City & Community

Rethinking the Role of Racial Segregation in the American Foreclosure Crisis

Jonathan P. Latner*
Bremen International Graduate School of Social Science

Abstract

Racial segregation is an important factor in understanding the foreclosure crisis, but must be understood to operate in particular and specific ways. The primary, positive impact of segregation on foreclosure risk operates prior to loan origination through the differential access to loan quality by race. Afterward, the impact of segregation is negative. Drawing on a rare dataset of loans that combine loan performance and borrower characteristics, I use a competing risks proportional hazard model to examine the impact of race and racial segregation on risk of foreclosure among borrowers. Results indicate that Black segregation has a large, negative impact on foreclosure risk. Instead, the strongest positive contributor to foreclosure is the negative value of the home relative to the balance of the loan (i.e., "underwater," as measured by the put option), which is also the mechanism that explains most of the difference in the foreclosure rate by race. The negative impact of racial segregation on foreclosure risk is the result of a mismatch between cities with high levels of segregation and cities with large declines in home prices and related foreclosures.

INTRODUCTION

The United States entered into a recession in December 2007, which became the longest and worst economic crisis since the Great Depression. One important driver of the recession was a housing boom gone bust, often called the foreclosure crisis. This crisis witnessed the rapid and sustained increase in the foreclosure rate well above the historical average, such that by 2010, the national foreclosure rate rose to about 5%, much higher than the 1% average in the decades prior (Schwartz 2010, p. 76).

The foreclosure crisis was not monocausal, but a central component was the rise and fall of the subprime mortgage market (Gramlich 2007). Some analyses highlight the role of securitization (Gotham 2009) and globalization (Aalbers 2009), which created an artificial supply of dollars for home loans where there was no real demand. Others highlight the role and responsibility of risky borrowers, lenders, and loans (Ding et al. 2011), whereby lenders assisted borrowers in obtaining loans on homes that they could not afford. A complementary piece of the discussion regarding the foreclosure crisis is the

*Correspondence should be addressed to Jonathan P. Latner, jonlatner@gmail.com.

City & Community 16:4 December 2017
doi: 10.1111/cico.12253

role of residential racial segregation (Hyra et al. 2013), which is a measure of the geographic separation between minority and majority racial groups in a city, as defined by the metropolitan statistical area (MSA).

The role of racial segregation, which contributed to the more efficient targeting of specific populations and neighborhoods by race, as a cause of the foreclosure crisis is widely recognized and cited (see Reskin 2012). Rugh and Massey (2010) used data aggregated at the level of the metropolitan area to indicate that Black–White racial segregation has a positive, causal effect on the number and rate of foreclosures in a city. However, the data used did not allow one to distinguish the impact of segregation on foreclosure after loan origination from the impact of segregation on loan quality prior to loan origination. More recently, Rugh (2015) used loan-level data and found a positive impact of race, racial segregation, and the interaction between race and segregation on a borrower's risk of foreclosure after loan origination.

In this paper, I further examine the relationship between race, segregation, and foreclosure after loan origination to shed light on its multiple, and as it turns out, complex links to the foreclosure crisis. Using a competing risks hazard model and a national dataset of individual loans with borrower demographic information, I demonstrate that Black–White residential racial segregation is negatively related to a borrower's risk of foreclosure. Instead, the degree to which the ratio between the value of the home and the balance of the loan is negative (i.e., the degree to which a borrower is "underwater") is a far more important and positive risk factor. This ratio is known as the "put option," and is the mechanism that best explains differences in the foreclosure risk by race. Therefore, the impact of race on foreclosure is shown to be indirect, not direct. The results of this work extend our understanding of the role played by race and racial segregation in the foreclosure crisis and emphasize the need for additional focus on the factors that affect foreclosure, not just the factors that affect loan origination.

The central argument is that declines in home prices constitute a critical mechanism through which race and segregation affect foreclosure. While some may argue that the role of housing prices is not surprising, it gets less attention in policy debates, as well as within the discipline of sociology, relative to the broader role played by race and segregation. One primary explanation for this is data availability. As will be addressed in greater detail later, many of the datasets that are used to examine the relationship between race, segregation, and foreclosure do not include time-varying loan-performance characteristics, such as current balance of the loan or the value of the home, which affect a borrower's risk of foreclosure. Of the few datasets like this one, which contain variables on both loan performance and borrower demographic characteristics, there is no one, single, nationally representative sample that may be used to examine foreclosure risk across the universe of loan and borrower types. The inclusion of loan-level data results in findings that are consistent with previous research demonstrating a link between race, segregation, and foreclosure, but indicate the importance of housing price change in determining how these factors affect foreclosure risk.

RACE, SEGREGATION, AND FORECLOSURE

This study tests three hypotheses regarding the impact of race, segregation, and their interaction on a borrower's risk of foreclosure during the foreclosure crisis and its

aftermath, between 2007 and 2013. Based on the existing literature, we would expect race, segregation, and their interaction to have a positive impact on a borrower's risk of foreclosure. This paper proposes an alternative perspective. If we consider the factors that affect foreclosure risk after a loan has been acquired, such as housing prices, then the association between segregation and foreclosure is negative, and the association between race and foreclosure is indirect. The goal of this paper is to improve our understanding of the relationship between segregation, race, and foreclosure by comparing these two explanations.

Before moving forward, it is important to note that the existing evidence on the topic area is divided by two intersecting lines of research: time, which is demarcated by loan origination; and unit of analysis, which is either the individual or the city. While relevant literature exists across the intersecting lines of research, the evidence presented here only examines an individual's risk of foreclosure after loan origination. Prior to loan origination, current and historical segregation increases the risk of foreclosure for non-White borrowers living in segregated neighborhoods by allocating loan quality by race unequally (Calem et al. 2004; Williams et al. 2005; Bocian et al. 2008). After loan origination, the impact of segregation on a borrower's risk of foreclosure remains an open question.

The current understanding regarding the relationship between race, segregation, and foreclosure is based on three components. First, segregation creates minority-dominant neighborhoods (Massey and Denton 1993). Disadvantages like poverty and joblessness are thus spatially concentrated by race as a result of segregation, which leaves minority populations more vulnerable to exploitation in cities with high levels of racial segregation, in comparison to those cities with low levels of segregation. Second, borrowers in these minority-dominant neighborhoods are at greater risk of receiving lower quality, subprime loans (Stuart 2003). There is also evidence that the relationship between segregation and subprime lending in a city may also go in the other direction, as relative changes in the level of subprime lending in a city may act to increase racial segregation (Bond and Williams 2007). Third, the foreclosure rate of subprime loans is much higher than prime loans (Gramlich 2007). Connecting the three, segregation and foreclosure are positively associated through the mechanism of subprime lending (Rugh and Massey 2010), which also affects differences in the foreclosure rate by race (Bocian et al. 2011).

There is no single, agreed upon definition for either the term subprime nor subprime loan (Demyanyk and Van Hemert 2011). The term subprime is used to describe a combination of loan, borrower, and lender types with at least one of the following characteristics: a borrower with a high risk of default, low credit scores, a history of delinquent payments, and/or bankruptcy; a loan with high and adjustable interest rates, payment, fees, etc.; and a lender specializing in high-cost loans and high-risk borrowers. When one is referring specifically to subprime loans, it is the characteristics of the loan that are determinative, but borrower and lender characteristics represent a critical component of what comprises the broader term subprime. Examples exist where prime borrowers received subprime loans and subprime borrowers received prime loans (Courchane et al. 2004) and it is an open question in the literature as to whether risk of default lies with the subprime borrower, subprime loan, or subprime lender (Ding et al. 2011). The common theme across the various definitions of both subprime loans and borrowers is a high risk of foreclosure due to loan and/or borrower characteristics.

Differences in the foreclosure rate by race are not only the result of differential access to loan quality due to racial segregation, but also the interaction between segregation and race, even after controlling for borrower and loan characteristics (Rugh 2015). For example, Blacks had a higher risk of foreclosure than Whites in the "Rust Belt" states of the Midwest, which contain cities with higher levels of segregation, as compared to the "Sand Belt" states of the Western United States, which contain cities with lower levels of segregation. In other words, the impact of race on foreclosure is exacerbated by the impact of segregation on foreclosure (Pager and Shepherd 2008).

One problem with the positive relationship between segregation and foreclosure is that there is a mismatch between the places most affected by subprime lending and related foreclosures and the places most affected by Black/White racial segregation. In the early part of the 2000s, fast-growing metropolitan areas with high rates of housing price appreciation, like Las Vegas, Phoenix, and Miami, had rates of subprime originations that were two to three times higher than the national average and higher than cities like Baltimore, Detroit, and Milwaukee (Mayer and Pence 2008). After 2007, as the housing market began to collapse, rates of foreclosure were highest in cities with large increases in housing prices prior to 2007 and subsequently large declines after 2007 (Immergluck 2008). While high rates of subprime lending and subsequent foreclosures existed in predominantly non-White neighborhoods within highly segregated cities (Hwang et al. 2015), relative rates may not have achieved the same degree of importance as compared to cities in the Western United States or Florida, which have lower levels of Black/White racial segregation. Despite evidence that the foreclosure crisis began in highly segregated cities (Rugh and Massey 2010), an alternative possibility is that the relationship between racial segregation and foreclosure over the course of the Great Recession and its aftermath is negative, not positive.

The negative association between segregation and foreclosure contains two components. First, segregation and housing prices are positively related. Second, housing price change and foreclosure are negatively related. Connecting the two, housing prices are the mechanism through which racial segregation has a negative effect on a borrower's risk of foreclosure, which explains most the direct impact of race on foreclosure risk. The positive impact of race on a borrower's risk of foreclosure would thus be indirect, operating through housing price change. We test the different expectations regarding the positive or negative link between segregation and foreclosure, as well as the role of race on foreclosure, using the following three hypotheses.

Hypothesis 1. *It asserts that borrowers in cities with higher levels of Black–White racial segregation have a lower risk of foreclosure.*

Between 2007 and 2013, there was a positive correlation (0.2) between segregation and housing price change, which is measured by the Housing Price Index, as shown in Figure 1. However, there was also a large and downward change in housing prices across the country. As a result, a positive correlation of home price change and segregation means that cities with lower levels of racial segregation experienced a larger decline in home prices, compared to cities with higher levels of racial segregation. Given that a borrower's likelihood of foreclosure rises as home prices decline (Gerardi et al. 2007), those living in cities with lower levels of housing price decline, which are also cities with higher levels of segregation, are "protected" from foreclosure.

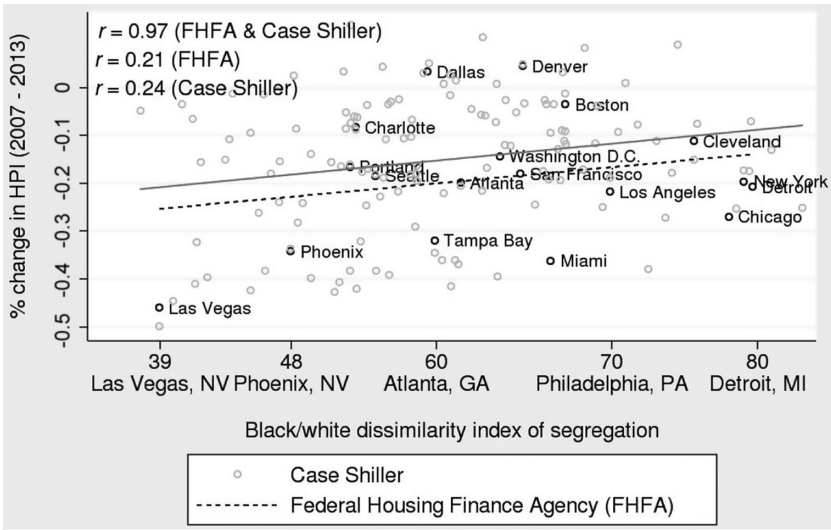

FIG. 1. Correlation between segregation and housing price index (HPI).

The underlying explanation for the positive relationship between segregation and home prices is the relationship between home prices and population growth between 2000 and 2007. There is a typology of metropolitan areas based on population growth, home price change, and foreclosure (Immergluck 2009), which correlates to patterns of racial segregation. Cities with low levels of racial segregation are generally newer and less industrial, with high rates of population growth, and are typically located in the West or the South, such as Las Vegas, Phoenix, and Miami. In contrast, cities with high levels of racial segregation are generally older and more industrial, with low or declining rates of population growth, and are located in the Northeast or Midwest, such as Milwaukee, Baltimore, and Detroit. Therefore, borrowers living in cities with high levels of segregation did not experience as much population growth or related home price change prior to 2007, nor subsequent declines in home prices and related foreclosures after 2007, when the crisis began.

Hypothesis 2. *It asserts that there are no racial differences in a borrower's risk of foreclosure.*

If home prices are the critical determinant of a borrower's risk of foreclosure, then differences in foreclosure should not depend on race. The reason is that a decline in home prices will increase all borrowers' risk of foreclosure because they are expected to be equally rational regardless of race (Ambrose and Capone 1998). This rationality postulate will be explored in more detail in the methods section below. If changes in home prices do explain all or most of the racial differences in terms of foreclosure risk, then the impact of race on foreclosure risk could still be present, but would be indirect, operating through the mechanism of housing price change.

Hypothesis 3. *It asserts that there is no interaction between race and segregation on a borrower's risk of foreclosure.*

Individuals who are non-White will not have a higher risk of foreclosure in cities with higher levels of segregation. The reason is that home prices are the critical determinant of a borrower's risk of foreclosure, and the degree to which a borrower is rational should not depend on geography, let alone race (as in Hypothesis 2), or their interaction.

In summary, the existing literature suggests that race, segregation, and their interaction all have a positive impact on a borrower's risk of foreclosure. However, another possibility is that segregation and foreclosure are negatively connected through the mechanism of home price change. Relatedly, the positive impact of race on a borrower's risk of foreclosure is indirect and operates through housing price change. Below, these competing expectations regarding the relationship between race, segregation, and foreclosure during the crisis and its aftermath are tested using the three hypotheses described above.

METHODS

To examine the factors that affect a borrower's risk of foreclosure, I use a competing risks proportional hazard model, which is an extension of a Cox proportional hazard model for a single risk. "Competing risks" refers to a borrower's risk of foreclosure against their competing possibility of prepayment. Modeling the joint effects of a borrower's competing outcome of foreclosure and prepayment is the standard approach for predicting foreclosure risk (Calhoun and Deng 2002).

Option theory is based on the assumption that a borrower faces one of three options in any given month: to make a payment, to not make a payment, or prepay the loan. Borrowers who make a payment are current on their monthly payments. Borrowers who do not make a payment in full are in default. While default and foreclosure are not identical, in this model, they are considered to be synonymous. Borrowers who prepay pay off the entire amount of the home loan. In any given month, these choices are mutually exclusive, competing options because by exercising one, a borrower forfeits the opportunity to exercise the other.

Some may argue that there is a fourth option, partial payment. While partial payment is still a payment, borrowers who pay only a partial amount are also in default. Default is defined in legal, contractual terms: A borrower who does not fulfill their payment obligation in that month is in default. Even if a loan is modified and monthly payments are reduced, then borrowers are still responsible for paying the reduced mortgage payment in full, each month. Failure to make a payment in full will result in the loan entering into default.

The option model is based on the following four assumptions (Vandell 1995). One, there are no transaction costs of refinancing, sale, or recuperating reputation (i.e., change in credit score). Second, it is assumed that a borrower is always able to access financing from other sources quickly and without cost, even when income is disrupted due to a negative life event (divorce, job loss, etc.). Therefore, a borrower will always make a payment when there is no financial incentive to default or prepay the loan. Also, a borrower will always prepay the loan when there is a financial incentive to do so. Third, the default and prepayment decisions are entirely the borrowers'. Therefore, the lender does not or cannot negotiate an alternative arrangement.

Fourth, the borrower is rational or "ruthless." For example, between 2007 and 2013, the period under study, interest rates declined at the same time that many homes

were worth less than the balance of the loan (i.e., "underwater"). Therefore, many borrowers had a financial incentive to both default and prepay the loan, even though they could not do both. The decision to prepay, default, or make a payment is primarily based on which option is more valuable, as discussed in the Variables section below.

It is immediately obvious that the four assumptions that option theory relies on do not hold in reality. First, there are always transaction costs in selling a home or refinancing a loan, which can be substantial, and there are always reputation costs on a credit rating for not making a loan payment, which can be both expensive and long lasting. Second, borrowers in financial distress are among those who are least able to access financing, especially in the case of job loss, where evidence of stable employment and income is often a precondition for a loan. Third, lenders are not absent from a borrower's decision to default or prepay their loan because it is the lenders who are able to reduce the probability of default by renegotiating the terms of a mortgage. Fourth, borrowers are not always rational, as many continue to make payments on their home even if it makes sense to walk away.

Despite the unbelievability of the option model's four main assumptions, its value is undeniable. The real world, after all, is characterized by informational failures, confusing contracts, misleading lenders, and barriers to good decision making under stressful circumstances. No empirical model can capture the world's fine-grained detail, nor some literal truth, but it can capture important, if incomplete, aspects of real-world decision making. While option theory does not explain the full spectrum of borrower behavior, it does incorporate the empirical fact that a borrower's risk of default increases as equity becomes more negative and a borrower's risk of prepayment increases as interest rates decline (Foote et al. 2008). Despite the assumptions, the power of the model is that it does a good job of explaining a borrower's risk of default and prepayment, as indicated by the predictive power of the put and call options (Deng et al. 2000).

A competing risks proportional hazard model is an extension of the Cox proportional hazard model for a single risk, which examines the probability or risk of foreclosure in a given month, conditional on the fact that a house is not in foreclosure at the beginning of that month. The competing risks model is similar to the Cox model, except that it indicates the probability of foreclosure in a given month, conditional on neither being in foreclosure *nor* prepaying at the beginning of that month. The difference between the two has to do with censoring, which is when the study period ends before loan outcomes are known. In the Cox proportional hazard model, all cases that do not experience foreclosure by the end of the study period are considered at risk of foreclosure. In the competing risks model, borrowers who do not experience foreclosure *or* prepayment are considered at risk because if a borrower prepaid the loan, they are no longer at risk of foreclosure. The competing risks model distinguishes between cases still at risk from those no longer at risk even if they did not experience foreclosure by the end of the study period.

The competing risks hazard model used here was developed by Fine and Gray (1999). To distinguish between the cause-specific hazards produced in the Cox model, Fine and Gray call the resulting estimates a *sub*hazard and denote it with an h-bar:

$$\overline{h}_1(t|x) = \overline{h}_{1,0}(t)\exp(x\beta), \tag{1}$$

where a vector of covariates x alters the baseline hazard function proportionately in exponential form. As a final step in the application of the model, all continuous variables are mean centered in order to provide a more meaningful interpretation of the results. If continuous variables were not mean-centered, then the resulting coefficients would be measured from the base case (i.e., if the Black–White segregation index is 0 or there is no segregation at all), which makes it difficult to interpret the results of the coefficients in a meaningful way.

While other models may be appropriate for different questions, the model used here best accounts for the competing option of prepayment in determining a borrower's risk of foreclosure. For example, a path model may help to illuminate the mechanism through which race and segregation affect foreclosure, either directly or indirectly (Baxter and Lauria 2000). Alternatively, a hierarchical linear model could be used to explore individual variation in foreclosure within neighborhoods within and among cities (Baumer et al. 2012). Despite the virtues of these and other models (for a comparison and explanation, see Calhoun and Deng 2002), the competing risks hazard model is considered to provide the most unbiased prospect for determining a borrower's risk of foreclosure (Yezer 2010).

DATA

To examine the impact of race and segregation on a borrower's risk of foreclosure, I use a dataset of home loans from across the United States that combines borrower demographic information with loan performance data. Loan performance data come from Corporate Trust Services (CTS), which includes information on the original balance, interest rate, credit score, and monthly loan status (current, delinquent, foreclosure, or prepayment), as well as a variety of other variables, and are publicly available from the web (www.ctslink.com).[1] Demographic data come from the Home Mortgage Disclosure Act (HMDA), which requires lenders to report the census tract of all loans issued, as well as the borrower's race, sex, and income at time of loan application, as indicated by the primary borrower. Independently, both are publicly available, but the merged data used in the analysis are not.

CTS is a service provided by Wells Fargo to other lending institutions for administering securitized loans. Securitization is a financial practice whereby loans with various levels of risk are combined into a single security and sold as a bond. As most subprime loans were securitized,[2] a majority (75%) of loans in the CTS sample contain at least one characteristic of a subprime loan.[3] As a result, the data used here comprise about one-third of all subprime loans issued during the peak of subprime lending.[4] According to Quercia and Ding (2009), who, along with White (2008), also use the same CTS data (albeit without matched borrower characteristics), the data are neither representative of all loans on single-family homes, most of which are prime,[5] nor all subprime mortgage loans. Instead, the data are representative of securitized loans, a majority of which are subprime, but a minority are prime.[6] The sample is a useful tool to examine the relationship between race, segregation, and foreclosure because subprime loans represented a majority of all foreclosures during the Great Recession (Ferreira and Gyourko 2015), the majority of subprime loans were securitized, and subprime loans were unequally distributed across race and geography.

Datasets that contain both loan performance and demographic characteristics are unique and rare, and only a few existing papers use such combined data (for other examples, see Munnell et al. 1993; Firestone et al. 2007). The CTS data were merged with HMDA data by the San Francisco Federal Reserve Bank using the following variables: loan number, origination date, loan amount, lien status (first or second), and loan purpose (purchase vs. refinancing). While HMDA data are publicly available, the HMDA data used here are restricted because it includes the actual loan number on which the data are merged together, not matched using the variables available in both datasets, as in other published journal articles (see Ghent et al. 2014 and Rugh 2015). Therefore, the data used here are the result of merging unrestricted, public data on loan performance with restricted, public data on borrower demographic information.

The study period is 84 months (7 years), from December 2006 through December 2013, where we observe a borrower's risk of foreclosure for each month on loans originating between 2004 and 2007, the peak of the subprime market.[7] The full study period actually contains two distinct periods, one from 2007 to 2010, which includes the foreclosure crisis, and another, from 2010 to 2013, which includes the aftermath; however, distinguishing one from the other does not alter the results, as detailed in the Sensitivity Analysis section of the Appendix A.2.

The analysis sample was created according to the following selection criteria. Of the nearly 5 million unique loans in the CTS dataset, more than 2.5 million, or 50%, were matched to the HMDA data for loans originating between 2004 and 2007. Due to computational limitations (the competing risks hazard model used takes nearly 36 hours to run), a random sample (20%) of matched loans was used,[8] leaving a sample size of 500,000 loans in the United States. However, I also rerun the analysis for each of the four Census regions (Northeast, Midwest, South, and West). Given the size of each individual region, it is possible to use a 50% sample for each region, prior to applying the selection criteria detailed below.

I apply the following filters to the 500,000 loans in the 20% sample of matched loans, which leaves 192,617 unique loans in the sample used for analysis. While the selection criteria exclude many observations, there are clear reasons for doing so and the results are largely unaffected by these decisions, as shown in Appendix A.2. Loans with missing race or income information were dropped because both are independent variables of interest, reducing the total number of loans by 11%. Also, I kept only loans that were first liens (i.e., mortgages) given to owner-occupied, single-family homes because these are distinct types of property owners and loans.[9] Mortgages issued to non-owner-occupied homes or non-single-family homes are primarily used by investors for manufactured or multifamily housing units. A first lien is the primary loan that is secured by the property. In the event of default, the first lien has first priority for repayment. This exclusion reduced the sample of loans by 30%.

I also dropped loans that enter into bankruptcy at any point in the study period because bankruptcy stops the foreclosure process, as lenders must now compete with each other for borrower repayment in court; this reduced the total number of loans by an additional 7%. Bankruptcy is important, but alters a borrower's risk of foreclosure in ways deserving of its own, separate exploration. By definition, borrowers that enter bankruptcy are less likely to prepay a loan because refinancing is no longer an option. At the same time, bankruptcy could make foreclosure either more or less likely. Foreclosure is more likely because secured loans, like mortgages, are prioritized in the bankruptcy process,

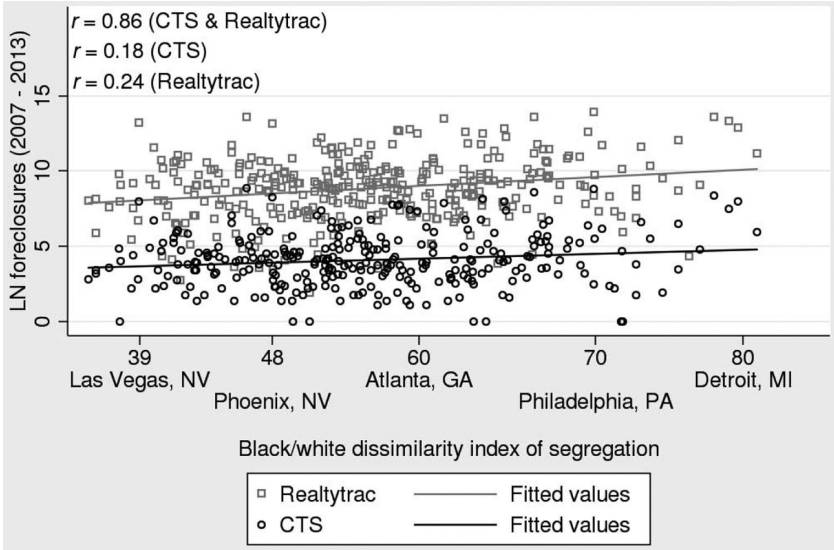

FIG. 2. Correlation between RealtyTrac and CTS foreclosures across metropolitan areas.

but foreclosure is less likely if homeowners use bankruptcy to reduce the cost of their mortgage. The inclusion of those who enter bankruptcy does not alter the main findings I present, as detailed in the Sensitivity Analysis section of Appendix A.2.

The sample size was further reduced by 12% to exclude loans with missing or unusual information (i.e., a credit score over 1,000). As a last step, I dropped loans with borrower income in the top and bottom 1% of the distribution. The final dataset contains 7,927,358 loan-month observations based on 192,617 unique loans and refers to the study period between December 2006 and December 2013.

To address the issue of the selection criteria, which excluded many loans, the sensitivity and robustness of the results to a variety of alternative variable and data specifications are described in detail in the Analysis section of Appendix A.2. Beyond the selection criteria, a primary concern is the 50% of all cases in the CTS data that do not match the HMDA data, in addition to the 11% of matched cases with missing race and income information. To address this issue, several models in the sensitivity analysis also include unmatched observations in addition to the matched ones.

Despite inclusion into the sample owing to loan type, as well as subsequent data cleaning and editing, the data are representative of foreclosures in metropolitan areas across the United States and are therefore appropriate for examining the relationship between segregation and foreclosure. As evidence, foreclosure data in the CTS are aggregated at the metropolitan level and compared to foreclosures in RealtyTrac between 2007 and 2013. RealtyTrac compiles data from all foreclosure filings made in county courthouses to create the largest source of information on foreclosures in the United States. Figure 2 plots the relationship between the number of foreclosures in a metropolitan area and the Black–White racial segregation measure in that metro area for both CTS and RealtyTrac data. The correlation between the CTS and RealtyTrac data is high (0.86), and both indicate a positive correlation (0.20) between foreclosures and racial segregation. In

summary, the CTS data are representative of foreclosures in the United States and are also in line with previous evidence regarding the positive relationship between segregation and foreclosure at the metro level (Rugh and Massey 2010).

VARIABLES

In addition to the dependent variable of foreclosure, there are four sets of independent variables: borrower, geographic, options, and loan-level characteristics. Of these, the variables of interest are race, income, segregation, and the options. The remaining control variables of interest (i.e., credit score, loan type, etc.) are mentioned, but described in detail in Appendix A.1. Table 1 presents descriptive statistics of the variables used in the model. The descriptive statistics are based on loans at their period of last observation because the data have a panel structure, where there are multiple observations of unique loans. Most variables are fixed over time, but a few vary over time and are indicated as such in the text.

The dependent variables are loan status, as indicated by foreclosure, prepaid, or current. Foreclosure refers to a loan that is real-estate owned (REO), the ultimate stage in the foreclosure process. Prepayment refers to a loan that is paid off early.[10] Current refers to a loan that is not REO or prepaid, meaning that loans that are current, delinquent, or have received notice of foreclosure are all considered current. The total, cumulative foreclosure rate over the entire five-year study period is 31%, the prepay rate is 42%, and the remaining 27% are current as of the end of the study period, December 2013.

The probability a borrower will default, prepay, or stay current on a loan is measured by what are called put and call options. The put option is the ratio of the value of the loan to the value of the home. The higher the put option, the less a home is worth (i.e., "underwater") and the higher the risk of default. The call option is the ratio of the current interest rate on the loan to the market interest rate. As the call option rises, the incentive to refinance the home loan at a lower interest rate increases, and the risk of prepayment rises. If neither option is "in the money," then the borrower will continue to make payments on the loan. The put and call options are time-varying variables and are defined in detail below.

The put option is the current balance of the home divided by the current value of the home in any given month, minus one. I subtract one in order to center the variable around 0, such that a negative put option means the value of a home is higher than the balance of its loan, while a positive put option means the value of a home is lower than the balance of its loan (i.e., a homeowner is "underwater"). The current value is equal to the original value of the home multiplied by a monthly, Zip-code-level home price index (HPI). The HPI value comes from Zillow.com, which uses public and private data from home sales to estimate sale prices on all homes. While no HPI is free of bias, Zillow is preferable because it offers a high correlation to other indices, is available at the Zip code level, publicly available, and, as a result, is used by real estate professionals and scholars (Mian and Sufi 2009). As the value of a home declines in relationship to the balance of its loan, the put option rises. The higher the value of the put option, the more it makes sense a borrower to stop paying the loan and go into default.

The average value of the put option at period of last observation was −2.2%, which indicates that the average borrower has a home that is 2.2% more than the value of the

TABLE 1. Descriptive Statistics

Variables (Expected Sign)	mean	SD	min	max
Race and income :				
White (Omitted)	0.578	0.494	0	1
Black (+)	0.138	0.344	0	1
Hispanic (+)	0.207	0.405	0	1
Other (Unknown)	0.077	0.267	0	1
Income ($10,000s) (−)	11.392	8.320	2	61
Index of segregation:				
Black (+)	60.529	11.159	22	81
Hispanic (+)	49.258	9.555	15	70
Asian (+)	44.739	6.186	20	72
Region:				
Northeast	0.138	0.345	0	1
Midwest	0.144	0.352	0	1
South	0.302	0.459	0	1
West	0.415	0.493	0	1
Put (default) option:				
[1]Put (default) option (+)	−2.165	36.102	−100	324
[1]Put option[2] (+)	1,308.002	3,258.597	0	105,244
Call (prepay) option:				
[1]Call (prepay) option (−)	7.562	41.811	−410	73
[1]Call option[2] (−)	1,805.340	3,567.680	0	168,100
Loan:				
<620 (Omitted)	0.253	0.434	0	1
620–679 (−)	0.300	0.458	0	1
680–719 (−)	0.182	0.386	0	1
>= 720 (−)	0.266	0.442	0	1
High Cost Loan (>= 300 BPS indicator) (+)	0.417	0.493	0	1
Loan to value (LTV) (+)	80.827	14.758	1	125
Purchase (vs. refinance) indicator (+)	0.469	0.499	0	1
[1]ARM (vs. FRM) indicator (+)	0.685	0.465	0	1
[1]Modification indicator (−)	0.145	0.352	0	1
Payment to income (PTI) > 31% indicator (+)	0.140	0.347	0	1
Dependent variables:				
Current	0.311	0.463	0	1
Prepay	0.423	0.494	0	1
Foreclosure	0.266	0.442	0	1

As of last period of observation.
[1]Indicates time-varying covariate.

loan. The minimum value of the put option (−100) means that the current value of a home is worth double the balance of its loan (e.g., a home with a current balance of $100,000 and a current value of $200,000). Here, there is no financial incentive to stop paying the mortgage because a borrower could sell the home, pay off the loan, and keep the remaining balance from the sale. The maximum value of the put option (324) means that the balance of the loan is worth 3.24 times the value of the home (e.g., a home with a current balance of $100,000 and a current value of $30,864). The put option is the standard way of examining the consequences of home price change on an individual borrower's risk of foreclosure.

The call option is the current market interest rate[11] divided by the current interest rate on the mortgage in any given month, subtracted by one to center the call option

around 0. A negative call option indicates that there is no financial incentive to refinance because borrowers would receive a higher interest rate than they are currently paying if they refinanced the loan at current rates. A positive call option indicates that there is a financial incentive to refinance because borrowers would receive a lower interest rate if they refinanced their loan at current rates, which reduces the total cost of the loan. As the call option rises, the better it would be to refinance the mortgage, the higher the risk of prepay, and the lower the risk of foreclosure.

The average value of the call option at period of last observation indicates that the value of the average interest rate on loans is 7.5% higher than the current market interest rate. By contrast, the minimum value of the call option (−410) means that the market interest rate is four times higher than the current interest rate on the loan. Technically, this ought to be good because the borrower's current interest rate is much lower than the market interest rate, but in this case, the loan is actually an ARM where the low teaser rates are about 1% and market rates are around 4%. An adjustable rate loan is a defining feature of a subprime loan because low introductory or "teaser" rates reduce monthly payments, making loans affordable that might otherwise be unaffordable when interest rates revert back to market rates after a period of time. It is therefore necessary to include mortgage type as a time-varying control variable, which I do.

The maximum value of the call option is 73, meaning that the value of the average interest rate on the loan is 73% higher than the current market interest rate (i.e., a current interest rate of 8.65% and a market interest rate of 5%). In this case, there is a financial incentive to refinance because borrowers would save money on their loan by prepaying their old loan and refinancing under current market rates. In keeping with the literature, the put and call options are squared because their effect on foreclosure is nonlinear and convex, meaning the larger the effect of a one-unit increase, the higher the option value (Ciochetti et al. 2002).

Race is aggregated into White, Black, Hispanic, and other. Borrowers are 58% White, 21% Hispanic, and 14% Black. Income is borrower income at the time of loan origination, as shown in $10,000s of dollars. Average income is $111,390, with clear differences in income by race: Compared to Whites, Hispanics earn 17% less and Blacks 31% less. I also include the interaction between race and income to capture any racial differences in the effect of income as an additional control variable.

If income is not misreported,[12] then it is high. According to HMDA data (Avery et al. 2010), average income for all borrowers in 2007 was $97,700 and $85,600 for borrowers with a high-cost (i.e., subprime) loan, as described in the Variables section. While this could present a potential threat to validity, subsequent analysis reveals that income for all borrowers would have to be negative in order to eliminate the main result, the negative impact of racial segregation on a borrower's risk of foreclosure, as detailed in Appendix A.2.

There are three segregation indices, one for each minority race, Black, Hispanic, and Asian. Of these, the particular variable of interest is the Black–White segregation index because Black–White segregation is shown to have a positive effect on foreclosures in an MSA (Rugh and Massey 2010). The segregation index is an index of dissimilarity, and measures the proportion, from 0 to 100 percent, one group would need to change census tracts to achieve an even distribution with Whites in an MSA. Data from the 2005–2009 American Community Survey (ACS) five-year Summary File are used to create the measure, which overlaps the years of loan origination (2004–2007)

and the study period (December 2006–December 2013).[13] The mean index of Black dissimilarity for the entire sample is 60, meaning that the average borrower lives in an MSA where 60% of Blacks would have to switch census tracts in order for Blacks and Whites to be equally distributed in that MSA. For reference, the MSA of Atlanta, Georgia, has a Black, Hispanic, and Asian dissimilarity index around the mean of the sample.

It is also standard protocol to include a set of control variables in the model when examining the factors affecting a borrower's risk of foreclosure (Kau et al. 1995; Deng et al. 2000): credit score (FICO), region, high cost loan (>300 basis points above the prevailing interest rate at time of loan origination), loan to value (LTV), loan purpose, payment-to-income (PTI), loan modification, and mortgage type (ARM vs. FRM). The control variables are described in Appendix A.1. Of these, only loan modification and mortgage type are time varying, the others are time invariant. Year of loan origination (2004–2007) is also included as a fixed effect. All control variables operate as expected and are shown in Table 3, e.g., an ARM has a higher risk of foreclosure than an FRM.

RESULTS

The results indicate that there is a negative association between segregation and foreclosure, which operates through the mechanism of home price change and explains most of the direct impact of race on foreclosure. A total of 10 models are presented, six using the national sample and then one for each of the four regions, separately. The findings indicate that the impact of Black–White racial segregation on a borrower's risk of foreclosure is negative, being non-White is positive, and the interaction between race and segregation is positive, but not enough to alter the main effects. However, it is the decline in home value in relationship to the value of the loan (i.e., the put option) that has the single largest positive impact on a borrower's risk of foreclosure. Results are consistent across the United States and in each of the four subregions, and are robust to a variety of alternative variable specifications and subsamples, as shown in Appendix A.2.

The results for the primary variables of interest are shown in Table 2, and the control variables are shown in Table 3. The coefficients and standard errors are presented in exponential form for ease of interpretation. A coefficient greater (or less) than 1 means the effect of that covariate increases (or decreases) the risk of foreclosure given the competing risk of prepayment with respect to the baseline (continuous variables are mean-centered, as stated in the Methods section). The models will be described below and Table 4 illustrates the results in terms of effect sizes to better understand the importance of the variables relative to one another, given their varying units of measurement.

The first model only controls for race, income, and their interaction. Blacks and Hispanics have a higher risk of foreclosure than Whites (51.9% and 97.6% higher, respectively) and, as income rises, the risk of foreclosure declines by 22% for each $10,000 increase in income. The second model controls only for segregation. As stated earlier, while all three segregation variables are included, only the Black–White variable is of interest, as it is the only one suggested to have a causal effect on foreclosures. A one-unit increase in Black dissimilarity is associated with a 1.5% decrease in a borrower's risk of

TABLE 2. Competing Risks Hazard Model

	(1) Race & Income b/se	(2) Segregation b/se	(3) Race, Income, & Seg b/se	(4) + Interaction b/se	(5) + Loan Controls b/se	(6) + Options (Main) b/se	(7) Northeast b/se	(8) Midwest b/se	(9) South b/se	(10) West b/se
Borrower:										
Black	1.519***		1.599***	1.567***	1.217***	1.038*	1.235***	1.233***	1.030	0.956*
	(0.023)		(0.024)	(0.024)	(0.019)	(0.017)	(0.040)	(0.026)	(0.016)	(0.019)
Hispanic	1.979***		2.063***	2.041***	1.652***	1.258***	1.380***	1.325***	1.319***	1.262***
	(0.021)		(0.022)	(0.023)	(0.019)	(0.015)	(0.041)	(0.033)	(0.019)	(0.013)
Other	1.310***		1.349***	1.334***	1.301***	1.221***	1.217***	1.294***	1.325***	1.147***
	(0.023)		(0.023)	(0.024)	(0.024)	(0.022)	(0.064)	(0.049)	(0.036)	(0.015)
Income ($10,000s)	0.978***		0.981***	0.981***	0.999	1.001	0.989***	0.995**	1.000	1.004***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Segregation:										
Black		0.985***	0.985***	0.985***	0.988***	0.988***	0.977***	0.986***	0.998	1.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Hispanic		1.008***	1.003***	1.003***	1.008***	1.003***	1.016***	1.005***	0.988***	1.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Asian		0.997***	0.999	0.999	1.000	1.012***	1.000	0.995*	1.022***	1.002*
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)
Race and segregation interaction:										
Black and Black segregation				1.007***	1.007***	1.005***	1.004	1.002	0.997	0.998
				(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.002)	(0.002)
Hispanic and Black segregation				0.996***	0.995***	1.002*	0.998	0.996	1.001	1.001
				(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)
Other and Black segregation				0.997*	0.995**	1.003*	1.012	0.985***	1.008*	1.002
				(0.001)	(0.001)	(0.002)	(0.007)	(0.003)	(0.004)	(0.001)

(*Continued*)

TABLE 2. Continued

	(1) Race & Income b/se	(2) Segregation b/se	(3) Race, Income, & Seg b/se	(4) + Interaction b/se	(5) + Loan Controls b/se	(6) + Options (Main) b/se	(7) Northeast b/se	(8) Midwest b/se	(9) South b/se	(10) West b/se
Put (default) option:										
Put option						1.038***	1.055***	1.046***	1.033***	1.040***
						(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Put option^2						1.000***	1.000***	1.000***	1.000***	1.000***
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Call (prepay) option:										
Call option						1.015***	1.026***	1.019***	1.016***	1.012***
						(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Call option^2						1.000***	1.000***	1.000***	1.000***	1.000***
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Observations	7,927,358	7,927,358	7,927,358	7,927,358	7,927,358	7,927,358	2,955,494	2,715,450	6,282,241	7,848,442
Number of Subjects	192,617	192,617	192,617	192,617	192,617	192,617	65,407	70,788	144,915	199,378
Number of Failures (Foreclosure)	50,874	50,874	50,874	50,874	50,874	50,874	8,755	21,065	35,711	62,607
Number of Competing (Prepay)	81,117	81,117	81,117	81,117	81,117	81,117	30,280	29,402	60,041	82,005
Control variables:										
Year of origination	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A	N/A
Race and income interaction	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan performance (LTV, ARM, etc.)	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients.
National models are run on a 20% sample. Regional models are run on a 50% regional sample.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 3. Competing Risks Hazard Model - Control Variables from Table 2

	(1) Race & Income b/se	(2) Segregation b/se	(3) Race, Income, & Seg b/se	(4) + Interaction b/se	(5) + Loan Controls b/se	(6) + Options (Main) b/se	(7) Northeast b/se	(8) Midwest b/se	(9) South b/se	(10) West b/se
Year of origination:										
2005	1.761***	1.827***	1.755***	1.755***	1.505***	0.896***	0.730***	0.803***	0.952*	0.951**
	(0.029)	(0.030)	(0.028)	(0.028)	(0.025)	(0.016)	(0.029)	(0.017)	(0.020)	(0.017)
2006	2.541***	2.696***	2.523***	2.521***	2.084***	0.762***	0.552***	0.678***	0.811***	0.782***
	(0.040)	(0.042)	(0.040)	(0.040)	(0.034)	(0.014)	(0.023)	(0.015)	(0.018)	(0.014)
2007	2.268***	2.324***	2.263***	2.262***	2.060***	0.621***	0.391***	0.573***	0.671***	0.640***
	(0.043)	(0.044)	(0.043)	(0.043)	(0.041)	(0.013)	(0.021)	(0.018)	(0.018)	(0.013)
Race and income interaction:										
Black and income	1.023***		1.023***	1.023***	1.008***	1.005*	1.029***	0.999	1.010***	1.012***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.003)
Hispanic and income	1.037***		1.037***	1.037***	1.022***	1.016***	1.025***	1.026***	1.018***	1.012***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.005)	(0.002)	(0.001)
Other and income	1.011***		1.010***	1.011***	1.005*	1.007***	1.015*	1.007	1.017***	1.006***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.006)	(0.003)	(0.002)
Region:										
Northeast (ommitted)										
Midwest	2.548***	2.682***	2.621***	2.611***	2.348***	1.648***				
	(0.052)	(0.056)	(0.055)	(0.055)	(0.050)	(0.036)				
South	1.845***	1.720***	1.538***	1.569***	1.690***	1.118***				
	(0.035)	(0.034)	(0.031)	(0.032)	(0.035)	(0.024)				
West	2.583***	2.263***	2.057***	2.073***	2.659***	1.668***				
	(0.047)	(0.044)	(0.041)	(0.042)	(0.055)	(0.035)				
Loan characteristics:										
FICO < 620 (ommitted)										

(*Continued*)

TABLE 3. Continued

	(1) Race & Income b/se	(2) Segregation b/se	(3) Race, Income, & Seg b/se	(4) + Interaction b/se	(5) + Loan Controls b/se	(6) + Options (Main) b/se	(7) Northeast b/se	(8) Midwest b/se	(9) South b/se	(10) West b/se
FICO (620–679)					0.951***	0.966**	1.154***	0.953**	1.010	0.974*
					(0.011)	(0.012)	(0.032)	(0.016)	(0.014)	(0.013)
FICO (680–719)					0.794***	0.814***	0.964	0.791***	0.868***	0.786***
					(0.012)	(0.013)	(0.036)	(0.020)	(0.016)	(0.012)
FICO (>= 720)					0.510***	0.563***	0.669***	0.509***	0.599***	0.575***
					(0.008)	(0.010)	(0.030)	(0.016)	(0.013)	(0.009)
High cost loan (>= 300 BPS indicator)					1.351***	1.296***	1.414***	1.239***	1.264***	1.357***
					(0.015)	(0.014)	(0.042)	(0.023)	(0.017)	(0.013)
Loan to value (LTV)					1.033***	1.005***	0.994***	0.984***	1.002**	1.009***
					(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Purchase (vs. refinance) indicator					1.187***	1.265***	1.105***	1.288***	1.309***	1.254***
					(0.012)	(0.013)	(0.028)	(0.020)	(0.017)	(0.012)
ARM indicator					1.492***	1.490***	1.330***	1.225***	1.466***	1.605***
					(0.017)	(0.018)	(0.036)	(0.021)	(0.019)	(0.019)
Modification indicator					1.164***	0.901***	0.800***	1.135***	1.025	0.781***
					(0.017)	(0.015)	(0.033)	(0.028)	(0.020)	(0.012)
PTI > 31% indicator					1.053***	1.024	1.058*	1.070**	1.019	0.996
					(0.014)	(0.014)	(0.029)	(0.022)	(0.017)	(0.013)

Exponentiated coefficients.
National models are run on a 20% sample. Regional models are run on a 50% regional sample.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 4. Marginal Effect on the Risk of Foreclosure from the Main Model

	United States				Northeast	Midwest	South	West
	mean	mean + 1 sd	$exp(\beta)$	β	β	β	β	β
Continuous variables (Marginal effect of 1 SD increase from the mean):								
Put option	2.508	38.24	3.048	1.114***	1.48***	1.278***	0.98***	1.166***
Call option	8.467	45.111	1.83	0.604***	1.033***	0.737***	0.658***	0.5***
LTV	80.74	95.49	1.073	0.071***	−0.082***	−0.241***	0.024**	0.139***
Income ($10,000s)	11.637	20.134	1.05	0.049***	−0.01	0.001	0.053***	0.07***
Asian segregation	44.869	51.078	1.075	0.072***	0.002	−0.03*	0.137***	0.013*
Hispanic segregation	49.586	59.15	1.025	0.024**	0.154***	0.05***	−0.112***	−0.004
Black segregation	60.88	71.97	0.886	−0.121***	−0.241***	−0.173***	−0.015	0.001
Categorical variables:								
ARM indicator			1.49	0.399***	0.285***	0.203***	0.383***	0.473***
High cost loan indicator			1.296	0.259***	0.347***	0.215***	0.234***	0.306***
Purchase to refinance			1.265	0.235***	0.1***	0.253***	0.269***	0.226***
Hispanic to White			1.258	0.229***	0.323***	0.281***	0.277***	0.233***
Black to White			1.038	0.037*	0.211***	0.21***	0.029	−0.045*
PTI > 31% indicator			1.024	0.023	0.057*	0.068**	0.019	−0.004
Modification indicator			0.901	−0.104***	−0.223***	0.126***	0.025	−0.248***
FICO >= 720 to FICO < 620			0.563	−0.575***	−0.401***	−0.675***	−0.513***	−0.553***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

foreclosure. The coefficients on the race variables are positive overall, but the coefficient is negative on the Black–White racial segregation variable.

The third, fourth, and fifth models add additional control variables for borrower, loan, and geographic characteristics. The third model includes only the variables for race, income, and segregation, which does not alter the interpretation of the coefficients found in model 1 or 2. The fourth model adds an interaction between race and segregation, which is positive, but does not alter the coefficients on the main race or segregation variables. The fifth model adds other important loan and borrower control variables (ARM, FICO, etc.). Adding borrower, loan, and geographic controls does not alter the main effects found in the first two models regarding the positive impact of race or the negative impact of segregation on foreclosure.

Finally, the sixth model using nation-wide data adds in the put and call options to model the joint impact of changes in housing prices and interest rates on foreclosure risk. The put option is positive, as expected, but the call option is also positive, not negative. Theoretically, risk of foreclosure should decline as the call option rises because these borrowers should prepay their loan by refinancing; however, the results suggest the opposite, that the risk of foreclosure rises as the call option rises. The reason is that during the study period, interest rates fell even as the availability of credit declined. As a result, only borrowers with excellent credit scores were able to get a loan, which would exclude most borrowers who got subprime loans in the first place. Adding the option variables reduces most, but not all, of the positive impact of race on foreclosure risk without altering the negative impact of segregation on foreclosure risk found in each of the models.

To better understand all the coefficients together, given the presence of both categorical variables, continuous variables with varying units of measurement, and their interactions, Table 4 presents the results in terms of their marginal effects. The marginal effect for each of the categorical variables is compared to the marginal effect of one standard deviation (SD) increase from the mean for each of the continuous variables.

Looking at the effect size of the continuous variables of interest, an SD increase in the put option (from −2 to 34, i.e., the value of a home is worth 2% *more* than the value of the loan vs. 34% *less* than the value of the loan) increases the risk of foreclosure by a factor of 3. An SD increase in the call option (from 8 to 49, i.e., the interest rate on the home loan is 8% higher than the current market interest rate vs. 49% higher) increases the risk by nearly a factor of 2. By contrast, an SD increase in Black–White racial segregation (from 61 to 72, i.e., from a city like Atlanta to Boston) decreases the risk of foreclosure by 12%. If we examine the raw beta coefficients, then the size of the positive coefficients on the put and call options is 10 and four times larger, respectively, than the size of the negative coefficient on Black–White racial segregation.

Looking at the effect size of the categorical variables of interest, Blacks have a 3.8% higher foreclosure risk than Whites, and Hispanics have a 25.8% higher foreclosure risk than Whites. Borrowers with an ARM have a 49% higher risk of foreclosure compared to borrowers with an FRM. Borrowers with a credit score below 620 (i.e., "subprime") have a 56.3% higher risk of foreclosure as compared to borrowers with a credit score above 720 (i.e., "excellent"). If we examine the raw beta coefficients, then the size of the coefficients on credit score and loan type is much larger than the size of the coefficients on being Black or Hispanic.

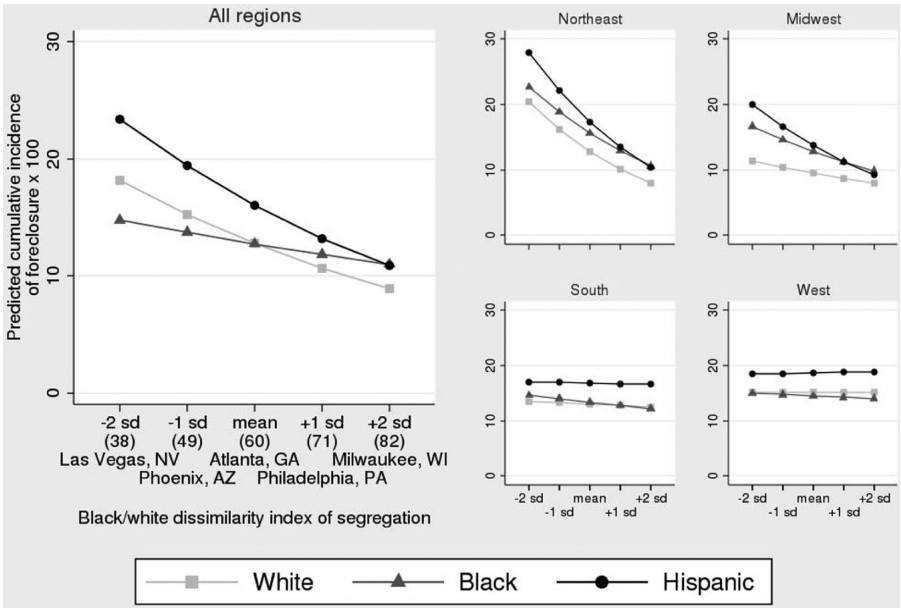

FIG. 3. Interaction between race and Black segregation.

DISCUSSION

I draw three main conclusions from the results. First, we may accept Hypothesis 1. The relationship between segregation and foreclosure is negative, not positive. Figure 3 presents the relationship between segregation and foreclosure risk in graphical form for both the United States and its four regions. Borrowers who live in metro areas with higher levels of Black–White racial segregation have a lower risk of foreclosure than borrowers who live in metro areas with lower levels of racial segregation. The negative impact of segregation on a borrower's risk of foreclosure appears to be the result of an inverse relationship between racial segregation and housing price decline, between 2007 and 2013, as shown in Figure 1. Declines in home prices were larger in cities with lower levels of racial segregation as compared to cities with higher levels of segregation.

Second, we may partially accept (or partially reject) Hypothesis 2. Racial differences in risk of foreclosure remain, but most of the difference in the foreclosure rate by race is explained by the full model. Figure 4 presents the relationship between race and foreclosure for both the United States and its four regions. Unadjusted, the national foreclosure rate is 50% higher for Blacks and 100% higher for Hispanics, as compared to Whites. The main model that controls for all covariates and interactions explains 97% of the difference between Blacks and Whites, and 81% of the difference between Hispanics and Whites. While racial differences in risk of foreclosure persist, the impact of race on foreclosure is mostly indirect and appears to operate through the mechanism of housing price change.

Third, we may partially accept (or partially reject) Hypothesis 3. There is an interaction effect between segregation and race, but it does not offset the main effects of segregation

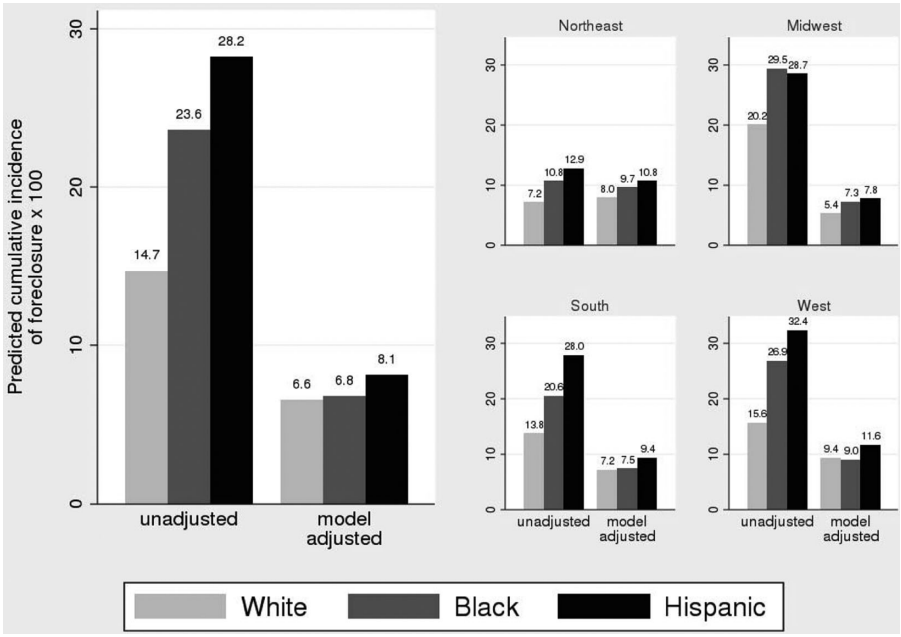

FIG. 4. Model-adjusted cumulative foreclosure rate by race.

or race. Figure 3 also graphs the relationship between foreclosure risk and the interaction of race and segregation. The interaction effect is positive, as Blacks living in highly segregated cities (+1 SD above the mean) have a higher risk of foreclosure than Whites who also live in highly segregated cities. However, the main effect remains, as Blacks living in cities with lower levels of segregation have a higher risk of foreclosure than Blacks living in cities with higher levels of segregation. Further, the coefficients on the interaction terms between race and segregation are insignificant in each of the four Census regions. The impact of race may be exacerbated by the impact of segregation, but neither of these appears to offset the importance of housing price change in determining a borrower's risk of foreclosure.

CONCLUSION

Segregation is an important factor in understanding the foreclosure crisis, but may operate in different ways depending on the stage of a home loan. The primary, positive impact of segregation on foreclosure risk operates prior to loan origination through the differential access to loan quality by race. After loan origination, the results presented here indicate that the impact of segregation is negative. Instead of segregation, it is home price decline that has the largest and positive impact on a borrower's risk of foreclosure, and explains most of the difference in the cumulative foreclosure rate by race, especially between Blacks and Whites.

The primary explanation for the negative impact of segregation on foreclosure risk found in the results is the inverse relationship between segregation and housing price

decline. While housing prices rose across the country in the period prior to the foreclosure crisis, housing prices rose most in cities with lower levels of segregation, as shown in Figure 1. Cities with high levels of racial segregation are older and more industrial, with stable or even declining population levels and related housing price changes. By contrast, cities with low levels of racial segregation are newer and less industrial, with rising population levels, and related housing price changes. After the foreclosure crisis began in 2007, housing prices declined the most in the cities where housing prices increased the most prior to the beginning of the crisis, cities with lower levels of segregation. Therefore, borrowers living in cities with high levels of segregation were more "protected" from large declines in housing prices.

Housing price change, as measured by the put option, is also the mechanism that connects race and foreclosure. While the final model, which included the option variables, explained 97% of the unadjusted difference in the foreclosure rate between Blacks and Whites and 81% of the unadjusted difference between Hispanics and Whites, this does not negate the role of race in foreclosure. Instead, it reveals how race indirectly affects foreclosure. Given declines in housing prices, borrowers responded in similar ways, regardless of race; however, housing price decline was not experienced equally by race. The degree to which housing price change explains differences in the foreclosure rate by race suggests that the impact of race on foreclosure risk is indirect, operating through the direct impact of housing price decline.

However, differences in the foreclosure rate by race persist, even after controlling for loan, borrower, and home price characteristics. The mechanisms that explain the impact of race on foreclosure after loan origination are less clear, and it is hard to disentangle the role of segregation as distinct from discrimination (Yinger 1995). One possible explanation is the role of racial differences in assets and wealth (Oliver and Shapiro 2006). While wealth is used in determining access to loan quality, it is not a variable available in the data used here. Wealth is critical for paying regular, fixed expenses, like a mortgage, in the presence of temporary or permanent declines in household income due to negative life events, like divorce and job loss. Racial differences in wealth mean that non-White borrowers cannot use personal savings to pay for a mortgage as much or as long as White borrowers can.

Another possible explanation is that racial differences have an effect on the ability of loan modifications to both reduce foreclosure risk and also achieve long-term financial stability (for a discussion of class differences in loan modifications, see Owens 2015). The ability of borrowers to either refinance or restructure the terms of a current loan in the face of mortgage delinquency or default is limited. It often requires the assistance of lawyers and loan counselors, who can be expensive. Further, a loan modification is not necessarily a positive outcome in itself (Collins et al. 2013). Many modifications reduce the monthly payment by extending the time on the mortgage, which makes the loan more expensive and costly in the long run. While loan modification is used as a control variable here, more work needs to be done to understand the role of loan modifications in ameliorating or perpetuating racial differences in foreclosure risk.

Differences in the underlying data make it difficult to compare the evidence presented here regarding the negative impact of segregation to the positive impact found in other studies (Rugh 2015). As stated earlier, the data used here are derived from securitized loans, which are primarily, but not exclusively, composed of subprime loans. Therefore, the data are not representative of all loan types, but rather securitized loans, most of

which are subprime. The majority of subprime loans were securitized, and subprime, securitized loans accounted for the majority of all foreclosures during the crisis. At the same time, sensitivity tests indicate that the results also hold with a subsample of prime loans that are also contained in the data. Further, the foreclosure data from the CTS are highly correlated to foreclosure data on all loans from RealtyTrac, as shown in Figure 2. Therefore, the data are well suited to examine the issues addressed in this paper.

Unfortunately, a nationally representative sample on home loans that combines both demographic and loan performance data does not exist. Therefore, the results presented here neither prove nor disprove previous findings. Instead, the mismatch suggests that there may be important differences in how the relationship between racial segregation and foreclosure depends on the structure of the data. In fact, one possibility is that cities with both high levels of racial segregation and rates of foreclosure may have had lower concentrations of subprime loans, as compared to other parts of the country (Mayer and Pence 2008). Additional research is needed to assess when and how residential segregation matters for foreclosure risk to better understand the complete relationship between the two.

While one must exercise caution before extrapolating the results presented here to make population-level conclusions, it is also possible to draw a few connections and implications from the analysis to broaden our understanding of the relationship between race, segregation, and foreclosure. A critical one is that our understanding of the relationship between segregation and foreclosure is based on cities with high levels of Black–White segregation, but the housing crisis primarily took place in cities with low levels of Black–White segregation. This is not to suggest that issues of inequality disappear, but that they may not look the same in different types of cities. Regional differences in inequality are evident if one compares the impact of being Black on a borrower's risk of foreclosure across each of the four Census Regions, the coefficient is positive and significant only in the Northeast and Midwest, but the foreclosure crisis was most concentrated in the Western and Southern parts of the United States. Therefore, future emphasis should be placed on doing more research on housing-related issues in cities with large and rising populations, such as Atlanta, Houston, and Los Angeles, which are perhaps comparably more representative of modern urban life in the United States, as compared to older cities with stagnating or declining populations, such as Boston, Chicago, or Detroit.

It is also important to remember that racial and/or geographic disparities in lending or foreclosure are not by themselves indicative of discrimination. Previous research indicates that including information on loan characteristics explains a large proportion of the difference in denial rates on loans by race (Munnell et al. 1993). The results presented here mirror that basic idea. Controlling for loan quality explains a large proportion of the difference in the foreclosure rate between Blacks and Whites. While racial differences remain, researchers must be diligent to include all relevant control variables and distinguish the impact of race and segregation on loan origination as distinct from their impact on foreclosure. The persistent issue is that the data that may be most useful in examining the relationship between race, segregation, and foreclosure, which also include all relevant control variables across the universe of borrower types and loan qualities, remain difficult to assemble and access. Therefore, more work ought to be done to expand data availability for those without connections to banking or regulatory institutions.

A final issue of concern to both researchers and policy makers is the role of housing prices as the mechanism through which race and segregation affect foreclosure. On the

one hand, the idea that housing prices constitute a dominant factor that affects foreclosure is not surprising; in fact, it would be unusual if housing prices were not important. On the other hand, the role of race and segregation on foreclosure appears to dominate in research on this topic area within the discipline of sociology, as well as larger policy debates. As suggested earlier, one of the reasons for this is data availability, but there is also a larger issue of concern when one under emphasizes the role of housing prices.

While the main focus of this paper is on foreclosure, there is also the issue of staying current on a loan that is underwater, which may never regain its original value. Given the importance of home ownership as a component of wealth, racial differences in loan status for homes that are underwater could exacerbate the current racial inequality in terms of wealth. Future work ought to supplement the quantitative evidence presented here with qualitative work to better understand racial differences in terms of who stays current and who defaults on home loans, extending the work of Owens (2015) on class.

The findings presented here add complexity to the prevailing understanding of the relationship between race, segregation, and foreclosure during the crisis. Black–White racial segregation has a negative impact on a borrower's risk of foreclosure, and the inclusion of housing price change explains most of the difference in the foreclosure rate by race. The results indicate that the role of home prices is a dominant factor that affects foreclosure risk. The less a home is worth with respect to the balance of the loan, the higher the risk of foreclosure. However, the importance of housing prices does not negate the role of race or segregation. Instead, the results suggest that race, segregation, and housing prices contribute to foreclosure risk in multifaceted ways.

Acknowledgements

The author wishes to acknowledge the following individuals for their help throughout the entire process (in alphabetical order): Jan Brülle, David Calnitsky, Matias Cocina, J. Michael Collins, Ted Gerber, Leah Gjertson, Pilar Gonalons-Pons, Gary Green, Carlotta Guistozzi, Erik Hembre, Jason Houle, Richard Latner, Timo Leper, Yifei Li, Kristina Lindemann, Heather O'Connell, Fabian Ochsenfeld, Ellen Pechman, Jacob Rugh, and Jody Schimek.

Notes

[1]Specifically, www.ctslink.com > residential securities (MBS) > other reports and files: shelf documents. Then, each of the following folders beginning with the phrase "Columbia Collateral File": "1999 deals & prior," "2000 thru 2004 deals," "2005 deals," "2006 deals," "2007 deals," "2008 deals." Raw data files are available within each "deal year" for each month. As recently as August 2017, ctslink.com has restricted the data available online to only include the most recent month of loan performance data within any given deal year. However, interested researchers may contact the author to gain access to the loan performance data from December 2006 through December 2015. Please note, the total file size is about 50 GB.

[2]As stated in (2011), "The Mortgage Market Statistical Annual (2007) reports securitization shares of subprime mortgages each year from 2001 to 2006 equal to 54%, 63%, 61%, 76%, 76%, and 75%, respectively."

[3]Of the 5 million unique loans in the full sample, 3.3 million of which were issued between 2004 and 2007, 2.5 million of which contain at least one characteristic of a subprime loan (LTV >= 700, FICO >= 720, a high cost loan, or an adjustable rate mortgage [ARM]).

[4]Estimates may differ on the number of subprime loans because definitions differ on what qualifies as a subprime loan. Further, different sources provide different figures. For example, using HMDA data, there

were 8.7 million "high-priced" loans issued between 2004 and 2007 (Avery et al. 2008; Mayer et al. 2009). At the same time, using loan performance data, a proprietary dataset on home mortgages, Chomsisengphet and Pennington-Cross (2008) find that 11.4 million subprime loans were issued between 2004 and July 2007.

[5]Most loans on single-family homes are prime loans, and subprime loans never constituted more than 20% of mortgage originations in any one year (Furlong and Krainer 2007).

[6]The results also hold with a sample of prime loans included in the data, as detailed in the Sensitivity Analysis section of Appendix A.2.

[7]The peak years of subprime loan origination were between 2004 and 2006, while 2007 was the lowest in number and amount since 2001 (Schwartz 2010; Mayer et al. 2009).

[8]Given that the data are panel with multiple observations of individual loans, the random sample is a 20% sample of loan clusters. The protocol is as follows: I keep one observation from each loan for each state, keep a 20% random sample from the remaining observations, merge the original dataset with the sample (m:1), and keep the matched observations. This follows the protocol suggested by STATA:

(www.stata.com/support/faqs/data-management/sampling-clusters/).

[9]First lien (i.e., mortgage) loans include loans used to refinance a mortgage, which is why refinancing is included as a control variable.

[10]In keeping with the literature, I also declare that a "short sale" is a foreclosure as opposed to a prepayment, which is how it is identified in the raw data. A short sale is a loan that is paid off, but it is comparably more like a foreclosure than a prepayment. Unlike prepayment, a short-sale has a negative impact on a borrower's credit score. Like foreclosure, a short-sale leaves the full amount of the loan unpaid. Designating a short-sale as a foreclosure alters the loan status of 6% of all loans, but the declaration of a short-sale as a foreclosure instead of a prepay does not alter the main findings (as shown in the Sensitivity Analysis in Appendix A.2).

[11]The current market interest rate is taken from the monthly federal interest rate for a 30-year fixed-rate mortgage (FRM), as determined by the Primary Mortgage Market Survey (PMMS), publicly available from Freddie Mac (http://www.freddiemac.com/pmms/pmms_archives.html).

[12]According to Blackburn and Vermilyea (2012), the average mortgage application overstated income between 15% and 20%, but "there is little support this overstatement played a substantial role in the subsequent mortgage defaults."

[13]US2010 Project, www.s4.brown.edu/us2010/Data/Download1.htm.

[14]Specifically, the coefficient on the main effect of Black segregation is −0.0123 and the interaction effect is −0.0002. Therefore, income for all borrowers must be negative for the effect of income to cancel the effect of segregation, specifically, −$62,000 ($0 = -0.0002(-61.5) - 0.0124$).

REFERENCES

Aalbers, Manuel B. 2009. "The Globalization and Europeanization of Mortgage Markets." *International Journal of Urban and Regional Research* 33:389–410.

Ambrose, Brent W., and Charles A. Capone. 1998. "Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions." *Real Estate Economics* 26:391–429.

Avery, Robert B., Neil Bhutta, Kenneth P. Brevoort, Glenn B. Canner, and Christa N. Gibbs. 2010. "The 2008 HMDA Data: The Mortgage Market during a Turbulent Year." *Federal Reserve Bulletin* 96:A169–A211.

Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner. 2008. "The 2007 HMDA Data." *Federal Reserve Bulletin* 94:A107–46.

Baumer, Eric P., Kevin T. Wolff, and Ashley N. Arnio. 2012. "A Multicity Neighborhood Analysis of Foreclosure and Crime." *Social Science Quarterly* 93:577–601.

Baxter, Vern, and Mickey Lauria. 2000. "Residential Mortgage Foreclosure and Neighborhood Change." *Housing P* 11:675–99.

Blackburn, McKinley L., and Todd Vermilyea. 2012. "The Prevalence and Impact of Misstated Incomes on Mortgage Loan Applications." *Journal of Housing Economics* 21:151–68.

Bocian, Debbie G., Keith S. Ernst, and Wei Li. 2008. "Race, Ethnicity, and Subprime Home Loan Pricing." *Journal of Economics and Business* 60:110–24.

Bocian, Debbie G., Wei Li, Carolina Reid, and Roberto G. Quercia. 2011. "Lost Ground, 2011: Disparities in Mortgage Lending and Foreclosures." Center for Responsible Lending. Chapel Hill, NC. Retrieved February 1, 2016, from http://www.responsiblelending.org/mortgage-lending/research-analysis/Lost-Ground-2011.pdf.

Bond, Carolyn, and Richard Williams. 2007. "Residential Segregation and the Transformation of Home Mortgage Lending." *Social Forces* 86:671–98.

Calem, Paul S., Jonathan E. Hershaff, and Susan M. Wachter. 2004. "Neighborhood Patterns of Subprime Lending: Evidence from Disperate Cities." *Housing Policy Debate* 15:603–22.

Calhoun, Charles A., and Yongheng Deng. 2002. "A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations." *Journal of Real Estate Finance and Economics* 24:9–33.

Capone, Charles A. Jr. 2001. "Research into Mortgage Default and Affordable Housing: A Primer." Prepared for the LISC Center for Home Ownership Summit, 2001.

Chomsisengphet, Souphala, and Anthony Pennington-Cross. 2008. "A Look at Subprime Mortgage Originations: 2000-2007." Federal Trade Commission, Consumer Information and the Mortgage Market. Conference titled, "An Economic Assessment of Information Regulation, Mortgage Choice, and Mortgage Outcomes." May 29, 2008.

Ciochetti, Brian A., Yongheng Deng, Bin Gao, and Rui Yao. 2002. "The Termination of Commercial Mortgage Contracts through Prepayment and Default: A Proportional Hazard Approach with Competing Risks." *Real Estate Economics* 30:595–633.

Collins, J. Michael, Maximillian D. Schmeiser, and Carly Urban. 2013. "Protecting Minority Homeowners: Race, Foreclosure Counseling, and Mortgage Modifications." *Journal of Consumer Affairs* 47:289–310.

Courchane, Marsha J., Brian J. Surette, and Peter M. Zorn. 2004. "Subprime Borrowers: Mortgage Transitions and Outcomes." *Journal of Real Estate Finance and Economics* 29:365–92.

Demyanyk, Yuliya, and Otto Van Hemert. 2011. "Understanding the Subprime Mortgage Crisis." *Review of Financial Studies* 24:1848–80.

Deng, Yongheng, John M. Quigley, and Robert Van Order. 2000. "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options." *Econometrica* 68:275–307.

Ding, Lei, Roberto Quercia, Wei Li, and Janneke Ratcliffe. 2011. "Risky Borrowers or Risky Mortgages: Disaggregating Effects Using Propensity Score Models." *Journal of Real Estate Research* 33:245–77.

Ferreira, Fernando, and Joseph Gyourko. 2015. "A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012." NBER Working Paper No. 21261.

Fine, Jason P., and Robert J. Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association* 94:496–509.

Firestone, Simon, Robert Van Order, and Peter Zorn. 2007. "The Performance of Low-Income and Minority Mortgages." *Real Estate Economics* 35:479–504.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen. 2008. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics* 64:234–45.

Furlong, Fred, and John Krainer. 2007. "The Subprime Mortgage Market: National and Twelfth District Developments." Annual Report, Federal Reserve Bank of San Francisco. Retrieved on February 1, 2016, from http://www.frbsf.org/files/2007annualreport.pdf.

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

Ghent, Andra C., Ruben Hernandez-Murillo, and Michael T. Owyang. 2014. "Differences in Subprime Loan Pricing across Race and Neighborhoods." *Regional Science and Urban Economics* 48:199–215.

Gotham, Kevin Fox. 2009. "Creating Liquidity out of Spatial Fixity: The Secondary Circuit of Capital and the Subprime Mortgage Crisis." *International Journal of Urban and Regional Research* 33:355–71.

Gramlich, Edward M. 2007. *Subprime Mortgages: America's Latest Boom and Bust.* Washington, DC: The Urban Institute Press.

Hwang, Jackelyn, Michael Hankinson, and Steven Kreg Brown. 2015. "Racial and Spatial Targeting: Segregation and Subprime Lending within and across Metropolitan Areas." *Social Forces* 93:1081–108.

Hyra, Derek S., Gregory D. Squires, Robert N. Renner, and David S. Kirk. 2013. "Metropolitan Segregation and the Subprime Lending Crisis." *Housing Policy Debate* 23:177–98.

Immergluck, Dan. 2008. "The Accumulation of Foreclosed Properties: Trajectories of Metropolitan REO Inventories during the 2007-2008 Mortgage Crisis." Community Affairs Discussion Paper No. 02-08, Federal Reserve Bank of Atlanta: Community Affairs, Atlanta, GA.

Immergluck, Daniel. 2009. *Foreclosed: High-Risk Lending, Deregulation, and the Undermining of America's Mortgage Market*. Ithaca, NY: Cornell University Press.

Kau, James B., Donald C. Keenan, Walter J. Muller, and James F. Epperson. 1995. "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment." *Journal of Real Estate Finance and Economics* 11:5–36.

Kelly, Austin. 2008. "'Skin in the Game': Zero Downpayment Mortgage Default." *Journal of Housing Research* 17:75–99.

Manski, Charles. 1989. "Anatomy of the Selection Problem." *Jouranl of Human Resources* 24:343–60.

Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid*. Cambridge, MA: Harvard University Press.

Mayer, Christopher, and Karen M. Pence. 2008. "Subprime Mortgages: What, Where, and to Whom?" 2008-29, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, DC.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* 23:27–50.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics* 124:1449–96.

Munnell, Alicia H., Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney. 1993. "Is Discrimination Racial or Geographic?" The Federal Reserve of Boston.

Oliver, Melvin L., and Thomas M. Shapiro. 2006. *Black Wealth, White Wealth: A New Perspective on Racial Inequality*, 2nd ed. New York: Routledge.

Owens, Lindsay. 2015. "Intrinsically Advantageous? Reexamining the Production of Class Advantage in the Case of Home Mortgage Modification." *Social Forces* 93:1185–209.

Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34:181–209.

Quercia, Roberto, and Lei Ding. 2009. "Loan Modifications and Redefault Risk: An Examination of Short-Term Impacts." *Cityscape: A Journal of Policy Development and Research* 11:171–94.

Reskin, Barbara. 2012. "The Race Discrimination System." *Annual Review of Sociology* 38:17–35.

Rugh, Jacob S. 2015. "Double Jeopardy: Why Latinos Were Hit Hardest by the US Foreclosure Crisis." *Social Forces* 93:1139–84.

Rugh, Jacob S., and Douglas S. Massey. 2010. "Racial Segregation and the American Foreclosure Crisis." *American Sociological Review* 75:629–51.

Schwartz, Alex F. 2010. *Housing Policy in the United States: An Introduction*. New York: Routledge.

Stuart, Guy. 2003. *Discriminating Risk: The U.S. Mortgage Lending Industry in the Twentieth Century*. Ithaca, NY: Cornell University Press.

Vandell, Kerry D. 1995. "How Ruthless Is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6:245–64.

White, Alan M. 2008. "Rewriting Contracts, Wholesale: Data on Voluntary Mortgage Modifications from 2007 and 2008 Remittance Reports." *Fordham Urban Law Journal* 36, 509–535.

Williams, Richard, Reynold Nesiba, and Eileen Diaz McConnell. 2005. "The Changing Face of Inequality in Home Mortgage Lending." *Social Problems* 52:181–208.

Yezer, Anthony M. 2010. "A Review of Statistical Problems in the Measurement of Mortgage Market Discrimination." Research Institute for Housing America.

Yinger, John M. 1995. *Closed Doors and Opportunity Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation Publications.

APPENDIX

CONTROL VARIABLES

FICO is a number from 300 to 850 that indicates the probability a borrower will repay a loan; the higher the number, the lower the probability of foreclosure. A credit score below 620 is considered to be subprime credit and a score above 720 is considered to be "excellent" (Capone 2001). In keeping with the literature, I categorize FICO into four

groups because the effect of credit on foreclosure is nonlinear. I expect that credit score has a negative effect on a borrower's risk of foreclosure, which is to say as categories of FICO rise, the risk of foreclosure falls.

The region indicates the location of the loan according to its Census region (Northeast, Midwest, South, and West). The distribution of loans across geography is as follows: 41% of loans are in the West, 30% are in the South, 15% are in the Midwest, and 14% are in the Northeast. Without controlling for region, then the effect of segregation might simply be explained by a select number of MSAs with large numbers of foreclosures at either end of the distribution of segregation. Controlling for region, the coefficient on segregation captures the average effect across the entire country.

High-cost loan is a dichotomous variable used by federal regulatory agencies to indicate that the interest rate on the loan at origination is 300 basis points, or three percentage points, higher than the current market interest rate. According to the high-cost indicator, 42% of the loans in this sample are subprime, and I expect these loans to have a higher risk of foreclosure in accordance with their subprime status.

LTV is the ratio of the loan to the value of the house at origination or, put another way, it is 100 minus the down payment. For example, if a borrower made a down payment of 20%, the LTV would be 80. The higher the LTV, the lower the amount of homeowner equity is at stake, and the greater a borrower's risk of foreclosure (Kelly 2008). Average LTV in the sample is 81. Of note, the maximum LTV is 125, and 10% of borrowers who used their loan to purchase a home have an LTV over 100. An LTV greater than 100 means that a borrower has negative equity in the home, which, along with ARM, is another defining feature of a subprime loan.

Loan purpose indicates whether a loan was used for a home purchase or refinancing, when a borrower replaces an existing loan with a new loan. Refinancing is often done to take advantage of more favorable mortgage interest rates and therefore reduces a borrower's risk of foreclosure relative to a loan used for home purchase. The dataset is nearly split in half between loans used for purchase and loans used for refinancing.

PTI is a dichotomous variable indicating whether a borrower's monthly payment and interest divided by monthly income at time of loan origination is greater than or equal to 31%. A PTI of 31% is the threshold above which the Federal Housing Authority deems a loan to be unaffordable and therefore at a greater risk of foreclosure. Fourteen percent of borrowers have a PTI greater than or equal to 31%.

Loan modification is a time-varying, dichotomous variable indicating whether a loan has been modified from its original terms to reduce the risk of foreclosure by providing temporary relief to the borrower. Fifteen percent of loans have been modified from their original terms.

Mortgage type is a time-varying, dichotomous variable indicating whether a loan is an ARM, where the interest rate on the loan floats with the market interest rate, or an FRM, where the interest rate remains fixed over the life of the loan. ARMs have a higher risk of foreclosure than loans that are fixed. Nearly 69% of the loans in the sample are adjustable rate mortgages.

SENSITIVITY ANALYSIS

I compare the sensitivity of the main findings across 13 alternate model specifications in order to measure the robustness of the results. The alternate specifications fall into two

main categories. The first category deals with the measurement of key variables and uses the data that match CTS data with HMDA data, as has been presented thus far, as shown in Tables A.1 and A.2.

Model 1 is the main model and is included for comparison. Model 2 replaces the dissimilarity index of Black segregation with an isolation index, another popular measure of segregation that indicates the proportion of one's own group members in a tract inhabited by the average own-group member. For example, a Hispanic isolation index of 60 indicates that the average Hispanic lives in a tract that is 60% Hispanic (e.g., Atlanta).

Model 3 modifies the competing risks analysis to isolate and examine the risk of 90+ days delinquent (i.e., "serious") on a home loan against the competing risk of prepayment. This distinguishes the risk of foreclosure from the risk of delinquency, as it is possible that the risk of one might not translate into the risk for the other, or vice versa. Model 4 uses a subset of the samples that are clearly prime loans, as defined in the table note. This allows one to test the sensitivity of the results by the quality of loans. While subprime loans are of great interest, as the foreclosure crisis was primarily the result of a collapse of the subprime market, prime loans comprise the vast majority of all loans issued.

Model 5 restricts the study period to 2007–2010, the time period of the foreclosure crisis, but excludes the aftermath. Model 6 defines a short sale as a prepayment, as it is in the raw data, instead of as a foreclosure as in the main model (1). Model 7 includes and controls for borrowers who enter into bankruptcy.

Model 8 adds an additional interaction to the main model between income and Black segregation in order to address the concern that the average income in the sample is too high. By itself, the fact that the sample income is too high is not a problem. If the sample is too high income in places with too little segregation to capture the true effect of segregation on a borrowers' risk of foreclosure, then the results might be compromised by selection bias. The coefficients on the main effect of Black segregation and the interaction effect, which reflects the effect of segregation at a hypothetical effect of zero income, are both negative. Therefore, income for all borrowers must be negative for the effect of income to cancel the effect of segregation.[14]

The second category of alternate specifications deals with sample exclusions and inclusions, as shown in Tables A.3 and A.4 . This category uses all the CTS data, including, but not exclusive to, the data that match HMDA. The matched and unmatched sample contains data that were previously dropped because of missing information on a borrower's race and income. I replace this missing information by assuming various configurations of race and income. For example, all the missing race observations are either Black, Hispanic, or White and the missing income observations are either high or low (three times larger or one-third smaller than the median income in a borrower's metro area, respectively). Testing the sensitivity of the results by making assumptions about observations with missing data is a method of overcoming issues of selection broadly based on the idea of bounded results proposed by Manski (1989).

The direction and magnitude of the coefficients are similar to the main model, where they are significant. The smaller, negative effect of being Black and Black segregation as well as the larger and positive effect of the put and call options are robust to a variety of alternative specifications.

Matched loans: Tables A.1 (independent variables of interest) and A.2 (control variables):

TABLE A.1. Sensitivity Analysis—w/Matched Loans

	(1) Main b/se	(2) Isolation b/se	(3) Serious b/se	(4) Prime b/se	(5) 2007–2010 b/se	(6) No Short Sale b/se	(7) Bankruptcy b/se	(8) Inc./Seg. int. b/se
Borrower:								
Black	1.038*	0.935	1.202***	0.889	1.115***	1.123***	0.968*	1.040*
	(0.017)	(0.033)	(0.015)	(0.080)	(0.021)	(0.021)	(0.014)	(0.017)
Hispanic	1.258***	1.275***	1.462***	1.336***	1.352***	1.331***	1.164***	1.257***
	(0.015)	(0.025)	(0.014)	(0.074)	(0.018)	(0.019)	(0.013)	(0.015)
Other	1.221***	1.147***	1.258***	1.327***	1.236***	1.214***	1.173***	1.219***
	(0.022)	(0.035)	(0.019)	(0.096)	(0.026)	(0.028)	(0.019)	(0.022)
Income ($10,000s)	1.001	1.002	1.003***	0.999	1.002	0.996***	1.000	1.001
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Segregation:								
Black	0.988***	0.995***	0.998*	0.993*	0.988***	0.985***	0.989***	0.988***
	(0.001)	(0.000)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	1.003***	0.997***	1.001	1.000	1.001	1.003***	1.001*	1.003***
	(0.001)	(0.000)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	1.012***	1.003***	1.000	0.995	1.011***	1.012***	1.010***	1.012***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Put (default) option:								
Put option	1.038***	1.038***	1.040***	1.037***	1.041***	1.037***	1.032***	1.038***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Put option^2	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***

(*Continued*)

TABLE A.1. Continued

	(1) Main b/se	(2) Isolation b/se	(3) Serious b/se	(4) Prime b/se	(5) 2007–2010 b/se	(6) No Short Sale b/se	(7) Bankruptcy b/se	(8) Inc./Seg. int. b/se
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Call (prepay) option:								
Call option	1.015***	1.015***	1.018***	1.039***	1.022***	1.017***	1.014***	1.015***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
Call option^2	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Race and segregation interaction:								
Black and Black segregation	1.005***	1.003***	1.007***	0.988	1.006***	1.001	1.003*	1.003*
	(0.001)	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic and Black segregation	1.002*	1.000	1.004***	1.010*	1.003**	1.001	1.000	1.002
	(0.001)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)
Other and Black segregation	1.003*	1.002*	1.002	1.014*	1.002	1.001	1.003**	1.003*
	(0.002)	(0.001)	(0.001)	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)
Number of observations	7,927,358	7,927,358	6,099,456	1,709,785	5,573,475	7,927,358	9,639,243	7,927,358
Number of subjects	192,617	192,617	188,687	35,246	192,700	192,617	218,833	192,617
Number of failures (foreclosure)	50,874	50,874	76,454	2,669	38,016	36,179	64,482	50,874
Number of competing (prepay)	81,117	81,117	78,089	20,380	71,467	95,812	82,117	81,117

Exponentiated coefficients.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE A.2. Sensitivity Analysis w/Matched Loans—Control Variables from Table A.1

	(1) Main b/se	(2) Isolation b/se	(3) Serious b/se	(4) Prime b/se	(5) 2007–2010 b/se	(6) No Short Sale b/se	(7) Bankruptcy b/se	(8) Inc./Seg. int. b/se
Year of origination:								
2005	0.896***	0.886***	0.881***	1.100	0.860***	0.863***	0.931***	0.897***
	(0.016)	(0.015)	(0.013)	(0.090)	(0.017)	(0.017)	(0.014)	(0.016)
2006	0.762***	0.750***	0.849***	1.001	0.670***	0.693***	0.810***	0.763***
	(0.014)	(0.013)	(0.012)	(0.081)	(0.014)	(0.014)	(0.013)	(0.014)
2007	0.621***	0.610***	0.781***	1.174	0.485***	0.497***	0.680***	0.622***
	(0.013)	(0.013)	(0.014)	(0.102)	(0.013)	(0.013)	(0.013)	(0.013)
Race and income interaction:								
Black and income	1.005*	1.005*	1.010***	1.006	1.008**	1.017***	1.004	1.007**
	(0.002)	(0.002)	(0.002)	(0.012)	(0.003)	(0.003)	(0.002)	(0.002)
Hispanic and income	1.016***	1.016***	1.016***	1.003	1.017***	1.017***	1.010***	1.015***
	(0.002)	(0.002)	(0.001)	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)
Other and income	1.007***	1.008***	1.007***	0.991	1.010***	1.010***	1.004*	1.008***
	(0.002)	(0.002)	(0.002)	(0.008)	(0.002)	(0.003)	(0.002)	(0.002)
Region:								
Northeast (ommitted)								
Midwest	1.648***	1.671***	0.745***	1.369**	1.595***	1.932***	1.603***	1.628***
	(0.036)	(0.038)	(0.012)	(0.144)	(0.040)	(0.051)	(0.031)	(0.036)
South	1.118***	1.167***	0.732***	1.103	1.051*	1.230***	1.137***	1.117***
	(0.024)	(0.024)	(0.010)	(0.101)	(0.025)	(0.031)	(0.021)	(0.024)
West	1.668***	1.637***	0.856***	1.878***	1.627***	1.651***	1.630***	1.663***
	(0.035)	(0.038)	(0.012)	(0.170)	(0.039)	(0.043)	(0.031)	(0.035)
Loan characteristics:								
FICO < 620 (ommitted)								

(*Continued*)

TABLE A.2. Continued

	(1) Main b/se	(2) Isolation b/se	(3) Serious b/se	(4) Prime b/se	(5) 2007–2010 b/se	(6) No Short Sale b/se	(7) Bankruptcy b/se	(8) Inc./Seg. int. b/se
FICO (620–679)	0.966**	0.965**	0.876***		0.979	0.961**	0.970**	0.966**
	(0.012)	(0.012)	(0.009)		(0.014)	(0.014)	(0.011)	(0.012)
FICO (680–719)	0.814***	0.813***	0.698***		0.803***	0.762***	0.842***	0.814***
	(0.013)	(0.013)	(0.009)		(0.015)	(0.014)	(0.012)	(0.013)
FICO (>= 720)	0.563***	0.560***	0.461***		0.530***	0.480***	0.617***	0.563***
	(0.010)	(0.009)	(0.006)		(0.011)	(0.010)	(0.009)	(0.010)
High-cost loan (>= 300 BPS Indicator)	1.296***	1.297***	1.177***		1.346***	1.318***	1.252***	1.296***
	(0.014)	(0.014)	(0.011)		(0.018)	(0.018)	(0.013)	(0.014)
Loan to value (LTV)	1.005***	1.004***	0.995***	1.025***	1.004***	1.003***	1.004***	1.005***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
Purchase (vs. refinance) indicator	1.265***	1.266***	1.130***	0.848***	1.318***	1.304***	1.281***	1.265***
	(0.013)	(0.013)	(0.010)	(0.040)	(0.016)	(0.016)	(0.012)	(0.013)
ARM indicator	1.490***	1.484***	1.262***		1.581***	1.380***	1.456***	1.490***
	(0.018)	(0.018)	(0.012)		(0.024)	(0.020)	(0.015)	(0.018)
Modification indicator	0.901***	0.896***	1.459***	2.661***	0.726***	0.905***	0.817***	0.902***
	(0.015)	(0.015)	(0.026)	(0.250)	(0.017)	(0.019)	(0.012)	(0.015)
PTI > 31% indicator	1.024	1.021	1.108***	1.031	1.030	1.010	1.039**	1.025
	(0.014)	(0.014)	(0.012)	(0.065)	(0.016)	(0.016)	(0.013)	(0.014)
Bankruptcy indicator							2.960***	
							(0.036)	
Segregation & income interaction								1.000*
								(0.000)

Exponentiated coefficients.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE A.3. Sensitivity Analysis—w/Matched and Unmatched Loans

	(1) Main b/se	(2) Black/ High Income b/se	(3) Black/ Low Income b/se	(4) Hispanic/ High Income b/se	(5) Hispanic/ Low Income b/se	(6) White/ High Income b/se	(7) White/ Low Income b/se
Borrower:							
Black	1.038*	1.069***	1.069***	1.103***	1.077***	1.065**	1.062***
	(0.017)	(0.012)	(0.013)	(0.028)	(0.019)	(0.025)	(0.017)
Hispanic	1.258***	1.381***	1.232***	1.125***	1.148***	1.329***	1.207***
	(0.015)	(0.021)	(0.018)	(0.013)	(0.013)	(0.017)	(0.015)
Other	1.221***	1.261***	1.210***	1.248***	1.205***	1.215***	1.186***
	(0.022)	(0.028)	(0.032)	(0.028)	(0.032)	(0.025)	(0.030)
Income ($10,000s)	1.001	1.000	0.998	1.000	0.997*	1.003***	0.999
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Segregation:							
Black	0.988***	0.989***	0.989***	0.990***	0.989***	0.990***	0.990***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	1.003***	1.002*	1.001	1.002***	1.002**	1.002*	1.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	1.012***	1.014***	1.014***	1.014***	1.014***	1.015***	1.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Put (default) option:							

(*Continued*)

TABLE A.3. Continued

	(1) Main b/se	(2) Black/ High Income b/se	(3) Black/ Low Income b/se	(4) Hispanic/ High Income b/se	(5) Hispanic/ Low Income b/se	(6) White/ High Income b/se	(7) White/ Low Income b/se
Put option	1.038***	1.037***	1.037***	1.037***	1.037***	1.037***	1.037***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Put option^2	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Call (prepay) option:							
Call option	1.015***	1.016***	1.016***	1.016***	1.016***	1.016***	1.016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Call option^2	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Race and segregation interaction:							
Black and Black segregation	1.005***	1.002*	1.002*	1.006***	1.006***	1.006***	1.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Hispanic and Black segregation	1.002*	1.003*	1.003*	1.002*	1.002	1.002*	1.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Other and Black segregation	1.003*	1.001	1.001	1.001	1.001	1.000	1.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of observations	7,927,358	10,251,341	10,249,914	10,251,341	10,249,914	10,251,341	10,249,914
Number of subjects	192,617	249,610	249,523	249,610	249,523	249,610	249,523
Number of failures (foreclosure)	50,874	59,402	59,324	59,402	59,324	59,402	59,324
Number of competing (prepay)	81,117	107,227	107,218	107,227	107,218	107,227	107,218

Exponentiated coefficients.
See authors note in the Appendix regarding small differences in the total number of observations in the high- and low-income models.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE A.4. Sensitivity Analysis w/Matched and Unmatched Loans—Control Variables from Table A.3

	(1) Main b/se	(2) Black/High Income b/se	(3) Black/Low Income b/se	(4) Hispanic/ High Income b/se	(5) Hispanic/ Low Income b/se	(6) White/ High Income b/se	(7) White/Low Income b/se
Year of origination:							
2005	0.896***	0.998	0.994	0.999	0.993	1.000	0.996
	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
2006	0.762***	0.887***	0.883***	0.889***	0.884***	0.889***	0.885***
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
2007	0.621***	0.730***	0.729***	0.731***	0.729***	0.732***	0.731***
	(0.013)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Race and income interaction:							
Black and income	1.005*	1.000	1.007***	1.003	1.007*	1.001	1.006*
	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Hispanic and income	1.016***	1.015***	1.020***	0.998	1.017***	1.013***	1.019***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Other and income	1.007***	1.005	1.008**	1.006*	1.008**	1.003	1.007**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Region:							
Northeast (ommitted)							
Midwest	1.648***	1.629***	1.634***	1.619***	1.637***	1.626***	1.622***
	(0.036)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
South	1.118***	1.102***	1.103***	1.122***	1.118***	1.108***	1.106***
	(0.024)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)
West	1.668***	1.583***	1.581***	1.621***	1.602***	1.586***	1.585***
	(0.035)	(0.027)	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)

(*Continued*)

TABLE A.4. Continued

	(1) Main b/se	(2) Black/High Income b/se	(3) Black/Low Income b/se	(4) Hispanic/ High Income b/se	(5) Hispanic/ Low Income b/se	(6) White/ High Income b/se	(7) White/Low Income b/se
Loan characteristics:							
FICO < 620 (ommitted)							
FICO (620–679)	0.966**	0.910***	0.907***	0.920***	0.910***	0.909***	0.909***
	(0.012)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
FICO (680–719)	0.814***	0.717***	0.715***	0.726***	0.716***	0.715***	0.717***
	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
FICO (>= 720)	0.563***	0.498***	0.497***	0.503***	0.496***	0.496***	0.498***
	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
High-cost loan (>= 300 BPS Indicator)	1.296***						
	(0.014)						
Loan to value (LTV)	1.005***	1.008***	1.008***	1.008***	1.008***	1.008***	1.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Purchase (vs. refinance) indicator	1.265***	1.286***	1.285***	1.298***	1.292***	1.285***	1.285***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)
ARM indicator	1.490***	1.638***	1.628***	1.652***	1.633***	1.635***	1.628***
	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Modification indicator	0.901***	0.952**	0.956**	0.954**	0.957**	0.953**	0.956**
	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)
PTI > 31% indicator	1.024	1.040*	1.023	1.039*	1.001	1.039*	1.035**
	(0.014)	(0.018)	(0.013)	(0.018)	(0.012)	(0.018)	(0.012)

Exponentiated coefficients.
See authors note in the Appendix regarding small differences in the total number of observations in the high- and low-income models.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1. Main model with all borrower, geography, loan characteristics, and interactions.
2. Use an isolation index of segregation.
3. Competing risk of 90+ days delinquent against the competing risk of prepayment.
4. Prime loan is a loan with a fixed-rate mortage, not a high-cost loan, a down payment >= 20%, and a credit score >= 720 5.6% of all loans meet all four criteria of a prime loan.
5. Study period from 2007–2010, i.e., the time period of the foreclosure crisis.
6. Does not account for short-sale, i.e., short-sale is counted as a prepayment.
7. Includes and controls for loans that enter into bankruptcy.
8. Include an interaction between income and segregation.

Matched and unmatched loans: Tables A.3 (independent variables of interest) and A.4 (control variables):

1. Main model with all borrower, geography, loan characteristics, and interactions.
2. Observations with missing race are Black, and observations with missing income are with high income (3 × median income of metro area).
3. Observations with missing race are Black, and observations with missing income are low income (1/3 median income of metro area).
4. Observations with missing race are Hispanic, and observations with missing income are high income.
5. Observations with missing race are Hispanic, and observations with missing income are low income.
6. Observations with missing race are White, and observations with missing income are with high income.
7. Observations with missing race are White, and observations with missing income are low income.

Note: The total number of observations in the models with both matched and unmatched loans (Tables A.3 and A.4) are not identical across models where loans with missing income are replaced with either low or high income. The reason is that when cases with missing income are replaced with low income (one-third smaller than the median income in a borrower's metro area), some of these cases contain loans with income that is below the 1st percentile. As stated in the Data section, I drop loans with borrower income in the top and bottom 1% of the distribution.