Title
Empirical Analysis of Twenty-First Century US Lending Markets and Wealth Inequality

Author
Liu, Rongchen

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

EMPIRICAL ANALYSIS OF TWENTY-FIRST CENTURY US LENDING MARKETS AND WEALTH INEQUALITY

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Rongchen Liu

March 2022

The Dissertation of Rongchen Liu
is approved:

Professor Ajay Shenoy, Chair

Professor Laura Giuliano

Professor Justin Marion

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Empirical Analysis of Twenty-First Century US Lending Markets and Wealth Inequality

by

Rongchen Liu

This dissertation studies issues related to public economics, lending markets, real estate, and wealth inequality. The first chapter examines the distortionary effects of federal mortgage repurchasing on lending patterns. Fannie Mae and Freddie Mac are restricted by law to purchasing loans with origination balances below a county-specific "conforming loan limit" (CLL). I examine the behavior of borrowers near the temporarily increased CLLs after the global financial crisis to understand the impact of mortgage repurchasing on lending patterns. I find a sharp bunching of the fraction of loans originated at the loan limit. Borrowers with bad credit histories disproportionately select into the program by offering a large enough down payment to ensure the loan size falls just below the cutoff. The default rate over the medium run of mortgages barely below the CLL is 2.17 percent higher than those barely above the CLL. This impact is largely driven by buyers with a previous home loan. Finally, to understand the distributional impacts of the federal home loan program, I examine differential sorting across the CLL by demographic and income characteristics. I find no evidence that racial minorities or people from "poorer zip codes" comprise a disproportionate share of those who manipulate their loan size to take advantage of the program.

The second chapter provides a test between a negative and positive selection model at a micro-level. The Stiglitz-Weiss model and the de Meza-Webb model make opposing predictions of the correlation between interest rate and default in the credit market. I employ loan-level data from a peer-to-peer online lending marketplace, Prosper, to investigate whether lowering interest rates improves or worsens the mix of applicants and repayment. My empirical result is generally consistent with the prediction of the de Meza-Webb model. I find that even after controlling for all observable characteristics, the pool of borrowers is still affected by some selections on unobservable information. The default rate statistically significantly increases as the interest rate drops for lower-rated borrowers.

The third chapter is a joint project with Nirvikar Singh and Anirban Sanyal. This chapter analyzes and quantifies how differences in the wealth levels of Black and White Americans relate to socioeconomic characteristics, including education, occupation, asset portfolio structures, inheritance and financial literacy, using data from the 2016 Survey of Consumer Finances. Some combination of inheritance, education, and occupation is significantly related to differences in wealth levels across races. However, education, homeownership, business ownership, and financial literacy are not, by themselves, pathways even to reducing wealth gaps, let alone eliminating them. Much of the wealth gap is related to unmeasured structural or systemic factors, rather than measured characteristics: this is estimated by a decomposition of overall wealth differences into those associated with characteristics or endowments and those with differential impacts across groups. Some of the empirical approaches in the estimates are relatively novel in the context of quantifying individual and systemic contributors to the racial wealth gap. Additionally, quantile regressions, which allow for different impacts of characteristics at different portions of the wealth distribution, enable some inferences about the role of class vs. race. The results reinforce the view that race matters for the wealth gap even after accounting for class.

Take pride in how far you've come.

Have faith in how far you can go.

But don't forget to enjoy the journey.

–Michael Josephson

Acknowledgments

I would like to acknowledge all the support, guidance, and friendship I have received over the past six years in Santa Cruz.

I would like to thank my main advisor Ajay Shenoy for his countless hours of advice on my research. Ajay has always been patient and willing to listen to my thoughts. He has always encouraged me to think behind the scenes, so that rather than stopping when I received an empirical result, I would continue to think about questions such as how the result helps people and society and why people should care about this result. I feel lucky to have had Ajay as my main advisor, and his motivation, enthusiasm, and dedication will continue influencing me in the future.

I would like to thank my dissertation committee members Laura Giuliano and Justin Marion. My communication with them always produced incredible insights. This dissertation would not have been possible without their invaluable feedback and input.

I would like to thank my co-authors Nirvikar Singh and Anirban Sanyal. It has been such a great experience to collaborate with them on two research projects, which helped extend my research interests.

Thanks are due to all of the faculty members in my department, as well as my colleagues and friends in Santa Cruz for all of their support and encouragement.

Finally, I would like to thank my parents and grandparents for their unconditional love and for everything they have done for me.

Chapter 1

The Distortionary Effects of Federal Mortgage Repurchasing on Lending Patterns: Evidence from Bunching at the Conforming Loan Limit

1.1 Introduction

Housing is the most important asset of the typical U.S. household. The federal government uses several policies to subsidize homeownership, especially in the wake of the global financial crisis. An important example is Fannie Mae and Freddie Mac[1] buying single-family mortgages with origination balances below a specific amount, known

[1] Fannie Mae and Freddie Mac are federally backed home mortgage companies created by the United States Congress. They don't originate mortgages. They buy and guarantee mortgages issued through lenders in the secondary mortgage market.

as the "conforming loan limit."[2] This program helps enhance the flow of credit to mortgage lenders and keeps eligible home loans cheaper for those purchasing a house. But this program separates the mortgage originator from the default risk bearer and creates a potential moral hazard problem in lenders' screening incentives.

I use thirty-nine county-level temporarily increased CLLs in high-priced California counties in 2008 to examine the impact of the consequences of the government-induced distortions on lending patterns. The two mortgage government-sponsored enterprises (GSEs)[3] purchase conforming loans that meet requirements from lenders. These loans are guaranteed by Fannie Mae and Freddie Mac, meaning they will make investors whole if the borrower goes into default. From the perspective of lenders, there is no downside to issuing a conforming loan, since these loans can be easily bundled and sold in the secondary market. Jumbo loans are not backed by Fannie Mae and Freddie Mac. If lenders keep a jumbo loan on the balance sheet, they bear both the interest rate and the default risk. Many lenders frown upon borrowers with a bad credit history who apply for jumbo loans because they believe that borrowers who experienced defaults in the past are highly likely to default again. But lenders are not supposed to refuse a conforming loan application by risky borrowers because lenders are likely to lose nothing when there is a default. This will distort a series of borrowers to bunch at the CLL. In other words, borrowers who are not allowed to take jumbo loans above the CLL will apply for the largest conforming loan instead.

[2]A conventional loan with a dollar amount no larger than the CLL is referred to as a conforming loan. When the CLL does not cover the loan amount, the loan is referred to as a jumbo loan.

[3]Freddie Mac and Fannie Mae

I test for systematic differences in the fraction of loans just below and above the new CLL by using public records of property transactions in high-priced California counties during the period 2005-2011. I find that the distribution of loan sizes varies systematically around the new CLL. Homebuyers who would otherwise take out a jumbo loan above the CLL now, because of this program, bunch at or slightly below the loan limit. I implement an event study analysis to examine the evolution of origination of loans near the new CLL. I observe an upward jump of the fraction of loans just below the new CLL, which continues to be substantial and significant in the post-treatment period. This analysis suggests that borrowers are maneuvering to get the largest sized loan possible through the program.

Next, I test what margins are manipulated by homebuyers taking out a loan under the new CLL. I plot the average loan to value (LTV) ratio against loan amount relative to the new CLL in each county to identify how leveraged people are around the threshold in the post-treatment period. I observe that borrowers just below the new CLL have a lower LTV ratio than those just above. The LTV ratio is one of the most important factors in determining the level of the interest rate charged to a loan. Borrowers with lower LTV ratios will be offered the lowest interest rates available. I investigate the consequences of distortions in the LTV ratio for interest rates. On average, the interest rates on conforming loans are approximately forty basis points lower than those on jumbo loans. Not surprisingly, I observe a discrete reduction of interest rates on the largest conforming loans, which is explained entirely by the distortion in how leveraged people are at the cutoff. In other words, the lowest interest rate of all is charged to homebuyers who manipulate the LTV ratio to get a loan just under the new CLL.

I then investigate who is bunching at the new CLL. To measure borrowers' creditworthiness, I link seventy-three million house purchase records and foreclosures to properties and buyers' names to identify buyers who had previously defaulted on a mortgage. I employ a difference-in-difference (DID) strategy to examine the difference in the share of loans taken out by buyers with past defaults near the new CLL. I find that borrowers with previous defaults are disproportionately more likely to jump over to the left side of the new CLL and line up just below. I also examine the negative preserved impact caused by the manipulation. I follow the performance of loans just below and above the new CLL for ten years, and observe that the subsequent default rate of conforming loans originated at the new CLL in the post-treatment period is 2.17 percent higher than that of jumbo loans barely above the new CLL. I classify buyers into first-time versus repeated by examining whether they own the property listed in their previous address. This helps me identify that the impact on the subsequent performance of loans barely below the new CLL is mainly driven by the behaviors of homebuyers with a previous mortgage.

Finally, I examine borrowers' income and demographics to shed light on the distributional consequences of federal mortgage repurchasing. Specifically, I investigate details, such as neighborhood characteristics, income, and race, of those whose loans constitute this bunching at the new CLL. Based on zip code, I calculate and link average median home value and average median household income to borrowers' previous addresses. I find that the bunching at the new CLL is driven by homebuyers who previously lived in high-priced or high-income zip codes. People from poorer zip codes are not the ones who can manipulate the loan size to disproportionately select into the program. Those findings are consistent with a preference for expensive properties among the buyers who bunch at the new CLL. Based on the previous neighborhood characteristics, borrowers who bunch at the threshold look similar to those who take out jumbo loans, but the bunchers are constrained by their bad credit history to take out a conforming loan. I use machine learning to train a Long Short Term Memory (LSTM) model to predict the race and ethnicity of homebuyers based on their full names, then I examine if the program has any differential impact by race. I find no evidence that people from any particular race group are statistically significantly overrepresented in the pool of borrowers bunching at the new CLL.

This paper connects several strands of existing literature. This paper contributes to literature analyzing the causes of the subprime mortgage crisis (Demyanyk et al., 2009; Gerardi et al., 2007; Mayer et al., 2009). It speaks to literature that discusses the connection between mortgage securitization and defaults (Bubb and Kaufman, 2009; Justiniano et al. 2015; Keys et al., 2010, Mian and Sufi, 2009; Piskorski et al., 2010; Rajan et al., 2015). This paper provides empirical evidence on how asymmetric information distorts contract choices (Brueckner, 2000; Harrison et al., 2004; Thompson, 2010) in the context of the residential real estate market. It also contributes to literature on how credit supply and interest rates affect borrowing (Adelino et al. 2012; Banerjee et al. 2015; De Meza and Webb, 1987; Karlan and Zinman 2010; Karlan and Zinman

2009; Stiglitz and Weiss, 1981). There is growing literature using the conforming loan limits to examine related topics. For example, Adelino et al. (2012) study the causal effect of easier access to credit on house prices, and they find an increase in the value of houses that are eligible for financing with a conforming loan. DeFusco and Paciorek (2017) investigate conventional loans that originated before the global financial crisis and estimate the interest rate elasticity of mortgage demand by measuring the degree of bunching in response to a discrete change in interest rates at the CLL. Consistent with DeFusco and Paciorek (2017), I also find a sharp spike in the fraction of loans originated in the bin barely below the loan limit and simultaneously a sizable region of missing mass above the limit. But I take a further step to investigate the characteristics of the buyers who are taking advantage of the program to buy big houses. I also look into what is the impact of the bunching at the loan limit.

The rest of the paper is structured as follows: Section 1.2 discusses policies within this program in detail, the strategic incentives of lenders and borrowers, and summarizes the expected outcome from the interactions of these agents. Section 1.3 discusses the data sets used in this paper. Section 1.4 presents the identification strategy and empirical results. Section 1.5 concludes.

1.2 The Program and Predicted Impacts

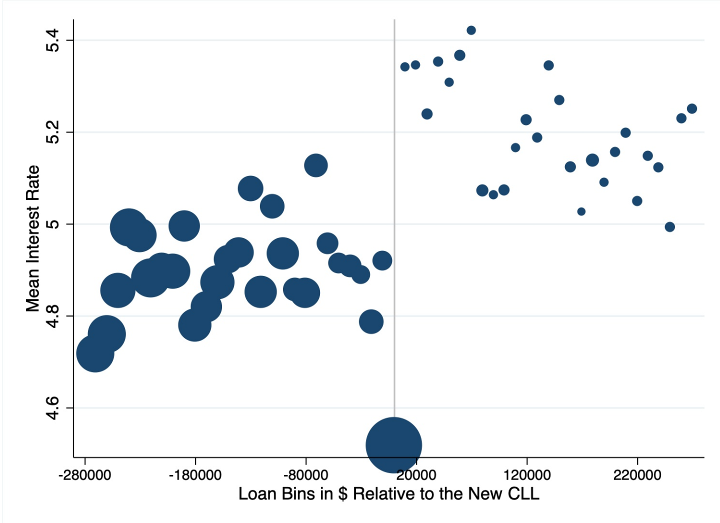
In this section, I describe the context of this study and the policies within the federal mortgage repurchasing program. I also discuss the strategic incentives of lendersand homebuyers resulting from mortgage repurchasing under the new CLL and predict the equilibrium outcome from the interactions of these participants.

1.2.1 Background

Individual mortgage balances are at their all time highest, and approximately 44 percent of U.S. consumers had a mortgage by 2020. The conventional loan dominates the U.S. mortgage market and it is difficult to get. Borrowers are required to have a minimum credit score of 620 and a maximum debt-to-income ratio of about 43 percent to qualify, which is the highest requirement of all mortgage products. Additionally, if borrowers make a down payment that is less than one-fifth of the purchase price of the property, or in other words, the mortgage's LTV ratio is greater than 80%, they have to pay for private mortgage insurance (PMI), which is equivalent to an increase in the interest rate. The average range of PMI premium rates is from 0.58 percent to 1.86 percent of the original amount of the loan. Freddie Mac estimates most borrowers will pay $30 to $70 per month in PMI premiums for every $100,000 borrowed.

To free up liquidity to lend more mortgages, Freddie Mac and Fannie Mae purchase conventional loans with origination balances below the conforming loan limit. The conforming loan limit is set each year by the Federal Housing Finance Agency (FHFA) based on the October-to-October changes in median home price within a metropolitan statistical area (MSA). Loans below this limit are known as conforming loans, while loans above this limit are jumbo loans. The interest rates on jumbo loans are higher than those on conforming loans. By examining the principle data, Figure 1.1 plots in-

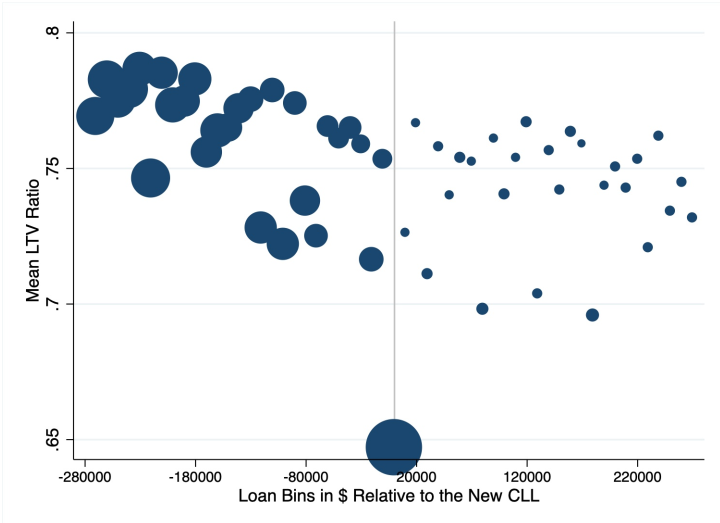
Figure 1.1: Average Interest Rate by Loan Amount



Notes: This figure plots the average interest rate as a function of loan amount relative to the new conforming loan limit for fixed-rate mortgages originated between April 1, 2008, and September 30, 2011. Each blue dot represents the mean interest rate within a $10,000 bin relative to the new conforming loan limit of each county. The gray line is where the new CLL locates.

terest rate against the difference between the loan amount and conforming loan limit for mortgages originated between April 1, 2008 and September 30, 2011. This figure helps identify the general magnitude of the interest rate differential to some extent. As can be seen, there is a clear discontinuity at the cutoff, and on average the interest rates on jumbo loans are approximately 40 basis points higher than those on conforming loans. What is interesting is that the lowest interest rate of all is charged to borrowers lining up in the last bin below the new CLL. To investigate why there is a discrete drop of average interest rate on the largest size of conforming loans, I plot a similar figure of LTV ratio. LTV ratio is one of the most important factors in determining the level of the interest rate charged to a loan. Borrowers with lower LTV ratios will be offered the lowest interest rates available. Figure 1.2, plots the average LTV ratio as a function of loan amount relative to the loan size limit using the same sample as Figure 1.1. It seems that the lowest interest rate at the new CLL is entirely explained by a distortion in how leveraged people are at the cutoff. People in the bin barely below the cutoff are the ones who can offer a larger down payment to borrow under the new CLL and benefit from the lowest interest rate.

Figure 1.2: Average LTV Ratio by Loan Amount



Notes: This figure plots the average LTV ratio as a function of loan amount relative to the new conforming loan limit for fixed-rate mortgages originated between April 1, 2008, and September 30, 2011. Each blue dot represents the mean LTV ratio within a $10,000 bin relative to the new conforming loan limit of each county. The gray line is where the new CLL locates.

For homebuyers, conforming loans are advantageous due to their low-interest rates. Buyers, however, still strongly desire the jumbo loan size because they are unable or unwilling to put a large enough down payment to buy an expensive property, they are less sensitive to an increase in the interest rate, or they are unaware of the lower-cost financing program. For lenders, conforming loans can easily be bundled and sold on the secondary market, which enhances the flow of credit and allows them to issue more loans to make profits. Jumbo loans are considered riskier loans since lenders bear both the interest rate and default risk, and it takes more time to liquidate a jumbo loan in the event the house forecloses.

A combination of rising home prices, loose lending practices, and an increase in subprime mortgages pushed up real estate prices to an unsustainable level. Average U.S. housing prices peaked in mid-2006, but foreclosures and defaults then crashed the housing market, and housing prices declined by over 20 percent during the global financial crisis. Over 10 percent of prime loans and 40 percent of subprime loans that originated during the housing boom period ended up in default within two years. The Housing and Economic Recovery Act of 2008 (HERA 2008) was designed to build confidence in Fannie Mae and Freddie Mac and inject capital into mortgage funding. A permanent formula of setting loan limits and temporarily increased loan limits for certain areas was published in this act. For a single-family property, the maximum temporary loan limit was 125 percent of the median house price for the highest-priced county in the property's MSA but no greater than $729,750.[4] This program became effective on April 1, 2008, and the baseline conforming loan limit of $417,000 remains in place.[5] The ceiling of loan limits for high-cost areas was reduced to $625,500 on October 1, 2011.

In this paper, I focus on housing and mortgage markets in California to study how the temporary increase in the CLL distorted people's behaviors. There are 58 counties in California, and 39 of them were assigned with increased conforming loan limits under this program. In Table 1.1, I list 19 low-cost counties where loan limits

[4]The exceptions are properties in Alaska, Hawaii, Guan, and the Virgin Islands.

[5]The nationwide $417,000 conforming loan limit started from October 1, 2005.

Table 1.1: High- vs. Low-Cost Counties in California

LCAs	HCAs	HCAs with the Max CLL
Butte	Alpine	Alameda
Colusa	Amador	Contra Costa
Del Norte	Calaveras	Los Angeles
Fresno	El Dorado	Marin
Glenn	Inyo	Monterey
Humboldt	Madera	Napa
Imperial	Mendocino	Orange
Kern	Merced	San Benito
Kings	Mono	San Francisco
Lake	Nevada	San Mateo
Lassen	Placer	Santa Barbara
Mariposa	Riverside	Santa Clara
Modoc	Sacramento	Santa Cruz
Plumas	San Bernardino	Ventura
Sierra	San Diego	
Siskiyou	San Joaquin	
Tehama	San Luis Obispo	
Trinity	Shasta	
Tulare	Solano	
	Sonoma	
	Stanislaus	
	Sutter	
	Tuolumne	
	Yolo	
	Yuba	

Notes: The first column shows 19 low-cost counties where the conforming loan limit remained at $417,000 since October 1, 2005. The second column shows 25 high-cost counties that were assigned with loan limits above $417,000 but below the maximum possible amount $729,750. The third column shows 14 high-cost counties that were assigned the highest conforming loan limit under the program.

remained unchanged, 14 extremely high-cost counties that were assigned with the maximum conforming loan limit, and 25 high-cost counties that were assigned with loan limits between the baseline level and the ceiling under this program. In this paper, I will focus on the 39 counties with new increased CLL.

1.2.2 Strategic Incentives of Lenders and Borrowers

As briefly mentioned in the first section, the two mortgage GSEs will purchase conforming loans from banks, which implies the originator of conforming loans is not the default risk bearer of them. Keys et al. (2010) point out that this kind of practice will adversely affect the screening incentives of lenders and cause much higher rates of defaults as compared to a similar risk profile group without the ease of securitization. When a borrower applies for a home loan, the lender has to take one of three actions: reject the borrower because she is not eligible to apply for this loan, screen the borrower carefully and then make a decision of approval or disapproval, or for some reason, approve a loan for the borrower without extra screening beyond what is in the credit request. Conforming loans are eligible to be purchased by GSEs and this practice will take the default risk off a lender's balance sheet. There is no incentive for lenders to investigate those to whom they are issuing conforming loans. Therefore, safe borrowers and risky borrowers are equally likely to be approved for a conforming loan. However, jumbo loans cannot be purchased by Fannie Mae and Freddie Mac. Because of this, borrowers have to undergo a lengthy process of investigation when they apply for a jumbo loan. This investigation makes it difficult for lower-quality borrowers to be approved for a jumbo loan because lenders know that keeping a bad-quality loan on the balance sheet is risky. It will hurt the lender if the borrower defaults on the loan, and it will be even worse if the borrower cannot provide valuable collateral. Lenders are supposed to strategically push a series of borrowers with bad risks to the left side of the conforming loan limit. This takes risk off while keeping loan volume high, which is profitable for lenders.

Borrowers taking out loans near the conforming loan limit must search for expensive properties within each county. Safe borrowers are equally likely to be approved for both conforming and jumbo loans, so they are more likely to take out a loan with the size they need. But it's difficult for lower-quality borrowers buying an expensive property to take out a jumbo loan. The margin they can manipulate to ensure they stay in the market is the size of the down-payment. To purchase an expensive home, they should borrow as much as they are allowed, which is the largest size of conforming loans and offer a large enough down-payment to ensure the loan size falls just below the CLL. While some bad borrowers are constrained by disposable income, they will either leave the market or look for smaller homes that are cheap enough to be purchased by a combination of a small share of cash and a conforming loan.

1.2.3 The Empirical Predictions

Fannie Mae and Freddie Mac purchase conforming loans to eliminate the risk premium and make loans cheaper. Therefore, I expect to see an increase in the share of loans that were previously not conforming after the implementation of the act. According to the discussion about strategic incentives of both the supply and demand side of the mortgage market, I expect to see a sharp spike of borrowers distorted by lenders from the right side taking out the largest size of conforming loans. And meanwhile, there should be a significant missing mass of borrowers above the threshold.

At equilibrium, borrowers with bad credit history having either a liquid asset or disposable income are overrepresented in the pool of people distorted by the lenders to bunch at the jumbo-conforming loan limit. In other words, high-risk borrowers are disproportionately less likely to show up just above the loan limit. Also, the leverage ratio of loans right at the cutoff should be significantly lower than others because borrowers are sorting across the loan limit by manipulating the size of the down-payment.

1.3 The Data

I use five data sets in this study, and two of them are obtained from Zillow. The primary data set Zillow's Transaction and Assessment Database (ZTRAX) contains more than 400 million detailed public records across U.S. counties, more than 20 years of deed transfers, mortgages, foreclosures, and also property characteristics as well as prior valuations for about 150 million parcels nationwide. Because Zillow has such expensive information, it is the perfect platform to use to examine the specific area of California counties. For each transaction in this data, I can observe the time of the purchase, the sales price of the property, the loan amount (if any), the type of mortgage product, the address of the property, the type of property, the buyers' full name, the buyers' previous address, etc. This data is continually growing, and the version I obtained is ZTRAX, 2018. In this version, I can observe all the information that I listed above from the year 1993 to 2018. Linking buyers' names to every piece of historical record associated with a property helps me to identify buyers who bought properties that were foreclosed and then focus on the behaviors of buyers who experienced property foreclosure before 2008. Also, I can follow buyers who took out a mortgage through this program to purchase a home but ended up defaulting on the loan by the end of my sample. I will mainly discuss these outcomes in the empirical analysis section, and the Zillow data helps me implement related investigations. Additionally, both buyers' and sellers' names and mailing addresses[6] are provided in records of deed transfers. To distinguish the impact on first-time vs. repeated homebuyers, buyers' previous addresses become an important piece of information. I check if homebuyers had a mortgage or a deed transfer associated with their previous address on record and define people who did not own a property before the program as first-time homebuyers.

The second data set, Zillow Home Value Index (ZHVI), provides information about median home value for all single-family residences at the zip code level. I calculate the average median home value for all zip codes in California and link it to buyers' previous addresses in ZTRAX. This allows me to estimate whether a buyer came from a high or low-priced zip code. Also, I use median household income by zip code from the American Community Survey Data 2006-2010 to do a robustness test. Both tests shed some light on the income level of a buyer.

The values of conforming loan limits before and after HERA 2008 are available from FHFA. Loan limits for loans issued to homebuyers in every California county in each calendar year can be observed in this data set. Since October 2005, the nationwide conforming loan limit was \$417,000, while limits for high-cost counties increased to \$729,750 in 2008 because of HERA 2008, and remained so until September 2011.

The last data set is the Florida Voter Registration Data (February 2017). This data provides the full name as well as the race and ethnicity of each voter. Using this FL voter registration data, I estimate an LSTM model to learn the relationship between the sequence of characters in a full name and race. Then I apply this model to buyers'

[6]For buyers in CA, about 3.498 percent of them did not provide a mailing address, and those people are excluded from this study.

full names in the California counties found in the Zillow data to predict buyers' race and ethnicity and investigate if individuals from a particular racial group are overrepresented in the pool of people bunching just below the new CLL. There are $13,044,043$ pieces of records in the FL data, 1.95 percent of them are Asian or Pacific Islanders, 16.71 percent are Hispanic, 14.21 percent are Non-Hispanic Black and 67.14 percent are Non-Hispanic White. However, no race or ethnic group constitutes a majority of California's population, and 15% of California residents are Asian or Pacific Islander, which is much larger than in Florida. Fortunately, the model can predict race from Asian names almost perfectly. Details of the model and outcomes are in Section 1.4.

1.4 The Empirical Results

Table 1.2 provides summary statistics of the sample with emphasis on loan amounts, the sales price of homes, and down payment ratios. This analysis uses observations from 39 high-priced counties in CA across the period October 1, 2005, to September 30, 2011. The number of loans declines after the financial crisis, and the market has witnessed a reduction in housing prices. There is also an increasing trend in the down payment ratio in both groups. Instead of taking out large loans, buyers tend to put down bigger down payments. I will explore the dimension of the down payment ratio in detail later in this section.

To visualize how borrowers react to the temporary increase in the CLL, I start with a McCrary Test for loans taken out by borrowers in all the 39 high-cost counties

Table 1.2: Summary Statistics

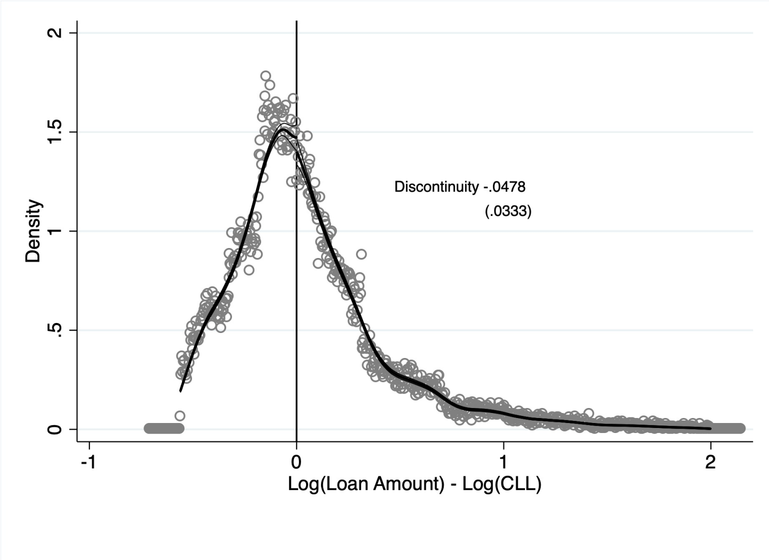
	HCAs with Max CLL		HCAs	
	Pre	Post	Pre	Post
Loan	560,499.4	450,352.7	367,610.7	264,572.8
	(333,053.9)	(346,634.1)	(198,254.9)	(184,015.4)
Price	781,732.5	668,100.4	493,057.4	368,067.9
	(524,773.1)	(566,838.4)	(291,470.4)	(293,564.9)
DP	.2559	.2841	.2347	.2495
	(.1442)	(.1519)	(.1516)	(.1465)
Obs	458,391	415,213	507,798	327,537

Notes: They are loan-level observations from 39 counties in CA. 14 counties are assigned with the max CLL in the post-treatment period and 25 counties are assigned with CLL between the baseline level and the ceiling. There are 30 months in the pre-treatment period and 42 months in the post-treatment period. Standard deviations are in parentheses.

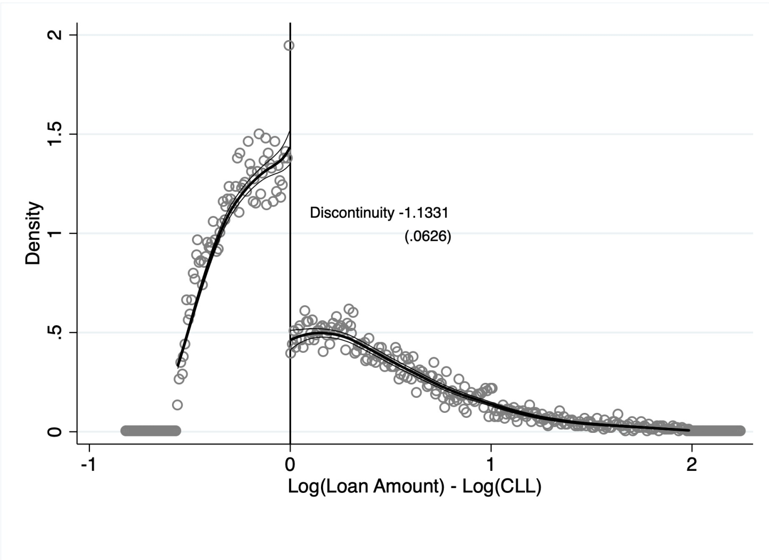
in California in the pre- vs. post-treatment period. CLL in both figure 1.3a and 1.3b refers to the increased loan limit for each county after the housing recovery act, and loans with dollar amounts below the baseline CLL are excluded from the test. As can be seen, in the post-treatment period, the estimated curve is strongly discontinuous at the loan limit in each county. Buyers select into the program by manipulating the size of loans, especially, many buyers taking out loans with size just below the cutoff, but fewer buyers take out loans with dollar amounts above the loan limit. The program does create a notch by making it relatively costly to take out a mortgage above a certain level. However, in the pre-treatment period, the estimated curve is smooth around the cutoff, and I do not observe any sorting behavior around the loan limit.

I restrict my attention to mortgages with sizes near the new CLL in each county and group loans into bins. Again, loans with dollar amounts lower than the baseline CLL are excluded from this study. I create 28 bins and the length of each bin is $20,000. For example, transactions that fall into Bin 0 in both Figure 1.4a and 1.4b

Figure 1.3: Density Test



(a) Density Test of Loans in the Pre-period



(b) Density Test of Loans in the Post-period

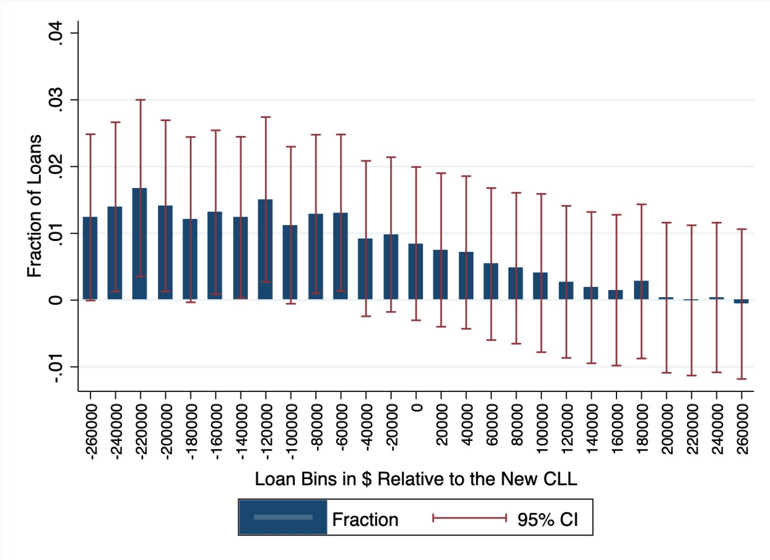
Notes: This is a McCrary Test for observations in all the 39 high-priced counties in the pre- vs. post-treatment period. CLL in both figures refers to the increased loan limit for each county after the program.

are those with loan amounts exactly equal to the new loan limit in each county or at most $19,999 lower than the new CLL. The horizontal axis in both Figure 1.4a and

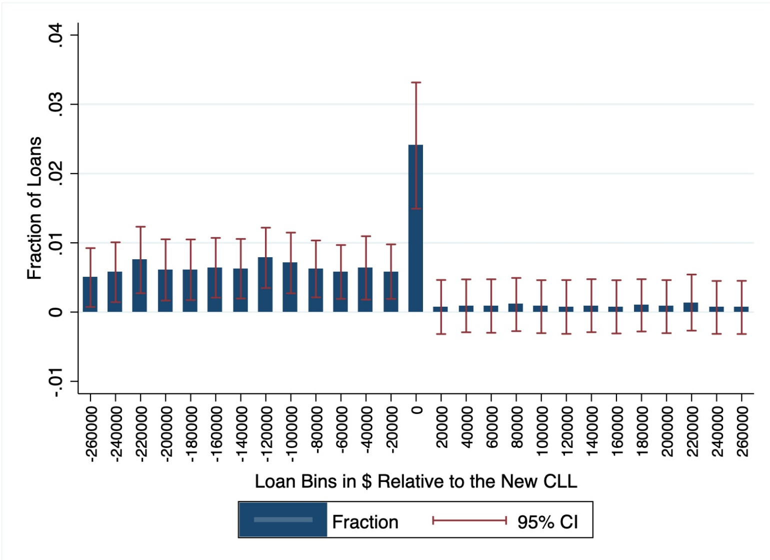
1.4b shows loan bins in dollar amount relative to the increased new loan limit in each county. The vertical axis shows the fraction of loans at the bin-year-county level. There is no observable discontinuity in Figure 1.4a, but the frequency distribution of loans in Figure 1.4b can identify the behavioral responses to the program. As can be seen, borrowers are bunching right below the new loan limit, while fewer borrowers tend to take out loans above the threshold. And to sum up, this program does have an impact on buyers' choice of financing.

Next, I design an event study to plot time paths for loan originations in three bins in the middle to better visualize how borrowers sort around the new CLL. Figure 1.5 shows event studies for the change of fraction of loans in Bin 0 vs. Bin 20000 as well as Bin 0 vs. Bin -20000. The fraction of loans is calculated as the number of loans at the bin-year-county level divided by the average number of loans of the pre-treatment period in that county. The "-1" on the horizontal axis refers to one year before the effective date of the program, and I am making a comparison of the fraction of loans in each year to that "-1" year. In both Figure 1.5a and 1.5b, the navy line represents the fraction of loans in Bin 0. The green line in Figure 1.5a represents the fraction of loans in Bin 20000, while the one in Figure 1.5b represents the fraction of loans in Bin -20000. In both Figure 1.5a and 1.5b, I can observe a similar pre-treatment trend, and then there is an upward jump of the fraction of loans in Bin 0, which continues to be significant four years out. People who take out loans from Bin 0 are the ones disproportionately likely to be altering their behaviors just to qualify for this program, and people in Bin 0 are more likely to reflect the characteristics of the buyers who are

Figure 1.4: Frequency Distribution



(a) Frequency Distribution of Loans in the Pre-period



(b) Frequency Distribution of Loans in the Post-period

Notes: Bin width in both figures is \$20,000. Bin 0 is where the new CLL locates. Bin -280000 is omitted.

willing or able to alter their behaviors to get a conforming loan. Moreover, it seems that some people are not only selecting into the program but also to get the largest possible size of loans through the program. From this event study, it's easy to see that this program was distorting the lending market. I will continue to investigate the impact of this distortion and also identify the characteristics of buyers who are distorting their behaviors.

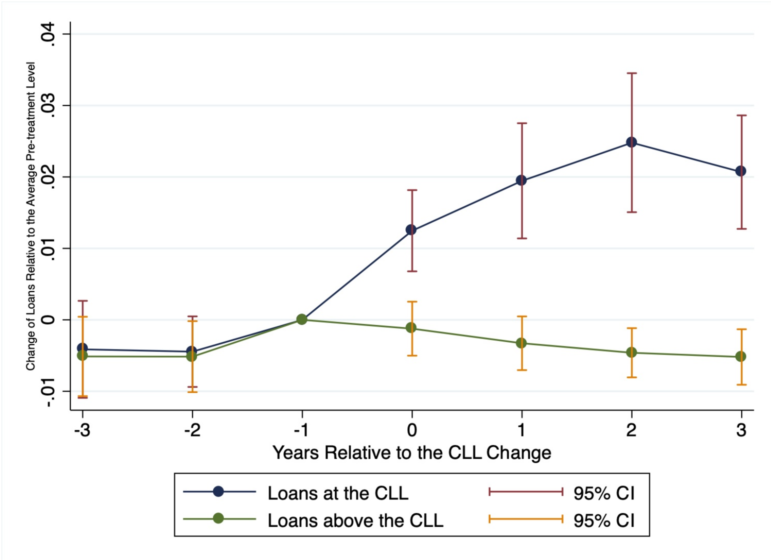
1.4.1 The Main Regression

I restrict the attention to observations near the new loan limit and use the following difference-in-difference specification to analyze different outcomes. There are several advantages of focusing on loans that fall into bins close to the threshold. First, this approach focuses on people's responses locally around loan sizes where CLL are likely to play an important role. Also, a localized approach gains precision by filtering out other possible shocks to borrowing and lending in somewhere else of the loan distribution. The unit of observation in regression equation (1.1) is a mortgage transaction.

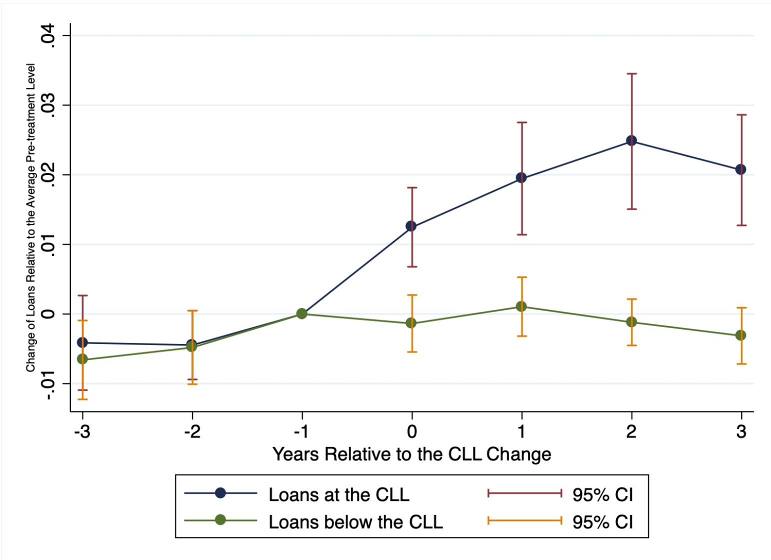
$$y_{it} = \alpha_0 + \alpha_1 UnderNewCLL_i * Post_t + \alpha_2 UnderNewCLL_i + \alpha_3 Post_t + \mu_t + \lambda_c + \varepsilon_{it} \quad (1.1)$$

The dummy variable $UnderNewCLL_i$ equals to 1 if "i" is a loan in bins covered under the program, and equals to 0 if the observation comes from bins above the new CLL. $Post_t$ is a time dummy, and it is 1 if "t" is a time after the effective date of the program. μ_t is the year fixed effect which allows me to control for state-wide evolution. λ_c is the county fixed effect which allows me to control for county-specific factors in the distribution of loan sizes. ε_{it} is the error term. I bootstrap cluster the standard errors

Figure 1.5: Event Study: Change of Loan Counts Relative to the Average Pre-treatment Level



(a) Bin 0 vs Bin 20000



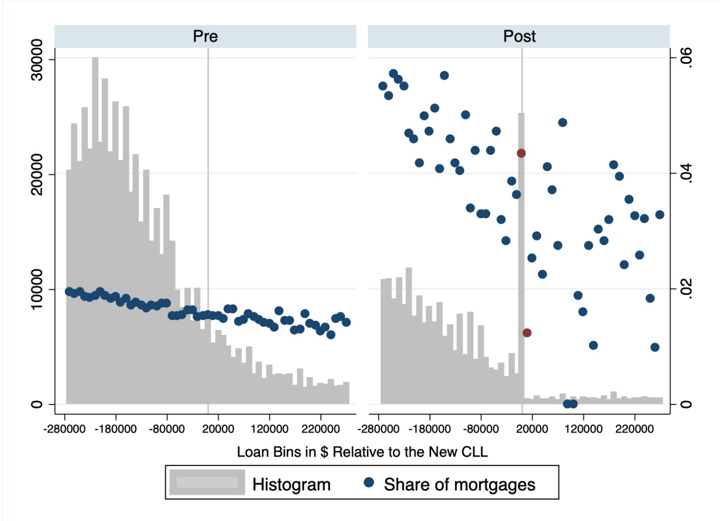
(b) Bin 0 vs Bin - 20000

Notes: "-1" year refers to one year before the effective date of the program, and the outcome in each year is compared to the "-1" year.

by county, which is the level at which the policy is assigned.

1.4.2 Selection of Borrowers and Performance of Loans

Figure 1.6: Percentage of Mortgages Taken Out by Home Buyers with Prior Default



Notes: This figure plots the percentage of loans taken out by homebuyers with historical default as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005, and March 31, 2008 (the left panel) vs. between April 1, 2008, and September 30, 2011 (the right panel). Each dot represents the share of loans taken out by buyers with prior default within a $10,000 bin relative to the new conforming loan limit of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

The dependent variable y_{it} refers to different outcomes that I am interested in. First, I examine if borrowers with prior defaults are disproportionately more likely to take advantage of this program. I define buyers with historical default as buyers who defaulted before the program to make sure that they are not buyers who defaulted on mortgages taken out through the program itself. For the first outcome, y_{it} equals 1 if "i" is a transaction made by a buyer with prior defaults. I group observations into bins with a size of $10,000 and plot the share of mortgages taken out by buyers with past default against loan amount relative to the new CLL for mortgages originated in the pre- vs. post-treatment period. Each dot in Figure 1.6 represents the percentage of mortgages taken out by buyers who defaulted before the program in each bin. The gray line is where the new CLL locates. The histogram shows the number of loans in each bin. In the post-treatment period, all the gray bars below the new CLL are much higher especially the one at the cutoff, which is consistent with the figures of the McCrary test. As compared to the left panel, the mean in the post-treatment period is getting larger. On average, the share of loans taken out by buyers with prior defaults is higher after the program. Presumably, in the pre-treatment period, lenders are reluctant to originate jumbo loans to borrowers with bad credit history, because they do not want to take the risk of losing both principal and following installments when risky borrowers default again. While in the post-treatment period, loans below the new CLL are now becoming more liquid, and lenders are willing to approve loans to risky borrowers when they become conforming. More importantly, I find borrowers with historical defaults are disproportionately likely to jump over to the left side of the new CLL and bunch in last bin covered under the program. Those borrowers are the ones who are looking for expensive houses and who also have cash at that time to be able to manipulate the loan size by putting down larger down payments. While other borrowers who just want to purchase a relatively cheaper house have to take out a smaller conforming loan and find a smaller home if they have a bad credit history. This is not a perfect sorting because I observe some borrowers with prior default are taking out jumbo loans even after the implementation of the program. Possibly, lenders find that the property those buyers are purchasing will appreciate, allowing them to take the valuable property as collateral even in the worst case.

Table 1.3: Share of Loans Taken Out by Buyers with Previous Defaults, the LTV Ratio and the Subsequent Performance of Loans

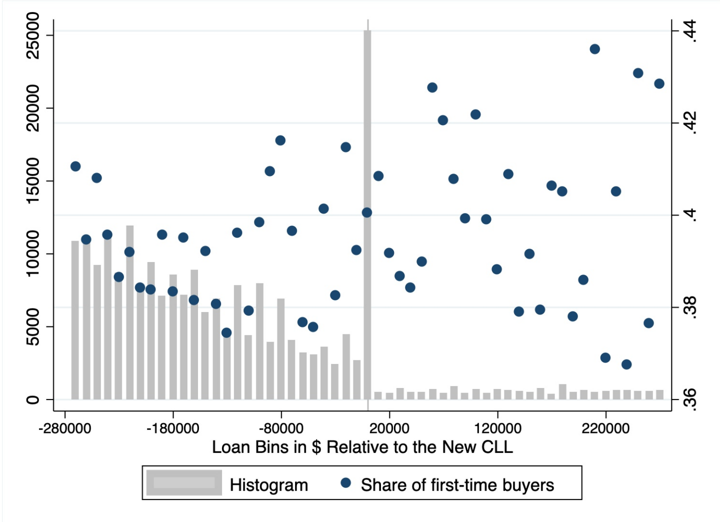
	Past Default	LTV	Repeat and First	Repeat	First
Under New CLL x Post	.0210***	-.0828***	.0217***	.0394***	.0107
	(.0063)	(.0067)	(.0080)	(.0116)	(.0110)
Under New CLL	.0001	.0002	-.0045	-.0196	.0011
	(.0003)	(.0046)	(.0060)	(.0129)	(.0073)
Post	.0104***	-.0361***	-.0720***	-.0737***	-.0664***
	(.0030)	(.0090)	(.0150)	(.0140)	(.0134)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Pre-mean	.0153	.7872	.0681	.0658	.0687

Notes: The dependent variable in the first column is a dummy that equals 1 if a transaction is made by a buyer with prior defaults. The outcome in the second column is the loan to value ratio of a transaction, and it is calculated as the loan size divided by the sales price of a property. The outcome in the third through fifth columns is a dummy that equals 1 if a loan ends up defaulting by 2018. The third column includes both first-time and repeated home buyers, the fourth column includes only repeated home buyers, and the fifth column includes only first-time home buyers. Standard errors are adjusted for bootstrap clusters in counties.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Next, I apply the regression equation (1.1) to observations falling in \$20,000 below or above the loan limit to investigate if people who are taking out the largest possible size of loans covered under the program are statistically significantly different from people who stick with jumbo loans. The regression results are presented in the first column of Table 1.3. This is suggesting that homebuyers who are more likely to take out the biggest possible size of mortgages covered under this program are systematically more likely to be the ones with bad credit history, and this is a big increase as compared to the pre-mean. The second column in Table 1.3 presents how leveraged borrowers are under the program. I conclude that borrowers who are taking out loans with a size just below the new CLL are more likely to reduce the LTV ratio than people who get the smallest possible size of jumbo loans. This is consistent with the discrete reduction at the loan limit in Figure 1.2 and this is also consistent with the prediction in section 1.2. The borrowers who disproportionately bunch at the new CLL are likely to apply for jumbo loans to buy some expensive properties, but lenders are willing to approve a conforming loan for them because of their bad credit history. The only possible way for them to purchase a big house is to take out the largest conforming loan available and pay for the remaining balance in cash.

One thing that needs to be noticed is first-time homebuyers are included in all the analyses discussed above, but they don't have a chance to default before this program because they don't have prior mortgages. Therefore, in the first column of Table 1.3, $y_{it} = 0$ for all the first-time homebuyers, but that does not imply they are safe buyers. They are completely different from repeated buyers who have not defaulted before the

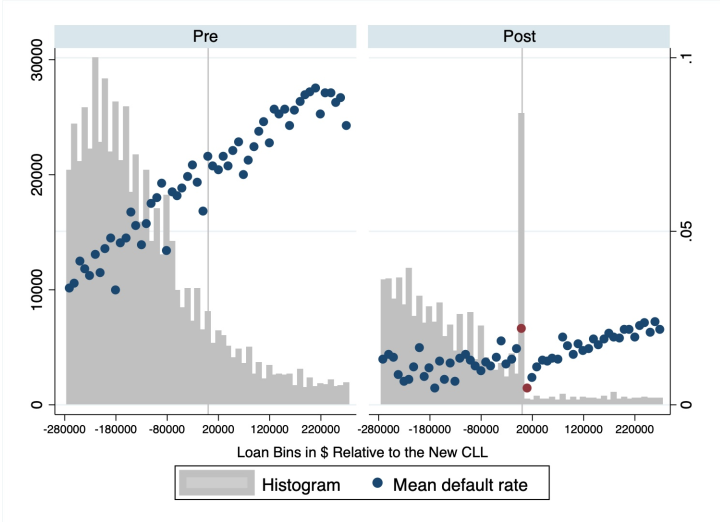
Figure 1.7: Probability of Being a First-time Homebuyer



Notes: This figure plots the percentage of loans taken out by first-time homebuyers as a function of loan amount relative to the new conforming loan limit for mortgages originated between April 1, 2008 and September 30, 2011. Each blue dot represents the share of loans taken out by first-time buyers within a $10,000 bin relative to the new conforming loan limit of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

program in creditworthiness. About 40 percent homebuyers are first-time taking GSE loans in my study period, and their responses to the program play an important role in this study. People might argue that the discontinuity at the threshold in Figure 1.6 can be driven by new homebuyers who do not have a prior default are more likely to take out a jumbo loan in the post-treatment period. To see if that is one possibility, I plot the share of loans taken out by a first-time homebuyer as a function of loan amount relative to the new CLL for observations in the post-treatment period. However, there is no discontinuity at the cutoff in Figure 1.7, which implies that the distortion at the threshold in Figure 1.6 can only be driven by the manipulation of the repeated homebuyers with bad credit history.

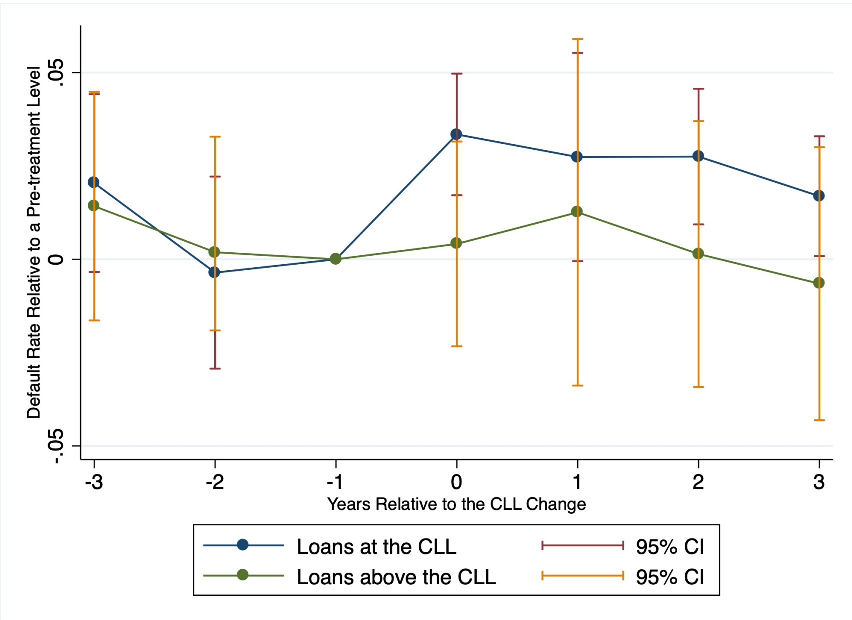
Figure 1.8: The Average Default Rate by Loan Amount



Notes: This figure plots the average default rate as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005, and March 31, 2008 (the left panel) vs. between April 1, 2008 and September 30, 2011 (the right panel). Each dot represents the mean default rate within a $10,000 bin relative to the new conforming loan limit of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

Historical default and LTV ratio are limited perspectives to estimate the creditworthiness of a borrower. The Zillow data does not provide rich information of borrower characteristics such as credit score and income, therefore, I am not able to estimate the creditworthiness of a random borrower from multiple aspects directly. But I can track the performance of loans taken out by a random borrower until the end of my sample. That helps to investigate what is the impact of the disproportionate concentration of borrowers with poor credit history at the loan limit, and also reveals whether borrowers with other unobservable bad risks select into the program by altering behaviors. I plot the average default rate as a function of loan amount relative to the new CLL for mortgages originated in the pre- vs. post-treatment period. In Figure 1.8, loans are grouped into bins with a size of $10,000, and each dot represents the mean default rate within a $10,000 bin relative to the new CLL of each county. Not surprisingly, the average default rate of loans originated in the post-treatment period is significantly lower than that of loans originated before the financial crisis. Before the program, there is no observable discontinuity at the cutoff. However, in the post-treatment period, I do observe a discontinuity at the new CLL and much of the behavior is coming from people shifting to be just below the loan limit. The subsequent default rate of loans with dollar amounts just above the new CLL is lower than that of the largest possible size of conforming loans. It seems like the distortions around the loan limit do have the negative impact of causing a higher subsequent default rate of conforming loans barely

Figure 1.9: Impact on Default Rate over Time



Notes: The sample includes loans $\pm\$20,000$ of the conforming loan limit in each county. "-1" year refers to one year before the effective date of the program, and the outcome in each year is compared to the "-1" year.

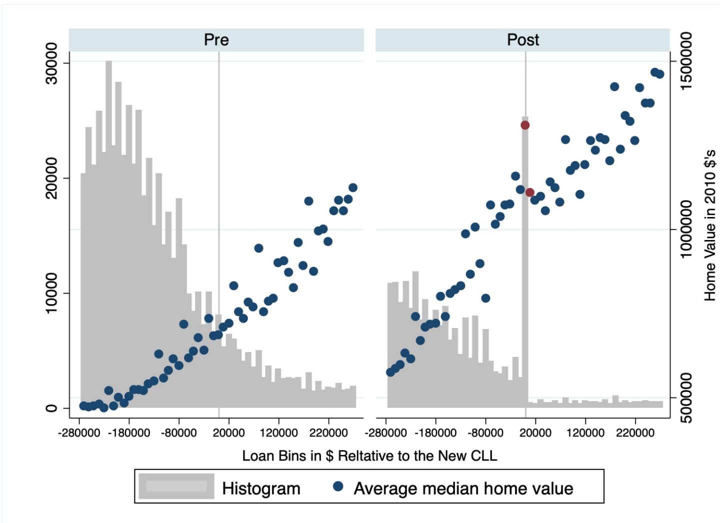
below the loan limit over the medium run. However, the histogram is suggesting at the same time this program does help a lot of people who lost their properties or who do not even own property to purchase a home by offering them reasonably priced loans. There is not enough evidence to prove this program negatively affects social welfare because of issuing too many loans to people with bad risks. Compared with the sharp spike in the number of loans originated in the bin at the loan limit, the increase in the subsequent default rate is a lot smaller.

To zoom in and better visualize the probability of mortgages taken out through the program but subsequently defaulting, I design an event study to plot the evolution of default rates of mortgages within $\pm\$20,000$ of the new CLL originated from 2005 to 2011. Figure 1.9 shows the time paths for the performance of loans with a size in that range. Loans with dollar amounts just above the new CLL are less likely to default over the medium run as compared to loans at or slightly below the new CLL, even though they have an almost identical trend before the program. When I apply regression equation (1.1) to this sample, I find that the default rate of the largest possible size of conforming loans under the program is on average 2.17 percent higher compared to jumbo loans just above the cutoff, and these regression results can be seen in the thrid column of Table 1.3. Although there is a statistically significantly decreasing trend of default rate on all loans in the post-treatment period, the sorting behaviors near the new CLL prevent the default rate of those conforming loans from a considerable reduction. As briefly mentioned before, I separate first-time homebuyers from repeated buyers to see if effects are preserved in each group. In the fourth column of Table 1.3, I identify all homebuyers who did not own any property before the program and exclude them from the regression. Using a sample of only repeated home buyers, I observe that the subsequent default rate of conforming loans barely below the loan limit is on average 3.94 percent higher as compared to jumbo loans barely above the cutoff in the post-treatment period. In the fifth column, I keep only transactions made by first-time homebuyers, but I do not observe a statistically significant impact on the subsequent default rate of loans barely below the CLL vs. those above in the post-treatment period. The overall impact on the subsequent default rate of loans near the new CLL is largely driven by the behaviors of the repeated homebuyers. It seems that borrowers who are taking out the largest size of conforming loans through the program who also have a poor credit history are the ones who end up subsequently defaulting. I keep enlarging the bandwidth and run the same regression for each group of borrowers. Consistent with the findings shown in Figure 1.8 presents, much of the behaviors I observe come from bins just below and above the cutoff. Not strikingly, the impact decreases as the bandwidth gets larger, and the regression results are presented in the Appendix.

1.4.3 Buyer Demographics

In this section, I examine the differential sorting across the new CLL by demographic and income characteristics to understand the distributional impacts of the federal mortgage repurchasing. I start by investigating where the borrowers bunching at the new CLL come from. I calculate the zip code level average median home value based on borrowers' previous addresses and plot it against the loan amount relative to the new CLL for mortgages originating in the pre- vs. post-treatment period. Figure 1.10 clearly shows that the bunching is driven by people who previously lived in high-priced zip codes. For robustness, I also calculated the zip code level average median household income based on borrowers' previous addresses and plot that as a function of the loan amount relative to the new CLL for loans originated in the pre- vs. post-treatment period. I observe a similar pattern in Figure 1.11 and there is also a sharp increase right at the threshold in the post-treatment period. I then plot the share of loans taken out by people from poor zip codes but now buying an expensive home in high-priced counties against the loan size relative to the new CLL. Poor zip codes are defined as zip codes with the average median home value being in the bottom quartile of

Figure 1.10: Average Median Home Value Based on Buyers' Previous Addresses

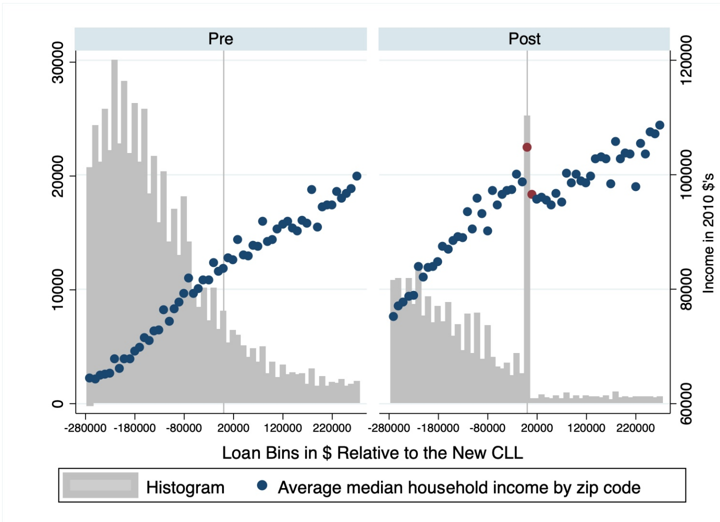


Notes: This figure plots the average median home value as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005 and March 31, 2008 (the left panel) vs. between April 1, 2008 and September 30, 2011 (the right panel). The average median home value is calculated by using zip code level median home values from 2017 to 2020 (deflated in 2010 $'s). Each dot represents the average median home value based on homebuyers' previous addresses within a $10,000 bin relative to the new CLL of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new CLL. The gray line is where the new CLL locates.

California. I observe smooth distribution in Figure 1.12 in both the pre-treatment and post-treatment periods. All three figures mentioned above are consistent with the idea that people who bunch at the threshold after the implementation of the program are likely to have the cash for a larger down payment. People who previously lived in poorer zip codes are no more likely to manipulate loan size to take advantage of the program given that they are less likely to have the cash based on their prior neighborhoods.

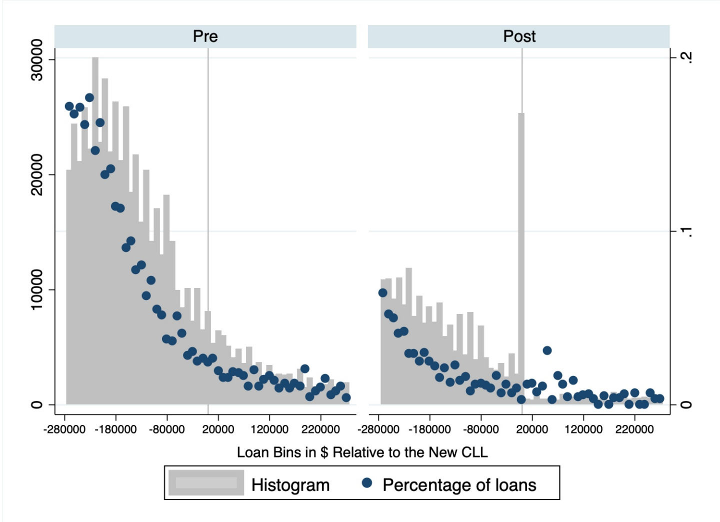
Table 1.4 presents the regression results by applying equation (1.1) to observations just below and above the new CLL. The outcome variable in the first column is the log of the zip code level median home value. The median home value is calculated by using ZHVI data from 2017 to 2020. To make the pre- and post-values more comparable,

Figure 1.11: Average Median Household Income Based on Buyers' Previous Addresses



Notes: This figure plots the average median household income as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005, and March 31, 2008 (the left panel) vs. between April 1, 2008 and September 30, 2011 (the right panel). The average median household is calculated by using zip code level median income from 2006 to 2010 (in 2010 $'s). Each dot represents the average median income based on homebuyers' previous addresses within a $10,000 bin relative to the new CLL of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new CLL. The gray line is where the new CLL locates.

Figure 1.12: Percentage of Loans Taken Out by Buyers from Poor Zip Codes



Notes: This figure plots the share of loans taken out by buyers from poor zip codes in CA as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005 and March 31, 2008 (the left panel) vs. between April 1, 2008 and September 30, 2011 (the right panel). Each blue dot represents the share of loans taken out by buyers from poor zip codes within a $10,000 bin relative to the new conforming loan limit of each county, and poor zip codes are defined as zip codes with the average median home value being in the bottom quartile of California. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

the median home value is measured in the same units, 2010$'s. The outcome variable in the second column is the log of the zip code level median household income. The median household income is coming from the ACS data from 2006-2010 but all in 2010 $'s. Consistent with what can be seen in the figures, the coefficients of the interaction term in both the first and second columns suggest that homebuyers who bunch at the new CLL in the post-treatment period are more likely to be economically advantaged. And the coefficient of the interaction term in the third column implies that borrowers who are distorting their behavior to take out the largest possible size of conforming loans are no more likely to be from the bottom quartile zip codes.

Finally, I investigate if this program has any differential impact by race. Specif-

Table 1.4: Average Median Home Value and Household Income around the New CLL and Share of Buyers from the Poor Zip Codes

	(1)	(2)	(3)
UnderNewCLL x Post	.1785***	.0978***	-.0097
	(.0395)	(.0162)	(.0062)
Under New CLL	-.0398**	-.0243***	.0030
	(.0175)	(.0090)	(.0038)
Post	.6012***	.1490***	-.0174
	(.0890)	(.0403)	(.0118)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Pre-mean	13.3186	11.3119	.0224

Notes: The sample in each column is loans in just $20,000 below or above the new CLL. The outcome in the first column is the log of zip code level average median home value (in 2010 $'s). The outcome in the second column is the log of zip code level average median household income (in 2010 $'s). The outcome in the third column is a dummy that equals 1 if a transaction is made by a buyer from poor zip codes, and poor zip codes are defined as zip codes with the average median home value being in the bottom quartile of California. Standard errors are adjusted for bootstrap clusters in counties.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

ically, I examine if there are people from one particular race group getting a conforming loan. The Zillow data does not provide information about the buyer's race or ethnicity, but the buyer's full name is associated with every single property. With this information I then train an LSTM model on 1,000,000 randomly sampled names from the Florida Voter Registration Data to learn the relationship between the sequence of characters in a voter's name and the race and ethnicity. I then apply the model to names in the Zillow data. I break voters' names into bi-chars and exclude all the infrequent and super frequent bi-chars, and then pad sequences to make sure that they are the same size. I train an LSTM model using Keras and TensorFlow on those sequences, then estimate the model and fit it for 12 epochs.

Table 1.5 presents the performance of the LSTM model. On average, the pre-

Table 1.5: Performance of the LSTM Model

Race	Precision	Recall	F1	Support
Asian and Pacific Islander	0.91	0.95	0.93	49,867
Black	0.76	0.75	0.75	50,070
Hispanic	0.89	0.85	0.87	50,075
White	0.69	0.70	0.70	49,988
Weighted Average	0.81	0.81	0.81	200,000

cision, recall, and f1-score of this model are about 0.81. But the accuracy of predicting the Asian or the Hispanic is much higher as compared to that of predicting the Black or the White given that Asian and Hispanic names are usually much more distinctive. As can be seen from the confusion matrix (Figure 1.13), about 19 percent of Black names are mistakenly predicted to be White names, and about 19 percent of White names are

Figure 1.13: Error Matrix of the LSTM Model

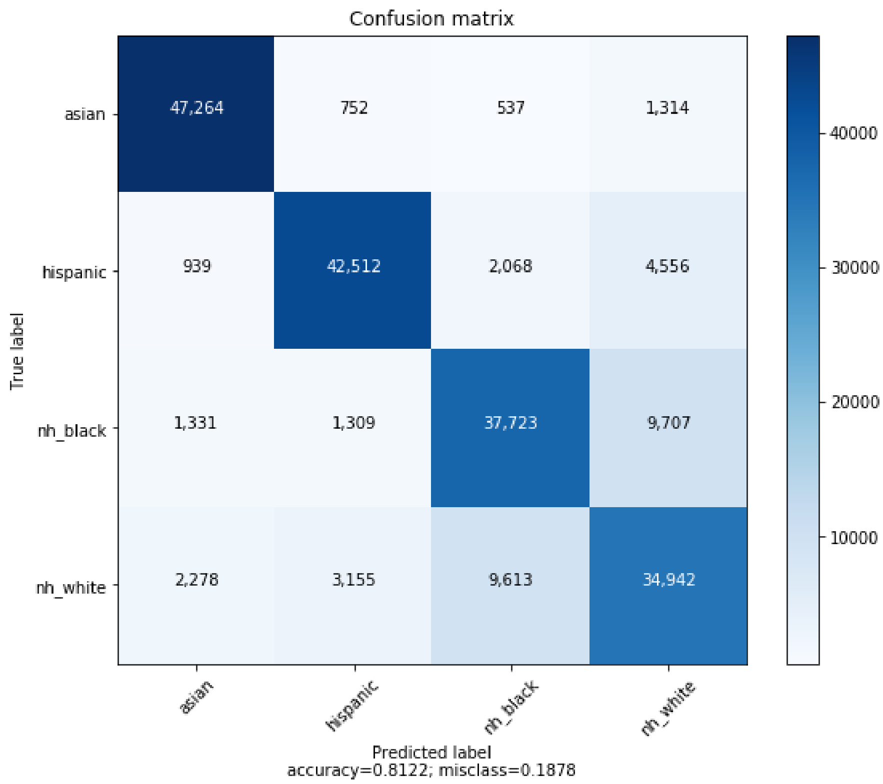
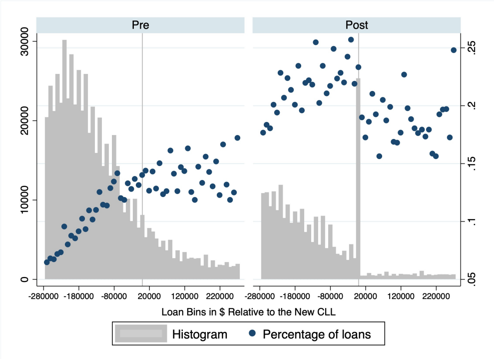
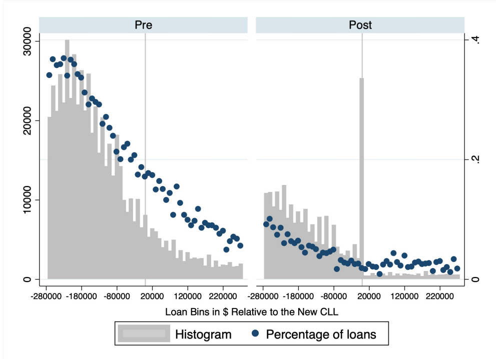


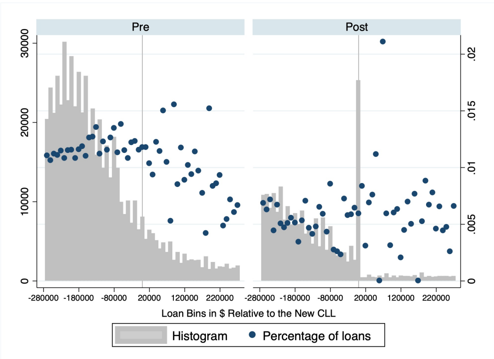
Figure 1.14: Percentage of Loans Taken out by Buyers from Different Race Groups



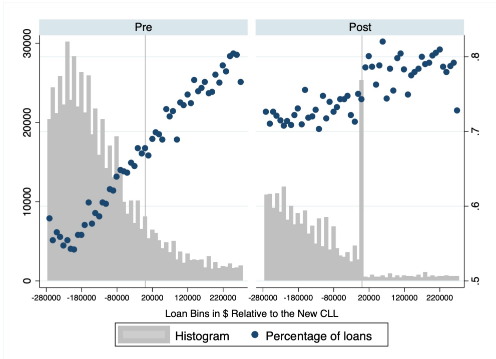
(a) Percentage of Loans Taken Out by Asians and Pacific Islanders



(b) Percentage of Loans Taken Out by Hispanics



(c) Percentage of Loans Taken Out by Blacks

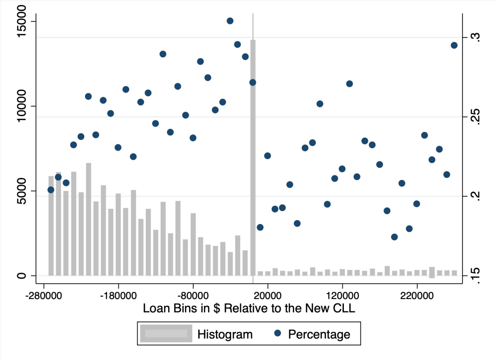


(d) Percentage of Loans Taken Out by Whites

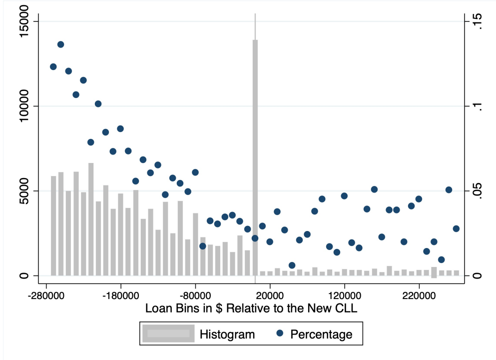
Notes: Each subfigure plots the share of loans taken out by buyers from one particular race group as a function of loan amount relative to the new conforming loan limit for mortgages originated between October 1, 2005 and March 31, 2008 (the left panel) vs. between April 1, 2008 and September 30, 2011 (the right panel). Each blue dot represents the share of loans taken out by buyers from one specific group of race or ethnicity within a \$10,000 bin relative to the new conforming loan limit of each county. Each gray bar represents the number of mortgages within a \$10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

mistakenly predicted to be Black names. About 9 percent of White names are predicted to be Hispanic names. This poses some concern about the accuracy of the model. There are also other concerns about using voter registration data from Florida to predict the race and ethnicity of people in California since the name pattern in Florida might be different from that in California. For example, the Asian population is smaller in Florida

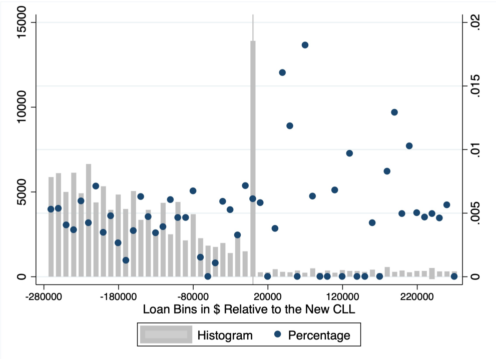
Figure 1.15: Percentage of Loans Taken Out by Repeated Buyers from Different Race Groups



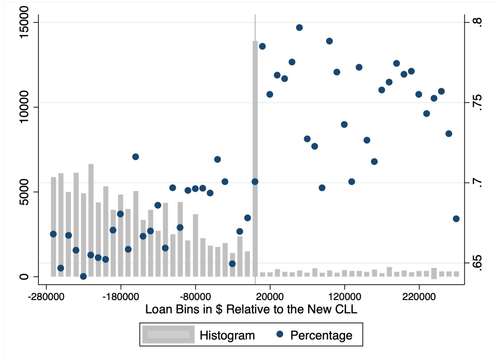
(a) Percentage of Loans Taken Out by Asians and Pacific Islanders



(b) Percentage of Loans Taken Out by Hispanics



(c) Percentage of Loans Taken Out by Blacks



(d) Percentage of Loans Taken Out by Whites

Notes: Each subfigure plots the share of loans taken out by repeated buyers from one specific group of race or ethnicity as a function of loan amount relative to the new conforming loan limit for mortgages originated between April 1, 2008 and September 30, 2011. Each blue dot represents the share of loans taken out by repeated buyers from one specific group of race or ethnicity within a $10,000 bin relative to the new conforming loan limit of each county. Each gray bar represents the number of mortgages within a $10,000 bin relative to the new conforming loan limit. The gray line is where the new CLL locates.

than in California. And, even though there is a large Hispanic population in Florida, researchers might be concerned that Mexican names are different from Cuban names in some way. Moreover, not every single voter, especially Black and Hispanic individuals, will be registered[7]. These are all possible limitations of this predicting model, therefore,

[7]Sood and Laohaprapanon (2018)

more work should be done to improve its accuracy.

Table 1.6: Share of Buyers with a Specific Predicted Race and Ethnicity

	Asian	Hispanic	Black	White
Under New CLL x Post	.0378	-.0050	-.0007	-.0321
	(.0302)	(.0104)	(.0041)	(.0301)
Under New CLL	.0016	.0041	.0007	-.0064
	(.0060)	(.0081)	(.0017)	(.0084)
Post	.0627	-.1827***	-.0038	.1238*
	(.0515)	(.0309)	(.0047)	(.0679)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pre-mean	.1341	.1754	.0110	.6795

Notes: Each column presents regression results using observations from $\pm\$20,000$ of the CLL. The outcome in the second column equals 1 if a transaction is made by an Asian or Pacific Islander. The outcome in the third column is equal to 1 if a transaction is made by a Hispanic. The outcome in the fourth column is equal to 1 if a transaction is made by a Black. The outcome is the last column equals 1 if a transaction is made by a White. Standard errors are adjusted for bootstrap clusters in counties.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

I apply the LSTM model to buyer names from the Zillow data and exclude names that I could not classify into any of the four groups (Asian, Hispanic, Black, or White) to analyze how buyers with different races and ethnicities behave or are treated under the program. Figure 1.14 plots the share of loans taken out by a homebuyer from a specific race group against the loan amounts relative to the new CLL. Asian individuals were more likely to take out conforming loans in the post-treatment period, but there is no significant bunching of conforming loans taken out by one particular race group at the threshold. Since I prove that the manipulation is mostly coming from repeated homebuyers, I also exclude first-time buyers and plot the share of loans taken out by repeated homebuyers from a particular race group against the loan size relative to the new CLL for loans originated in the post-treatment period (Figure 1.15). But all the figures make clear that the bunching is not driven disproportionately by a particular race group. When I apply the regression equation (1.1) to observations close to the new CLL, I also find no evidence that borrowers from a particular race group comprise a disproportionate share of those who manipulate their loan size to take advantage of the program. The regression results are presented in Table 1.6. The pre-mean percentage of Hispanic people is much lower than that of the California population, which implies Hispanic people may want loans but were screened out in some way, or they are eligible for the loans but they cannot find a home in the place that they want to buy, or they just do not buy such an expensive home [8].

1.5 Discussion

I propose a novel approach that infers the distortionary effects of federal mortgage repurchasing on lending patterns from the change in the frequency distribution of loans. I use an event study analysis examining thirty-nine temporarily increased CLLs in high-priced California counties in 2008 and show how mortgage repurchasing by Fannie Mae and Freddie Mac affects the fraction of loans near the threshold of the program. I find borrowers sort across the new CLL to bunch at or slightly below the cutoff. Second, I test what margins are manipulated for borrowers to take out a conforming loan just under the threshold. I find that borrowers bunching at the loan limit are statistically significantly less leveraged than others. In other words, the lowest interest rate is charged to homebuyers who manipulate the LTV ratio to get a loan just

[8] I also examine the impact on the probability of subsequent default rate under the program for borrowers in each race group or who are coming from poor zip codes. The regression results are presented in the Appendix.

under the new CLL. Next, I use a difference-in-difference analysis to investigate who is bunching at the loan limit. The main estimates show that borrowers with previous defaults are disproportionately more likely to jump over to the left side of the new CLL and line up at or slightly below the cutoff. The subsequent default rate of conforming loans originated at or slightly below the new CLL in the post-treatment period is 2.17 percent higher than that of jumbo loans barely above the new CLL. The increase in the subsequent default rate is mainly driven by homebuyers with a previous mortgage. Finally, I look into borrowers' demographic details, such as the neighborhood characteristics, income, and race, of those whose loans constitute the bunching at the loan limit. I find that the bunching is driven by people who previously lived in wealthier zip codes. I find no evidence that racially or economically-disadvantaged borrowers comprise a disproportionate share of those who manipulate their loan size to take advantage of the program.

In this paper, I show some negative impacts of federal mortgage repurchasing by examining the subsequent performance of loans near the threshold of the program. More importantly, I prove that the program does help a large fraction of homebuyers who previously lost their property find a new place to live. In the next step of this project, a theoretical framework will be provided to demonstrate the strategic incentives of lenders and borrowers. I also expect to derive a mathematical equilibrium outcome from the interactions between the two agents. Second, in section 1.4.3, no empirical evidence shows that borrowers from a particular race group are overrepresented in the pool of people bunching at the threshold, but I do observe that Asian borrowers are more likely to take conforming loans while White borrowers are more likely to take jumbo loans according to Figure 1.14 and 1.15. This could have different causes, and race discrimination in mortgage lending is likely to be one of them. More research will be done to investigate this issue.

Chapter 2

Negative Selection or Positive Selection: Evidence from Peer-To-Peer Online Lending

2.1 Introduction

Adverse selection due to information asymmetry is salient in the credit market. Starting from Stiglitz and Weiss (1981), there is large growing theoretical literature examining adverse selection in the credit market. But because it is hard to identify information asymmetries in the real world, empirical investigations have lagged behind. Interestingly, both theoretical research and thin empirical research bring up mixed predictions and implications. This project will focus on two well-known cases in the credit market, which yield opposing outcomes and policy implications. Stiglitz and Weiss's (1981) model assumes that a lender is faced with projects having the same mean return but cannot ascertain the riskiness of a project. When the interest rate charged by the credit market is higher, an adverse selection problem causes projects that are poor from the lender's viewpoint to drive out good projects. Borrowers in the loan pool are investing in projects with a lower average probability of success, and consequently, the default rate rises as the interest rate increases. On the other hand, de Meza and Webb (1987) argue that it would be more likely that lenders just know the actual outcome that a project yields if it succeeds, but not the probability of success. They predict that asymmetric information causes good projects to draw in bad projects, specifically, when the lender charges borrowers lower interest rates than before, borrowers in the loan pool are investing in projects with a lower probability of success, so the default rate increases as the interest rate decreases. It is the distribution of unobservable factors that forces the two models to make opposing predictions of the direction of selection. I will define the first case as a negative selection since increasing interest rates in the credit market worsens the mix of applicants. And I define the second case as a positive selection because, in this model, the quality of the loan pool and the interest rate charged by the lender move in the same direction.

I generally test between predictions of the Stiglitz-Weiss model and the de Meza-Webb model by using a natural experiment on a peer-to-peer online lending marketplace, Prosper. On January 13, 2011, Prosper announced that they have lowered interest rates across the board for all new loans. I find that the number of applications submitted by lower-rated applicants increased after the event. Also, for lower-rated borrowers, compared with 12.4 percent, which is the mean probability of default before lowering interest rates, this event increased the default rate by 11.8 percent, and this increase is statistically significant. While for higher-rated borrowers, the probability of default decreases by 1.1 percent, but this is statistically insignificant. Then I take a further step to analyze how the unobservable selection of interest rates contributes to the change of the loan pool and the change in default rate. I decompose the total effect of lowering interest rates on the default rate into three channels: a direct effect, i.e. the normal change in the default rate that occurs under perfect information, an indirect effect driven by observable selection, and an indirect effect driven by unobservable selection. The Stiglitz-Weiss model and the de Meza-Webb model have made opposing predictions of the direction of the last channel. In the real world, it is impossible to observe the unobservables. Fortunately, I can identify the first two channels thanks to the rich loan-level data provided by Prosper. As a result, after controlling for the confounding direct effect and observable selection effects, I am allowed to use the sign of the total effect to estimate the selection driven by unobservable factors.

Suppose that the lender is trading with borrowers under perfect information. When the lender charges borrowers with higher interest rates, borrowers are expected to pay back more. As a result, increasing interest rates make borrowers more likely to default. Alternatively, decreasing interest rates makes a borrower's life much easier since she is supposed to pay less. And consequently, she would be less likely to default than before. So the direct effect of lowering interest rates on the default rate is weakly positive, which is the opposite of the sign of the total effect.

To determine the direction of the observable selection effect, I compare between the pool of borrowers who submitted applications for credits before the event and the ones who requested credits after the event to see whether one group of people exhibits inferior or better characteristics as compared to the other. According to the outcomes of the balancing test of observable characteristics in pre- and post-period for either lower-rated or higher-rated applicants, I find no statistically significant change in either of them. Therefore, I conclude that the selection on observable information within either risk group is close to zero.

I conclude that there is a positive selection on unobservable information for lower-rated borrowers, in other words, the interest rate and the unobservable quality of borrowers move in the same direction. Lowering interest rates makes good projects draw in bad ones and raises the default rate. This empirical evidence is consistent with the prediction of the de Meza-Webb model. For higher rated borrowers, even though in total I observe that the default rate decreases, I am not able to conclude that it is consistent with any prediction of these two models. First, the change is statistically insignificant and the estimation is noisy. Second, it is probably just because the scoop of selection for higher-rated borrowers is too narrow and selection on unobservable information is too small; and in that situation, even though there is a positive selection on unobservable information, it is too small to cancel the direct effect. Since I am not able to make a convincing conclusion about the higher-rated borrowers, I will focus on examining the financial behaviors of lower-rated borrowers in this project.

To rule out the possibility that the number of lower-rated applications increases because applicants who request credits right before the event withdraw the previous applications after hearing about the news of lowering interest rates and apply for new loans in the post-period, I check the probability of withdrawing an application one week before and after the event, ten days before and after the event, and two weeks before and after the event. I find that in each of these windows, there is no statistically significant large enough change in the probability of withdrawing an application driven by the event. Also, to rule out the possibility that investors are more likely to fund lower-rated borrowers after the event (since even though interest rates for all new loans decrease, still lower-rated borrowers are charged with high-interest rates by the platform), I check the probability of getting funded before and after the event. I find that there is no statistically significant change in investors' funding behaviors. After controlling for all observable information and other possible channels, I confirm that it could only be the selection on unobservable information that raises the default rate.

In recent years, there are a growing number of empirical projects studying the problem of adverse selection in the credit market. Some researchers estimate the presence of adverse selection in the credit market by employing data from developing countries. Karlan and Zinman (2009) use randomized field experiments in which they schedule small, high-interest, short-term, uncollateralized credit with fixed monthly repayment offers to poor farmers in South Africa. By observing people's selection into different contracts and tracking their repayments, they don't find strong empirical evidence of adverse selection. Klonner and Rai (2007) use data from an Indian financial institution in which interest rates are determined by competitive bidding to test if riskier borrowers are willing to pay higher interest rates than safer borrowers. They find strong evidence of adverse selection as predicted by Stiglitz and Weiss (1981). Einav, Jenkins, and Levin (2012) analyze a research question similar to mine. They investigate how different elements of credit offers affect the quality of the borrower pool and the subsequent prospects for repayment in the auto loan market. But they are interested in a group of subprime consumers with low incomes or poor credit histories because the goal of their project is to understand high-risk lending. Agarwal et al. (2007) analyze if borrowers self-select into loan contracts that are designed to reveal information about their risk level and if lenders still face adverse selection problems conditional on borrowers' choices of contract type. They find even after controlling for borrowers' contract choice and other observable risk characteristics, lenders continue to face adverse selection problems because of private information, which is consistent with the prediction of Stiglitz and Weiss (1981). Edelberg (2004) studies the significance of unobservable default risk in the mortgage and automobile loan market by comparing average interest rates and collateral requirements across groups of borrowers with different observable information. She finds strong and robust evidence of adverse selection as predicted by Stiglitz and Weiss (1981), in other words, higher-risk borrowers pay higher interest rates.

The current project is contributing to the literature in the following aspects. First, many papers examine the policy change on Prosper. However, they tend to employ large research windows. They ignore the fact that there exist confounding time effects and the policy change they are interested in might not be the only major policy change in their research windows. For example, Meyer (2014) as well as Wei and Lin (2016) use the same data as I do and they both employ large research windows to investigate the effects of the pricing regime change on December 19, 2010. But they all ignore the interest rate adjustment in the post period of their research windows, and as a result, all the effects they found are attributed to just the mechanism change. In this research, I am restricting the sample within twenty-three days before and after the event to make sure that I am providing an accurate estimation of the effect of lowering interest rates on the default rate. In addition, at the end of the project, I employ data from other years to detrend the data in my sample period. I try to tease out the potential confounding time effects and make more precise predictions about the effect of the policy. Second, there is a paper that analyzes the credit market imperfections in the context of developing countries with plausible research design, and loans are originated to very poor farmers by charging them with generally high-interest rates. But empirical evidence found in developing countries is not sufficient to explain what is happening in a country like the U.S. In the developing context, the distribution of income is random, and the repayment is supposed to be largely affected by seasons, harvest, etc. But I am doing this research in the context of a developed country. What's worth mentioning is that Prosper does not even accept subprime applicants to request loans on its platform. So people in this project are relatively wealthier and lower-risk. Third, there is paper in the context of developed countries but they focus more on subprime borrowers or a group of borrowers with huge heterogeneity and most empirical studies found evidence in favor of the Stiglitz-Weiss model. But I am proving that the de Meza-Webb model is sometimes more plausible at a micro-level.

The paper proceeds as follows. In Section 2.2, I demonstrate basic setups and derive predictions of both the Stiglitz-Weiss model and the de Meza-Webb model. In Section 2.3, I introduce empirical settings including the business model of Prosper and a detailed description of data. Section 2.4 is the heart of this project. In this section, I explain the identification strategy that I use to analyze each possible channel through which lowering the interest rate affects the default rate. After controlling for the confounding direct effect and observable selection effects, I present the regression results and determine which of the theoretical predictions is more plausible at a micro-level. Section 2.5 concludes.

2.2 Model

In this section, I will start by describing the assumptions of both the Stiglitz-Weiss model and the de Meza-Webb model, then obtain and simplify the predictions that I will test in the empirical part of this current project.

2.2.1 Assumptions

The models in Stiglitz and Weiss (1981) and de Meza and Webb (1987) have some common crucial assumptions. They build models in a two-period world. In the first period, risk-neutral entrepreneurs choose to borrow and invest. In the second period, investment outcomes are realized, and entrepreneurs pay back their debts (if they can) and consume whatever is left. Each entrepreneur invests in just one project, which either succeeds or fails. All borrowers have limited liability. If a project fails, the lenders cannot extract further payments from the entrepreneur. To activate a project requires an investment of K, while the entrepreneur is endowed with $W < K$ so that if a project is undertaken, the entrepreneur must additionally finance $B = K - W$ in the credit market.

Assume that the interest rate charged to the entrepreneur in the credit market is r and the rate of return for a safe outside option is ρ. An increase in r discourages some entrepreneurs from investing. However, Stiglitz and Weiss (1981) and de Meza and Webb (1987) make different predictions of the distribution of projects. Stiglitz and Weiss (1981) argue that borrowers may be investing in riskier projects now so that the default rate increases as r increases. But de Meza and Webb (1987) point out that when r increases, borrowers will invest in projects with an higher average probability of success, and then the default rate falls as r increases.

2.2.2 Stiglitz and Weiss (1981)

In this case, they assume that the bank is able to distinguish projects with different mean returns, so they confine themselves to the problem of a bank facing projects having the same mean return. However, the bank is not able to ascertain the riskiness of a project. If p is a project's probability of success, then lower p corresponds to greater risk in the sense of mean preserving spreads. Assume that for each project, there is a probability distribution of returns $\tilde{R}$, and they write the distribution of returns as $F(\tilde{R}, p)$[1]. Then they say that the individual defaults on her loan if the return $\tilde{R}$ is

[1]For $p_1 < p_2$ if $\int_0^\infty \tilde{R} f(\tilde{R}, p_1) d\tilde{R} = \int_0^\infty \tilde{R} f(\tilde{R}, p_2) d\tilde{R}$, then for $y \geq 0$ $\int_0^y F(\tilde{R}, p_1) d\tilde{R} \geq \int_0^y F(\tilde{R}, p_2) d\tilde{R}$

insufficient to pay back the promised amount, if

$$\tilde{R} \leq B(1+r) \tag{2.1}$$

Conditional on activating a project, the net return to the borrower is

$$\pi(\tilde{R}, r) = \max[\tilde{R} - (1+r)B, 0] \tag{2.2}$$

This is a convex function of $\tilde{R}$, so the expected profits increase with risk, i.e. decrease in p. Consider the borrower with the marginal project $\hat{p}$, that is, the project for which (2.3) holds with equality.

$$\Pi(r, \hat{p}) = \int_0^\infty \max[\tilde{R} - (1+r)B, 0] dF(\tilde{R}, \hat{p}) = 0 \tag{2.3}$$

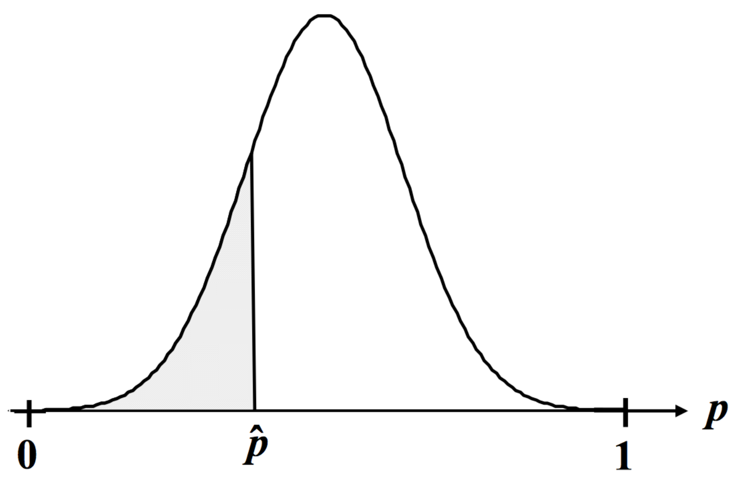
If the marginal borrower has the probability of success $\hat{p}$, only borrowers with $p < \hat{p}$ apply for credit. In other words, the borrower with the marginal project is the safest in the loan pool, and all other borrowers are in the “risky tail” of the distribution as in Figure 2.1. The marginal borrower is less likely to default and should get a lower interest rate than others. Unfortunately, she cannot be distinguished from others and thus gets a high-interest rate that is probably appropriate for the average riskiness of the loan pool. This leads to the fact that the marginal borrower’s incentive to stay in the loan pool is too low. In equilibrium, the adverse selection of interest rates could make the mix of applicants become worse and riskier projects would be undertaken. If

r rises, then $\hat{p}$ must decreases, since

$$\frac{\mathrm{d}\hat{p}}{\mathrm{d}r} = \frac{B\int_{(1+r)B}^{\infty} dF(\tilde{R},\hat{p})}{\frac{\partial \Pi}{\partial \hat{p}}} < 0 \tag{2.4}$$

The lower $\hat{p}$ means that there is a smaller chance that applicants' projects in the loan pool will be successful, so that the default rate goes up as r increases.

Figure 2.1: Projects Undertaken in the Stiglitz and Weiss Case



2.2.3 De Meza and Webb (1987)

In this case, all projects have the same actual outcome R^G if they succeed but differ in the probability of success (p). The lenders have no prior information on the characteristics of individual entrepreneurs, and they only know the distribution of characteristics of the population of entrepreneurs. Under these assumptions, now the expected profit to a borrower who applies for credit is given by

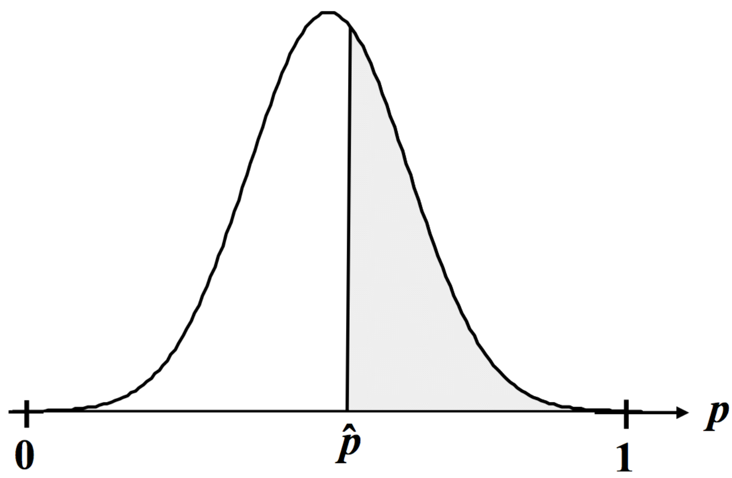
$$p[R^G - (1 + r)B] \tag{2.5}$$

which is increasing in p. The borrower will accept the project if

$$p[R^G - (1 + r)B] \geq (1 + \rho)W \tag{2.6}$$

And the marginal borrower is the applicant with the project whose probability of success is $\hat{p}$, which is the project for which (2.6) holds with equality. If r rises, then according to (2.5) and (2.6) $\hat{p}$ must rise. And the higher $\hat{p}$ means that there is a greater chance that an entrepreneur applicant's project will be successful. So in the de Meza-Webb model, the marginal borrower is the riskiest in the loan pool, and projects that are undertaken are those with $p \geq \hat{p}$, i.e. in the "safe tail" of the distribution as in Figure 2.2. If lenders can identify the marginal borrower, they would offer her a high-interest rate since it is relatively likely that she will default. But this borrower is indistinguishable from the other safer projects and thus gets an inefficiently low-interest rate. As a result, the marginal borrower's incentive to stay in the loan pool is too high. In equilibrium, there is over-investment, and too many projects of a given return will be funded. If the lender decreases interest rates for borrowers, more riskier projects would be drawn in the pool, which means that on average there is a lower chance that a project will be successful. Only if r rises, then the mix of applications becomes better, and consequently, there is a negative correlation between interest rate and default.

Figure 2.2: Projects Undertaken in the de Meza and Webb Case



2.2.4 Decomposition

I believe that in the real world when the interest rate in the credit market changes, it will affect the default rate through three different channels: a direct effect, i.e. changes in the repayment burden under perfect information; an indirect effect driven by selection on observable information; and an indirect effect driven by selection on unobservable information. If X is a vector of borrowers' observable characteristics, θ represents some unobservable factors and $d(\cdot)$ is the default rate. The total effect of adjusting the interest rate is supposed to be decomposed in the following way:

$$\frac{\mathrm{d}d(r,X,\theta)}{\mathrm{d}r} = \frac{\partial d(r,X,\theta)}{\partial r} + \frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r} + \frac{\partial d(r,X,\theta)}{\partial \theta}\frac{\partial \theta}{\partial r} \tag{2.7}$$

The direct effect, i.e. $\frac{\partial d(r,X,\theta)}{\partial r}$, is that lowering the interest rate will make a borrower's life much easier, thus it is relatively unlikely that she will default. Alternatively, a borrower facing an increased interest rate is expected to pay back more in the future, intuitively she is more likely to default than before. An indirect selection effect driven by observable characteristics i.e., $\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r}$ indicates that when changing the interest rate, borrowers with some specific types of observable factors may self-select into the market and drive the default rate to a specific direction. In terms of the predictions of Stiglitz and Weiss (1981) and de Meza and Webb (1987), it is through the mechanism of unobservable selections that the two models are different from each other. They take as a given both the direct effect and the selection effect driven by observable characteristics. In the real world, I am going to demonstrate that my context is sufficiently similar that these first two channels are not going to confound my attempt to use the total effect to figure out how interest rate adjustment affects the default rate through the mechanism of unobservable selections, i.e. $\frac{\partial d(r,X,\theta)}{\partial \theta}\frac{\partial \theta}{\partial r}$. I am going to show that even after controlling for all observable characteristics, there is still an unobservable selection on interest rates. Moreover, I will prove which prediction of the two models is more plausible at a micro-level in the real world.

2.3 Empirical Setting

2.3.1 Prosper

Prosper was founded in 2005 as the first peer-to-peer online lending marketplace in the United States. Since then, Prosper has facilitated more than $14 billion in loans to more than 880,000 people. Borrowers apply for a fixed-rate, fixed-term loan through Prosper, and investors invest in the loans and earn returns. Prosper handles all loan servicing on behalf of the matched borrowers and investors. Before being shut down by the US Securities and Exchange Commission (SEC) in 2008, Prosper functioned as like a social website. Borrowers posted profiles, complete with pictures and biographies. Lenders formed groups and made lending recommendations to each other. Interest rates were set by Dutch Auctions. In November 2008, the SEC issued a cease and desist order to Prosper, then Prosper changed their model and reopened to sell SEC-registered securities in July 2009. In December 2010, Prosper eliminated the auction regime and started a new model of posted pricing. For this analysis, I will consider only listings submitted after the elimination of the auction regime.

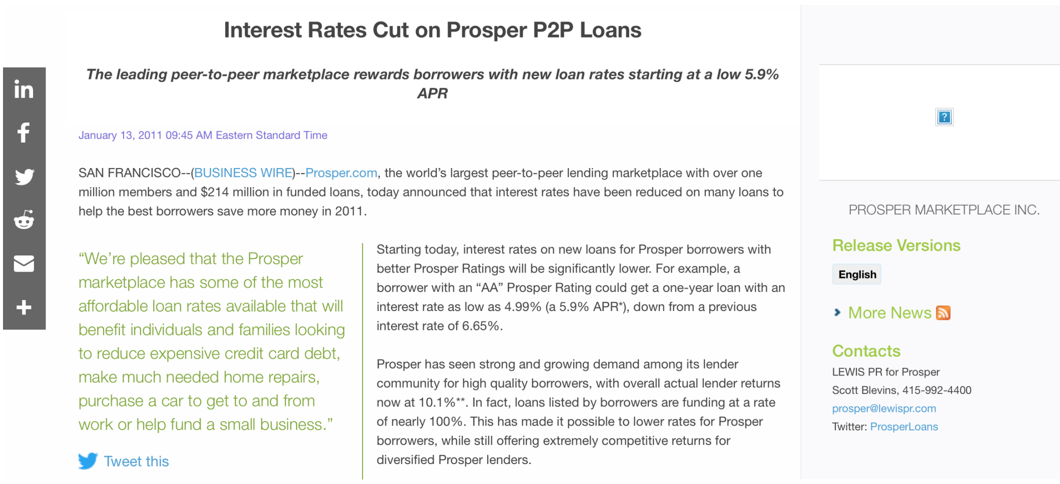
2.3.2 The Process

A potential borrower goes to Prosper to complete a loan application form and to apply for credits. The platform will either accept or reject the application. If the platform accepts this applicant, it will assign a score called "Prosper score" to her. Prosper score is a custom risk score built to assess the risk of Prosper borrower listings. It was built based on Prosper's historical data. It ranges from 1 to 10, with 10 being the best. Then the platform will predict the likelihood an applicant will default on a loan according to the applicant's Prosper score and her FICO score. After that, Prosper will assign a Prosper Rating to the borrower based on the estimated loss rate. In my sample period, the six possible letter scores, from the best to the worst, are AA, A, B, C, D, and E. Applicants within the same category are assigned the same interest rate. If the applicant accepts the offer, she will create a listing web page to describe basic information about the application as well as detailed information about herself, such as the requested amount, a short description of her credit history and self-reported employment data, etc. And investors then browse available listings displayed on the platform and specify a dollar amount for their investments.

2.3.3 The Event

On January 13, 2011, Prosper announced that they have lowered interest rates across the board for all new loans to keep up the momentum and ensure their rates remain competitive. Figure 2.3 is a screenshot of their official press release. Table 2.1 shows the differences between the old and new rates for each category.

Figure 2.3: The Official Press Release Announcing Prosper's Interest Rate Adjustment



Interest Rates Cut on Prosper P2P Loans

The leading peer-to-peer marketplace rewards borrowers with new loan rates starting at a low 5.9% APR

January 13, 2011 09:45 AM Eastern Standard Time

SAN FRANCISCO--(BUSINESS WIRE)--Prosper.com, the world's largest peer-to-peer lending marketplace with over one million members and $214 million in funded loans, today announced that interest rates have been reduced on many loans to help the best borrowers save more money in 2011.

"We're pleased that the Prosper marketplace has some of the most affordable loan rates available that will benefit individuals and families looking to reduce expensive credit card debt, make much needed home repairs, purchase a car to get to and from work or help fund a small business."

Tweet this

Starting today, interest rates on new loans for Prosper borrowers with better Prosper Ratings will be significantly lower. For example, a borrower with an "AA" Prosper Rating could get a one-year loan with an interest rate as low as 4.99% (a 5.9% APR*), down from a previous interest rate of 6.65%.

Prosper has seen strong and growing demand among its lender community for high quality borrowers, with overall actual lender returns now at 10.1%**. In fact, loans listed by borrowers are funding at a rate of nearly 100%. This has made it possible to lower rates for Prosper borrowers, while still offering extremely competitive returns for diversified Prosper lenders.

PROSPER MARKETPLACE INC.

Release Versions

English

More News

Contacts

LEWIS PR for Prosper

Scott Blevins, 415-992-4400

prosper@lewispr.com

Twitter: ProsperLoans

Table 2.1: The Interest Rate Adjustment for Each Rating

Ratings	Old Rate	New Rate	Difference
AA	8.00%	6.55%	1.45%
A	10.95%	9.55%	1.40%
B	14.90%	12.99%	1.91%
C	21.40%	19.99%	1.41%
D	29.50%	26.99%	2.51%
E	32.20%	31.99%	0.21%

Notes: This table shows the differences between the old and new rates for listings with 3-year maturity applied by first-time applicants.

2.3.4 Data

I use Prosper listings data and Prosper loans data to test between predictions of the two models. Prosper listings data contains all applications of loans submitted by the applicants, no matter if the listing was successfully funded or not. For each listing, I can observe information including the interest rate, the requested amount, the term of the loan, the listing status, each borrower's credit data, each borrower's employment information, etc. Prosper loans data contains only loans successfully funded and originated. For each originated loan, I can observe information such as the interest rate, repayment status, etc. To control for confounding time effects and to make sure that the interest rate adjustment is the only major policy change during this period, I focus on a short research window. I restrict the sample to a small bandwidth from December 20, 2010 to February 03, 2011, i.e. 23 days before and after lowering the interest rate. There are 1,120 listings, i.e. 1,120 applications of loans were submitted during this sample period. In this project, I drop all the borrowers who have at least one prior loan with the platform and I am only focusing on the first-time applicants because Prosper has already observed repeated borrowers' previous financial behaviors with the platform and thus obtained further detailed information of this group of people. Also, repeated borrowers might feel good or bad about applying for credits on the platform according to their previous experiences, and thus determined to stay in or exit the market regardless of the interest rate adjustment. What's more, Prosper uses a different algorithm of assigning interest rates to repeated borrowers. In my sample period, first-time applicants are the majority and about 82 percent applicants are first-time (see Figure 2.4). There are listings with either 1-year, 3-year, or 5-year maturity. I only keep listings with 3-year maturity to ensure that I am not comparing loans with different maturities to one another. Also, loans with different periods of maturity will be assigned with different interest rates, for example, if the term of the loan is longer, the interest rate must be higher. 3-year listings are the majority, nearly 94 percent of the total listings in my sample period are listings with 3-year maturity. After dropping off disqualified observations, 847 listings are left.

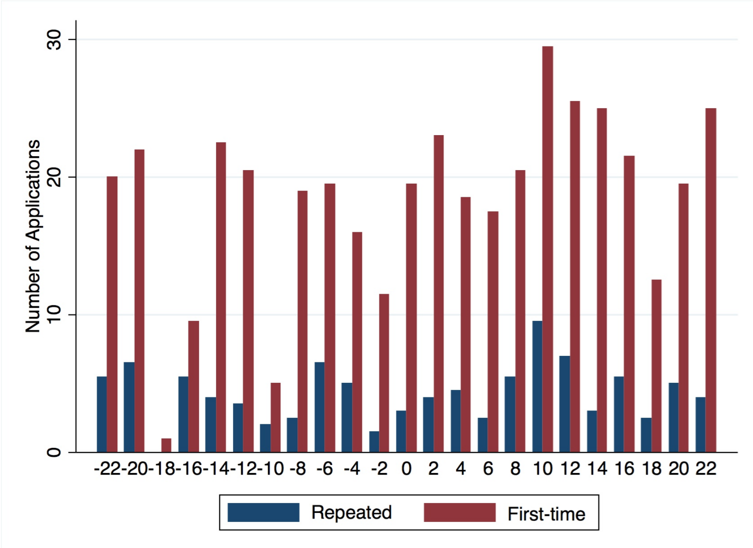
2.4 Regressions and Results

2.4.1 The Total Effect

I start with estimating the total effect of lowering the interest rate on the default rate by employing the following regression:

$$Default_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Low_i + \beta_3 Post_t * Low_i + \beta_4 X_{it} + \epsilon_{it} \tag{2.8}$$

Figure 2.4: Number of Applications Submitted by Repeated Applicants and First-time Applicants around Lowering the Interest Rate



Notes: This figure shows how the number of applications submitted by repeated applicants and first-time applicants changes in the research window. Applications had a 3-year maturity and were submitted by either repeated or first-time applicants. All applications were submitted from December 20, 2010 to February 3, 2011. Days away from the event are displayed on the horizontal axis. I am using a 2-day bin, where bin 0 includes January 11, 2011 and January 12, 2011, bin -1 includes January 9, 2011 and January 10, 2011, and bin 1 includes January 13, 2011 and January 14, 2011.

$Default_{it}$ indicates whether a loan is past due. It is a binary variable which equals 1 regardless if a loan is partially past due or fully past due.[2] $Post_t$ is a time dummy which equals 1 after Jan 12, 2011 and equals 0 otherwise. Low_i indicates whether a loan is categorized by Prosper into ratings below B. Borrower characteristics include all borrower credit variables available, stated income, months employed, and homeownership. β_3 explains the total effect of lowering interest rates on default for lower-rated borrow-

[2]Prosper will attempt to contact borrowers via email, phone, and letter to collect any past due payments. Investors are not allowed to contact borrowers directly. Prosper gives all borrowers 15 days to make payments with no penalty. Interest will continue to accrue on the loan daily, but after the 15-day grace period, borrowers would be charged a late fee. Prosper will hand off the borrower to a collection agency if no payment has been made after 30 days.

ers. Table 2.2 presents the estimation results of the regression above. 847 qualified applications were submitted by applicants between December 20, 2010 and February 3, 2011, and 498 of them were successfully funded and accepted by the borrowers. I analyze the probability of default around lowering the interest rate by using the 498 loans, and I find that even after controlling for all observable characteristics there is still a statistically significant increase in the default rate driven by the event for lower-rated borrowers. And compared with the average default rate for lower-rated borrowers in the pre-period, 12.4 percent, this is a large increase.

Table 2.2: Default Rate around the Event Lowering the Interest Rate

	(1) Default	(2) Default
Post	-0.017	-0.011
	(0.027)	(0.028)
Low	0.066**	0.018
	(0.032)	(0.053)
Post x Low	0.121***	0.111**
	(0.045)	(0.047)
Borrower X's	No	Yes
Observations	498	498
R-Squared	0.051	0.068
Pre-mean (AA-B)	0.058	0.058
Pre-mean (C-E)	0.124	0.124

Notes: The sample is restricted to applications that were submitted by applicants from December 20, 2010, through February 3, 2011 and eventually successfully funded and originated by the platform. Loans are restricted to 3-year loans for borrowers who do not have a prior loan with the platform. Default indicates whether a loan is past due. *Post* is a time dummy that equals 1 after Jan 12, 2011 and equals to 0 otherwise. Low indicates whether a loan is categorized by Prosper as lower-rated (C-E). Borrowers' characteristics include all borrower credit variables available, stated income, months employed, and homeownership. Robust standard errors are in parentheses.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Earlier I decomposed the total effect of lowering interest rates on the default rate ($\frac{\mathrm{d}d(r,X,\theta)}{\mathrm{d}r}$) into three channels: a direct effect ($\frac{\partial d(r,X,\theta)}{\partial r}$), an indirect selection effect on observable information ($\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r}$), and an indirect selection effect on unobservable information ($\frac{\partial d(r,X,\theta)}{\partial \theta}\frac{\partial \theta}{\partial r}$). According to outcomes in Table 2.2, $\frac{\mathrm{d}d(r,X,\theta)}{\mathrm{d}r} < 0$ even after controlling for all observable information. In the following, I demonstrate which of the three channels actually drives the sign of the total effect.

2.4.2 The Direct Effect

I start by analyzing the direct effect of lowering the interest rate on the probability of default. Alternatively, I could define this channel as the causal effect of lowering the interest rate on default. Assume that now investors on the platform are trading with borrowers under perfect information: When the platform charges borrowers higher interest rates, borrowers are expected to pay back more. Therefore, increasing interest rates makes borrowers more likely to default. Alternatively, decreasing interest rates makes a borrower's life much easier since she is supposed to pay back less. As a result, she would be less likely to default than before. So the direct effect of lowering interest rates on default is weakly positive, i.e. $\frac{\partial d(r,X,\theta)}{\partial r} > 0$, which is opposite to the sign of the total effect.

2.4.3 The Selection on Observable Information

The indirect selection effect on observable information, $\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r}$, could be further decomposed into two parts: $\frac{\partial d(r,X,\theta)}{\partial X}$ and $\frac{\partial X}{\partial r}$. The first part could be interpreted as whether knowing the observable characteristics conditional on prosper ratings is useful or not and the second part could be interpreted as whether borrowers with certain types of characteristics select into the platform because of the interest rate adjustment. I investigate the sign of the second part by comparing the mean of each available observable characteristic in the pre-period with those in the post-period for the group of higher-rated applicants and the group of the lower-rated ones.

Table 2.3: Summary Statistics for Higher-Rated Applications

	Pre	Post	P-value
Amount	9,343	10,232	0.174
	(4,968)	(5,534)	
~FICO	757	759	0.498
	(21.497)	(20.478)	
Prosper Score	8.700	8.808	0.254
	(0.681)	(0.833)	
Experience	17	17	0.523
	(7.173)	(8.158)	
Revolving	16.430	20.867	0.357
	(18.013)	(50.027)	
DTI	0.111	0.114	0.792
	(0.076)	(0.103)	
Utilization	0.343	0.343	0.983
	(0.218)	(0.270)	
Current Lines	9.242	9.493	0.680
	(4.128)	(5.513)	
Open Lines	8.325	8.514	0.735
	(3.896)	(4.972)	
Inquiry	0.625	0.521	0.363
	(1.030)	(0.841)	
Public Record	0.117	0.116	0.996
	(0.393)	(0.322)	
Delinquencies	0.692	1.240	0.270
	(2.675)	(4.857)	
Income	6.330	9.564	0.267
	(5.155)	(31.508)	
Months Employed	6.342	7.836	0.117
	(5.959)	(8.899)	
Homeowner	0.558	0.610	0.400
	(0.499)	(0.490)	
Observations	120	146	

Notes: This table shows summary statistics for only higher-rated (AA-B) applications as categorized by Prosper. Listings have a 3-year maturity. Applicants are all first-time. All applications were submitted between December 20, 2010 and February 3, 2011. Standard Deviations are in parentheses.

Table 2.3 summarizes the means for all the observable characteristics before and after lowering interest rates for applications submitted by higher-rated applicants. In total, 26 more applicants came to the market after the event. The amount is measured in dollars, and it is the amount of the loan requested by the applicant. The reason why I compare the average requested amount before and after the event is that there are papers that show that the probability of default for any particular borrower increases as the amount borrowed increases.[3] I was thinking about whether applicants tend to apply for more credits since they are charged with lower interest rates than before. People may argue that it could be the case that borrowers apply for higher amounts than before and thus the default rate in the post period is raised. But this possible channel is ruled out by the fact that there is no statistically significant increase in the requested amount. The FICO score is approximated as the mean of the applicant's 20-point range. The Prosper score is a custom risk score built to assess the risk of Prosper borrower listings. It was built based on data collected from more than 40,000 loans issued since 2006. It ranges from 1 to 10 with 10 being the best. The credit experience is measured as the number of years since the borrower's first credit line was opened. The revolving balance refers to the total outstanding balance that the borrower owes on open revolving credit accounts and it is in thousands of dollars. Debt-to-Income DTI ratio is the sum of the borrower's monthly debt payments divided by their monthly income, capped at 10. Utilization is the sum of balances owed on open bankcards divided by the sum of their credit limits. Current lines are the number of credit lines that the borrower is paying on time. Open lines are the number of credit

[3]Stiglitz 1970, 1972

lines open. Inquiry denotes the number of credit inquiries within the last 6 months. Public record indicates the number of bankruptcies, liens, and judgments within the past 10 years. Delinquencies indicate the number of delinquencies within the last 7 years. Income refers to reported monthly labor income and is measured in thousands of dollars. Months employed refers to the number of months that the borrower has held their current job. Homeowner indicates the probability that a borrower owns a residential property. There is no statistically significant change in each variable after lowering the interest rates. As a result, conditional on viewing the group of higher-rated applicants, I do not observe applicants with some specific types of characteristics who select into the market because of the lower interest rates. So I conclude that $\frac{\partial X}{\partial r}$ is close to zero, and consequently, the whole channel $\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r}$ is zero for higher-rated applicants.

Then I did the same analysis for just the lower-rated applicants. Table 2.4 summarizes the means for all the observable characteristics before and after lowering interest rates for applications submitted by lower-rated applicants. The number of lower-rated applicants increased by 79 after the event. And there is no statistically significant change on any observable variable. So for lower-rated borrowers, $\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r}$ is also zero. In other words, within the group of lower-rated borrowers, there is no selection on observable information.

Table 2.4: Summary Statistics for Lower-Rated Applications

	Pre	Post	P-value
Amount	5,651	5,583	0.757
	(2,695)	(2,617)	
~FICO	703	702	0.958
	(36.231)	(38.783)	
Prosper Score	6.295	6.206	0.331
	(1.043)	(1.122)	
Experience	18	17	0.186
	(8.392)	(8.600)	
Revolving	23.908	23.535	0.922
	(44.878)	(46.086)	
DTI	0.185	0.179	0.861
	(0.454)	(0.392)	
Utilization	0.514	0.538	0.379
	(0.333)	(0.314)	
Current Lines	8.474	8.470	0.992
	(5.203)	(5.084)	
Open Lines	7.458	7.485	0.945
	(4.657)	(4.582)	
Inquiry	1.056	1.209	0.185
	(1.279)	(1.449)	
Public Record	0.367	0.330	0.659
	(0.921)	(1.021)	
Delinquencies	3.506	3.852	0.162
	(9.301)	(9.599)	
Income	5.556	6.004	0.301
	(3.757)	(6.034)	
Months Employed	7.042	7.696	0.321
	(7.883)	(7.847)	
Homeowner	0.454	0.436	0.669
	(0.499)	(0.497)	
Observation	251	330	

Notes: This table shows statistics for only lower-rated (C-E) applications, as categorized by Prosper. Listings are with a 3-year maturity. Applicants are all first-time. All applications were submitted between December 20, 2010 and February 3, 2011. Standard Deviations are in parentheses.

2.4.4 Other Potential Channels

For the lower-rated borrowers, I have ruled out the possibility that the channel of direct effect or the channel of selection on observable information drives the sign of the total effect of lowering interest rates on default. Other possible channels are likely to affect the probability of default when lowering interest rates. Earlier I dropped all the borrowers with at least one prior loan with the platform. People may argue that the number of applications submitted by repeated borrowers probably changes because of the interest rate adjustment, and this change will potentially affect the estimation of repayment behaviors. Figure 2.4 shows the number of applications submitted by repeated applicants versus the first-time applicants both before and after the event.

Table 2.5: Loans and Listings Behaviors of Repeated Borrowers

	(1) Default	(2) Default	(3) Applications
Post	-0.026	-0.030	-0.085
	(0.026)	(0.032)	(0.375)
Low	0.006	0.059	0.125
	(0.040)	(0.049)	(0.481)
Post x Low	0.081	0.068	0.665
	(0.055)	(0.053)	(0.774)
Borrower X's	No	Yes	No
Observations	167	167	92
R-Squared	0.030	0.134	0.036
Pre-mean (AA-B)	0.026	0.026	2.438
Pre-mean (C-E)	0.031	0.031	2.563

Notes: This table shows how the default rate and the number of applications change around the interest rate adjustment. The sample is restricted to applications that were submitted by repeated borrowers between December 20, 2010 and February 3, 2011. Loans are restricted to 3-year loans for borrowers that have at least one prior loan with the platform. Post is a time dummy, which equals 1 after Jan 12, 2011 and equals 0 otherwise. Low indicates whether a loan is categorized by Prosper as lower-rated (C-E). Borrower characteristics include all borrower credit variables available, stated income, months employed, and homeownership. Column (1) - (2) focus on the default rate and default indicates whether a loan is past due. Column (3) focuses on the number of applications and it is measured at the day-risk level. Robust standard errors are in parentheses.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

In total, there are few repeated borrowers in each bin, specifically, there are on average five repeated borrowers per bin. And there is no significant change in the number of applications submitted by repeated borrowers around the event. I take a further step to collapse the number of applications submitted by repeated borrowers by day and group risk, i.e. whether it is a higher or lower rated application. Then

I regress the daily risk level of applications on the same event dummy, risk dummy, and their interaction that I used in equation (2.8). Table 2.5 presents the outcomes. I find that there is no statistically significant evidence showing that more higher-rated or lower-rated repeated borrowers select into the platform because of the event. Also, I investigate the repayment behavior of repeated borrowers. Whether controlling for observable variables or not, there is no statistically significant evidence showing that repeated borrowers changed their probability of default after the event. Therefore, dropping off repeated borrowers should not affect the empirical results of selection or repayment.

Table 2.6: Applications and Loans around the Interest Rate Adjustment

	(1) Applications	(2) Fund	(3) Fund	(4) Made	(5) Made
Post	0.922	0.007	0.007	0.033	0.040
	(0.833)	(0.032)	(0.033)	(0.059)	(0.058)
Low	5.199***	-0.045	-0.053	-0.071	-0.117
	(1.467)	(0.032)	(0.049)	(0.054)	(0.078)
Post x Low	3.165	-0.020	-0.032	-0.017	-0.026
	(1.969)	(0.042)	(0.043)	(0.072)	(0.072)
Borrower X's	No	No	Yes	No	Yes
Observations	92	847	847	847	847
R-Squared	0.391	0.007	0.042	0.006	0.057
Pre-mean(AA-B)	5.714	0.925	0.925	0.625	0.625
Pre-mean(C-E)	10.913	0.880	0.880	0.554	0.554

Notes: This table shows how the number of applications, the probability of successfully funding loans, and the probability of accepting offers changed around the interest rate adjustment. The sample includes applications from 23 filing dates before and after the event. Column (1) focuses on the number of applications, which is measured at the day-risk level. Columns (2)-(3) focus on the probability of successfully funding loans, and Prosper defines an application with a percentage of at least 70 percent as a successfully funded application. Columns (4)-(5) focus on the probability of accepting offers. Post is an indicator for days on or after January 13, 2011 and Low is an indicator for ratings below B. Borrower characteristics include all borrower credit variables available, stated income, months employed, and homeownership. Applications are restricted to 3-year loans for borrowers who do not have a prior loan with the platform. Pre-mean (AA-B) is the mean of the dependent variable in the pre-period for higher-rated borrowers, and pre-mean (C-E) is the mean of the dependent variable in the pre-period for lower-rated borrowers. Robust standard errors are in parentheses.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

People may point out that the probability of getting funded likely changed after the event, because even though interest rates drop for all new loans, interest rates for lower-rated applications are much greater than those of the higher-rated ones, and investors tend to fund loan applications submitted by lower-rated borrowers to make higher returns. I check the probability of getting funded around the event (see Table 2.6), but I find even after considering all observable information, there is no statistically significant change in investors' funding preferences. Lower-rated applications are generally less likely to be funded and this holds true even after the event. What's more, there is no convincing evidence showing that either low-risk or high-risk borrowers tend to be more reluctant to accept offers after lowering interest rates, because I find no statistically significant evidence showing that the probability of issuing loans changed after the event.

People may also argue that applicants who submitted applications before the event rescind the applications right after the event and reapply for credits to enjoy relatively lower interest rates. And people may be concerned that this is the reason why I observe an increase in the number of applications after the event. To rule out that possibility, I check the probability of withdrawing an application submitted two weeks before and after the event, ten days before and after the event, and one week before and after the event. If that possibility does exist, I would observe a larger probability of withdrawing as I gradually shrink the window. From Table 2.7, there is no statistically significant evidence showing that applicants are taking time to withdraw and reapply for credits although they would have lower interest rates after the event.

Table 2.7: The Probability of Withdrawal around the Interest Rate Adjustment

	±7 days	±7 days	±10 days	±10 days	±14 days	±14 days
Post	-0.046	-0.030	0.008	0.011	-0.040	-0.041
	(0.085)	(0.088)	(0.066)	(0.066)	(0.057)	(0.058)
Low	-0.077	-0.012	0.018	0.139*	0.006	0.113*
	(0.078)	(0.104)	(0.059)	(0.076)	(0.055)	(0.068)
Post x Low	0.144	0.146	0.037	0.050	0.036	0.037
	(0.104)	(0.106)	(0.081)	(0.081)	(0.071)	(0.072)
Observations	246	246	389	389	541	541
R-Squared	0.012	0.044	0.005	0.053	0.002	0.031
Pre-Mean (AA-B)	0.148	0.148	0.158	0.158	0.194	0.194
Pre-Mean (C-E)	0.196	0.196	0.193	0.193	0.179	0.179

Notes: This table shows how the probability of withdrawal changed around the event. The samples in Columns (1)-(2) are applications submitted by applicants one week before and after the event. The samples in Columns (3) -(4) are applications submitted by applicants ten days before and after the event. The samples in Columns (5)-(6) are applications submitted by applicants two weeks before and after the event. Applications are restricted to 3-year loans for borrowers who do not have a prior loan with the platform. Pre-mean (AA-B) is the mean of the dependent variable in the pre-period for higher-rated borrowers, and pre-mean (C-E) is the mean of the dependent variable in the pre-period for lower-rated borrowers. Robust standard errors are in parentheses.

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

2.4.5 Detrending

According to Column (1) in Table 2.6, there is a statistically insignificant increase in the number of applications submitted by lower-rated borrowers after lowering interest rates. Standard errors are large and the estimation is noisy. People may worry that there are three major holidays (Christmas, New Year, and Martin Luther King, Jr. Day) in my research window and the number of applications is highly likely to vary on weekdays compared to weekends. Even though I am using a short research window, it is useful to detrend data to remove time effects that might cause some kind of distortion. I use data from one year before and five years after my research window to detrend data around the event. I am investigating whether I can observe a change in the number of applications driven by the event after controlling for holiday fixed effects, day of week fixed effects, and year fixed effects. Since I observe a strong nonlinear upward trend in the number of applications in the raw data, instead of directly using the change in the number of applications, I calculate the percentage change in the number of applications. Column (1) in Table 2.8 shows the percentage change in the number of applications in my research window. After detrending the data around the event, the regression results are consistent with the estimation when just using the raw data from my sample period. I do observe that after lowering the interest rate there is an about 50 percent increase in the number of applications submitted by lower-rated borrowers, also the standard errors are much smaller than before, but it is still not statistically significant. Even after detrending the data, the estimation is still noisy and affected by some unobservable factors in the platform.

Table 2.8: The Number of Applications around the Event No Detrend vs. Detrend

	(1)	(2)	(3)	(4)
Post	0.922	0.636	-12.043	0.262
	(0.833)	(0.850)	(18.275)	(0.168)
Low	5.199***	6.895***	-6.712	0.045
	(1.467)	(1.328)	(17.746)	(0.158)
Post x Low	3.165	1.469	9.139	0.198
	(1.969)	(1.870)	(21.122)	(0.221)
Borrower X's	No	No	No	No
Observations	92	82	92	92
R-Squared	0.391	0.473	0.008	0.1289

Notes: The number of applications is measured on the day-risk level. Column (1) shows how the daily-risk level number of applications changed around the event. Column (2) shows the number of applications around the event after taking out applications submitted on holidays (December 25, 2010; December 26, 2010; January 1, 2011; January 2, 2011; and January 18, 2011). In order produce the analysis in Columns (3)-(4)of this table, I generate 7 groups: group -1 is from December 20, 2009 to February 3, 2010, group 0 (my sample period) is from December 20, 2010 to February 3, 2011; ...; and group 5 is from December 20, 2015 to February 3, 2016. Column (3) shows how the number of applications change for group 0 after detrending the data in group 0 by employing data from the other 6 groups. Column (4) shows the percentage change in the number of applications for group 0 after detrending the data in group 0 by employing data from the other 6 groups. In Columns (1)-(2), Post is a time dummy that equals 1 after Jan 12, 2011, and equals 0 otherwise. In Columns (3)-(4), Post is a time dummy that equals 1 after Jan 12 of each group and equals 0 otherwise. Low indicates whether a loan is categorized by Prosper as lower-rated (C-E). Loans are restricted to 3-year loans for borrowers who do not have a prior loan with the platform. Robust standard errors are in parentheses.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

It might be inappropriate to take out data given that I have a short window with a relatively small sample size. However, there are few applications around Christmas and New Year and many applications submitted on Martin Luther King, Jr. Day in the research window. Christmas and New York are major holidays in the pre-period while Martin Luther King, Jr. Day is in the post-period. So I take out applications submitted on holidays to see if the estimation on the number of applications is still consistent with the previous result.[4] According to the coefficients displayed in Column

[4]I take out observations on December 25, 2010; December 26, 2010; January 1, 2011; January 2, 2011;

(2), the effects on the daily risk level number of applications driven by the policy would not be affected by borrowers' significant abnormal behaviors on holidays.

2.4.6 The Selection on Unobservable Information

Recalling the decomposition in Section 2.2 of this project, I have shown that for lower-rated borrowers the sign on the left-hand side is strictly negative. And for the right-hand side, I start with assuming that $\frac{\partial d(r,X,\theta)}{\partial r}$ is weakly positive, and I prove that there is no statistically significant change in the distribution of observable characteristics conditional that one is a lower-rated borrower, i.e. $\frac{\partial d(r,X,\theta)}{\partial X}\frac{\partial X}{\partial r} = 0$. Therefore, the sign of the selection on unobservable information, i.e. $\frac{\partial d(r,X,\theta)}{\partial \theta}\frac{\partial \theta}{\partial r}$ has to be negative and large enough to cancel the direct effect. I conclude that, for lower-rated borrowers, after controlling for the direct effect, the selection on observable information, and all the other confounding effects, my empirical evidence is generally consistent with the views held by the de Meza-Webb model. In other words, there is a negative correlation between the interest rate and the default rate, and the effect is driven by selection on unobservable information.

2.5 Conclusion

In this project, I have used an experiment on Prosper to investigate the selection of interest rates and repayment in online credit markets. This experiment involved

and January 18, 2011, because by plotting the number of applications, I found too few observations on the first 4 dates and too many observations on the last one lowering interest rates for all new loans after January 13, 2011. The experiment causes lower-rated borrowers to be more likely to default. By decomposing the total effect of lowering interest rates on default, I find that the increase in default is driven by the selection on unobservable information. My empirical evidence is generally consistent with what has been predicted by de Meza and Webb (1987). Even though the well-known prediction of Stiglitz and Weiss (1981) adverse selection model has had considerable influence on finance, development, and macroeconomics, policymakers are expected to test the micro assumptions behind the macro models before they make policy changes based on these.

Chapter 3

Understanding and Quantifying Aspects of the US Black-White Wealth Gap

3.1 Introduction

Racial inequality, particularly for Black Americans, is a fact of life in the United States. Manifestations of this inequality include, most painfully, the treatment of Black Americans in various components of the criminal justice system and in civil society more generally. Economic conditions are another major aspect of inequality and certainly contribute to and exacerbate other problems faced by Black Americans. Access to housing, jobs, and education are all part of the complex of economic inequalities in the United States. While Black-White inequality of incomes is not trivial and has been slow to improve, the problem is even worse for wealth, which depends on the accumulation of savings over time, potentially over generations. Wealth inequality is greater and more persistent than income inequality. Arguably, this situation is a matter of national economic concern and not just a problem for marginalized or disadvantaged groups (Tippett et al., 2014).

Factors contributing to Black-White wealth inequality in the US have received scholarly attention for several decades, including documentation of the extent of the inequality (e.g., Parcel, 1982; Horton, 1992; Oliver and Shapiro, 2006; Conley, 1999; Keister, 2000a, b; Avery and Rendall, 2002; Shapiro, 2004). Many studies focus on one or more causes or proximate contributors to racial wealth inequality, including unequal access – often traceable to discriminatory practices – to housing, jobs and education. There are also analyses that emphasize differences in preferences, such as risk aversion, or differences in cultural and social norms, which can also affect wealth accumulation. Herring and Henderson (2016) provide a relatively recent and comprehensive survey of the literature, and in doing so, they divide explanations into two classes: (1) cultural and behavioral and (2) structural and unequal ownership opportunity. However, as they go on to point out and discuss, these classes of explanations are not mutually exclusive, nor is it always possible to place particular explanations neatly into one category or the other. As Herring and Henderson point out, claims of behavioral or cultural causes, such as attitudes to saving or to ownership of risky assets, can often be traced to structural factors. For example, apparent risk aversion in asset choices – a behavioral or cultural explanation – may actually be the result of a lack of information, high transaction costs and low wealth levels, all structural factors. It is also important to recognize that cultural norms and behavioral factors may be malleable rather than fixed. Indeed, treating them as ascriptive characteristics of a particular race is a problematic scholarly approach.

While almost all studies recognize the multifaceted nature of Black-White wealth inequality in the US, many do concentrate on a specific potential contributor to this inequality, such as homeownership, education, or asset portfolio choices. In that process of concentration, the complicated factors that determine this inequality can be lost. This problem has led Darity et al. (2018) to describe what they label as "10 myths" about closing the Black-White wealth gap – suggested policy responses that view the problem in monocausal terms – and counter these oversimplified perspectives. For example, various studies highlight disparities in education levels, homeownership rates, or financial decision-making skills, and these studies implicate policies that can correct these disparities and can lead to significant reductions in the wealth gap. The problem, of course, is that individual factors may be intertwined in how wealth is determined, and several causes may need to be tackled together to actually make a difference. It is also possible that major contributors to the wealth gap lie outside this set of observable variables and even concerted policy responses may not have significant impacts. In particular, structural inequalities may be the result of individual preferences for discrimination, collective norms, or other societal characteristics that are difficult to measure or to change through government policies.

Of course, empirical analyses such as those of Keister (2000b), Herring and Henderson (2016) and many others attempt to uncover the multiple contributors to Black-White wealth inequality in the US. Our analysis is in that spirit, using regression analysis to examine multiple contributing factors to the racial wealth gap. It is closest in methodology to Herring and Henderson and uses the same data source. Our innovations and contributions compared to their study are as follows. We use the 2016 Survey of Consumer Finances rather than the 2013 data. We use a more appropriate nonlinear specification instead of a linear estimation that potentially gives too much weight to higher wealth households because of the highly skewed distribution of wealth. Most importantly, we quantify the contribution of different variables to the wealth gap by using our regression results to estimate conditional and positional effects rather than making unconditional comparisons that can be misleading. Studies such as Herring and Henderson do this as well, but we would argue that our methodology provides more accurate quantitative estimates.

To illustrate the methodology with the case of homeownership, in the data, homeownership rates are about 60 percent higher for White Americans than for Black Americans. Conditional on homeownership, the average home owned by a White household is worth about \$308,000 versus \$180,000 for a Black household. But this comparison does not control for other characteristics. Once one does so through multivariate regression, the estimated addition to wealth associated with homeownership is about \$200,000 for an average White household and about \$81,000 for an average Black household. However, if a Black household has the other measured characteristics of an average White household, homeownership does not contribute \$200,000 to their net worth, but only \$104,000. This disparity can be interpreted as an indicator of structural inequalities, and is indicative of how policies targeting single causes of wealth disparities may be insufficient. In another example, we find that while having a college degree is associated with higher wealth for both groups, the wealth impact is much smaller for Blacks, and the absolute wealth disparity increases with a college degree after controlling for other household characteristics in a multivariate regression.

In addition to these calculations based on individual factors, we use a modified version of the Blinder-Oaxaca decomposition to separate out the contributions of endowments or characteristics to wealth disparities versus structural factors that are reflected in the regression coefficients. This approach was originally applied to Black-White income differences by Winsborough and Dickinson (1971) and subsequently popularized by Blinder (1973) and Oaxaca (1973) in the context of gender wage gaps. Darity et al. (1995) uses such decompositions to examine both gender and race/ethnicity in income gaps. We use a modified version of this method for our nonlinear specification. This decomposition sheds further light on the problem of policies to address the racial wealth gap, indicating that even a joint attack on multiple sources of inequality (homeownership, education, asset ownership patterns, and so on) may not be sufficient, unless the processes that translate these characteristics into wealth are equal across groups.

A final contribution of our paper is to examine the role of class to some extent. While this has been discussed by many scholars, our specific contribution is to perform quantile regression analysis, which allows one to see how different factors such as education vary in their impacts over the range of the wealth distribution. As might be expected, there are such differences, but not for all characteristics. Importantly, the variation across the wealth distribution often differs for Black and White households, indicating that there are disparities that are more likely to be the result of race rather than class.

In our analysis, the use of cross-section data means that questions of causality and dynamics have to be treated with due caution. Nevertheless, the patterns that we uncover are informative, and, as illustrated above, our regression analysis goes beyond simple comparisons of different characteristics across races, and our specific results include several innovations. Longitudinal data, such as from the Panel Study of Income Dynamics (PSID), allows for tracking the accumulation of wealth by households over time, and the PSID has been used in several empirical analyses, but there are difficulties caused by changes in the survey methodology, and some other challenges of the data (e.g., Shapiro et al. 2013). Many studies of wealth inequality have therefore used the SCF, which is conducted every three years and collects a rich variety of data on household wealth and on characteristics that can have a bearing on wealth levels.

The rest of the paper is organized as follows. The next section describes our dataset and its main characteristics, in particular the summary statistics for variables of interest. This is followed by a section describing our empirical strategy, which has three parts: regression analysis of the wealth data along with the income data; an overall decomposition of wealth differences; and quantile regressions to examine variation over the wealth distribution. We use regression analysis to calculate quantitative contributions of various individual or household characteristics to the Black-White wealth gap. These results are presented in the fourth section. We find that a combination of inheritance, education, and occupation is significantly related to differences in wealth levels between Black Americans and White Americans, but there is no single indicator, such as level of education, that can be changed to make a major dent in the wealth gap. In fact, in most cases, such changes are not even predicted to reduce the gap. One of our new results is that differences in financial literacy, as measured in the SCF, are not a useful explanator of the racial wealth gap. The regression results from section 3.4 are used in section 3.5 to quantify the effects of different individual characteristics such as homeownership and having a college degree. We show that changing individual characteristics like these, while holding other factors constant, does not imply a reduced wealth gap in the cross-section data.

Overall, structural factors or unmeasured quality are also likely reasons for the wealth gap, since much of it is not explained by the overall difference in characteristics, as estimated by a modified Blinder-Oaxaca decomposition. That analysis is presented in section 3.6. The Blinder-Oaxaca approach decomposes the mean gap between groups into a component explained by differences across groups in the means of the independent variables and a component due to differences in the effects of the independent variables. Section 3.7 discusses some of the results from quantile regression estimates, which allow for differences in impacts at different points of the wealth distribution. This allows for a consideration of the extent to which the differences in racial groups are associated with class as proxied by position in the wealth distribution. In section 3.8, we provide an overall discussion and interpretation of our results. Sometimes, the results of empirical exercises such as this one are interpreted in terms of broad categories such as culture, behavior, or societal structures, even labeling regression models as "post-racial" (reflecting restrictive assumptions about the impact of characteristics such as education on wealth) or allowing for “structural racism.” While cautioning against making broad interpretations, our results are consistent with the view that the playing field for wealth accumulation is not level for Black Americans versus White Americans. In particular, differences in pathways to education and employment, as well as different starting points, are important factors in the racial wealth gap. The final section of the paper is a summary conclusion.

3.2 Data

The Survey of Consumer Finances (SCF) is conducted every three years, primarily sponsored by the US Federal Reserve Board.[1] A nationally representative data set is achieved by combining a geographically based random sample with a separate sampling of wealthy families, which is derived from tax returns. We use the 2016 survey for our analysis. Overall, this data set comprises detailed survey responses from just over 6,000 households. Respondents indicate their racial group, and from the overall sample, we select only households that identify as Black or White. Furthermore, we trim two kinds of outliers. Following McKernan et al. (2014), we remove the upper and lower 0.25 percent of the wealth distribution from the sample. Second, we remove households headed by individuals younger than 26 or older than 79. The dataset uses multiple imputations to fill in missing values for some households, and we use the

[1]The actual survey is conducted by the National Opinion Research Center at the University of Chicago. Also, the US Treasury Department cooperates with the US FRB in the survey.

recommended procedure for combining these imputations into single observations for regression analysis.[2] Our analysis therefore proceeds with 4,400 observations, of which 3,631 are classified as White households and the remaining 769 are Black. Due to the survey methodology, the ratio of Black to White households in the sample is reflective of the US population ratio (about 1 to 5).

While the survey collects a very large set of information about household finances and characteristics, we focus on a relatively small number of variables, that we think are sufficiently representative of important factors in shaping the wealth distribution. In other words, we are seeking a relatively parsimonious analysis of the relationship between household characteristics and wealth levels. The household characteristics variables used in our analysis are listed in Table 3.1, which separately reports means and standard deviations of the Black and White subsamples. Our regression analysis is conducted by analyzing these subsamples separately, which avoids imposing any implicit restrictions on how household characteristics affect wealth across the two groups. Averages for the entire sample are not reported in Table 3.1, but are simply sample-size-weighted averages of the two columns in the table. Most of the variables used are 0-1 characteristics, and the numbers in the table in such cases are proportions. For example, 26.8 percent of our sample consists of households with a female head, but there is a substantial difference between Blacks (47.1 percent) and Whites (21.8 percent). The

[2] Full details of sampling methods and imputation procedures can be found at the SCF website, https://www.federalreserve.gov/econres/scfindex.htm. We compared our results using the appropriate weighted combination of imputations with those for a single imputation, and there were no significant differences. For some parts of the analysis we had to use single imputations because of restrictions on the estimation method.

exceptions to the 0-1 rule are the variables $FinLit$, which is the number of financial literacy questions answered correctly; Age, which is reported in years; and Kids, which is the number of children in the household. The financial literacy questions were added to the survey for the first time in 2016, and consist of three questions meant to test knowledge of asset choices and risk, interest rate compounding, and inflation (Bricker et al., 2017).

Two classes of variables in Table 3.1 have multiple categories, those for education and for occupation, and there is an omitted category in the table. In the case of education, the omitted category is those with less than a high school diploma.[3] In the case of occupation, the omitted category is those who do not have a job, which is a somewhat diverse group, including the unemployed, retirees, students and homemakers.[4] For both these categories, the relevant marginal impact calculation in the regression analysis will be with the omitted category. This will be clear in our discussion of the results.

We also summarize the wealth variables used in the analysis in Table 3.2. For our regression analysis, we focus on net worth, namely, the difference between total assets and liabilities as the outcome variable, but Table 3.2 also reports wealth components. For Net Worth, Assets and Debts, the means are calculated and reported based on the entire sample for each racial group. However, for the three reported components of assets (house, stocks and business), the reported means are conditional on holding that type of asset.To illustrate, for the 5.2 percent of Black households in the sample

[3]Hence, this percentage is 8.1 for Whites and 16.6 percent for Blacks in the sample.

[4]Since this grouping is so broad and heterogeneous, we experimented with breaking down this category further, but it did not affect our results.

Table 3.1: Summary Statistics by Race

	(1)	(2)
	Mean/ SD	
	Black	White
Female head of HH	0.468	0.219
	(0.499)	(0.414)
Bankruptcy	0.038	0.032
	(0.192)	(0.175)
Spending exceeded income	0.199	0.141
	(0.399)	(0.348)
Have stock	0.051	0.176
	(0.219)	(0.381)
Have business	0.069	0.168
	(0.254)	(0.373)
Have home	0.460	0.746
	(0.499)	(0.436)
Receive inheritance	0.083	0.262
	(0.275)	(0.440)
Have pension	0.473	0.653
	(0.500)	(0.476)
Financial Literacy	1.961	2.315
	(0.862)	(0.815)
HS/GED only	0.287	0.248
	(0.453)	(0.432)
Some College	0.318	0.272
	(0.466)	(0.445)
College and Above	0.229	0.399
	(0.421)	(0.490)
Managerial/Professional	0.231	0.334
	(0.422)	(0.472)
Technical/Sales/Services	0.279	0.193
	(0.449)	(0.395)
Other Job	0.142	0.167
	(0.349)	(0.373)
Age	50.030	52.736
	(14.350)	(14.395)
Children	0.894	0.730
	(1.188)	(1.074)
Observations	771	3635

Note: Numbers in parentheses are standard deviations.
Source: Constructed from SCF 2016 dataset

that own stocks (Table 3.1), the average value of their holdings is $32,630.69 (Table 3.2), while for the 17.5 percent of White households in the sample that own stocks (Table 3.1), the average value of their holdings is $230,962.70 (Table 3.2). Overall, the average

Table 3.2: Net Worth and Components by Race

	Black	White
Net Worth	118,253.60	712,069.40
	(486,625.50)	(1,973,488.00)
Assets	177.358.00	828,207.40
	(527,964.00)	(2,047,296.00)
Debts	59,104.42	116,137.90
	(102,071.10)	((217,516.40)
Houses	180,118.60	308,456.00
	(211,017.90)	(471,631.70)
Business	312,534.90	774,480.20
	(1,294,649.00)	(2,034,949.00)
Stocks	32,630.69	230,962.70
	(93,177.41)	(689,096.80)

Note: Means for Houses, Business and Stocks are conditional on owning assets in that class. Numbers in parentheses are standard deviations.
Source: Constructed from SCF 2016 dataset

ratio of Net Worth for Whites versus Blacks is a little more than 5 to 1. Conditional on ownership, the inequality in the three reported components of wealth is relatively greater for stocks, and relatively less for business assets and for houses. However, this inequality measure needs to be combined with the inequalities in proportions to have a full picture. For example, the Black homeownership percentage is about 60 percent of that for Whites, and Black owners' homes are on average worth about 60 percent of White owners' homes, so, averaged over the whole sample, for Blacks, homes are only about 36 percent of their mean aggregate asset value, as compared to homes for Whites. Bricker et al. (2017) provide a detailed descriptive overview of various features of the 2016 SCF data, including comparisons to the results of the previous survey.

3.3 Empirical Strategy

The empirical analysis has three parts. In the first part, we estimate OLS regressions for each of the Black and White subsamples. These regressions are used to calculate the individual marginal contributions of key variables related to education, occupation, and asset composition of wealth, with the goal of quantifying impacts that are due to characteristics or endowments (e.g., level of education), versus those that can be associated with the racial group (the differing regression coefficients across groups). Because these calculations are performed in the context of regression estimates, they are more informative than just comparing sample characteristics such as means or medians. The precise nature of our calculations may be somewhat novel in the context of quantifying contributors to the US Black-White wealth gap.

The second part of the empirical analysis is a modified Blinder-Oaxaca decomposition, which extends the idea behind the individual variable calculations to the overall regression, estimating the separate contributions of endowments and coefficients to the Black-White wealth gap. More specifically, the Blinder-Oaxaca approach decomposes the mean gap between groups into two components, one explained by differences across groups in the means of the independent variables and another that is due to overall differences in the effects of the independent variables, as captured in the regression coefficients. A modification of the standard approach to this kind of decomposition is helpful in our case, because our dependent variable is a nonlinear transformation of net worth: we use the method of Kaiser (2016) for this analysis. This modification allows for the decomposition to be measured directly in dollar terms, rather than in transformed units. The details of the methodology are presented along with the results, in section 3.6.

Finally, we estimate quantile regressions for each of the Black and White samples. This allows us to get at issues of class, at least partially, since we can compare marginal impacts at various points of the wealth distribution, both for each group, and also across groups. Rather than predicting the conditional mean, as in standard least-squares regressions, quantile regression methods use conditional quantiles, the median being the most prominent example (Herring and Henderson, 2016; Thompson and Suarez, 2015). Our analysis uses the deciles of the net worth distribution, similar to Maroto (2016), who analyzes earlier SCF data, though the details of her implementation are different.

In our initial regressions, the dependent variable is the natural logarithm of wealth, measured as total net worth, but with a shift amount equal to the absolute value of the minimum net worth in the sample. This is necessary to avoid dropping observations with negative net worth.[5] It is important to repeat the caution that causal inferences in such a cross-section regression exercise have to be treated with care, since almost all the right-hand side variables are more or less endogenous. The processes by which wealth levels affect portfolio choices, such as whether or not to hold stock, are relatively straightforward. For example, at low wealth levels, a household may prioritize

[5] A popular alternative is the inverse hyperbolic sine transformation, which does not require a data-dependent choice of a shift value, but for this particular data set and analysis, our approach is simpler and more intuitive.

safer financial assets, or non-financial assets which provide shelter or a livelihood.[6] In other cases, such as the level of education attained, one can also posit a fairly direct positive impact from wealth to education, especially for acquiring a college degree or advanced degree, since those can require substantial outlays of financial resources.[7]

In the case of other variables, the linkages are more complicated. While being poor may not directly affect whether a household has a female head, and since labor market gender inequalities suggest that female-headed households will tend to be poorer, both wealth and family structure are jointly determined by complex circumstances, including marriage patterns and even differential incarceration rates across racial groups. In another example, age is exogenous, and does not directly determine wealth, but older people in the cross-section have had more years to accumulate savings, and are likely to be richer. Perhaps the variable that is most clearly predetermined within the set of variables in Table 3.1 is whether the household had received an inheritance or not. This variable is at least partly an indicator of the wealth that was transferred from the preceding generation.

3.4 OLS Regression Results

Turning to the regression results, presented in Table 3.3, we get the typical result that households with a female head have lower wealth, in this case controlling for

[6] Our approach, using only indicators of specific asset holdings, leads to a different conclusion than Thompson and Suarez (2015), who note that "Wealth differences between black and white families are completely due to different asset holdings."

[7] Of course, this relationship can be confounded in a cross-section by differences in access to and use of student loans.

all the other included characteristics. The value of the negative impact of this characteristic is smaller in magnitude for Black households, but this is in percentage terms, so is possibly a result of these households having much lower wealth on average, as the baseline. Therefore, this relative magnitude can be interpreted as an indicator that other (non-female-headed) Black households are relatively more disadvantaged compared to corresponding White households. It is also true that there is a much higher percentage of female-headed households in the Black sample than the White sample, implying more widespread disadvantage.[8] Broadly similar effects are seen for the characteristics of having declared bankruptcy and having spending exceeding income: the negative impact for Blacks is smaller in magnitude, but not proportionately to the average wealth disparity across the groups. Again, in both cases, the proportion of Blacks with this characteristic is higher than for Whites (though not as dramatically as it is for female-headed households).

[8] The results should definitely not be viewed as underplaying the economic challenges faced by black women, and it has to be emphasized that regression coefficients are average effects – what is happening in the tail of the distribution can be what matters for welfare judgments. A useful discussion based on PSID data is Zaw et al. (2017).

Table 3.3: Wealth Regressions by Race

	log(Net worth + k)	
	Black	White
Female head of HH	-0.119***	-0.214***
	(0.033)	(0.030)
Bankruptcy	-0.171**	-0.408***
	(0.080)	(0.067)
Spending exceeded income	-0.098**	-0.144***
	(0.038)	(0.034)
Have stock	0.022	-0.001
	(0.284)	(0.125)
Have business	-0.210	0.249*
	(0.270)	(0.142)
Have home	0.176	0.090
	(0.127)	(0.100)
Receive inheritance	0.104	0.103***
	(0.069)	(0.036)
Have pension	0.063*	0.078***
	(0.034)	(0.027)
Financial Literacy	0.011	0.112***
	(0.018)	(0.015)
HS/GED only	0.022	0.021
	(0.047)	(0.048)
Some College	0.086*	0.107**
	(0.048)	(0.048)
College and Above	0.058	0.386***
	(0.064)	(0.055)
Managerial/Professional	0.060	0.168***
	(0.060)	(0.048)
Technical/Sales/Services	0.029	0.032
	(0.044)	(0.038)
Other Job	0.006	0.059
	(0.053)	(0.041)
Age	0.004***	0.013***
	(0.002)	(0.002)
Children	0.028**	0.051***
	(0.014)	(0.012)
Age x have stock	0.007	0.011***
	(0.005)	(0.002)
Age x have business	0.014**	0.006**
	(0.006)	(0.003)
Age x have home	0.004*	0.008***
	(0.002)	(0.002)
College and above x Managerial	-0.028	0.037
	(0.085)	(0.057)
College and above x Receive inheritance	0.310***	0.117**
	(0.118)	(0.054)
Constant	11.822***	11.140***
	(0.110)	(0.104)
Observations	769	3,625
Adj R-Squared	0.458	0.529

Standard errors in parentheses, value of k = 173255.90

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Having holdings of stock, or owning a home or business, are all indicators of the asset portfolio of the household. We also interact each of these indicators with age, to reflect the fact that older households are likely to have higher wealth in these categories, and hence overall – of course, it does not make sense to include the asset values directly, since they are components of the dependent variable. Stock ownership and business ownership are uncommon in the sample, but the relative proportions of White ownership to Black ownership are both very high. The coefficient of stock ownership is not significantly different from zero for either group, but the interaction terms of stock ownership and age indicate that stock holding matters for wealth for older White Americans but not so clearly for older Black Americans (the latter coefficient is positive but not statistically significant). Business ownership, on the other hand, has a very different relationship to net worth for Black Americans than for White Americans – White business owners are considerably wealthier than their Black counterparts, though the advantage reduces somewhat with age, since the Black interaction term coefficient is somewhat larger.[9]

With respect to home ownership, a variable that is often the focus of analyses of Black-White wealth gaps, the proportional advantage of White Americans versus Black Americans in rates of homeownership is not as high as for stock ownership or business ownership, but the absolute difference in rates of ownership is higher. However, based

[9]Because of the use of a shift term and the logarithmic transformation of net worth in the regression, the interpretation of the coefficient requires some care. The absolute marginal impact is not the coefficient itself, but it is increasing in wealth. The percentage marginal impact, however, is decreasing in wealth. All of these complications are handled in our subsequent calculations comparing the effect on net worth of, say, business ownership for Black Americans vs. White Americans.

on the coefficients of homeownership and those of the interaction terms with age, one can say that the connection between home ownership and wealth is not as dissimilar for the two racial groups as it is for the other two types of asset.

By itself, the role of having received an inheritance is somewhat ambiguous: the magnitude of the coefficient of this indicator is similar for Black Americans and White Americans, but it is estimated less precisely for Black Americans, and is not statistically significantly different from zero. However, there is a wide disparity in the proportions of Black and White households that have received inheritances (Table 3.2).[10] We also discuss the role of inheritance further in the context of education, where its impact shows up in a different way.

Turning to demographic characteristics, age has the expected positive coefficient for both groups, but the baseline coefficient (without considering age interaction terms) for White households is over three times that for Black households. In other words, if we compare households without stock holdings, or without business or home ownership, older White households in the cross-section have accumulated more wealth than their Black counterparts. Since this is a cross-section, what is being captured is past wealth accumulation, rather than contemporaneous differences. For homeowners and stockholders, this age effect is reinforced, as indicated by the interaction terms, and only for business owners does the relative age effect switch, although in that case the baseline for business owners indicates a large wealth advantage for White business-owner

[10] Again, it is also the case that the nonlinear specification means that, for a given coefficient value, the dollar wealth difference will be higher at higher wealth levels, and the average White household is much richer than the average Black household.

households, which is much greater than any differential age effect.

Another demographic variable, the number of children in the household, is also positively associated with wealth in these regressions. In this case, the causality arguably operates in the opposite direction, since children do not contribute to wealth, whereas wealthier households are plausibly more likely to have more children. Of course, there are other factors at work beyond the capacity to support children, including "quality" and time-allocation decisions, and in dynamic contexts, the number of children per household falls as societies become wealthier, so there is no simple explanation for this particular cross-section result.

The educational achievement and occupational variables provide some striking results. Having a high school diploma or equivalent is not associated with a net worth that is significantly higher, compared to not having completed high school. A significant positive relationship exists for having some college versus the baseline of not having completed high school, and the coefficients are similar in magnitude for Black and White households. For households where the household head has a college degree or advanced degree, the results are interesting. The coefficient for White households is quite high, and statistically significant. However, for Black households, it is much smaller in magnitude, and it is insignificant. However, if one takes account of the interaction of this education term with having an inheritance, the difference in the coefficients narrows considerably. This is suggestive of a scenario in which Black Americans and White Americans do not have access to college education of similar quality, for financial or social reasons. This is not directly observable, but may be proxied by the more equal returns of Black Americans who have received an inheritance, versus White Americans in the same category.

In the case of occupational characteristics, White Americans are much more likely to be in managerial or professional occupations, and according to the regression results, this is associated with higher wealth for White Americans, but that is not true for Black Americans. On the other hand, for both groups, there is no significant relationship in the regressions between wealth level and simultaneously having a college degree and being in a managerial or professional occupation, as indicated by the coefficients of the interaction term. This contrasts with the results for the interaction of having a college degree with receiving an inheritance, and is consistent with the latter reflecting an overcoming of financial constraints to education, whereas the former is an indicator of the working of the job market.

The 2016 SCF was the first to include questions meant to measure financial literacy. The regression results show that financial literacy is positively related to wealth for Whites, but not for Blacks. Therefore, this provides some new evidence against claims that Blacks have lower wealth because they are less equipped to manage money. Indirect claims of this nature that use differences in asset portfolio structures have been criticized for neglecting the different levels of wealth and resulting choice constraints for Blacks versus Whites. In any case, ascribing racial wealth differences to insufficiently astute financial decision-making seems to be a problematic exercise, both conceptually and empirically.

A final point to note from our wealth regressions is that the constant terms are quite similar across Black and White households. We remark on this because Herring and Henderson (2016), using a different specification (linear in wealth), but also working with SCF data,[11] find a lower constant term for Black Americans and interpret it as evidence of structural racism, since it is not "explained" by the combination of right-hand side variables. However, differences in the constant term are not necessary to make that exact argument: differences in coefficients and marginal impacts of characteristics such as business ownership or educational achievement levels can also be indicative of such structural inequities. However, even there, the processes that drive such results are not necessarily transparent. We discuss some possible issues of interpretation in the next section.

Since wealth is the result of a process of cumulative earning and saving (combined with exogenous shocks such as an inheritance, health problems, or macroeconomic fluctuations), it is useful to compare the wealth regressions with corresponding income regressions. To facilitate this comparison, we estimate exactly the same specification for the income regressions, reported in Table 3.4, even though some of the variables are less natural to include in this case.[12]

---

[11] Their analysis used 2013 data. We were able to reproduce their results almost exactly, with that data, to benchmark our analysis.

[12] For an analysis of the relationship between income and wealth accumulation, see Aliprantis and Carroll (2019) and Aliprantis et al. (2018).

Table 3.4: Income Regressions by Race

	log(Income + 1)	
	Black	White
Female head of HH	-0.452***	-0.499***
	(0.060)	(0.043)
Bankruptcy	-0.069	-0.026
	(0.140)	(0.095)
Spending exceeded income	-0.110	-0.241***
	(0.068)	(0.049)
Have stock	0.088	0.458***
	(0.543)	(0.172)
Have business	-0.334	0.314
	(0.434)	(0.200)
Have home	0.440**	0.290**
	(0.218)	(0.145)
Receive inheritance	-0.086	0.052
	(0.121)	(0.051)
Have pension	0.534***	0.451***
	(0.058)	(0.037)
Financial Literacy	0.049	0.081***
	(0.031)	(0.022)
HS/GED only	0.167**	0.047
	(0.083)	(0.067)
Some College	0.286***	0.190***
	(0.084)	(0.069)
College and Above	0.494***	0.543***
	(0.114)	(0.078)
Managerial/Professional	0.315***	0.610***
	(0.106)	(0.068)
Technical/Sales/Services	0.374***	0.436***
	(0.078)	(0.055)
Other Job	0.520***	0.478***
	(0.090)	(0.058)
Age	0.005*	0.008***
	(0.003)	(0.003)
Children	0.116***	0.097***
	(0.024)	(0.017)
Age x have stock	0.006	-0.001
	(0.010)	(0.003)
Age x have business	0.008	-0.003
	(0.009)	(0.004)
Age x have home	-0.004	0.000
	(0.004)	(0.003)
College and above x Managerial/Professional	0.316**	-0.049
	(0.146)	(0.081)
College and above x Receive inheritance	-0.002	0.035
	(0.207)	(0.077)
Constant	9.414***	9.277***
	(0.192)	(0.158)
Observations	769.000	3,625.000
Adj R-Squared	0.459	0.336

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As one might expect, the inheritance indicator has no relationship to income, whereas it has a positive relationship with wealth. Furthermore, the interaction of age with asset portfolio indicators is not significant, although age is positively associated with income levels. There are some striking differences in the relationships of education and occupation to income versus the relationships to wealth. At lower levels of education, the coefficients are now significant, and slightly higher for Black Americans than for White Americans, though this ranking changes for those with a college degree or advanced degree. Nevertheless, and as one might expect, the relationship of education to income is stronger and more equal than the relationship of education levels to wealth.[13] At the same time, the pattern of coefficients measuring the relationship of income to occupation, and the interaction of having a college degree and being in a managerial or professional occupation, is suggestive of a narrower pathway to financial success for Black Americans than for White Americans. This is parallel to the results of the wealth regressions, although from a different perspective.[14]

Another parallel between the wealth and income regressions is the pattern of coefficients of the financial literacy measure. In both cases, there is a positive relationship for White households, but not for Black households. Both cases are suggestive

[13]Even without formal calculations, it should be clear that the benefits of higher incomes through higher education, even in cases where they are proportionately greater for Black Americans than for White Americans (itself not established in the data), can do little or nothing by themselves to reduce existing wealth inequalities. In particular, at the lower end of the income distribution, incomes are unlikely to permit significant saving and wealth accumulation.

[14]Hamilton et al. (2011), using Current Population Survey (CPS) and American Community Survey data from the US Census, make the case that occupational segregation is an important contributor to the lower wages of black men in the US. Jones and Schmitt (2014) provide additional evidence of discriminatory labor market outcomes for black college graduates, using CPS data. They also summarize field experiments by economists which have demonstrated discrimination against those with "black-sounding" names, even in the absence of personal contact.

of differences in opportunity for White Americans and Black Americans: for Black Americans, this very simple measure of human capital is not correlated with financial outcomes, once other characteristics are controlled for, but it does matter positively for White households. In particular, this reinforces the point that differentials in financial literacy are not indicators of differences in the quality of financial decision-making.

Finally, we should note that two of the right-hand side variables in the income regression almost certainly reflect reverse causality. Both having a home and having a pension or pension plan are likely to be a consequence of a higher income, rather than contributing factors to higher income. In the case of a pension plan, we are conjecturing that jobs or occupations that provide such benefits are likely to be more remunerative on average. Although being retired or having a rental property could possibly support the direction of causality from the presence of the characteristic to higher income, these are not typical or prevalent characteristics of the sample households.

3.5 Quantifying Contributions to the Wealth Gap

As noted in the introduction, studies such as those of Keister (2000b), Herring and Henderson (2016), and many others recognize the multiple factors that shape Black-White wealth inequality in the US. However, a precise quantification of how these different factors contribute to the overall wealth gap is often not carried out. In this section, we provide this quantification for a subset of the factors that were included in the regression analysis described in the previous section. Later in the paper, after presenting the results of additional empirical approaches, we will further discuss these quantitative estimates.

Specifically, using the regression estimates reported in Table 3.3, we calculated the differences associated with changes in single characteristics, holding other household characteristics constant, at an "average" level. These are reported in Table 3.5, and provide the basis for the following discussion. To illustrate the interpretation of the numbers in the table, consider the impact of having a college degree or above, versus the baseline of no high school diploma. A household with the characteristics of the average Black household in the sample is predicted to have only $16,160 in additional wealth associated with this additional education, if their wealth is predicted from the sample of Black households. Even if they have the characteristics of the average White household, this only predicts an additional $21,603 in wealth on top of the $16,160.

Table 3.5: Wealth Difference Estimates

	Black Household Regression		**White Household Regression**	
	Black Average	White Average	Black Average	White Average
College degree and above vs. No HS diploma	16,160	37,763	107,420	186,812
Received inheritance vs. not	39,978	70,914	34,094	67,276
Managerial/professional vs. not working	11,403	14,400	44,730	78,878
Owning stocks vs. not	93,679	131,649	178,754	308,161
Owning a home vs. not	81,361	104,248	128,359	199,975
Owning a business vs. not	129,708	187,700	194,781	330,574
Inheritance & College degree and above vs. no inheritance & no HS diploma	118,208	155,474	183,573	292,832
College & Business ownership vs no HS Diploma & no business	35,518	47,979	247,053	386,857
Business & Stock Ownership vs No Business & no stock	145,753	209,720	177,235	297,910

Note: Estimated from SCF 2016 dataset.

But for White households, the impacts of a college education are an order of magnitude greater. For a White household with other characteristics equal to those of an average Black household, having a college degree is associated with an expected gain of $107,420 in wealth. And having the characteristics of an average White household adds a further $79,392 in wealth to that gain. Therefore, in this comparison, having a college degree versus a high school diploma increases the absolute wealth disparity between Black Americans and White Americans.

From Table 3.5, we see similar large disparities for ownership of stocks or a business, homeownership, and being in a managerial or professional occupation. In each case, having that characteristic has a much larger impact for a White household versus a Black household, when both have identical characteristics otherwise, those of the average Black household. The "White premium" in each case is at least 50 percent, and sometimes much greater. Furthermore, for each of these characteristics, if the other characteristics of the household change from the Black average to the White average, the percentage gain is much higher for this hypothetical White household than it is for the corresponding Black household. For the Black household, the percentage gain from having other characteristics match those of an average White household is less than 50 percent for each of these four variables, whereas it is much greater than 50 percent for the White household that has average White characteristics versus average Black characteristics.

By contrast to this general pattern, the disparity of impacts between Black and White households is not present for the case of receiving an inheritance versus not getting one. However, that characteristic is the single exception among the cases reported in Table 3.5. Furthermore, the proportion of Black Americans who do receive an inheritance is much lower than for White Americans. This does suggest that direct increases in wealth have more equal effects across the races, rather than changes that have to work their way through the economic system. For illustrative purposes, we also calculate and report the marginal impact of changing two characteristics together, e.g., owning a business and owning stock, versus owning neither. In these cases as well, we can observe large differences between Black and White households when they have the other characteristics of an average Black household, and these differences increase if the other characteristics change to those of an average White household. Qualitatively, these calculations support the perspective that there is no single change in the situation of Black households that would translate into narrowing wealth gaps.

It is also helpful to benchmark these numbers against comparisons that do not control for other characteristics. This can be done by comparing the numbers in Table 3.5 with those in Table 3.6, which reports sample means for Black and White households, with and without particular characteristics. For example, the average Black household that owns a home has about $243,000 higher net worth than a Black household that does not own their home. This is about three times the marginal impact of homeownership for a Black household, when all its other characteristics are held constant as those of an average Black household, or more than double the marginal impact evaluated at the characteristics of an average White household. Of course, the larger difference in means in Table 3.6 reflects that fact that other characteristics are not being held constant, making it unsuitable for assessing the impact of hypothetical policy changes. Note that, in any of these calculations, the disparity in rates of possessing the characteristic (e.g., homeownership) is not relevant, since comparisons are being made with and without the characteristic.[15]

Table 3.6: Average Wealth by Characteristics

	Black	White
Own Home	258,582.50	994,388.00
Not	15,804.66	151,372.70
Own stock	556,403.40	2,091,193.00
Not	104,591.90	500,234.20
Own business	646,984.20	2,145,517.00
Not	88,843.79	505,240.30
College Degree and above	271,987.50	1,455,631.00
No HS Diploma	84,408.43	330,678.90
Received inheritance	309,978.40	1,235,434.00
Not	110,995.70	618,671.80
Managerial/ Professional	259,930.10	1,273,600.00
Not working	87,701.25	532,384.70
Inheritance and College Degree	533,026.70	2,015,566.00
No Inheritance and no HS Diploma	76,501.41	268,979.00

Note: Estimated from SCF 2016 dataset.

A different kind of benchmarking can be done using the income regressions reported in Table 3.4. The same exercise as for the wealth regressions that was reported in Table 3.5, is carried out for the income regressions (Table 3.7). For many of the individual characteristics, there are no appreciable differences between Black and White households, whether evaluated at the Black average household characteristics or at White average. This is true even when allowing for the fact that magnitudes of income are mostly somewhat lower than those of wealth.[16] The exceptions to this characterization are being in a managerial or professional occupation, and the combination of business ownership either with having a college degree or with owning stocks.

[15]A separate issue is that large changes in homeownership or possessing a college degree could have additional effects through market equilibrium adjustments, but these are arguably of second order concern, given the existing disparities. Falling returns to a college degree as they become more common are not a reason for failing to correct these disparities.

[16]Median incomes in the data are about \$39,000 for Blacks and \$70,000 for Whites, whereas the means are about \$60,000 and \$125,000 respectively. In the case of net worth, however, whereas the White median and mean are about \$183,000 and \$712,000 respectively, which are healthy multiples of income, the Black median and mean are about \$19,500 and \$118,000. The median wealth is lower than the median income, whereas the mean wealth is double the mean income.

Thus, while income differences do not have as stark structural inequalities as wealth differences, they are not absent, and there is nothing in the data that would indicate that higher incomes for Black Americans would have any effect on the wealth gap.

Table 3.7: Nonlinear Blinder–Oaxaca Decomposition

Method	**Difference**	**Percentage**
Using Black coefficients		
Difference in characteristics (E)	234718.00	34.71
Difference in coefficients (C)	441592.20	65.29
Using White coefficients		
Difference in characteristics (E)	538615.60	79.64
Difference in coefficients (C)	137694.50	20.36

Note: Estimated from SCF 2016 dataset.

3.6 Aggregate Decompositions

The previous section analyzed differences in wealth as related to single characteristics, or combinations of characteristics, while controlling for all other household characteristics. It quantified the magnitude and nature of the wealth gap, and its sources, with respect to single characteristics. For example, the wealth gain associated with a college education (versus no high school diploma) was $16,160 for an average Black household and $186,812 for an average White household (Table 3.5). This gap of

\$170,652 can be decomposed in two ways. The additional gain for a Black household with characteristics of an average White household would be \$21,603. The remainder of the gap is then \$149,049, and might be interpreted as the gain from being in the socioeconomic position of an average White household. One could also perform this decomposition in the reverse sequence. The average Black household with a college education would gain a further \$91,260 (\$107,420 minus \$16,160) if they also had White "positionality,"[17] and the remainder of the gap, \$79,392, would be attributed to improvements in other, measured, characteristics associated with the average White household.

The decompositions proposed by Blinder (1973) and Oaxaca (1973) extend the above logic to a consideration of the entire regression equation, so that the changes are associated with simultaneous adjustments of characteristics and of coefficients. While this decomposition approach has been popular in applications such as gender-wage gaps, it has not been used much in analyzing the Black-White wealth gap.

To illustrate, consider the regression equation

$$Y_{ig} = X_{ig}\beta_g + \epsilon_{ig} \; \forall i \in \{W, B\} \tag{3.1}$$

The Blinder-Oaxaca decomposition in this case is given by

[17] We use this term in the sense of the social and political context that shapes identity, rather than the complementary idea that identity shapes one's understanding of that context.

$$\bar{Y}_W - \bar{Y}_B = \bar{X}_W\hat{\beta}_W - \bar{X}_B\hat{\beta}_B = (\bar{X}_W - \bar{X}_B)\hat{\beta}_B + \bar{X}_W(\hat{\beta}_W - \hat{\beta}_B) \tag{3.2}$$

The logic of this decomposition is as follows. The first term measures the overall impact on the dependent variable if an average Black household instead has the characteristics of an average White household. The second term measures the impact on the dependent variable of a household with the characteristics of an average White household shifting from Black to White positionality. Therefore, this decomposition is analogous to our first illustration for the case of the single characteristic of having a college education.

As in the case of the single characteristic, an alternative decomposition is also possible at the aggregate level. The expression is as follows.

$$\bar{Y}_W - \bar{Y}_B = \bar{X}_W\hat{\beta}_W - \bar{X}_B\hat{\beta}_B = (\bar{X}_W - \bar{X}_B)\hat{\beta}_W + \bar{X}_B(\hat{\beta}_W - \hat{\beta}_B) \tag{3.3}$$

In either case, we can think of the first term in the decomposition as the "explained" portion of the gap, since it is attributable to differences in observed characteristics. The second term reflects possible systemic inequalities, which could also encompass unobserved differences in quality of individual characteristics (e.g., just having a college degree may not provide information on the the subjects studies, the quality of the education, resulting access to social and employment networks, and other factors that could have an impact on wealth). Given the more "favorable" coefficients for White households, the "explained" term of the decomposition will invariably be a smaller proportion of the total wealth gap. One could also conceivably average the two decompositions, but it is useful to calculate and report both separately.

One important issue is that the standard decomposition does not provide direct information on the breakdown of the difference in the variable of interest (here, net worth) if the specification is nonlinear. Most applications of the Blinder-Oaxaca decomposition have a dependent variable that has been log-transformed, and the calculations are made using this transformed variable. The nonlinearity issue was first tackled systematically, primarily for discrete dependent variables, by Fairlie (2005) and Bauer and Sinning (2008). Kaiser (2016) provides a detailed approach to handling the case where the dependent variable is subject to a logarithmic transformation, and we adapt that approach, which uses a weighted Poisson quasi-maximum likelihood estimator, in our decomposition.

The decomposition expressions are now in terms of the original dependent variable, and are as follows.

$$\Delta_B^{NL} = \Big[\mathbf{E}_{\beta_B}(Y_{iB}|X_{iB}) - \mathbf{E}_{\beta_B}(Y_{iW}|X_{iW})\Big] + \Big[\mathbf{E}_{\beta_B}(Y_{iW}|X_{iW}) - \mathbf{E}_{\beta_W}(Y_{iW}|X_{iW})\Big] \quad (3.4)$$

$$\Delta_W^{NL} = \Big[\mathbf{E}_{\beta_W}(Y_{iW}|X_{iW}) - \mathbf{E}_{\beta_W}(Y_{iB}|X_{iB})\Big] + \Big[\mathbf{E}_{\beta_W}(Y_{iB}|X_{iB}) - \mathbf{E}_{\beta_B}(Y_{iB}|X_{iB})\Big] \quad (3.5)$$

In the first equation, the first term estimates the impact of shifting an average Black household to the characteristics of an average White household, but retaining the coefficients of the Black new worth regression. The dependent variable is net worth in its original units, and not the log-transformed value. Hence, implementing these decompositions will provide dollar-denominated values.

The results of both decompositions are presented in Table 3.7. The total average Black-White gap in net worth is $676,310. In the first decomposition, just over one third of this gap is attributable to measured characteristics or endowments, and the rest to unobserved or systemic factors. As discussed earlier, this decomposition uses the positionality of Black households as the baseline. Alternatively, if we switch to using the coefficients of the White household regression, then almost four-fifths of the total wealth gap is attributed to measured characteristics or endowments. One can think of these estimates as lower and upper bounds, while a simple average of the two decompositions would yield an estimate of about 57 percent of the gap being attributable to measured characteristics, with the remainder unexplained, or due to unmeasured factors, at the household and/or systemic level.[18]

3.7 Quantile Regression Analysis

Quantile regressions estimate conditional quantiles (cut points of the probability distribution dividing it into equal probability intervals). For example, as an alternative to OLS, which estimates the conditional mean, a quantile regression can be

[18] For completeness, we also estimated the conventional decomposition directly based on the log specification. Those results are very close, giving estimates of about 54 percent and 78 percent attributable to measured characteristics, respectively, when the different assumptions about positionality are used, with a simple average of about 66 percent.

used to estimate the conditional median. To illustrate, Herring and Henderson (2016, Table 3.3) compare results from an OLS regression, predicting mean net worth, and a quantile regression for median net worth. The coefficients are mostly similar in magnitude and statistical significance, and differences in magnitudes reflect differences in how the two methods respond to outliers and the shape of the distribution.[19] In our analysis, we use quantile regressions at the deciles of the distribution of net worth, allowing us to track how the different household characteristics affect net worth at different points of the wealth distribution. We can, at least to some extent, relate our results to class differences, to the extent that they are reflected in position in the wealth distribution: this will provide an important additional dimension to the understanding of the Black-White wealth gap.

Since the complete results of the quantile regressions involve nine sets of regression coefficients for each group, the tables of regression results are relegated to the Appendix. Instead, we present some salient results graphically, to be able to discuss patterns of impacts over the range of the wealth distributions, Each graph plots coefficients for each group across deciles, along with confidence intervals, for a specific characteristic or combination of characteristics. It is important to keep in mind that the same decile for different groups corresponds to very different levels of net worth. For example, as we noted earlier, median net worth for White households is about $163,000

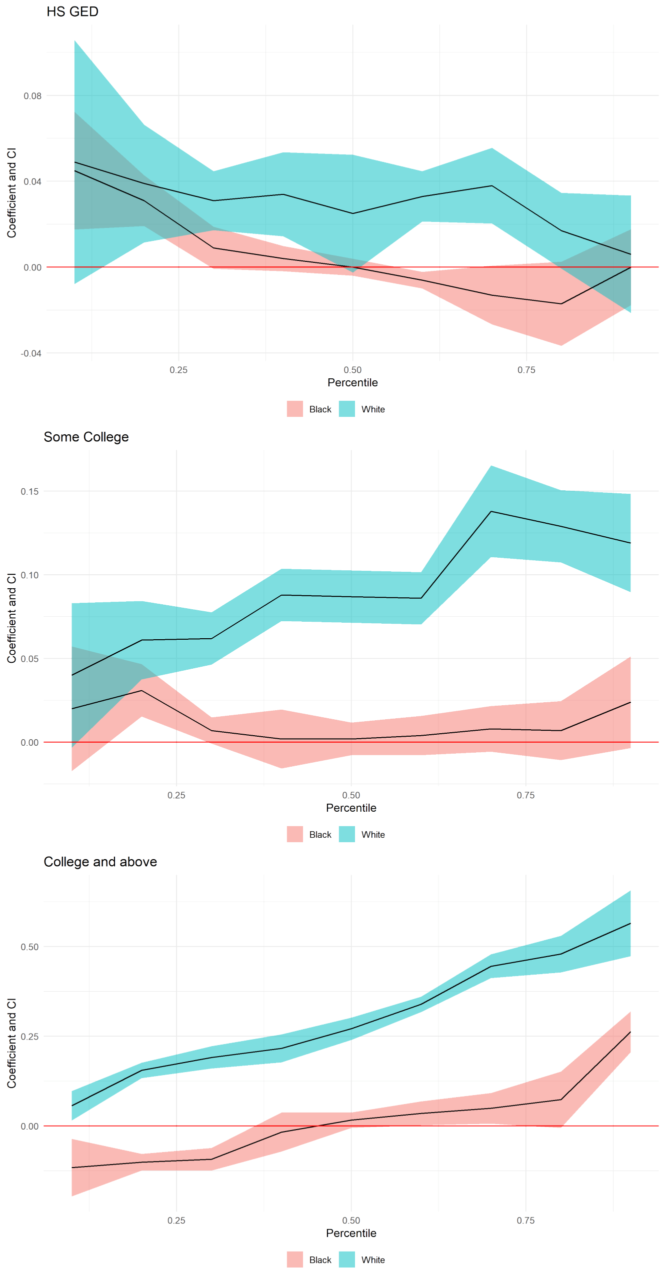
[19]Because those authors use specifications that are linear in wealth, the difference between mean and median is heightened, and the shape of the net worth distribution makes the mean regression very sensitive to outliers. In this case, the quantile regression on the median may have considerable advantages. Our use of log-transformed net worth makes the issue somewhat less important. Maroto (2016) estimates quantile regressions for deciles, but with a linear specification. She also combines these estimates with the standard Blinder-Oaxaca decomposition by decile.

greater than for Black households. Indeed, median net worth for White households corresponds to somewhere above the 80th percentile in the distribution of net worth among Black households. We will use such benchmarks in comparing the results across deciles for the two groups, as well as comparing how impacts of different characteristics on net worth vary over each group's individual distribution.

First, consider the three educational attainment variables. The OLS regressions indicate that, compared to the baseline of not having completed high school, having a high school diploma or equivalency does not have a statistically significant impact on the wealth of Black households or White households. In fact, the magnitudes of the point estimates are also small, so economically insignificant. The quantile regressions suggest a different picture. From the top panel of Figure 3.1, we see that the magnitude of the impact of a high school diploma is greater than the OLS estimate for the middle half of the White net worth distribution. Furthermore, these magnitudes are significantly greater than those for the middle of the Black net worth distribution. A similar conclusion follows if we compare the impact at the middle of the White distribution with the upper deciles of the Black distribution, since these represent similar levels of net worth.

In the case of households with some college, but not a degree, the quantile regression results are once again somewhat different, and illuminating (Figure 3.1, middle panel). The OLS regression results suggest that having some college experience has a positive relationship to wealth, once that is statistically significant (only at the 10 percent level for Black households) and similar across the two groups. But the quan-

Figure 3.1: Quantile Regressions – Education Levels



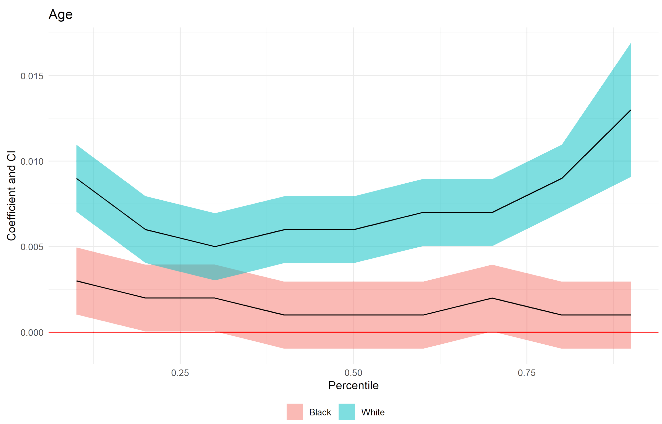
tile regressions suggest a different story. The relationship of this level of educational attainment to net worth is not significant for Black households, but is so for White households. Furthermore, the difference between the groups is significant. Finally, the relationship for White households is stronger in the higher deciles of the distribution. In sum, this level of educational attainment has a completely different connection to the wealth distribution for the two groups.

The highest level of educational attainment in our formulation, a college degree or above, offers yet another lesson. The OLS results (Figure 3.1, bottom panel) indicate no relationship of this category of educational attainment with wealth for Black households, but a very strong one for White households. The quantile regression estimates are broadly consistent with this picture, but with some added nuance. First, for both groups, the relationship is stronger as one moves up the wealth distribution. Second, this positive "slope" of the relationship is somewhat lower for Black households than for White households. Third, the level of the relationship is everywhere lower for Black households, and, except at the upper end of the distribution, is not statistically significantly different from zero. Finally, the magnitude of the impact for Black households at the upper end of their wealth distribution is similar to the magnitude for White households at their median. This compares the groups at similar absolute levels of net worth, and hints at causes that are related to class and to access to a particular subset of the economy from which many Black households may be excluded, even if they have a college degree. There are, of course, alternative explanations possible, and these are discussed in the next section.

Next, consider the relationship of age to wealth. The baseline coefficient, assuming no ownership of a home, or business or stocks, is significantly higher for White

Americans than it is for Black Americans, at every decile of the wealth distribution Figure 3.2). Also, the association of age with wealth is higher for richer White Americans, as would be expected if they have been more successful over their years of wealth accumulation. However, this effect is missing for the Black wealth distribution. It is also noteworthy that the coefficients are higher even when one compares deciles that reflect similar wealth levels. At the median of the White wealth distribution, the coefficient is significantly higher than at the top deciles of the Black wealth distribution.

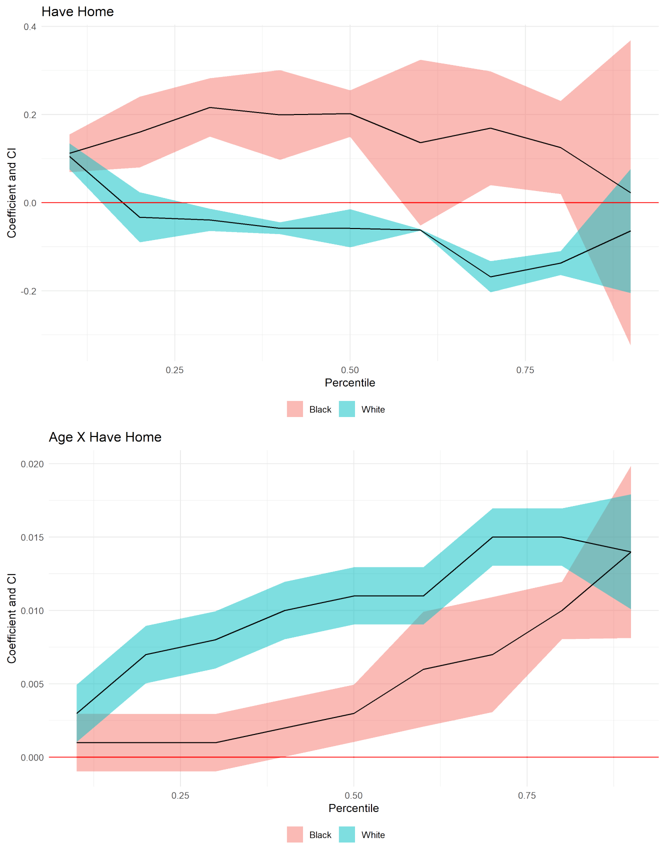
Figure 3.2: Quantile Regressions – Age



For the three asset types – owning a home, business or stocks, we consider each in turn, but discuss the baseline coefficients along with the age interactions. The top panel of Figure 3.3 displays the baseline relationship of homeownership to wealth, but it should be borne in mind that this is at an implied age of zero. Hence, the coefficient should be adjusted for the minimum age in the sample, or for whatever age is being considered. The bottom panel of Figure 3.3 displays the relationship of homeownership interacted with age, which provides the information needed to make the desired adjustments. For example, a Black household at the 40 percentile with a respondent age of 30 has an implied homeownership coefficient of approximately 0.199 + 30x0.002 = 0.259. Correspondingly, a White household with the same age and same percentile for that group's wealth distribution has an implied coefficient of -0.058 + 30x0.010 = 0.242, which is only slightly lower. If, instead, the household age is 50, then the respective measures are 0.299 and 0.442, which represents a much greater impact of homeownership for White households with this age and position in the wealth distribution. Moving up the wealth distribution for either group, for any given age, the implied coefficients actually decrease, but the age interaction effect increases. Since the age effect is larger for White households, with the exception of the 90th percentiles, a broad general implication is that the wealth benefits of homeownership are strongest for older White households.

Homeownership is quite prevalent for both groups, but business ownership and stock ownership are less common, with the proportions of White households being relatively large compared to Black households versus the case of homeownership. In the case of business ownership (Figure 3.4), the baseline impacts are almost always greater for White households than for Black households, over the entire wealth distribution. For young Black households, in the lower deciles of the wealth distribution, business ownership is actually associated with lower wealth. This is not the case for White households. By contrast to the case of homeownership, the age premium of business ownership is similar for Blacks and Whites over the different deciles of the wealth distributions. Interestingly, in the case of stock ownership (Figure 3.5), the baseline coefficients and

Figure 3.3: Quantile Regressions – Homeownership



age interaction terms are not too dissimilar for Black and White households. Thus the associations of stock ownership and wealth are less different for the two groups than are the differences for the homes and businesses. This is consistent with an economic system where the set of financial assets available to the Black and White households is the same, but not the set of homes or of businesses. On the other hand, rates of ownership are more disparate for stocks than for homes or businesses, so the structural inequality appears to operate more at that level.

Finally, we consider several other characteristics in the context of differential

Figure 3.4: Quantile Regressions – Business Ownership

impacts over the wealth distribution. Figure 3.6 displays the coefficients for the case of having a pension. While the average coefficients for Black and White households are fairly similar (Table 3.3), the quantile regressions show that the group differential is much larger for households in lower deciles. By contrast, the average difference between Black and White households is very large in the case of those in managerial and professional occupations (Table 3.3), and the quantile regression results (Figure 3.7) indicate that this differential is actually higher for the upper deciles of the group wealth distributions. Finally, the financial literacy measure displays a similar pattern (Figure

Figure 3.5: Quantile Regressions – Stock Ownership

3.8). There is a large average differential between Black and White households, in terms of its relationship to wealth, and the quantile regressions indicate that the difference is greater in the upper deciles of the wealth distributions.

3.8 Discussion

By now a large number of empirical studies, many using SCF data, have analyzed the factors that help in understanding the wealth gap between Blacks and Whites in the US. While studies such as Keister (2000b), Herring and Henderson (2016) and

Figure 3.6: Quantile Regressions – Pensions

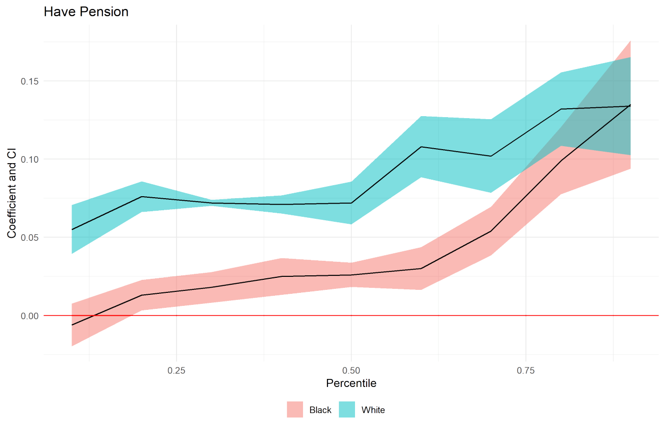
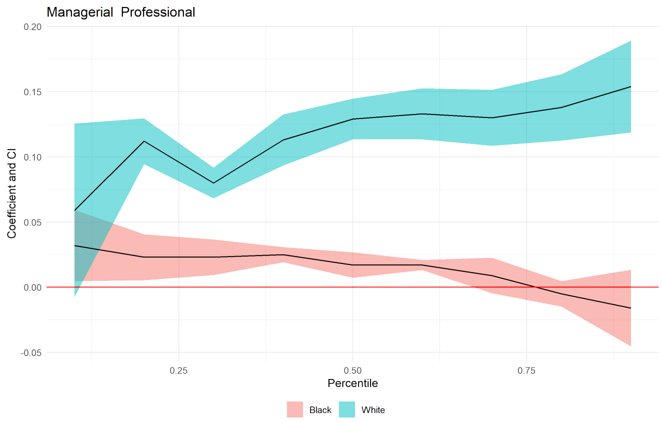
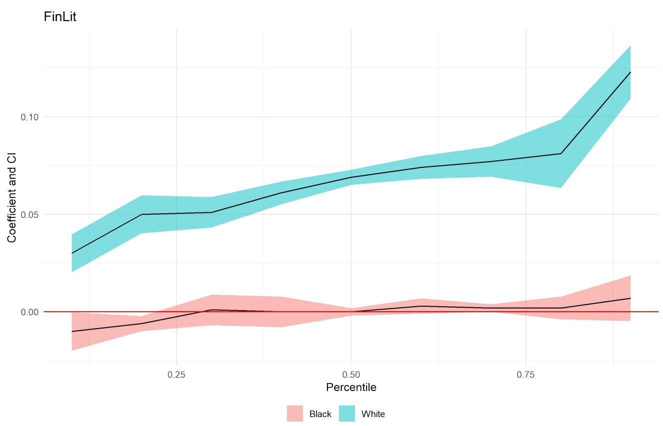


Figure 3.7: Quantile Regressions – Managerial and Professional Occupations



many others have examined the multiple structural and systemic factors that contribute to this wealth gap with as comprehensive a lens as possible, some studies do focus on specific characteristics such as homeownership or education, or somewhat more general factors such as asset ownership patterns. While these focused studies typically control for factors other than the one of interest, they can can create the impression of oversimplifying the challenge of how to address this particular aspect of wealth inequality. For Darity et al. (2018), this is such a strong concern that they offer a refutation of single-

Figure 3.8: Quantile Regressions – Financial Literacy



cause single-solution approaches in terms of ten myths. They rely on their own data analysis as well as evidence from various other studies, including several that employ various regression techniques with SCF data.[20] We first discuss our findings in relation to five items on the list of Darity et al. (2018) relating our findings to other studies as well. Our contribution here is to provide a uniform treatment of these five factors. Then we go on to discuss our empirical contributions in a broader context.

Educational attainment (Myth 1). Our results are very consistent with various studies that find unequal impacts of higher education on wealth, to the disadvantage of Blacks. Sometimes, conclusions are stated in terms of rejecting the strong hypothesis that college education eliminates wealth differences (e.g., the title of Emmons and Ricketts, 2017: "College Is Not Enough: Higher Education Does Not Eliminate Racial and Ethnic Wealth Gaps"). But, as our empirical analysis reinforces, there is

[20]Another synthesis that systematically refutes various singular claims about the racial wealth gap, including Latino/(a)s as well as Blacks, is Taub et al. (2017). Their table of contents describes their scope: Attending college does not close the racial wealth gap; Raising children in a two-parent household does not close the racial wealth gap; Working full time does not close the racial wealth gap; Spending less does not close the racial wealth gap.

no clear evidence that education even narrows the gap.[21] This can also be understood from our income regressions, which suggest that higher education allows for something closer to parity in incomes, but that parity does not provide any basis for catching up in wealth accumulation. Our results on the differential role of occupational choices as related to wealth levels are also suggestive of a narrower educational pathway to career and financial success for Black Americans than for White Americans.

While some evidence questions the value of a college education in terms of its financial returns, it would be facile or even misleading to conclude that it does not matter from a societal perspective. In particular, equal access to higher education should be a goal independently of financial returns from college degrees. In this context, the evidence suggests that family transfers of all kinds can be important. As an indicator of one type of such transfers, our results show that having an inheritance has a significant positive association with wealth levels, and that the strength of this effect is similar across the two racial groups. But it is much stronger for Black households when interacted with having a college degree. Being able to finance a college education without large or expensive student loans can be very important for wealth accumulation. Indeed, Meschede et al. (2017) document directly that this factor matters for explaining the

[21] Indeed, Emmons and Ricketts' actual empirical analysis, which uses regression analysis for SCF data from 1992 to 2013, is more nuanced. They also observe stark differences in trends for Black and White college graduates over the period of their analysis, "In particular, the median Black college-graduate family in 2013 had 56 percent less wealth than the median Black college-graduate family in 1992, ... (... adjusted for inflation). Meanwhile, the median White college graduate family in 2013 had 86 percent more wealth than the median White college-graduate family in 1992" (p. 8). Another important study, which includes a consideration of employment attributes, although it does not employ regression analysis, is Hamilton et al. (2015), which uses data from the US Census Survey of Income and Program Participation (SIPP), rather than SCF data. It is significant that different data sources yield consistent conclusions. Jez (2017), in commenting on some of Emmons and Ricketts' findings, offers some additional possibilities pertaining to nonlinearity of effects, and unobservable quality of colleges.

Black-White wealth gap.[22]

Homeownership (Myth 2). Along with education, homeownership receives the most attention in discussions about factors behind the Black-White wealth gap in the US. Again, any claim that equalizing rates of homeownership can eliminate the Black-White wealth gap is far too strong, but our results indicate something less hopeful, namely, that the benefits of homeownership are skewed toward White households. Specifically, controlling for all other characteristics, the wealth increment for a White household from owning a home is substantially greater than for a Black household with the same characteristics. This differential is on top of factors that may deter Black households from home ownership, as reflected in lower ownership rates.[23] These factors include discrimination in lending, as well as even more pernicious behaviors such as implicit and explicit segregation operating through both government and private institutions. If Black homeowners are restricted to lower quality housing stock in less attractive neighborhoods, that would explain the large differences in wealth contributions between the two racial groups that we estimate in our regressions. In any case, it is clear that just increasing home ownership rates by itself will do little or nothing to close the existing wealth gap.

[22]Another study with similar conclusions is Nam, et al. (2015), which uses PSID data. A separate issue, but one that can make it harder for Blacks to catch up with Whites, even if access is made moe equal, is that the returns to college appear to be falling for all races and ethnicities in the US: see Emmons, Kent and Ricketts (2019).

[23]Of course, if the returns to home ownership are lower for Black households, they may rationally choose not to buy a home. Discrimination in the housing market is a longstanding and persistent problem in the US, with a large academic and policy literature: see Darity, et al. (2018) and references therein. Choi, et al. (2019) provide a recent, geographically disaggregated, analysis of differences in Black-White home ownership rates, including credit scores, marital status and income as important explanatory variables. They do not address discrimination directly, though they refer to previous studies, and they note that more segregated Metropolitan Statistical Areas have higher proportions of White homeowners.

Financial Literacy (Myth 5). Our result here is not definitive, because it relies on a very specific and relatively narrow measure of financial literacy. Nevertheless, since any measure of this characteristic is difficult to come by, it is novel as well as quite striking. Since the 2016 SCF was the first to include questions designed to measure basic financial literacy, there have been few, if any, direct tests of the claim that financial literacy can help with wealth accumulation. Our regressions show that, controlling for other factors, there is no evidence that financial literacy is positively associated with wealth for Blacks, but it is for Whites, suggesting that there are deeper factors at work that are not directly observed in the data. Our results therefore provide quantitative evidence that supports the broader analysis of Hamilton and Darity (2017), which conceptually critiques claims that lack of financial literacy is a contributor to racial wealth gaps, as well as providing some evidence to support that critique.

Entrepreneurship (Myth 6). As in the case of education and home ownership, our results confirm that, controlling for other factors, owning a business is associated with higher wealth, but the increment is much greater for Whites than for Blacks with the same measurable characteristics. Therefore, on average, entrepreneurship may help with wealth accumulation, but there is no evidence that it contributes to closing the racial wealth gap. One also has to emphasize that business ownership rates are quite low as percentages of the population, so this is not a likely to be a pathway to wealth accumulation for the vast majority of the population. In some sense, though in a less extreme manner, focusing on entrepreneurship for wealth accumulation is similar to arguing that there is, or can be, equality in the dimension of wealth-building by pointing out the success of Black entertainers or athletes.[24]

Family Structure (Myth 10). There is a large difference between the proportion of female-headed households for Blacks and Whites. However, in the wealth regressions, the negative impact of being a female-headed household, when other factors are controlled for, is greater for Whites than for Blacks. Hence, while the higher proportion of female-headed households in the Black population contributes to the average wealth gap, conditional on that characteristic holding, this aspect of family structure does not further contribute to the wealth gap. This lack of racial disparity in the negative effect of being a female-headed household is an interesting phenomenon, in its contrast to the benefits of positive characteristics being skewed toward White households. It is also not the case that the number of children has any significant implications for wealth differences between Blacks and Whites. While our regression specification is quite different, these results are consistent with those of Emmons and Ricketts (2017), who conclude that (p. 30). "The contribution of family-structure variables to explaining racial and ethnic wealth gaps is negligible." Note that Lerman (2017) offers a slightly different perspective, confirming that family structure does influence wealth, but also acknowledging that it does not explain much of the wealth gap between Black Americans and White Americans, or in changes in that wealth gap from 2001 to 2013.

A common theme in our discussion of some of the Darity et al. (2018) "myths"

[24]This focus on celebrities is Myth 9 in Darity et al.'s list. Other myths that our empirical analysis is not able to address are: relying more on black businesses (Myth 3), saving (Myth 4), emulating successful minorities (Myth 7), and "soft skills" and "personal responsibility" (Myth 8). The analysis of Chakravorty et al. (2016), which emphasizes the joint roles of education, job-skill matching and labor market access as contributing factors for the economic success of many Indian Americans, can also be interpreted as a counter to Myth 7.

is that factors that potentially need policy attention in the context of the Black-White wealth gap require deeper analysis. Simply increasing homeownership rates or college education rates will not translate into significant reductions on the wealth gap without attention to quality disparities and systemic issues of discrimination. Furthermore, some of these interventions are connected, and cannot be tackled independently. Residential neighborhood disparities translate into inequalities in schooling quality and affect access to high quality higher education. All this is well known. One of our contributions is to reinforce this perspective by quantifying the disparities in impact of simple and single-factor interventions (Table 3.5). The manner in which we have done so, from our regression estimates, is also relatively new in this literature. It is also worth remarking that our results provide evidence that financial interventions that work without the mediation of unequal economic or social structures have more equal impacts. For example, we do not find a disparity of impacts that disfavors Black households in the case of receiving an inheritance versus not getting one. Similarly, Black households who do hold stocks in their asset portfolios exhibit positive associations of this characteristic with wealth levels that are comparable to White households. By contrast, this is not true of business ownership or homeownership. These results are supportive of the idea that wealth gaps are best addressed through direct, equalizing financial interventions. The proposal for "baby bonds" made by Hamilton and Darity (2010), and further analyzed and supported in Zewde (2019) is a clear example of this approach.

We also estimated income regressions to parallel the wealth regressions. While the causality for characteristics such as homeownership and stock ownership likely goes from income to these asset choices, rather than in the other direction, we maintained the same specification. A comparison of the results for the two types of regressions indicated that education level has a positive impact on income that is relatively equal for Black and White households, very different than the case of wealth. Overall, the income regressions are less indicative of structural differences in the impacts of characteristics, which suggests that the larger differences in impacts for wealth are reflective of the unequal conditions of past accumulation of wealth. This reinforces the idea that simply leveling the playing field going forward is inadequate as an approach to correcting wealth disparities that are the result of many years of unequal conditions of accumulation, and therefore equalizing financial transfers such as baby bonds are a more relevant policy.

Another contribution of our empirical analysis is to carry out a Blinder-Oaxaca-style decomposition of overall wealth differences, where the decomposition separates out what can be attributed to differences in household characteristics or endowments, versus what can be attributed to differences in the processes or system that determine wealth, as reflected in differences in the regression coefficients. Although this latter component can also reflect unmeasured quality differences, these may also be a function of structural inequalities. If we average the two methods of carrying out the decomposition, our estimate is that over 40 percent of the wealth difference is associated with structural factors that go beyond measured household characteristics. Note that we use a relatively new approach to this decomposition that accounts for the nonlinear specification of the wealth regressions, and produces estimates denominated in the original dollar units.

A final contribution of our analysis is the use of quantile regressions to allow for differences in effects of various characteristics on wealth over the different parts of the wealth distribution. In particular, we obtain estimates at different deciles, and these allow one to get a sense of how various characteristics are associated with higher wealth at various wealth levels. One can then compare impacts for Black and White households at the same deciles, or at different deciles that correspond to similar wealth levels. Very broadly, we see that there are class-type effects for both Black Americans and White Americans, for example as reflected in greater benefits of a college education at higher wealth levels. For some characteristics, these distributional effects are stronger for White Americans than for Black Americans, but the level differences in effects are present throughout the wealth distribution. In other words, the inequalities in wealth have a racial component that is not explainable by an appeal to "class," just as it is not explainable by appealing to differences in education or asset portfolio composition or financial literacy, or any other such factor. Race matters for wealth inequality.

3.9 Conclusion

Our results reinforce the perspective that there is no single or simple explanation of wealth disparities between Blacks and Whites. In particular, focusing individually on levels of education, homeownership, business ownership, financial literacy, or family structure does not provide a convincing picture of the determinants of wealth inequality in this case. This is true even when one controls for various other characteristics. Our results do support a perspective that the pathway to wealth accumulation is much narrower for Black Americans than for White Americans, with greater hurdles, on top of the starting point itself being unequal. This is inferred from the differences in marginal impacts of education, occupational choices, asset ownership of various types, and financial literacy. All of these results are consistent with a complex of structural and systemic factors being behind Black-White wealth inequalities, rather than single-factor explanations, especially ones that appeal to various versions of "personal responsibility." This may seem obvious, but, perhaps somewhat surprisingly, such explanations are still being discussed in the academic literature.

Although our empirical analysis uses cross-section data and does not directly establish any causal connections, our quantitative decompositions of the contributors to the wealth gaps, both measured and unmeasured are still informative. Additionally, the quantile regressions provide an indication of how impacts of various characteristics vary over the wealth distribution. A comparison of these within-group variations with differences between the racial groups leads to the conclusion that race matters, even after allowing for these inferred class effects.

Within the cross-section framework, the role played by having received an inheritance is suggestive of the importance of intergenerational wealth transfers. This is not surprising, of course, in the context of wealth. Setting aside issues of causality, there is a contrast in the observed empirical patterns between the positive association of inheritance (effectively a wealth transfer) and wealth, and processes such as education, employment and asset ownership, which are embedded within complex socioeconomic structures. In future work we will extend our analysis to multiple cross-sections of SCF data, using synthetic cohorts (e.g., McKernan et al., 2014) to understand some aspects of these wealth dynamics more precisely.

Bibliography

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. “Credit Supply and House Prices: Evidence from Mortgage Market Segmentation.” *NBER Working Paper Series*, 2012.

Agarwal, Sumit, Brent W. Ambrose, Souphala Chomsisengphet, and Chunlin Liu. “Asymmetric Information in Dynamic Contract Settings: Evidence from the Home Equity Credit Market.” *Working Paper*, 2007.

Agarwal, Sumit, Itzhak Ben-David, and Vincent W. Yao. “Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real-Estate Market.” *Management Science*, 2015.

Akerlof, George A. “The Market for ’Lemons’: Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics*, 84, no.3 (1970): 488-500.

Aliprantis, Dionissi, and Daniel R. Carroll.“What Is Behind the Persistence of the Racial Wealth Gap?” *Economic Commentary.*Federal Reserve Bank of Cleveland, 2019.

Aliprantis, Dionissi, Daniel R. Carroll, and Eric R. Young. "The Dynamics of the Racial Wealth Gap." *FRB of Cleveland Working Paper*, 2019.

"American Community Survey 5-Year Data 2010." U.S. Census Bureau, 2011.

An, Xudong, and Vincent W. Yao. "Credit Expansion, Competition, and House Prices," 2016.

Ausubel, Lawrence M. "Adverse Selection in the Credit Card Market." *Working Paper*, 1999.

Avery, Robert B., and Michael S. Rendall. "Lifetime inheritance of three generations of whites and blacks." *American Journal of Sociology* 107, no.5 (2002):1300–1346.

Bai, Bin, Jun Zhu, and Laurie Goodman. "A Closer Look at the Data on First-Time Homebuyers." *Urban Institute*, 2015.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics*, 7, no.1 (2015): 1-21.

Bauer, Thomas K., and Mathias Sinning. "An extension of the Blinder–Oaxaca decomposition to nonlinear models." *Advances in Statistical Analysis*, 92, no.2 (2008):197–206.

Bhutta, Neil, and Benjamin J. Keys. "Eyes Wide Shut? The Moral Hazard of Mortgage Insurers During the Housing Boom." *NBER Working Paper Series*, 2018.

Blinder, Alan S."Wage discrimination: reduced form and structural estimates." *Journal of Human Resources*, 8,no.4 (1973): 436–455

Boadway, Robin, and Michael Keen. "Financing New Investments under Asymmetric Information: A General Approach." *CIRPEE Working Paper*, 2004.

Bricker, Jesse, Lisa J. Dettling, Alice Henriques, Joanne W. Hsu, Lindsay Jacobs, Kevin B. Moore, Sarah Pack, John Sabelhaus, Jeffery Thompson, and Richard A. Windle. "Changes in US family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin*, 103, no.3 (2017): 1–42.

Brueckner, Jan K. "Mortgage Default with Asymmetric Information." *Journal of Real Estate Finance and Economics*, 20, no.3 (2000): 251-274.

Bubb, Ryan, and Alex Kaufman, "Securitization and Moral Hazard: Evidence from a Lender Cutoff Rule." *Research Department Public Policy Discussion Papers.* Federal Reserve Bank of Boston, 2009.

Cardon, James H., and Igal Hendel. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 2001, 408-427.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134, no.3 (2019): 1405-1454.

Chakravorty, Sanjoy, Devesh Kapur, and Nirvikar Singh. "The Other One Percent: Indians in America." *New York: Oxford University Press*, 2016.

Chiappori, Pierre-André, and Bernard Salanié. "Testing for Asymmetric Information in Insurance Markets." *The Journal of Political Economy*, 108, no. 1 (2000): 56-78.

Choi, Junghyun, Alanna McCargo, Michael Neal, Laurie Goodman, and Caitlin Young. "Explaining the Black-White Homeownership Gap: A Closer Look at Disparities across Local Markets." *Housing Finance Policy Center, Research Report, Washington, DC: Urban Institute*, 2019.

Christensen, Peter, Ignacio Sarmiento-Barbieri, and Christopher Timmins. "Housing Discrimination and the Toxics Exposure Gap in the United States: Evidence from the Rental Market." *Review of Economics and Statistics*, 2020.

"Conforming Loan Limits." *URL: https://www.fhfa.gov/DataTools/Downloads*

Conley, Dalton. "Being black, living in the red: Race, wealth, and social policy in America." *Berkeley, CA: University of California Press*, 1999.

Darity Jr., William, David Guilkey, and William Winfrey. "Ethnicity, race, and earnings." *Economics Letters*, 47, no.3 (1995): 401–408.

Darity Jr., William, Darrick Hamilton, Mark Paul, Alan Aja, Anne Price, Antonio Moore, and Caterina Chiopris. "What We Get Wrong About Closing the Racial Wealth Gap." *Samuel DuBois Cook Center on Social Equity*, 2018.

DeFusco, Anthony A., and Andrew Paciorek. “The Interest Rate Elasticity of Mortgage Demand: Evidence from Bunching at the Conforming Loan Limit.” *American Economic Journal: Economic Policy*, 2017, 210-240.

De Meza, David, and David C. Webb. “Too Much Investment: A Problem of Asymmetric Information.”, *The Quarterly Journal of Economics* 102 (1987): 281-292.

De Meza, David, and David C. Webb. “Advantageous Selection in Insurance Markets.” *RAND Journal of Economics*, 2001, 249-262.

Demyanyk, Yuliya, and Otto Van Hemert. “Understanding the Subprime Mortgage Crisis.” *Review of Financial Studies*, 2009.

Edelberg, Wendy. “Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets.” *FEDS Working Paper*, no.09 (2004).

Einav, Jenkins, and Levin. “Contract Pricing in Consumer Credit Markets.” *Econometrica*, 2012, 1387-1432.

Emmons, William R., and Lowell R. Ricketts. “College Is Not Enough: Higher Education Does Not Eliminate Racial and Ethnic Wealth Gaps.” *Federal Reserve Bank of St. Louis Review*, 99, no.1 (2017): 7-40.

Emmons, William R., Ana H. Kent, and Lowell R. Ricketts.“Is College Still Worth It? The New Calculus of Falling Returns.” *Federal Reserve Bank of St. Louis Review*, 101, no.4 (2019): 297-329.

Fairlie, Robert W. "An extension of the Blinder–Oaxaca decomposition technique to logit and probit models." *Journal of Economic and Social Measurement*, 30, no.4 (2005): 305–316.

Finkelstein, Amy, and Kathleen McGarry. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *The American Economic Review*, 2006, 938-958.

Freedman, Seth, and Ginger Jin. "Learning by Doing with Asymmetric Information: Evidence from Prosper.com." *NBER Working Paper*, 2009.

Gerardi, Kristopher, Adam H. Shapiro, and Paul Willen. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." *Federal Reserve Bank of Boston Working Paper*, no.07-15 (2007).

Graves, Alex, and Jürgen Schmidhuber. "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures." *Neural Networks* 18 (2005): 602-610.

Hamilton, Darrick, Algernon Austin, and William Darity, Jr."Whiter Jobs, Higher Wages: Occupational segregation and the lower wages of black men." *Economic Policy Institute Briefing Paper*, no. 288 (2011).

Hamilton, Darrick, and William Darity, Jr. "Can 'Baby Bonds' Eliminate the Racial Wealth Gap in Putative Post-Racial America?" *The Review of Black Political Economy*, 37, (2010): 207–216.

Hamilton, Darrick, William Darity, Jr., Anne E. Price, Vishnu Sridharan, and Rebecca Tippett. “Umbrellas Don’t Make it Rain: Why Studying and Working Hard Isn’t Enough for Black Americans.” *The New School*, 2015.

Hamilton, Darrick, and William Darity, Jr. “The Political Economy of Education, Financial Literacy, and the Racial Wealth Gap” *Federal Reserve Bank of St. Louis Review*, 99, no.1 (2017): 59-76.

Harrison, David M., Thomas G. Noordewier, and Abdullah Yavas. “Do Riskier Borrowers Borrow More?” *Real Estate Economics*, 32, no.3 (2004): 385-411.

Herring, Cedric, and Loren Henderson. “Wealth Inequality in Black and White: Cultural and Structural Sources of the Racial Wealth Gap.” *Race and Social Problems*, 8, no.1 (2016): 4–17.

Horton, Hayward D.“Race and wealth: A demographic analysis of black ownership.” *Sociological Inquiry*, 62, no.4 (2007): 480–489.

House, Christopher L. “Adverse Selection and the Financial Accelerator.” *Journal of Monetary Economics*, 2006, 1117-1134.

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F.P. Luttmer, and Kelly Shue. “Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?” *AFA 2011 Denver Meetings Paper*, 2009.

Jez, Su J. “Not All College Is Equal When It Comes to Wealth and Race.” *Federal Reserve Bank of St. Louis Review*, 99, no.1 (2017): 45-52.

Jones, Janelle, and John Schmitt. "A College Degree is No Guarantee." *Center for Economic Policy Research, Washington, DC*, 2014.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. "Credit Supply and the Housing Boom." *Federal Reserve Bank of New York Staff Reports*, no.709 (2015).

Kaiser, Boris. "Decomposing differences in arithmetic means: a doubly robust estimation approach." *Empirical Economics*, 50, no.3 (2016): 873-899.

Karlan, Dean, and Jonathan Zinman. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *The Review of Financial Studies*, 23, 1 (2010): 433-464.

Karlan, Dean, and Jonathan Zinman. "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment." *Econometrica*, 77, 6 (2009): 1993-2008.

Keister, Lisa A. "Race and wealth inequality: The impact of racial differences in asset ownership on the distribution of household wealth." *Social Science Research*, 29, no. 4 (2000): 477–502.

Keister, Lisa A. "Wealth in America." *New York: Cambridge University Press*, 2000.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. "Did Securitization Lead to Lax Screening." *The Quarterly Journal of Economics*, 125, no.1 (2010): 307-362.

Klonner, Stefan, and Ashok Rai. “Adverse Selection in Credit Markets: Evidence from a Policy Experiment.” *Department of Economics Working Paper, Williams College* 01 (2007).

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. “The Housing Market(s) of San Diego.” *NBER Working Paper Series*, 2012.

Lerman, Robert I. “Do Family Structure Differences Explain Trends in Wealth Differentials?” *Federal Reserve Bank of St. Louis Review*, 99, no.1 (2017): 85-102.

Liskovich, Inessa, and Maya Shaton. “Borrowers in Search of Feedback: Evidence from Consumer Credit Markets.” *Finance and Economics Discussion Series.* Federal Reserve Board, Washington, D.C., 2017.

Maroto, Michelle. “Growing Farther Apart: Racial and Ethnic Inequality in Household Wealth Across the Distribution.” *Sociological Science*, 3 (2016): 801-824.

Mayer, Christopher, Pence Karen, and Shane M. Sherlund. “The Rise in Mortgage Defaults.” *Journal of Economic Perspectives*, 23, no.1 (2009): 27-50.

McCrary, Justin. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics*, 2008.

McKernan, Signe-Mary, Caroline Ratcliffe, Eugene Steuerle, and Sisi Zhang. “Disparities in Wealth Accumulation and Loss from the Great Recession and Beyond” textitAmerican Economic Review: Papers and Proceedings, 104, no.5 (2014): 240–244.

Meschede, Tatjana, Joanna Taylor, Alexis Mann, and Thomas Shapiro. "'Family Achievements?': How a College Degree Accumulates Wealth for Whites and Not for Blacks" *Federal Reserve Bank of St. Louis Review*, 99, no.1 (2017): 121-138.

Meyer, Ana Lourdes Gomez Lemmen. "Pricing Mechanisms in Peer-to-Peer Online Credit Markets." *Department of Economics Working Papers.*Stanford University, 2014.

Mian, Atif, and Amir Sufi. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics*, 124, no.4 (2009):1449-1496.

Nam, Yunju, Darrick Hamilton, William A. Darity, Jr., and Anne E. Price. "Bootstraps Are for Black Kids: Race, Wealth, and the Impact of Intergenerational Transfers on Adult Outcomes" *The Samuel Dubois Cook Center on Social Equity at Duke University*, 2015.

Oaxaca, Ronald. "Male–female wage differentials in urban labor markets." *International Economic Review*, 14, no.3 (1973): 693–709.

Oliver, Melvin, and Shapiro, Thomas. "Black Wealth/White Wealth: A new perspective on racial inequality" *New York and London: Routledge Publishers*, 2006.

Parcel, Toby L. "Wealth accumulation of black and white men: The case of housing equity." *Social Problems*, 30,no.2 (2014): 199–211.

Parlour, Christine A., and Uday Rajan. "Competition in Loan Contracts." *The American Economic Review*, 2001, 1311-1328.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis." *Journal of Financial Economics*, 97, 3 (2010): 369-397.

Rajan, Uday, Amit Seru, and Vikrant Vig. "The Failure of Models that Predict Failure: Distance, Incentives, and Defaults." *Journal of Financial Economics*, 115 (2015): 237-260.

Riddiough, Timothy J., and Howard E. Thompson. "Commercial Mortgage Pricing with Unobservable Borrower Default Costs." *Real Estate Economics*, 12 (1993): 265-291.

Rothschild, Michael, and Joseph Stiglitz. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, 90, no.4 (1976): 629-649.

Shapiro, Thomas M. "The hidden cost of being African American: How wealth perpetuates inequality." *New York: Oxford University Press*, 2004.

Shapiro, Thomas M., Tatjana Meschede, and Sam Osoro. "The Roots of the Widening Racial Wealth Gap: Explaining the Black-White Economic Divide." *Institute on Assets and Social Policy*, 2013.

Sood, Gaurav. "Florida Voter Registration Data." Harvard Dataverse, 2017. https://doi.org/10.7910/DVN/UBIG3F.

Sood, Gaurav, and Suriyan Laohaprapanon. “Predicting Race and Ethnicity From the Sequence of Characters in a Name,” 2018.

Stiglitz, Joseph E. “Perfect and Imperfect Capital Markets.” *paper presented to the New Orleans meeting of the Econometric Society*, 1970.

Stiglitz, Joseph E. “Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs.” *Bell J. Econ*, 1972, 458-482.

Stiglitz, Joseph E., and Andrew Weiss. “Credit Rationing in Markets with Imperfect Information.” *The American Economic Review*, 71, no.3 (1981): 393-410.

Taub, Amy, Laura Sullivan, Tatjana Meschede, and Thomas Shapiro. “The Asset Value of Whiteness: Understanding the Racial Wealth Gap.” *Demos*, 2017.

Thompson, James R. “Counterparty Risk in Financial Contracts: Should the Insured Worry about the Insurer?” *The Quarterly Journal of Economics*, 125, no.3 (2010): 1195-1252.

Thompson, Jeffrey P., and Gustavo A. Suarez. “Updating the Racial Wealth Gap, Finance and Economics Discussion Series” *Divisions of Research and Statistics and Monetary Affairs*, Federal Reserve Board, Washington, D.C., 2015.

Tippett, Rebecca, Avis Jones-DeWeever, Maya Rockeymoore, Darrick Hamilton, and William Darity, Jr. “Beyond Broke: Why Closing the Racial Wealth Gap is a Priority for National Economic Security” *Center for Global Policy Solutions*, 2014.

Wei, Zaiyan, and Mingfeng Lin. "Market Mechanisms in Online Peer-to-Peer Lending." *Management Science*, 2016

Winsborough, H. H., and Peter Dickinson. "Components of Negro-White Income Differentials." *Center for Demography and Ecology*, University of Wisconsin, 1971.

Zaw, Khaing, Jhumpa Bhattacharya, Anne Price, Darrick Hamilton, and William Darity, Jr. "Women, Race & Wealth, Research Brief Series." *Samuel DuBois Cook Center on Social Equity and Insight Center for Community Economic Development*, 1, (2017).

Zewde, Naomi. "Universal Baby Bonds Reduce Black-White Wealth Inequality, Progressively Raise Net Worth of All Young Adults." *The Review of Black Political Economy*, 47, no.1 (2020): 3–19.

"Zillow Home Value Index." *URL: https://www.zillow.com/research/data/.*

"Zillow's Transaction and Assessment Database." 2018. https://www.zillow.com/research/ztrax/.

Appendix A

Chapter 1

A.1 The Performance of Loans Using Larger Bandwidth

Table A.1: The Impact on Performance of Loans Taken out through the Program (Using Larger Bandwidth)

	First and Repeated		Repeated		First	
	±\$100,000	±\$60,000	±\$100,000	±\$60,000	±\$100,000	±\$60,000
Under New CLL x Post	.0155**	.0054	.0168	.0099	.0135*	.0005
	(.0072)	(.0065)	(.0113)	(.0102)	(.0077)	(.0075)
Under New CLL	-.0185***	-.0062	-.0182***	-.0058	-.0184***	-.0060
	(.0051)	(.0039)	(.0063)	(.0088)	(.0051)	(.0044)
Post	-.0571***	-.0618***	-.0553***	-.0572***	-.0625***	-.0632***
	(.0139)	(.0189)	(.0156)	(.0209)	(.0139)	(.0168)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-mean	.0715	.0709	.0670	.0661	.0725	.0720

Notes: The sample of the first, third, and fifth columns are loans within \$100,000 below or above the new CLL. The sample of the second, fourth, and sixth columns are loans within \$60,000 below or above the new CLL. The outcome in each column is a dummy that equals 1 if a transaction ends up with default by 2018. Standard errors are adjusted for bootstrap clusters in counties.

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.2 The Default Rate of Loans Taken Out by Borrowers from a Particular Race or Income Group

Table A.2: The Impact on Performance of Loans within Different Group of Buyers

	(1)	(2)	(3)	(4)	(5)
Under New CLL x Post	-.0696	.0525*	.0202	-.0219	.0019
	(.2933)	(.0299)	(.0266)	(.0337)	(.0095)
Under New CLL	.0147	-.0241	.0351*	.0219	.0138
	(.0689)	(.0219)	(.0187)	(.0337)	(.0124)
Post	.0434	-.0741***	-.1043***	-.1197***	-.0544***
	(.3771)	(.0286)	(.0374)	(.0399)	(.0152)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Pre-mean	.0921	.0571	.1001	.1206	.0569

Notes: Each column presents regression results using observations from $\pm\$20,000$ of the CLL. Col (1) includes only people from the bottom quartile zip codes of California, Col (2) includes only Asian people or Pacific Islanders, Col (3) includes only Hispanic people, Col (4) includes only Black people, and Col (5) includes only White people. The sample size in either Col (1) or Col (4) is too small, and parameters could not be estimated in some bootstrap replicates and standard errors estimates include only complete replications. Standard errors are adjusted for bootstrap clusters in counties.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Appendix B

Chapter 3

Table B.1: Quantile Regressions: Ln(Wealth+k) – Black Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log of (Wealth+k) for Black								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
Female head of HH	-0.009	-0.022***	-0.025***	-0.027***	-0.032***	-0.039***	-0.048***	-0.068***	-0.118***
	(0.007)	(0.007)	(0.006)	(0.004)	(0.003)	(0.002)	(0.007)	(0.010)	(0.015)
Bankruptcy	-0.080***	-0.116***	-0.148***	-0.111***	-0.110***	-0.059**	-0.045	-0.024	-0.009
	(0.008)	(0.019)	(0.032)	(0.023)	(0.026)	(0.024)	(0.033)	(0.029)	(0.049)
Spending exceeded income	-0.066***	-0.053***	-0.045***	-0.050***	-0.042***	-0.046***	-0.038***	-0.035***	-0.025
	(0.005)	(0.008)	(0.005)	(0.005)	(0.005)	(0.009)	(0.002)	(0.009)	(0.016)
Have stock	-0.308***	-0.171	-0.183***	-0.255	-0.371***	-0.165	-0.308**	0.004	0.725
	(0.060)	(0.143)	(0.047)	(0.160)	(0.124)	(0.188)	(0.121)	(0.371)	(0.896)
Have business	-0.962***	-0.566***	-0.246***	-0.184*	-0.401	-0.447*	-0.253**	-0.104	-0.279
	(0.059)	(0.042)	(0.041)	(0.100)	(0.292)	(0.264)	(0.115)	(0.202)	(0.203)
Have home	0.112***	0.160***	0.216***	0.199***	0.202***	0.136	0.169**	0.125**	0.022
	(0.022)	(0.041)	(0.034)	(0.052)	(0.027)	(0.096)	(0.066)	(0.054)	(0.177)
Receive inheritance	-0.034*	0.047***	0.086***	0.083***	0.074***	0.130***	0.136***	0.112***	0.086***
	(0.018)	(0.015)	(0.013)	(0.010)	(0.016)	(0.022)	(0.018)	(0.012)	(0.031)
Have pension	-0.006	0.013**	0.018***	0.025***	0.026***	0.030***	0.054***	0.099***	0.135***
	(0.007)	(0.005)	(0.005)	(0.006)	(0.004)	(0.007)	(0.008)	(0.011)	(0.021)
# of fin lit questions answered correctly	-0.010**	-0.006***	0.001	-0.000	0.000	0.003	0.002*	0.002	0.007
	(0.005)	(0.002)	(0.004)	(0.004)	(0.001)	(0.002)	(0.001)	(0.003)	(0.006)
HS/GED only	0.045***	0.031***	0.009*	0.004	-0.000	-0.006**	-0.013*	-0.017*	-0.000
	(0.014)	(0.006)	(0.005)	(0.003)	(0.002)	(0.002)	(0.007)	(0.010)	(0.009)
Some College	0.020	0.031***	0.007**	0.002	0.002	0.004	0.008	0.007	0.024*
	(0.019)	(0.008)	(0.004)	(0.009)	(0.005)	(0.006)	(0.007)	(0.009)	(0.014)
College and Above	-0.116***	-0.101***	-0.093***	-0.017	0.016	0.035**	0.049**	0.073*	0.263***
	(0.041)	(0.012)	(0.016)	(0.028)	(0.011)	(0.017)	(0.022)	(0.040)	(0.029)
Managerial/Professional	0.032**	0.023**	0.023***	0.025***	0.017***	0.017***	0.009	-0.005	-0.016
	(0.014)	(0.009)	(0.007)	(0.003)	(0.005)	(0.002)	(0.007)	(0.005)	(0.015)
Technical/Sales/Services	-0.003	0.006	0.019***	0.026***	0.024***	0.032***	0.034***	0.037***	0.040
	(0.006)	(0.010)	(0.005)	(0.001)	(0.002)	(0.005)	(0.008)	(0.009)	(0.026)
Other Job	0.018	0.019***	0.015**	0.027***	0.033***	0.033***	0.047***	0.032***	0.017
	(0.011)	(0.005)	(0.006)	(0.007)	(0.008)	(0.006)	(0.010)	(0.009)	(0.034)
Age	0.003***	0.002***	0.002***	0.001***	0.001***	0.001***	0.002***	0.001***	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Kids	0.005	0.010***	0.008***	0.006***	0.009***	0.012***	0.013***	0.012***	0.014***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)
Age x have stock	0.012***	0.010***	0.011***	0.012***	0.016***	0.011***	0.014***	0.010	-0.004
	(0.001)	(0.002)	(0.001)	(0.004)	(0.002)	(0.004)	(0.002)	(0.006)	(0.015)
Age x have business	0.023***	0.014***	0.010***	0.010***	0.016**	0.022***	0.018***	0.016***	0.021***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.008)	(0.005)	(0.002)	(0.005)	(0.005)
Age x have home	0.000	0.001	0.000	0.002**	0.003***	0.006***	0.007***	0.010***	0.014***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)
College and above x Managerial/Professional	-0.235***	-0.060***	-0.008	-0.078***	-0.040	-0.029	0.091***	0.126***	0.093
	(0.027)	(0.015)	(0.021)	(0.022)	(0.028)	(0.036)	(0.021)	(0.046)	(0.073)
College and above x Receive inheritance	0.480***	0.307***	0.204***	0.178***	0.181*	0.286**	0.210***	0.173*	0.294***
	(0.041)	(0.018)	(0.018)	(0.033)	(0.097)	(0.127)	(0.031)	(0.100)	(0.034)
Constant	11.831***	11.923***	11.971***	12.005***	12.026***	12.025***	12.030***	12.076***	12.168***
	(0.023)	(0.007)	(0.018)	(0.014)	(0.006)	(0.016)	(0.011)	(0.015)	(0.046)
Observations	769	769	769	769	769	769	769	769	769

Standard errors in parentheses,value of k = 173255.90

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Quantile Regressions: Ln(Wealth+k) – White Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log of (Wealth+k) for White								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
Female head of HH	-0.076***	-0.059***	-0.116***	-0.136***	-0.135***	-0.136***	-0.145***	-0.176***	-0.218***
	(0.010)	(0.009)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.008)
Bankruptcy	-0.226***	-0.240***	-0.214***	-0.248***	-0.296***	-0.325***	-0.385***	-0.403***	-0.351***
	(0.022)	(0.015)	(0.014)	(0.004)	(0.019)	(0.011)	(0.024)	(0.010)	(0.012)
Spending exceeded income	-0.123***	-0.116***	-0.120***	-0.121***	-0.117***	-0.080***	-0.106***	-0.092***	-0.057***
	(0.012)	(0.009)	(0.011)	(0.007)	(0.007)	(0.012)	(0.007)	(0.012)	(0.021)
Have stock	-0.045	-0.170***	-0.231***	-0.248***	-0.285***	-0.300***	-0.276***	-0.346***	-0.001
	(0.050)	(0.043)	(0.015)	(0.035)	(0.040)	(0.019)	(0.098)	(0.069)	(0.214)
Have business	0.234**	0.110**	-0.168***	-0.212**	0.038	0.133***	0.168***	0.250***	0.132
	(0.098)	(0.048)	(0.044)	(0.086)	(0.064)	(0.038)	(0.033)	(0.060)	(0.136)
Have home	0.105***	-0.033	-0.039***	-0.058***	-0.058***	-0.062***	-0.168***	-0.137***	-0.064
	(0.015)	(0.029)	(0.013)	(0.007)	(0.022)	(0.001)	(0.018)	(0.014)	(0.072)
Receive inheritance	0.027***	0.056***	0.035***	0.040***	0.089***	0.104***	0.123***	0.132***	0.203***
	(0.008)	(0.012)	(0.005)	(0.012)	(0.012)	(0.005)	(0.008)	(0.016)	(0.017)
Have pension	0.055***	0.076***	0.072***	0.071***	0.072***	0.108***	0.102***	0.132***	0.134***
	(0.008)	(0.005)	(0.001)	(0.003)	(0.007)	(0.010)	(0.012)	(0.012)	(0.016)
# of fin lit questions answered correctly	0.030***	0.050***	0.051***	0.061***	0.069***	0.074***	0.077***	0.081***	0.123***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.009)	(0.007)
HS/GED only	0.049*	0.039***	0.031***	0.034***	0.025*	0.033***	0.038***	0.017*	0.006
	(0.029)	(0.014)	(0.007)	(0.010)	(0.014)	(0.006)	(0.009)	(0.009)	(0.014)
Some College	0.040*	0.061***	0.062***	0.088***	0.087***	0.086***	0.138***	0.129***	0.119***
	(0.022)	(0.012)	(0.008)	(0.008)	(0.008)	(0.008)	(0.014)	(0.011)	(0.015)
College and Above	0.056***	0.155***	0.191***	0.216***	0.271***	0.339***	0.445***	0.479***	0.565***
	(0.021)	(0.011)	(0.016)	(0.020)	(0.016)	(0.011)	(0.017)	(0.026)	(0.047)
Managerial/Professional	0.059*	0.112***	0.080***	0.113***	0.129***	0.133***	0.130***	0.138***	0.154***
	(0.034)	(0.009)	(0.006)	(0.010)	(0.008)	(0.010)	(0.011)	(0.013)	(0.018)
Technical/Sales/Services	0.037**	0.030***	0.002	0.024***	0.041***	0.051***	0.027***	0.014**	0.056***
	(0.015)	(0.011)	(0.005)	(0.004)	(0.005)	(0.008)	(0.006)	(0.006)	(0.015)
Other Job	0.044**	0.055***	0.018***	0.051***	0.040***	0.045***	0.036***	0.025**	0.090***
	(0.020)	(0.004)	(0.006)	(0.007)	(0.007)	(0.010)	(0.010)	(0.013)	(0.022)
Age	0.009***	0.006***	0.005***	0.006***	0.006***	0.007***	0.007***	0.009***	0.013***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Kids	0.037***	0.034***	0.028***	0.036***	0.038***	0.040***	0.047***	0.054***	0.052***
	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)
Age x have stock	0.008***	0.011***	0.013***	0.015***	0.017***	0.018***	0.018***	0.020***	0.016***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.004)
Age x have business	-0.000	0.005***	0.011***	0.013***	0.010***	0.010***	0.011***	0.012***	0.018***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Age x have home	0.003***	0.007***	0.008***	0.010***	0.011***	0.011***	0.015***	0.015***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
College and above x Managerial/Professional	0.121***	0.033***	0.043***	0.074***	0.099***	0.120***	0.113***	0.123**	0.081
	(0.025)	(0.012)	(0.015)	(0.013)	(0.021)	(0.016)	(0.016)	(0.049)	(0.055)
College and above x Receive inheritance	0.115***	0.140***	0.205***	0.197***	0.181***	0.130***	0.117***	0.164***	0.164***
	(0.019)	(0.015)	(0.011)	(0.014)	(0.017)	(0.029)	(0.019)	(0.038)	(0.036)
Constant	11.291***	11.450***	11.611***	11.550***	11.562***	11.556***	11.588***	11.606***	11.501***
	(0.043)	(0.019)	(0.010)	(0.010)	(0.023)	(0.019)	(0.025)	(0.064)	(0.062)
Observations	3,625	3,625	3,625	3,625	3,625	3,625	3,625	3,625	3,625

Standard errors in parentheses,value of k = 173255.90

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$