisis.

We list the variables used in our analysis, their definitions, and descriptive statistics in Table 1. The data set has some unique features. The Home Mortgage Disclosure Act (HMDA) does not require mortgage companies to report credit scores for individual borrowers, the branch of the company that sold the loan, or the name of the lending officer. Researchers analyzing lending discrimination commonly make do without such measures. Our data set, on the other hand, not only includes three credit scores for each borrower, but also the branch office that originated the loan and the name of the lending officer. These variables allow us to use fixed effects and

Table 1: This table lists the variables used in the analysis, the definitions of those variables, and descriptive statistics where appropriate.

Variable name	Definition	Min.	Mean	Max.
BRANCH	Branch that originated the loan	NA	NA	NA
OFFICER	Officer that sold the loan	NA	NA	NA
DATE	Date of the transaction	8/1/00	NA	10/31/00
GOVERNMENT	1=FHA loans	0.000	0.1330	1.000
OVERAGE	Overage in points	-5.000	0.4776	6.000
AGE	Borrower age, in years	19	39.8	80
INDIAN	1=INDIAN; 0=otherwise	0.000	0.004	1.000
ASIAN	1=ASIAN; 0=otherwise	0.000	0.038	1.000
BLACK	1=BLACK; 0=otherwise	0.000	0.050	1.000
HISPANIC	1=HISPANIC; 0=otherwise	0.000	0.058	1.000
WHITE	1=WHITE; 0=otherwise	0.000	0.850	1.000
SEX	1=MALE; 0=otherwise	0.000	0.736	1.000
EQUIFAX	Equifax credit score	495	705.78	829
TRANSUNION	TransUnion credit score	492	710.27	829
EXPERIAN	Experian credit score	476	707.98	831

clustering in the following analyses.

Additionally, we have a suite of variables that mortgage companies must report to the federal government under the HMDA. These variables include the date of application, the race of the borrower, the gender of the borrower, the age of the borrower, and the type of loan (product).

54.2% of the borrowers in our sample have positive overages, which means that over half of the sample paid more for their loan than the price quoted to the lending officer by the bank.[8] Of those with positive overages, the mean overage is 0.962 with a standard deviation of 0.85. White borrowers comprise 85% of the sample, while black and Hispanic borrowers collectively comprise 10.8% of the sample (each about 5%). Asian and South Asian Indians make up less than 4% of the sample. A cross-tab of

[8] 36.4% have an overage of zero, and 9.4% have an underage.

overages by race (trichotomized into "under," "none," and "over") is in Table 2. Slightly over half of white borrowers have positive overages (53%) while over three-quarters of black (79%) and Hispanic (78%) borrowers have positive overages. A χ^2-test returns a p-value of 0 indicating that overage and race may not be independent of one another.

Table 2: A cross-tab of overages (trichotomized into "under," "none," and "over") by race. A χ^2-test of independence returns a p-value of 0.

Overage	White	Black	Hispanic	Asian
Under	0.135	0.100	0.000	0.075
None	0.334	0.114	0.225	0.453
Over	0.531	0.786	0.775	0.472

Loan type is related to overages. The relationship, however, is driven completely by one particular kind of mortgage. The data set includes nine categories of loans such as adjustable rate mortgages, government mortgages, second mortgages, construction mortgages, etc. Agency or conformable loans meet guidelines set forth by Fannie Mae and Freddie Mac. Agency mortgages comprise 65% of the sample. Government loans are those administered under various programs run by The Federal Housing Administration. Such loans are easier to qualify for and have lower down payments requirements. Government-type mortgages comprise 13.3% of the sample. (The other categories contain fewer than 100 observations each.) The government category is of interest because 85% of these loans carry positive overages, as opposed to 53% of agency loans.[9]

[9] A χ^2-test returns a p-value of 0, indicating that loan type and overage may not be independent of one another. The other loan type categories behave as expected. Only 28.4% of second loans have overages, as second mortgages are taken out by those who have been through the process at least once before.

Moreover, a strong relationship exists between government-type loans and race. 47% of blacks and 50% of Hispanics have government-type loans while only 16% of whites do. We can see these patterns clearly in a three-dimensional table of overage by race by loan type (see Table 3). 83% of whites with government loans have positive overages compared with 91% of blacks and 100% of Hispanics ($n = 40$). 51% of whites with agency loans have positive overages compared with 79% of blacks and 75% of Hispanics.[10]

Table 3: A three-way cross-tab of overages by loan type (agency or government) by race. Minorities, on average, pay higher overages than whites for both agency and government-type loans.

	Overage	Agency	Government
White	Under	0.160	0.095
	None	0.333	0.079
	Over	0.507	0.826
Black	Under	0.083	0.000
	None	0.125	0.091
	Over	0.792	0.909
Hispanic	Under	0.000	0.000
	None	0.250	0.000
	Over	0.750	1.000

The relationship between government-type mortgages and overages and race is not the result of the differences between the mortgages but, rather, the characteristics of borrowers who take out such mortgages. FHA mortgages, for example, are insured by the FHA, generally require lower down payments (which can also take the form of gifts), and a portion of the closing

[10]While it is clear that loan type plays a significant role, it is unclear whether product type is a pre- or post-treatment variable. In the analyses to follow, we treat it both ways. We include product type in regressions, but we do not match on it. All analyses, however, have been done both ways, and the presence or absence of product type among the control variables does not change any of our conclusions.

costs can be included in the loan amount. Furthermore, borrowers pay an insurance premium, as conventional loan borrowers do, but this premium is not canceled when equity grows.[11] In the analyses to follow, we condition on government-type loans but, in reality, we use this binary variable as a crude proxy for loan size, a variable to which we do not have access. Borrowers with government-type loans borrow far less than those with conventional loans, and loan officers may wish to extract additional money, in the form of overages, from such borrowers to make up for lower fees.

5 Empirical analysis

Here, we report results from our empirical analysis. We show estimated regression coefficients, evidence of robustness across a variety of specifications, and a discussion of inferential threats including omitted variables and selection issues. Two major findings emerge. The first is that credit scores are unrelated to overage size. This finding is perhaps less than surprising as we do not expect risk of default to be related to overages. We expand on this point in the discussion. The second is that race is related to overage size across all specifications.

The results from five different multiple regressions are in Table 4.[12] The consistency of the effects across the regressions is remarkable. In all five regressions, the effects of being black, being Hispanic, or having a government loan are positive, substantive, and highly statistically significant. The

[11] These are the rules that were still in place as of 2002. Some rules may have changed in the ensuing years.

[12] We also estimated specifications that included interaction terms. No estimates changed direction or significance.

effects of the three credit scores, on the other hand, are essentially zero and do not even approach statistical significance.[13]

Table 4: Multiple regression results. Overage is the dependent variable in each case. The specifications vary by controls and standard errors (robust v. clustered by branch office). The fourth regression includes fixed effects (by branch), and the fifth regression uses a robust estimator. Credit scores are divided by 100.

	OLS	OLS	OLS	Fixed	Robust
	Robust SEs	Robust SEs	Cluster: branch	by branch	
Intercept	0.4538	0.4546	0.4546		0.1823
	(0.050)	(0.050)	(0.097)		(0.027)
Black	0.4337	0.4285	0.4285	0.4065	1.0152
	(0.143)	(0.143)	(0.116)	(0.109)	(0.056)
Hispanic	0.5143	0.5110	0.5110	0.5085	0.1613
	(0.162)	(0.161)	(0.258)	(0.107)	(0.055)
Government	0.6425	0.6392	0.6392	0.6635	0.4659
	(0.070)	(0.071)	(0.162)	(0.653)	(0.032)
Equifax	-0.0142	-0.0145	-0.0142	-0.0110	-0.0088
	(0.018)	(0.022)	(0.021)	(0.018)	(0.010)
TRW		-0.0093	-0.0083	-0.0079	-0.0052
		(0.024)	(0.013)	(0.017)	(0.009)
Experian		0.0103	0.0086	0.0154	0.0090
		(0.018)	(0.019)	(0.017)	(0.009)
N	1905	1905	1905	1905	1905
Missing	235	235	235	235	235
SEE	0.8268	0.8271	0.8271	0.7985	0.3984
R^2	0.095	0.096	0.096	0.368	0.092

Columns one and two of Table 4 contain estimates from ordinary least squares linear regressions with heteroscedastic-consistent standard errors. The regression in column one contains only a single measure of credit, while column two contains three such measures to demonstrate that no multicollinearity problem exists between the three measures.[14] Column three

[13] The credit scores have been divided by 100 to increase interpretability.

[14] The correlations between the three credit scores are roughly 0.87, which is generally

contains regression estimates with the standard errors clustered by branch. The loans in the sample originated from 46 different branches, with some producing almost 200 loans during third quarter (branch 230: 196) and other branches producing far fewer loans. Many of the observations are therefore not independent, and clustered standard errors are appropriate.[15] While some standard errors have increased in size, no estimates have changed significance. Similarly, we might imagine that some branches, by virtue of being in different areas, might routinely assess overages differently. Column four contains the results of a regression with branch fixed effects. The results are substantively indistinguishable from the previous results.

Across the first four regressions, blacks on average pay 40% of a point more than white borrowers. Hispanics pay 50% of a point more, and FHA borrowers pay 60% of a point more.[16] In terms of the example given in Section 3, blacks would pay on average $1200 more, Hispanics $1500 more, and FHA borrowers $1800 more. Black borrowers with FHA loans could well have a overage that is a full point larger than white borrowers with conventional loans.

Contrast these effects with those of the various credit scores. A 100-point increase in an Exquifax credit scores is associated with a 1-2% of a point decrease in overage size. In terms of our example, a borrower with a credit score of 700 would pay on average $30-60 less than a borrower with a credit

---

not high enough to produce serious multicollinearity with this many observations. The condition number of the full design matrix is 1, well below problematic levels (Belsley et al., 1980).

[15] There are 46 branches, so clustering by branch does not generate problems associated with a small number of clusters.

[16] Conditioning on conventional loans made no difference to the analysis, and the variable itself was non-significant in every specification.

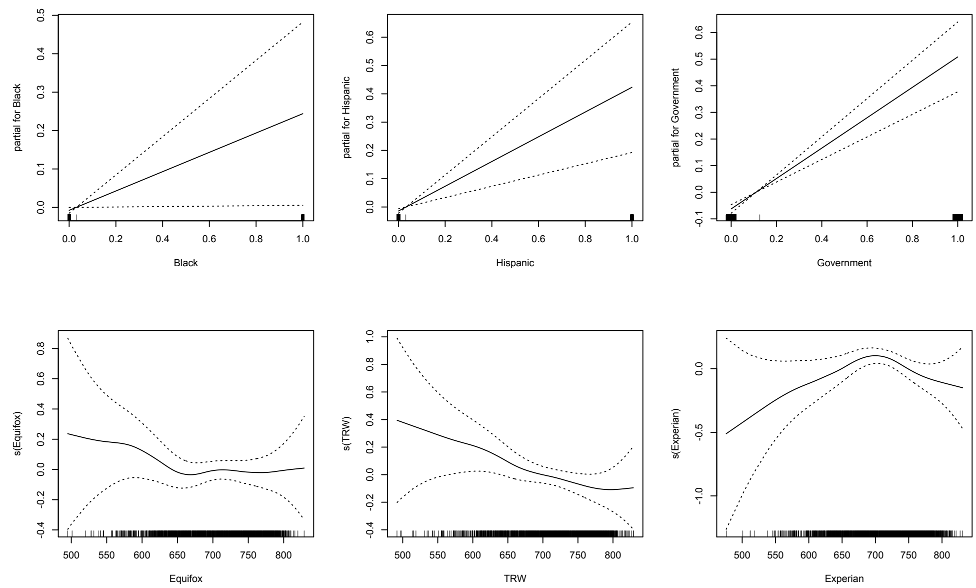
score of 600. The estimated coefficient on Experian credit scores is positive but, as noted earlier, the standard errors are relatively large indicating that these effects are just bouncing around zero. Once the loan approval decision has been made, the risk of default is unrelated to overage size.

Regression diagnostics indicate that influential observations and nonlinearity may pose a threat to our estimates and inferences. We therefore turn to robust estimators and nonlinear estimators. Column five of Table 4 contains results from a robust regression estimated using Yohai et al.'s (1991) MM-estimator, which combines both high efficiency and high breakdown point (resistance to outliers). The results vary somewhat from the other regressions, but no estimate changes direction or significance. The estimated coefficient on being Hispanic shrinks, and the estimated coefficient on being black increases.[17]

We might imagine that credit scores are nonlinearly related to overages. Figure 1 contains the results of a generalized additive model run on the same set of covariates as the previous regressions. The three lower plots are graphs of the effects of the various credit scores. None of them appear to be related substantively to overages. In each case, there is very little movement as credit scores increase (note the small range of the y-axis), and even where the confidence bands are tightest, zero is included. On the other hand, the confidence intervals for black, Hispanic, and government do not include zero (the three upper plots). The effect of being black appears to be somewhat smaller than what we saw in the linear regressions, and this result is echoed

[17] These changes are related to high leverage observations that are down-weighted by the MM-estimator.

Figure 1: Generalized additive model results. The dependent variable is overages. The effects of the six right-hand side variables are displayed in graphical form with confidence bands (2 standard errors) and rug plots.



in the one of the matching analyses to follow. The effects of being Hispanic and having a government loan remain unchanged.

As a final robustness check, we assess the effect of being a minority in a less model-dependent manner by performing a matching analysis using three different methods: one-to-one, genetic (Sekhon, 2011), and propensity score (Rosenbaum and Rubin, 1983b).[18] We also use three different treatments: being Hispanic, being black, and being minority (the sum of Hispanic and black).[19] For each treatment and method, we match on age and the three credit scores.[20] The results from the one-to-one and genetic matching are

[18]There is no reason to prefer one method to another.
[19]Minority is used as a treatment to increase the number of treated observations.
[20]Including the possibly post-treatment variable government-type loan makes no sub-

Table 5: Treatment effect estimates from matching done three ways (one-to-one, genetic, propensity score) with Abadie-Imbens standard errors. The variable minority is black plus Hispanic.

Treatment	Estimate	AI SE	# Treated Obs	Matched # Obs
One-to-One				
Minority	0.3463	0.1273	114	1655
Hispanic	0.5596	0.2017	54	1655
Black	0.3172	0.1695	60	1655
Genetic				
Minority	0.3388	0.1304	114	1655
Hispanic	0.5444	0.1943	54	1655
Black	0.3349	0.1815	60	1655
Propensity Score				
Minority	0.6443	0.1313	114	1655
Hispanic	0.4267	0.2076	54	1655
Black	0.2648	0.1846	60	1655

quite similar to each other and remarkably similar to the previous analyses. The effects of being a minority and being Hispanic are both of the expected size and significance. The effect of being black just misses significance at conventional levels. The propensity score analysis increases the estimated effect of minority and decreases the effect of black, but the general inference that minorities pay higher overages on average than whites remains unaffected.[21]

Discussion

We demonstrated above that credit scores are unrelated to overages.

stantive difference.

[21] Table 7 in the Appendix contains balance statistics from the genetic matching using Hispanic as the treatment. (Results from using black or minority as treatments are similar.) The matching improves the balance on all five variables. In each case, the standard mean difference decreases significantly, and the t-tests indicate that we fail to reject the null hypotheses of no difference after matching. These results suggest that the matching is credible.

We also demonstrated that being black, Hispanic, or having a government-backed loan does seem to be related to overages. We still need to address two serious threats to our inferences: remaining omitted variables and selection bias.

The omitted variable problem is mitigated somewhat in our study due to the unique features of our data set. First, our dependent variable is the overage, as opposed to the approval decision, and unobserved variables in mortgage studies most often relate to the latter (the risk of default). The Boston Fed Study, for example, was criticized for not including, among others, "whether data could not be verified" and "whether the applicant's credit history meets loan policy guidelines for approval" (Ross and Yinger, 2002, 109). Variables related to the risk of default theoretically should affect the decision to approve the mortgage loan, but not any potential overage. The fact that we found no relationship between credit scores, a major determinant of default risk, and overages bolsters our claim. Second, although we cannot condition on the potentially important variable loan size, we can condition on an imperfect proxy, government-type loans. Our proxy has the expected large and positive effect, but it does not swamp the effect of race.

Another inferential threat is that selection bias may mask the effect of credit scores, which would in turn diminish the effects of race. Prospective borrowers with relatively low credit scores who are approved for loans are likely to have higher interest rates on average. An overage assessed atop an already high interest rate may make the loan unaffordable. If such borrowers select out of the sample (i.e., decide against taking the loan), we may underestimate the effect of credit scores on overages. To put it another way, the distribution of overages for relatively low-credit borrowers would be truncated at the point where the loans become prohibitively expensive. The same principle may be at work at the approval stage. Potential borrowers with low credit scores who were not approved for loans do not appear in the data. These borrowers may have had large overages if they had been approved, and their absence from the data set again may lead us to underestimating the effect of credit.

Fortunately, such selection bias is unlikely to be affecting our estimates. There are two reasons. The first lies in the nature of overages. Recall that in many cases overages are pure profit for both the loan officer and the lending institution. It is in no one's interest to drive away approved borrowers with restrictive pricing. The loan officer should maximize the overage without driving the customer away. Such borrowers are likely then to appear in the data set. A slightly different logic is at work with unapproved borrowers. In this situation, borrowers are not approved for loans generally because the risk of default is too high. If these low-credit borrowers were approved, they would have high interest rates, and it is unlikely that they could be assessed large overages simply because they would not be able to afford them.

The arguments just outlined may not convince all readers who may prefer that we include the omitted variables or run the selection model. We would if we could. We do not have to rely, however, exclusively on arguments. The sensitivity analysis to follow addresses both threats. For example, it is now commonplace to understand selection bias an omitted variables problem, where the inverse Mills ratio is the variable missing from the second-stage equation. If either the missing inverse Mills ratio or any other omitted variable would change our results, evidence will show up in the sensitivity analysis.

6 Sensitivity analysis

The collective results from our *prima facie* analysis suggest that being a minority borrower means paying on average somewhere between 40% and 60% of a point in overages more than white borrowers. The loan officers in our sample, however, were privy to information not contained in our unique data set, and omitted variable bias remains a possibility. In addition to a better measure of loan size and an estimate of the inverse Mills ratio, we would like a measure of financial sophistication as savvy consumers presumably know to shop around and thus reduce their chances of paying an overage. Financially sophisticated customers might at least demand to see the rate sheet being used by the lending officer. Some of the effect of financial sophistication is no doubt captured by credit score. Financially unsophisticated consumers are unlikely to have sound credit, given the credit scores reflect not only the availability of credit, but also the judicious use of past credit. Credit scores also correlate highly with education (Cole et al., 2012). Nonetheless, we feel that it is important to assess how much of an impact an unobserved variable such as financial sophistication might have on our main finding.

The goal of our sensitivity analysis is understanding how much omitted variable bias is required to nullify the effect of being Hispanic, say, on overages.[22] Formal sensitivity analysis goes back to Cornfield et al. (1959) and was developed by Rosenbaum and Rubin (1983a) and Rosenbaum (1988). The method we use comes from Hosman et al. (2010), which is related to Imbens (2003).

Consider a regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \Gamma * \text{Hispanic} + \delta * Z + \mathbf{u},$$

where $\mathbf{y}$ is a $N \times 1$ vector of overages, $\mathbf{X}$ is an $N \times (k-2)$ matrix of control variables (age, product type, credit scores, etc.), $\mathbf{u}$ is an $N \times 1$ vector of errors, Γ is the coefficient on the treatment, being Hispanic, and δ is the coefficient on the putative omitted variable, Z. The regressions we report in Section 5 are all a variant of

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma * \text{Hispanic} + \mathbf{u},$$

where the unknown variable Z is omitted. The sensitivity analysis quantifies how large an effect Z must have, when included in the regression, before the estimated coefficient on Hispanic, $\hat{\gamma}$, becomes zero.[23]

The bias on $\hat{\gamma}$ caused by a possibly omitted variable Z is a function of Z's confounding with being Hispanic and Z's effect on the dependent variable, overage. The Hosman et al. (2010) method generates sensitivity

[22] We focus on the effect of being Hispanic because the empirical analysis shows that its estimated effect is somewhat larger than that of being black. Space concerns keep us from providing sensitivity analysis for all the independent variables, but the results for being black are substantively similar.

[23] We write of a single omitted variable, Z, for convenience. We can, however, think of Z as being a combination of two or more omitted variables without doing damage to the argument.

intervals for $\hat{\gamma}$ that are a function of these two effects. Confounding is measured by the t-statistics from a regression of Z on the other regressors. We denote confoundedness of Z with the treatment of interest (being Hispanic, in this case) as t_Z. Z's effect on the dependent variable is measured by the proportionate reduction in unexplained variance when Z is included in the regression,

$$\rho^2_{y\cdot z|\mathbf{x}} = \frac{(1 - R^2_{\text{no } Z}) - (1 - R^2_{\text{with } Z})}{(1 - R^2_{\text{no } Z})}.$$

Note that neither the t-statistics nor $\rho^2_{y\cdot z|\mathbf{x}}$ are used for inferential purposes. Both values simply describe the relationships between the possibly omitted variable Z and either the treatment or the dependent variable.

Hosman et al. (2010) prove that the omitted variable bias can be written as a product of the two effects described above and the standard error on $\hat{\gamma}$,

$$\hat{\gamma} - \hat{\Gamma} = \text{SE}(\hat{\gamma}) t_Z \rho_{y\cdot z|\mathbf{x}},$$

provided $R^2_{y\cdot z\mathbf{x}} < 1$ and t_Z is finite. They go on to prove, under the same conditions, that the same statistics can be used to express the effect of omitting Z on the standard error

$$\text{SE}(\hat{\Gamma}) = \text{SE}(\hat{\gamma}) \sqrt{1 + \frac{1 + t_Z^2}{\text{df} - 1}} \sqrt{1 - \rho^2_{y\cdot z|\mathbf{x}}},$$

where $\text{df} = n - \text{rank}(\mathbf{X}) - 1$, the residual degrees of freedom after Y is regressed on X and Hispanic. Taken together, these results allow the specification of a union of interval estimates

$$\hat{\Gamma} \pm q\,\mathrm{SE}(\hat{\Gamma}) : |t_Z| \leq T, \rho^2_{y\cdot z|\mathbf{x}} \leq R$$

for any nonnegative limits T and R.[24] The union is the collection of $\hat{\Gamma}$ values falling into the interval after adding the omitted variable Z.

Table 6: 95% sensitivity intervals for Hispanic with the unobserved variable's treatment confounding hypothesized to be no greater than the treatment confounding of the 4 deliberately omitted variables below. The decrease in unexplained variance is hypothesized to be no greater than either 1% or 5%.

Variable	t_Z	$100 \cdot \rho^2$	5%	1%
Age	1.50	0.29	(0.231, 1.015)	(0.250, 0.996)
Loan type	3.80	10.30	(0.171, 1.075)	(0.222, 1.024)
Equifax	1.70	0.18	(0.225, 1.021)	(0.247, 0.999)
TRW	1.60	0.24	(0.230, 1.016)	(0.249, 0.996)

Z is unobserved. How then should a researcher choose values for t_Z and $\rho^2_{y\cdot z|\mathbf{x}}$? Hosman et al. (2010) suggest benchmarking: treating the observed covariates one at a time as being the unobserved covariate Z and collecting values for t_Z and $\rho^2_{y\cdot z|\mathbf{x}}$ to use as guides. The downside to this procedure is that we can only benchmark against covariates that exist in the data set and, in our case, the number of covariates is limited. That being said, we have access to covariates with small effects, such as age, and large effects, such as loan type, that provide us with a reasonable range of benchmarks.

Table 6 lists 4 variables taken from our data set and treated as the unobserved covariate Z. Loan type is confounded with being Hispanic, but neither credit score is. Age is uncorrelated with being Hispanic. None of the variables moves $\rho^2_{y\cdot z|\mathbf{x}}$ more than 1 percentage points. (Loan type is a

[24]See Hosman et al. (2010) for derivations and proofs.

nominal variable that has 9 categories, i.e. 9 binary variables, which explains its larger value.)

The intervals in Table 6 are for $\rho^2_{y\cdot z|\mathbf{x}} \leq 0.01$ or 0.05, which are both conservative bounds. None of the intervals at the 1% level include 0; none of the intervals even approaches zero. The same is true at the 5% level. We conclude that unless an omitted variable exists that is confounded with being Hispanic at the same level as having a government-type loan and is a stronger predictor of large overages than having a government-type loan, we can be confident that being Hispanic has a real effect on overages.[25] Thus, even if financial sophistication or loan size or the inverse Mills ratio were an omitted variable, it would not change our conclusions unless these variables had a larger impact than having a government-type loan. In each case, we have reasons to believe that the effects would not be that large. We have, for example, an imperfect proxy for loan size and, given that most borrowers had no idea that they were even being charged an overage, it is unlikely that financial sophistication would have such effects.

7 Discussion

One means of interpreting these findings is through the lens of the law. There are two ways that a lending institution can run afoul of anti-discrimination laws: disparate treatment and disparate impact. The Federal Deposit Insurance Corporation's *Side by Side: A Guide to Fair Lending* defines disparate treatment as a lender treating an applicant differently based on a prohibited

[25] We also ran a sensitivity analysis on the propensity score matching. The results are in the Appendix and are substantively similar.

category such as age, sex, race, or religion. Disparate impact occurs when a policy or practice applied equally to all applicants has a disproportionate adverse impact on applicants in a protected group. The difference lies in motive: disparate treatment requires lenders to discriminate intentionally, and disparate impact can be the result of purely neutral policies.

Whether our results show disparate treatment or disparate impact depends on how convincing the reader finds the sensitivity analysis. If the reader is reasonably sure that no omitted variable exists that would significantly alter our results, then we have shown disparate treatment. We recognize that even readers who find our results troubling might hesitate before arriving at that conclusion. In response, we attempted to compare black and Hispanic loan officers to white loan officers directly. Following Anwar and Fang (2006), who compare the behavior of police officers of different races, if white and minority loan officers behave in similar ways, then they are mostly likely profit-maximizing. We used the list of surnames occurring more than 100 times available from Census.gov to identify the race of each loan officer in our sample. This procedure has some error associated with it so we used Facebook and LinkedIn to find online photographic evidence of race. The error was thus reduced considerably. Unfortunately for our analysis, however, there are simply too few minority loan officers making too few loans to minority borrowers in the year 2000 for justifiable statistical analysis.

If the reader does not find the sensitivity analysis sufficiently convincing, then we cannot claim to have shown disparate treatment because we will not have demonstrated that minority borrowers paid larger overages as a result of intentional racial discrimination. Not all is lost, however. We will still have shown that the mortgage brokerage company from which we obtained our sample was guilty of disparate impact. Even if each lending officer was charging the most he or she could regardless of race, minority borrowers were disproportionally affected, and that is classic disparate impact.[26] These different conclusions highlight the profound impact that omitted variables can have on discrimination research, and the important role that sensitivity analysis can play in sorting out the issues.

8 Conclusion

Race-based discrimination in mortgage pricing is of concern to both scholars and policymakers. Finding clear evidence, however, has proven problematic due to omitted variable bias. Empirical results indicating discrimination are routinely met with claims that the findings could be compromised by unmeasured confounders. The resulting ambiguity has hampered, though not completely derailed, attempts at affecting change.

Certainly, the fear of omitted variable bias has haunted the study of mortgage discrimination. We mitigate the issues raised by potential confounders by considering mortgage pricing (overages) and employing a formal sensitivity analysis. In doing so, we show that credit scores and, by extension, default risk, are unrelated to overages. Conversely, we show that being black or Hispanic is strongly associated with positive overages. In short, we

[26]In a court of law, a plaintiff would have to meet an additional hurdle concerning the business necessity of the policies, and whether the same business goals could be met with less problematic policies.

find the kind of evidence of discrimination that many scholars believe exists, but have difficulty demonstrating unequivocally.

Our analysis covers one lender during one quarter of 2000, but the results are nonetheless clear and troubling. Further, we suspect that our findings are not unique: Although Bank of America has banned the use of overages, there remain calls for mortgage lending that is not incentivized by selling overages out of concern for continued racial discrimination (van den Brand, 2015). We hope that our results contribute meaningfully to the discussion.

Appendix

Table 7: Balance Statistics from Genetic Matching: Treatment=Hispanic. The matching improves the balance on all five variables. In each case, the standard mean difference decreases significantly, and the t-tests indicate that we fail to reject the null hypotheses of no difference after matching.

	Before	After
Age		
Mean Treatment	35.148	35.148
Mean Control	39.905	35.204
Std Mean Diff	-47.053	-0.5495
T-test p-value	0.002	0.920
Government		
Mean Treatment	0.4259	0.4259
Mean Control	0.1106	0.4259
Std Mean Diff	63.194	0.000
T-test p-value	0.0003	1.000
Equifax		
Mean Treatment	665.04	665.04
Mean Control	685.59	666.19
Std Mean Diff	-26.289	-1.469
T-test p-value	0.0706	0.6451
TRW		
Mean Treatment	617.41	617.41
Mean Control	689.23	618.13
Std Mean Diff	-38.24	-0.3855
T-test p-value	0.0072	0.5351
Experian		
Mean Treatment	667.19	667.19
Mean Control	683.12	666.19
Std Mean Diff	-24.427	1.5334
T-test p-value	0.1016	0.7758

Table 8: 95% sensitivity intervals for Hispanic (using propensity scores) with the unobserved variable's treatment confounding hypothesized to be no greater than the treatment confounding of the 4 deliberately omitted variables below. The decrease in unexplained variance is hypothesized to be no greater than either 1% or 5%.g

Variable	5%	1%
Age	(0.191, 1.004)	(0.207, 0.989)
Loan type	(0.311, 1.112)	(0.324, 1.099)
Equifax	(0.235, 0.991)	(0.234, 0.992)
TRW	(0.216, 1.035)	(0.232, 1.019)
Experian	(0.210, 0.987)	(0.215, 0.982)

References

Anwar, S. and H. Fang (2006, March). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review 96*(1), 127–151.

Ayres, I. (2001). *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination.* Chicago: The University of Chicago Press.

Becker, G. S. (1993, April 19). The evidence against banks does not prove bias. *Business Week*.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* Hoboken, NJ: John Wiley and Sons.

Black, H. A., T. P. Boehm, and R. P. DeGannaro (2003). Is there discrimination in mortgage pricing? the case of overages. *Journal of Banking & Finance 27*(6), 1139–1165.

Cole, S., A. Paulson, and G. K. Shastry (2012). Smart money: The effect of education on financial behavior.

Cornfield, J., W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder (1959, January). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute 22*(1), 173–203.

Courchane, M. and D. Nickerson (1997). Discrimination resulting from overage practices. *Journal of Financial Services Research 11*(1-2), 133–151.

Garrison, T. (2014, November 13). First-time homebuyers still face challenge with mortgage process. *HousingWire*.

Glaser, J. (2014). *Suspect Race: Causes and Consequences of Racial Profiling*. New York: Oxford University Press.

Goenner, C. F. (2010). Discrimination and mortgage lending in boston: The effects of model uncertainty. *Journal of Real Estate Finance and Economics 40*(3), 260–285.

Guttentag, J. (2010, March 27). Why mortgage lenders charge overages, and why they may stop. *The Washington Post*.

Han, S. (2011). Creditor learning and discrimination in lending. *Journal of Financial Services Research 40*(1-2), 1–27.

Hanson, A., Z. Hawley, H. Martin, and B. Liu (2016). Discrimination in mortgage lending: Evidence from a correspondence experiment. *Journal of Urban Economics 92*, 48–65.

Harney, K. R. (1993, August 15). Your mortgage: Loan 'overage' charges come under fire. *Los Angeles Times*.

Heckman, J. and P. Siegelman (1993). The urban institute audit studies: Their methods and findings. In M. Fix and R. J. Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America*, pp. 187–258. Washington, DC: The Urban Institute Press.

Heckman, J. J. (1998, Spring). Detecting discrimination. *The Journal of Economic Perspectives 12*(2), 101–116.

Hosman, C. A., B. B. Hansen, and P. W. Holland (2010). The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *The Annals of Applied Statistics 4*(2), 849–870.

Imbens, G. W. (2003, May). Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review 93*(2), 126–132.

JCHS (2008). The state of the nation's housing 2008. Technical report, Joint Center for Housing Studies of Harvard University.

Kau, J. B., D. C. Keenan, and H. J. Munneke (2012). Racial discrimination and mortgage lending. *Journal of Real Estate Financial Economics 45*, 289–304.

Knowles, J., N. Persico, and P. Todd (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy 109*(1), 203–229.

Munnell, A. H., G. M. B. Tootell, L. E. Browne, and J. McEneaney (1996). Mortgage lending in boston: Interpreting hmda data. *American Economic Review 86*(1), 25–53.

Neumark, D. (2012, October). Detecting discrimination in audit and correspondence studies. *The Journal of Human Resources 47*(4), 1128–1157.

Pope, D. G. and J. R. Syndor (2011). What's in a picture? evidence of discrimination from prosper.com. *Journal of Human Resources 46*(1), 53–92.

Ravina, E. (2012). Love & loans: The effect of beauty and personal characteristics in credit markets.

Rosenbaum, P. and D. Rubin (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B 45*(2), 212–218.

Rosenbaum, P. R. (1988, September). Sensitivity analysis for matching with multiple controls. *Biometrika 75*(3), 577–581.

Rosenbaum, P. R. and D. B. Rubin (1983b, April). The central role of the propensity score in observational studies for causal effect. *Biometrika 70*(1), 41–55.

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

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software 42*(7), 1–52.

Steverman, B. and D. Bogoslaw (2008, October 18). The financial crisis blame game. *Bloomberg Business*.

Taylor, W. (2011-2012, Fall). Proving racial discrimination and monitoring fair labor compliance: The missing data problem in nonmortgage credit. *Review of Banking and Financial Laws 31*, 199–264.

van den Brand, J. (2015, June 8, 2015). Freedom at last from discriminatory lending: A fairer future with technology. *Huffington Post*.

Yohai, V., W. Stahel, and R. Zamar (1991). A procedures for robust estimation and inference in linear regression. In W. Stahel and S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics, Part II.* New York: Springer-Verlag.