



# Mortgage Lending Discrimination Across the U.S.: New Methodology and New Evidence

Manthos D. Delis<sup>1</sup> · Panagiotis Papadopoulos<sup>2</sup>

Received: 23 December 2016 / Revised: 16 January 2018 / Accepted: 8 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Is there discrimination in mortgage-loan origination and pricing? If so, does the level of discrimination differ before and after the eruption of the subprime crisis? Using data from 6.5 million loan applications from 2004 through 2013, we propose a novel approach aiming to substantially lower the notorious omitted-variable bias of the Home Mortgage Disclosure Act (HMDA) database and identify the level of racial, ethnic, and gender discrimination in mortgage lending across the United States. In stark contrast with previous studies, we find, on average, very little discrimination in loan origination. Although discrimination increases somewhat after 2007, its probability remains well below 1%. In contrast, we find that white (non-Hispanic) applicants pay a lower spread on the originated loans by 0.37 (0.11) basis points, a result that almost entirely comes from the pre-crisis period.

**Keywords** Mortgage lending · Racial and ethnic discrimination · HMDA

**JEL classification** G21 · J15

## 1 Introduction

This paper revisits the issue of mortgage-lending discrimination, which has vital implications for egalitarianism in general and individuals' social and economic welfare in particular. We specifically address two old questions about whether race, gender, and ethnicity affect (1) the probability of banks granting a new mortgage loan and/or (2) the spread that banks charge on

---

✉ Manthos D. Delis  
m.delis@montpellier-bs.com

Panagiotis Papadopoulos  
p.papadopoulos@surrey.ac.uk

<sup>1</sup> Montpellier Business School - Montpellier Research in Management, 2300 Avenue des Moulins, 34080 Montpellier, France

<sup>2</sup> Surrey Business School, University of Surrey, Guildford GU2 7XH, UK

newly originated loans. We also ask a new question: What is the role of the subprime crisis with respect to these issues? The main novelty of our work, relative to the extant literature, rests in improvement of identifying discrimination practices when using countrywide U.S. loan-level data, along with placing the role of the subprime crisis at the heart of our analysis. Answering these questions in the most precise way possible is especially relevant for policy-makers and the general public, because housing is, *inter alia*, a core part of human well-being and a well-known part of the American dream (Pager and Shepherd 2008).

Existing literature on identifying discriminatory practices in the mortgage loan market considers both discrimination in a bank's decision to originate the loan or not (henceforth called loan origination discrimination) and discriminatory practices in the pricing of loans (henceforth called loan pricing discrimination). The premise is that minority applicants, mostly African-Americans or Hispanics, are potentially discriminated against relative to non-minority applicants (i.e., whites and non-Hispanics) with respect to loan origination and/or pricing. The most prominent branch of the existing literature augments information from the Home Mortgage Disclosure Act (HMDA) with information on specific loan applicants and property values, etc., and it shows that white applicants have at least an 8% higher chance of receiving a loan relative to African-Americans, *ceteris paribus* (e.g., Munnell et al. 1996; Hubbard et al. 2012; Wheeler and Olson 2015).

Identifying discriminatory practices was one important reason for the creation of the HMDA database. Unfortunately, the HMDA does not provide crucial information on several dimensions of the bank's decision-making process regarding loan origination and pricing. The missing information includes, but is not limited to, the loan-to-value ratio as well as the applicant's wealth and creditworthiness. Thus, inference on discriminatory practices from research that uses only the HMDA database will very likely overstate the level of discrimination significantly and thus be invalid. For this reason, researchers and policy-makers tend to examine specific subsets of the HMDA—usually in local markets, where information on these basic omitted variables is available.

This approach can also be criticized, however, along two dimensions. First, the list of important omitted variables can be very idiosyncratic at the individual level. In fact, when discriminatory practices are alleged the involved financial institutions defend their operating methods based on this argument (see the numerous lawsuit cases filed by the National Association for the Advancement of Colored People 2007<sup>1</sup>). Essentially, this idiosyncratic omitted-variable problem means that any augmentation of an HMDA subset might still lack the information required to clear the endogeneity of the variables characterizing, for instance, race or ethnicity. Second, even in the unlikely event that researchers and policy-makers had access to augmented local data, and these data were sufficient to overcome the omitted-variable bias for these local markets, inference across the full gamut of mortgage loans and the entire United States cannot be made. As a result, regulators find discriminatory practices

<sup>1</sup> Between 2007 and 2009 the National Association for the Advancement of Colored People (NAACP) filed numerous class action lawsuits against a number of the United States' largest lenders for discriminatory lending practices (Sen 2012) including, among others, Wells Fargo & Co, HSBC Finance Corporation, CitiMortgage, SunTrust Mortgage, JP Morgan, First Horizon, Ameriquest Mortgage Company, Fremont Investment & Loan, Option One Mortgage Corporation, WMC Mortgage Corporation, Long Beach Mortgage Company, BNC Mortgage, Accredited Home Lenders, Bear Stearns Residential Mortgage Corporation, Encore Credit, First Franklin Financial Corporation and Washington Mutual, Inc. The class action lawsuits alleged that the financial institutions violated the Fair Housing Act, the Equal Credit Opportunity Act, and the Civil Rights Act of 1866 engaging in systematic, institutionalized racism and discriminatory practices in home mortgage lending. A more detailed analysis on the aforementioned lawsuit cases is available at: <http://www.naacp.org>.

very hard to identify, let alone penalize and reduce. In addition, new legislation on discrimination might be hampered by the lack of specific evidence.

This paper's main goal is to introduce a new empirical strategy that significantly reduces the omitted-variable bias in identifying mortgage-loan discrimination (both in loan origination and in loan pricing) using information solely from the HMDA, and thus across the entire United States, for a very large number of loan applications (more than 6 million from 2004 through 2013).

Our method runs in two stages. In the first stage, we estimate separate equations for the requested (i.e., the amount in the loan application) and granted (i.e., the amount in the loan application if the loan is originated; zero if the loan is not originated) loan amounts, with regional (census-tract) fixed effects, year fixed effects and, in some sensitivity tests, individuals' income as the explanatory variables. Subsequently, we take the difference of the residuals between the two equations, which includes the unobserved reasons behind banks' willingness to originate a loan or not relative to the applicant's reasons in applying for the loan. Among these unobserved reasons is also the potential effect of race and gender. In the second stage, we estimate a binary model on the bank's decision whether or not to grant the loan, using observable characteristics such as race, ethnicity, and gender as independent variables along with the difference in the residuals from the first stage. Essentially, this means that we extract the effects of race, ethnicity, and gender from the remaining unobserved component that is also included in the binary model.

Our results indicate that white applicants had approximately a 0.39% higher probability of being granted a loan than African-Americans from 2004 through 2013, and non-Hispanic applicants had a 0.25% higher probability than Hispanic applicants. These estimates, although statistically significant, are quite low and show limited discrimination on average. Further, our estimates are significantly lower compared with, for example, most previous studies' estimates of 8% or higher loan approval for whites and non-Hispanics. Our findings support the suggestion that omitted-variable bias is very idiosyncratic and cannot be easily captured, even in data sets augmented with additional applicant characteristics. The equivalent estimates for gender are even lower and economically unimportant.

In forming our two-stage model, we discuss the potential bias from imprecise estimation of the first stage, as well as bias from applicants' income misreporting. We carry out a number of sensitivity tests to remedy these sources of measurement error, including the use of the model of Erickson et al. (2014) and of data only from regions with low probability of income misreporting (Piskorski et al. 2015).

Moreover, by splitting our sample to the periods before and after 2007, we find an economically small, albeit notable, increase in the estimated level of racial and ethnic discrimination after 2007. Specifically, before the crisis, white (non-Hispanic) applicants have a 0.23% (0.14%) higher probability of being granted a loan compared with African American (Hispanic) applicants, whereas after the crisis, the foregoing probability increases to 0.66% with respect to African Americans and 0.54% with respect to Hispanics. Therefore, we find that the subprime crisis increased discrimination, even though these estimates remain far lower than those of previous studies.

Subsequently, we turn to discrimination in loan pricing. We show that, when all other factors are equal, white applicants pay 0.37 basis points lower spreads on their mortgage loans compared with African Americans and non-Hispanic applicants pay 0.11 basis points lower spreads compared with Hispanics. The difference between genders is negligible. The significant difference in spreads between minority and non-minority applicants is observed almost entirely during the pre-crisis period. Together with the findings on discrimination in loan

origination, we suggest that before the financial crisis, discrimination in loan origination was very limited but some pricing discrimination did occur. After 2007, discrimination in loan pricing disappeared, most likely as a result of the credit crunch.

In a nutshell, we find that using as a regressor the difference between the unobserved reasons behind banks' willingness to provide a specific loan amount or not and the applicant's reasons in applying for that loan amount, controls for substantial information in a simple model of the bank's decision to grant the loan or not. Using this control, lowers the probability of discrimination of minority groups from about 7–8% to less than 1%. Thus, we posit that, even though not panacea, our approach lowers the omitted-variable problem to "more acceptable" levels, highlights that discrimination is much more limited than previously thought, and allows drawing better conclusions from the HMDA data and the entire gamut of U.S. loan applications. These are the contributions of research.

The following section discusses the theoretical considerations on mortgage-lending discrimination, the identification methods and empirical estimates of the most prominent existing literature, and the subprime crisis' potential role in banks' lending practices. Along these dimensions, Section 2 specifies our paper's three research questions. Section 3 presents our data set and discusses in detail our empirical approach. Section 4 presents the empirical results, and Section 5 provides policy implications based on our findings and concludes the paper.

## 2 Theoretical considerations and empirical facts

Almost 50 years after the 1968 enactment of the Fair Housing Act,<sup>2</sup> and more than a generation after the Equal Credit Opportunity Act (ECOA) of 1974, the Home Mortgage Disclosure Act (HMDA) of 1975, and the Community Reinvestment Act (CRA)<sup>3</sup> of 1977, lending discrimination remains a key issue in the social, political, and research agenda. Several early studies posit that, in the past, mortgage lenders discriminated against disadvantaged groups of borrowers and that much of the discrimination was actually part of their lending policy (Munnell et al. 1996; Ladd 1998).

The antidiscrimination laws imply that the lenders should use only objective information on a loan's expected return to decide whether to originate the loan, and they should avoid using the applicant's membership in a protected group as part of their decision. ECOA clarifies that the term "protected group" encompasses a neighborhood's racial composition.

Turner and Skidmore (1999) summarize the possible key forces that lead to discrimination.<sup>4</sup> Assuming that lenders are profit-maximizers, lending discrimination in its purest form implies the unlikely event that lenders are willing to forgo profits. In more plausible forms of lending discrimination, however, economic interest occurs as a possible explanation when lenders believe that personal characteristics, such as race or ethnicity, are reliable proxies for factors that affect credit risk and cannot be easily observed. In this case, lenders have the economic motive to deny credit to minority applicants based on an average approximation of that

<sup>2</sup> The Fair Housing Act was enacted in 1968 by the Office of Fair Housing and Equal Opportunity in the U.S. Department of Housing and Urban Development (HUD) to prohibit discrimination based on race (Hubbard et al. 2012).

<sup>3</sup> The Home Mortgage Disclosure Act (HMDA) and the Community Reinvestment Act (CRA) were enacted by Congress in 1975 and 1977, respectively, to monitor lending institutions' practices toward minorities and low-income borrowers and neighborhoods (Hubbard et al. 2012).

<sup>4</sup> Empirically distinguishing between these forces is very hard because of a lack of specialized data.

minority's creditworthiness and not based on the economic characteristics of individual applications. Such discrimination is usually termed "statistical" or "profit-motivated".

A more direct form of discrimination relates to prejudice (Quillian 2006) which occurs when lenders consider minority borrowers as inferior and do not want to interact with them. Such prejudice is usually referred to as "taste-based discrimination." Notably, if local home owners and depositors of local banks prefer not to have minorities as neighbors, the local banks might be negatively inclined against minority applicants. Yinger (1986) and Ondrich et al. (2003) provide evidence for such behavior on the part of real estate agents. Cultural affinity is another possible reason for taste-based discrimination, which occurs when white loan officers exert less effort to help minority borrowers meet underwriting criteria compared with non-minority borrowers.

The literature on lending discrimination is very large and can be distinguished between loan origination discrimination and loan pricing discrimination.<sup>5</sup> We do not review this literature here because detailed reviews are already available.<sup>6</sup> We do briefly note, however, some of the relevant history behind the academic interest in lending discrimination (mostly racial discrimination) and the most influential findings as a means to guide our own research.

## 2.1 Discrimination in loan origination

In the 1990s, academic interest in mortgage lending discrimination exploded, mainly as a result of the HMDA database's amendment in 1989. The HMDA database provides information on individual mortgage loans to assist, *inter alia*, in determining whether financial institutions are serving their communities' housing needs and in identifying possible discriminatory lending practices. The 1989 amendment made extra information (on the outcome of loan applications, the location of the property, the applicant's race and gender, etc.) available to the public and allowed more-coherent studies of discriminatory practices in loan origination.

However, the lack of data on crucial information considered during the underwriting process, such as credit scores, loan-to-value (LTV) ratios, debt-to-income (DTI) ratios, etc., still hampered empirical identification of a causal effect running from racial and gender characteristics to the loan origination decision. Munnell et al. (1996) took advantage of unique data collected by the Federal Reserve Bank of Boston in 1990 (38 variables, including risk of default, loan characteristics, personal characteristics, etc.) and amended the HMDA to circumvent the omitted-variable bias. Their findings showed that mortgage loan applicants who belong to a minority group have approximately an 8% higher probability of rejection than white borrowers. This effect is, of course, economically quite potent, despite the fact that the addition of the new variables significantly reduces the estimated level of discrimination.

Munnell et al. (1996), however, did not clear the picture on the actual level of discrimination. Specifically, a number of studies show that even the augmented list of variables used by Munnell et al. is insufficient (Horne 1997) and that racial discrimination is either lower (e.g.,

<sup>5</sup> There is also a relatively smaller number of studies that focus on geographical redlining against minority neighbourhoods (Holmes and Horvitz 1994; Turner and Skidmore 1999) or the default rates of minority applicants (Ferguson and Peters 1995).

<sup>6</sup> Ladd (1998), LaCour-Little (1999), Turner and Skidmore (1999), Ross and Yinger (2002, 2006), and Ross (2005) provide an exhaustive review of the earlier literature on mortgage lending discrimination. For more recent studies, see Pager and Shepherd (2008), Hubbard et al. (2012), Sen (2012), and Wheeler and Olson (2015) and references therein.

the separate studies by Day and Liebowitz 1998, and Faber 2013, estimate it to be about 2.8%) or non-existent (e.g., Harrison 1998). More-recent studies (most notably Hubbard et al. 2012 and Wheeler and Olson 2015) find that Munnell et al.'s estimates are more or less accurate.

The debate on the existence of discrimination in loan origination leads us to revisit an old empirical question, which we formulate as follows:

Question 1: Is there racial, ethnic, and/or gender discrimination in the origination of loans in the United States?

## 2.2 Discrimination in loan pricing

The second type of discrimination concerns loan pricing. The post-2000 literature argues that loan-pricing discrimination became increasingly important mainly because of the subprime boom that took place in the early to mid-2000s (e.g., Williams et al. 2005; Faber 2013; Ghent et al. 2014). The premise is that instead of excluding minority applicants, lending institutions opened up credit to African-Americans and Hispanics, but minorities had to pay higher loan spreads (Rugh and Massey 2010).

In 2004, the HMDA amended its database with partial loan pricing information. Specifically, lenders must report the spread (difference) between a loan's annual percentage rate (APR) and the rate on Treasury securities of comparable maturity—but only for loans with spreads above designated thresholds. Thus, rate spreads are reported for some (relatively high-risk) mortgage loans. Consequently, the last 15 years have seen a boom in research on loan-pricing discrimination, using either the HMDA database alone (e.g., Avery et al. 2005) or the HMDA enhanced with additional information (e.g., Courchane 2007; Bocian et al. 2008; DeLoughy 2012; Ghent et al. 2014; Bayer et al. 2014). Most of this literature highlights the existence of significant loan-pricing discrimination against African American and Hispanic applicants.

A small part of the literature on loan-pricing discrimination considers the role of gender. Sen (2012) suggests that female applicants who visit the same lenders as identically situated male applicants receive approximately the same amount of high-cost loans. In a similar vein, Haughwout et al. (2009), and Cheng et al. (2011, 2014) downplay the importance of gender in mortgage lending discrimination, suggesting that any differences in mortgage rates either are attributable to the different “shopping behavior” of female applicants (they are more likely to choose lenders by recommendation instead of searching for the lowest rates) or exist only among African-American female applicants.

Based on the foregoing discussion, we formulate our second research question as follows:

Question 2: Is there racial, ethnic, and/or gender discrimination in the pricing of loans in the United States?

## 2.3 The role of the subprime crisis

Almost a decade after the global financial crisis began, a third and novel question emerges: does discrimination in mortgage lending, either in loan origination or loan pricing, differ before and after the subprime crisis? This question is based on the theoretical assumption

that discriminatory policies might be exacerbated during crisis periods, which see an overall increase in economic and social uncertainty, decrease in trust, and increase in anti-social behavior.

More specifically, historical experience and academic literature shows that a financial crisis constitutes a serious challenge to protecting fundamental rights for specific groups of people. This challenge is particularly serious for the most marginalized people, who are also the most vulnerable because they probably already suffer from discrimination. For example, the International Labour Office (2011), the Agency for Fundamental Rights of the European Union (2010), and UNESCO (2009) note that a crisis feeds many types of direct or indirect discrimination because of rising economic or social inequality and insecurity. Further, the same organizations argue that during periods of turmoil, policymakers tend to de-prioritize policies targeting discrimination.

In that line, lending institutions might increasingly decide on granting mortgage loans during crisis periods by using racial and ethnical stereotypes. As in the broader social spectrum, crisis periods create severe insecurity and uncertainty in the banking sector. Given that informational asymmetries are higher during crises, banks might be less willing to take risks and be more suspicious that minority applicants lack the ability to repay a loan. Consequently, banks would reject these applicants to minimize the risk of loan default.

Correspondingly, lenders might embrace elements that reflect the role of race or ethnicity in other related markets (Reskin 2012), which affects minority borrowers' capacity to repay loans. For instance, it has been documented that minorities face higher unemployment rates and greater decreases in wealth during economic downturns.<sup>7</sup> Some lenders might thus perceive that minority applicants are more likely to be unemployed or see a substantial decrease in their wealth during (or shortly after) economic downturns, and these lenders might place a higher probability of default on loan applications by minority applicants. If this perception does not follow solely from the economics of the decision to supply the loan, such lender behavior implies increased statistical discrimination during economic downturns. In fact, according to Avery et al. (2010), mortgage lending to African-Americans and Hispanics fell more quickly than the U.S. average from 2007 to 2008.

The role of the subprime crisis in discrimination, especially with respect to loan origination, has not yet been given due consideration. Based on the theoretical considerations analyzed in this subsection, we expect that, if anything, discrimination in loan origination increased after the subprime crisis erupted. The subprime crisis' role in loan-pricing discrimination is much more nebulous. On one hand, following the same reasoning with loan origination, the subprime crisis might increase loan-pricing discrimination. On the other hand, if discrimination in loan origination increases, then loan-pricing discrimination might fall because minority applicants are not granted a loan at all.

<sup>7</sup> According to the Pew Research Centre (2011), from 2008 to 2010, unemployment increased among the white population by 4.7% (from 4.9% to 9.6%), among the African-American population by 7.9% (from 9.4% to 17.3%), and among the Asian population by 5.2% (from 3.2% to 8.4%). Further, African-American and Hispanic households lost 53% and 66% of their wealth, respectively, between 2005 and 2009. During the same period, white households' wealth dropped by 16%. Also, Hoynes et al. (2012) provide evidence that the effect of the 2007–2009 recession in the labor market differs across demographic groups, with African-American and Hispanic workers being more affected. Experience from earlier crises suggests that these differences are stable through all the recessionary periods for the last three decades.



Considering our theoretical arguments, we reiterate our third research question about the role of the subprime crisis in mortgage lending discrimination as follows:

Question 3: Has the level of racial, ethnic, and/or gender discrimination in loan origination or loan pricing intensified since the subprime financial crisis?

### 3 Data and econometric methodology

Our study aims to make progress on the identification of lending discrimination using the full gamut of mortgage loan applications across the United States, not only for specific regions or subsamples with superior data availability. In doing so, we provide a method that can serve as a policy tool for researchers, regulators, and policymakers. To fulfill this aim, we use applicant-level information solely from the HMDA for the period 2004 through 2013. Data before 2004 do not include information on ethnicity and thus limit the study on ethnicity-based discrimination. Also, the specific time period encompasses the years associated with the subprime crisis of 2007 and allows us to answer Question 3.

The HMDA provides information on a number of financial and demographic characteristics of mortgage loans. Specifically, it includes information on the requested loan amount (in thousands of dollars), loan type (conventional, insured by the Federal Housing Administration, guaranteed by Veterans' Administration or provided by the Farm Service Agency or Rural Housing Service, etc.), loan purpose (home purchase, home improvement, or refinancing), whether or not a preapproval was requested, the owner's occupancy status (whether or not the owner plans to occupy the home as a principal dwelling), property type (whether the application was for a one- to four-family dwelling, manufactured housing, or multifamily dwelling), loan decision details (whether the mortgage was originated, the application was denied by the financial institution or withdrawn by the applicant, etc.), lien status (whether the loan is securitized or not), and the application's denial reason(s) if the loan was not approved.

Further, the HMDA has information on the lender's identification number and supervisory/regulatory agency code, as well as information on the applicant's and co-applicant's race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White), ethnicity (Hispanic/Latino or not), gender, and income (gross annual income in thousands of dollars). Finally, the property's location information includes the identification number of the Metropolitan Statistical Area (MSA), the state and county codes, and the census tract number.

To ensure comparability across the loans, we homogenize our data set by using only applications for conventional loans for home purchases. Further, we limit the sample to securitized (by a first or a subordinate lien) loans on owner-occupied (as a principal dwelling), one- to four-family properties. Our goal is to examine whether minorities have equal housing opportunities as borrowers who belong to non-minority groups. We thus are interested only in home-purchase loans (rather than refinancing or home improvement loans), because they increase individuals' access to homeownership. For the same reason, we focus on loans for owner-occupied as a principal dwelling properties and not for second or vacation homes.

We also remove from the analysis observations with missing data for any of the foregoing variables. The final restriction we make is to exclude applications that did not result in either originations or denials. Thus, we rule out any other type of action taken (e.g., applications withdrawn by the applicant or approved but not accepted) because the decisions made for these



applications are not under the lender's control. Following our cleansing procedure, we have a sample of 6,452,279 loan applications for the 2004–2013 period. For descriptive statistics (along with a summary of variable definitions) see Table 1.

Our identification method relies on a number of regressions on the usual “origination-denial” decision. The model's general form is as follows:

$$gr_i = f(a_j, y_i, \mathbf{x}_i, u_i), \quad (1)$$

where  $gr_i$  is a binary variable that takes the value one if the loan is originated and zero if the loan is denied by the bank;  $a_j$  are the  $j$  parameters to estimate;  $y_i$  is the applicant's income;  $\mathbf{x}$  is a vector of variables capturing the observed applicant's characteristics; and  $u_i$  is the stochastic disturbance.

The problem with directly estimating Eq. (1), using the full gamut of the loan applications in our final sample, is of course the lack of crucial underlying factors and information on borrowers' creditworthiness, including (but not limited to) credit score variables (e.g., FICO score), LTV ratio, DTI ratio, and down payment size. These criteria are privately observed by the lending institutions and play a crucial role when a bank assesses an applicant's riskiness and decides whether or not to originate the loan. In turn, these criteria can be severely correlated with variables such as race, ethnicity, and gender, and their exclusion from Eq. (1) leads to overestimation of the effect of minority characteristics.

In this paper, we make progress on this omitted-variable bias by proposing a new identification method in two stages. The first stage corresponds to Eqs. (2), (3), and (4). In Eq. (2), we estimate the quantity of loan amount demanded as a function of the reported applicant's income, as follows:

$$L_i^D = a_0^D + a_1^D y_i + a_2^D d_k + u_i^D \quad (2)$$

where  $L_i^D$  is the dollar-denominated amount of the loan demanded by borrower  $i$ ,  $y_i$  is the dollar-denominated income for borrower  $i$ ,  $d_k$  is a dummy variable equal to 1 for borrowers belonging to the  $k^{\text{th}}$  census tract,  $a_0^D$  and  $a_2^D$  are parameters to be estimated by OLS and  $u_i^D$  is the residual associated with borrower  $i$  in this equation.

**Table 1** Variables description and summary statistics

Name	Description	Mean	St. Dev.	Min.	Max.
Action type	Dummy variable equal to one if the loan is originated and zero if the loan is denied by the bank	0.86	0.34	0	1
Spread	The loan's spread (originated loans only)	5.07	1.64	1.50	21.45
Loan amount	Natural logarithm of the loan amount	4.97	0.84	2.40	8.81
Loan supply	Natural logarithm of an originated loan's amount	4.30	1.88	0	8.81
Race	Dummy variable equal to one if the applicant is white and zero if the applicant is African-American	0.92	0.27	0	1
Ethnicity	Dummy variable equal to one if the applicant is Hispanic or Latino and zero otherwise	0.12	0.32	0	1
Gender	Dummy variable equal to one if the applicant is male and zero if the applicant is female	0.70	0.46	0	1
Income	Natural logarithm of the applicant's gross annual income	4.33	0.67	0	7.60

The important factor for our identification approach is  $u_i^D$ , which includes all the unobserved (to the researchers) components of an applicant's decision to apply for a specific loan amount, besides his or her income and the fixed effects. For example,  $u_i^D$  includes the LTV ratio, applicants' perceptions about banks' reaction to his or her FICO score, the DTI ratio relative to the loan amount requested, etc. In fact, the list of unobserved variables can be very idiosyncratic and is non-exhaustive, thus even specialized (and confidential) data sets with more detailed information are unable to fully measure these variables.

Subsequently, we regress the loan amount supplied by the bank, which equals the demanded loan amount for originated loans and zero when the loan is denied by the bank, on applicants' income and the same set of fixed effects. This equation takes the following form:

$$L_i^S = a_0^S + a_1^S y_i + a_2^S d_k + u_i^S \quad (3)$$

where  $L_i^S$  is the dollar-denominated amount of the loan supplied to borrower  $i$  (which can be equal to zero in case the loan is not granted),  $a_0^S$  and  $a_2^S$  are parameters to be estimated by OLS and  $u_i^S$  is the residual associated with borrower  $i$  in this equation.

Equivalently to Eq. (2), the term  $u_i^S$  captures the unobserved (to researchers) information that lending institutions use to assess each applicant's riskiness and decide whether or not to grant the requested loan amount. This information includes, for example, FICO scores and LTV ratios, but it is obviously not the same with the unobserved variables  $u_i^D$  used by the applicants to reach a decision regarding the quantity of the demanded loan amount. Importantly,  $u_i^S$  also includes any discriminatory behavior of banks against applicants' race, ethnicity, and gender.

Next, we take the difference of the residuals between Eqs. (2) and (3), as follows:

$$u_i^D - u_i^S = u_i^h \quad (4)$$

To the extent that Eqs. (1) and (2) are robustly estimated,  $u_i^h$  is the difference between the applicant's model for deciding how much money to request for a mortgage (conditional on his/her income and fixed effects) and the bank's business model conditional on the applicant's income and the fixed effects. This difference is precisely the reasons leading the bank to deny a loan when an individual has applied. If the applicant knew that the bank would deny the loan with certainty, then he/she would not have applied. If the difference is small, then the probability of granting the loan increases. The difference still includes any discriminatory practices concerning specific loan applications, and thus the stochastic component  $u_i^h$  also encompasses the potential effect of discrimination. In addition,  $u_i^h$  encompasses the different valuation of, for instance, FICO scores, LTV ratios, and DTI ratios (among other unobserved variables) by lenders compared with borrowers.

In the second stage of our approach, we introduce  $u_i^h$  in the usual binary model of loan originations given by Eq. (1), rewritten here as follows:

$$gr_i = c_0 + c_1 y_i + c_2 d_r + c_3 d_g + c_4 d_e + c_5 d_k + c_6 (u_i^D - u_i^S) + \varepsilon_i \quad (5)$$

where  $gr_i$  is a dummy variable equal to 1 if borrower  $i$  receives a loan (that is, if  $L_i^S > 0$ ),  $d_r$ ,  $d_g$ , and  $d_e$  are dummy variables capturing race, gender and ethnicity,  $c_0 \dots c_6$  are parameters to be estimated, and  $\varepsilon_i$  is the residual associated with borrower  $i$  in this equation.

By including  $d_r$ ,  $d_g$ , and  $d_e$  in Eq. (5), we essentially subtract these observed variables from the residual  $u_i^h$ . Then,  $u_i^h$  in Eq. (5) includes only the difference in the unobserved variables between the model of the applicant (conditional on his or her income) and the bank's business model (conditional on the applicant's income), leaving the effect of discrimination, purified from unobserved variables, to be captured by coefficients  $c_2$  to  $c_4$ .

Notably, we include in Eqs. (2), (3) and (5) year fixed effects to control for any temporal variation in our dependent variable that is common to all applicants and census-tract fixed effects to control for local housing and socioeconomic characteristics that might shape the bank's decision.<sup>8</sup> Given that including fixed effects is essential, and non-linear estimators such as the logit model face convergence difficulties with such a high number of fixed effects (at least in our sample), we use a linear probability ordinary least squares (OLS) model to estimate Eq. (5).<sup>9</sup>

Our analysis so far assumes that  $u_i^D$  and  $u_i^S$  (and thus also their difference) are estimated with no statistical bias. There are two potential sources of such bias. The first concerns the reported increase in the income misreporting (or fraud) in the U.S. mortgage market from 2002 through 2006 (Financial Crimes Enforcement Network 2006). If this increase produces a serious systematic (upward) measurement error in our income variable, we should expect that our coefficient estimates in Eqs. (2), (3) and (5) are biased. We should also note, however, that given the inclusion of income in both Eqs. (2) and (3), this bias should cancel out when taking their difference in Eq. (4). Thus, the only concern is whether the bias from the inclusion of income in Eq. (5) is systematically correlated with *race*, *gender*, or *ethnicity* (e.g., African-American applicants systematically misreport income more than white applicants).

We do take a number of steps, however, as remedies for the potential presence of measurement error. The first and simplest is to estimate Eqs. (2) and (3) without income and examine the sensitivity of the results. If results do not change, then the measurement error is indeed cancelled out in Eq. (4), and the only remaining error concerns the potential correlation between income misreporting and the three discrimination-related variables in Eq. (5). Then, the second step is to use the errors-in-variables model of Erickson et al. (2014), in which the regression coefficients are identified using information from the higher-order cumulants of the observed variables.<sup>10</sup> An alternative approach, particularly useful for identifying income fraud, is to use mortgage applications only in zip codes with mortgage-loan misrepresentation values at the lowest 5% or 10% centile of the distribution of the misrepresentation measure compiled by Piskorski et al. (2015). We further discuss these approaches below.

The second source of statistical bias could result from correlation between the discrimination dummies ( $d_r$ ,  $d_g$ , and  $d_e$ ) and  $u_i^D$  and  $u_i^S$ . However, we note after estimation that the highest

<sup>8</sup> The census tract is an area roughly equivalent to a neighborhood established by the Bureau of Census for analyzing populations. Census tracts generally encompass a population between 2500 to 8000 people, and represent the smallest territorial unit for which population data are available.

<sup>9</sup> There are a number of issues we examine to ensure that estimation with the linear probability model is not problematic. The two main problems of the LPM vis-à-vis the probit or logit models are heteroscedasticity and the out-of-bound (out of the 0 and 1 bounds) predictions (Wooldridge 2009). Heteroscedasticity is easily resolved with the estimation with robust standard errors. The out-of-bound problem is not that large in our sample, where for e.g. the baseline regression with the full sample, 3045 observations out of the 6,452,279 are below 0 and 127 are above 1. This is less than 0.05% of our sample. Thus, we anticipate very little econometric problems with the OLS estimation. Our best bet to actually show that this is indeed the case, would be to show that results from probit or logit models without fixed effects are similar to results from the LPM and no fixed effects. Unfortunately, even the simple logit or probit models do not converge in our data set.

<sup>10</sup> As suggested by Erickson et al. (2014), we examine the sensitivity of the estimated elasticities on a case-by-case basis when using the higher-order cumulants (from 3 to 8) of the mismeasured regressor, which in our case is the residuals from Eqs. (2) and (3).

observed pairwise correlation coefficient between these variables never exceeds 0.0645. Further, we show that orthogonalizing these variables produces minimal differences in the results.

The equivalent to Eq. (5) loan-pricing equation takes the form:

$$Spread_i = b_0 + b_2 d_r + b_3 d_g + b_4 d_e + b_4 y_i + u_i^S + \varepsilon_i' \quad (6)$$

where *Spread* is the difference between a loan's APR and the rate on Treasury securities of comparable maturity. Given that spreads are available only for loans with spreads above designated thresholds, the sample is significantly smaller compared with the equivalent of discrimination in loan origination.

Identification of Eq. (6) relies only on the inclusion of  $u_i^S$  because Eq. (6) is estimated using only originated loans. In other words, this is a de facto supply equation and does not concern the choice of individuals to apply for the loan. Thus, the bank, in setting the interest rate on the loan, is not concerned with the unobserved characteristics related to individuals' decisions to apply. The bank is only concerned with their own model, which is given by Eq. (3).<sup>11</sup> This makes using Erickson et al. (2014) method even more suitable in this case, so as to account for the measurement error in both Eqs. (3) and (6). We demonstrate, however, that changes in the results between the estimation of all models with OLS and with Erickson et al. (2014) method does not significantly change the main message of our estimates.

## 4 Empirical results

### 4.1 Discrimination in loan origination

Table 2 reports the baseline results from Eq. (5); the results from Eqs. (2) and (3) are available on request. In general, we find no significant differences whether using income in the first stage of the model (Columns I–III) or not (Columns IV–VI). Thus, we can infer that any significant measurement error related to the inclusion of income in Eqs. (2) and (3) cancels out by taking the differences in the residuals in Eq. (4).

The results for the full period (Column I) show that *race*, *gender*, and *ethnicity* bear a statistically significant coefficient at the 1% level. The economic significance of these effects, however, is quite lower than expected. Specifically, we find that white applicants have 0.39% higher probability to be granted a home mortgage loan than African-Americans, while non-Hispanic applicants have a 0.25% higher probability than Hispanic applicants. For *gender*, the coefficient shows that male applicants have a 0.06% higher probability to be denied a loan relative to female applicants, which economically is very small. As expected, the coefficient estimate on the first-stage residual differential,  $u_i^h$ , is highly statistically significant.

Evidently, our estimates are significantly lower than equivalent estimates from existing studies on mortgage lending discrimination. For example, Munnell et al. (1996) and Hubbard et al. (2012) find an approximately 8% higher probability that white applicants will be granted a loan compared with African-American applicants, and Wheeler and Olson (2015) find an equivalent 9.8%. Notably, if we estimate Eq. (5) without the stochastic component  $u_i^h$ , the coefficient on race becomes 0.128, on ethnicity −0.085, and on gender 0.002 (results in

<sup>11</sup> In fact, econometrically it is not even possible to include  $u_i^D$  in Eq. (6) because  $L_i^D = L_i^S$  and, if included,  $u_i^h$  drops out from the regression due to perfect collinearity.

**Table 2** Discrimination in loan origination: Baseline results. The table reports coefficient estimates and  $t$ -statistics (in parentheses) from the estimation of Eq. (5), given estimates from Eqs. (2)–(4). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year and census-tract fixed effects. Columns I and IV report the results for the period 2004–2013, Columns II and V the results for the pre-crisis period (2004–2007), and Columns III and VI for the crisis and post-crisis period (2008–2013). Columns I–III report the results when income is included in the first stage of the model, and Columns IV–VI when income is not included in the first-stage regressions

	Income is included in the first stage of the model			Income is not included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013	IV 2004–2013	V 2004–2007	VI 2008–2013
Race	0.0039*** (26.73)	0.0023*** (14.66)	0.0066*** (17.53)	0.0039*** (26.73)	0.0023*** (14.66)	0.0066*** (17.53)
Income	0.0441*** (811.59)	0.0413*** (630.51)	0.0471*** (479.96)	0.0151*** (273.93)	0.0142*** (214.24)	0.0161*** (162.56)
Gender	−0.0006*** (−9.26)	−0.0007*** (−9.86)	−0.0001 (−1.25)	−0.0006*** (−9.26)	−0.0007*** (−9.86)	−0.0001 (−1.25)
Ethnicity	−0.0025*** (−21.57)	−0.0014*** (−10.58)	−0.0054*** (−19.22)	−0.0025*** (−21.57)	−0.0014*** (−10.58)	−0.0054*** (−19.22)
First-stage residual differential $u_i^h$	−0.1974*** (−5910.3)	−0.1988*** (−5028.9)	−0.1929*** (−3140.6)	−0.1974*** (−5910.3)	−0.1988*** (−5028.9)	−0.1929*** (−3140.6)
Observations	6,452,279	4,788,873	1,658,207	6,452,279	4,788,873	1,658,207
Adjusted R-squared	0.9661	0.9651	0.9705	0.9661	0.9651	0.9705
Census Tract effects	YES	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

Table 12 in the Appendix). These estimates are very close to the estimates of past research that solely use the HMDA database to infer lending discrimination.<sup>12</sup> Also, the adjusted R-squared decreases from 96.6% in the model with  $u_i^h$  to 8.4% in the model without  $u_i^h$ . Given the foregoing, we must attribute the very large drop in the discrimination-related probabilities to the very significant reduction in omitted-variable bias achieved by including  $u_i^h$  in Eq. (5). In this respect, our approach contributes to the literature on lending discrimination and allows for more-accurate estimation of discriminatory practices using the full gamut of loan applications in the HMDA database across the United States.<sup>13</sup>

In the regressions of Columns II and III (or equivalently, in those of V and VI for the models without income in the first stage), we split our sample into the pre- and post-2007 periods. We find that during the pre-crisis period, white applicants have a 0.23% higher probability to be

<sup>12</sup> The estimates become even more economically significant if we remove the regional fixed effects, and they decrease by about 10% if we include bank fixed effects.

<sup>13</sup> To further compare our findings with the existing literature, and especially the seminal study of Munnell et al. (1996), we collect HMDA data for Boston (Suffolk County) in 1990. As in Munnell et al., we identify 1200 black and Hispanic applicants and choose a random sample of white applicants to produce roughly equal numbers of white and minority denials (10% rejection rate and 3300 applications by whites). Using that sample, we run a basic model, which includes race, income, gender, and census tract fixed effects as explanatory variables. We find a discrimination estimate of 11.4%. Recall that Munnell et al. find an estimate of 8% when adding 38 more variables, which are not available to us (data are confidential). Then, we repeat the analysis using our method and find an estimate of 0.9%. We report these results in Table 13 of the Appendix. Thus, we must conclude that there is still a lot of omitted-variable bias in the study by Munnell et al. (1996).

granted a mortgage loan than African-American applicants (Column II). As hypothesized in Section 2, this probability increases after the onset of the subprime crisis in 2007. In particular, during the post-2007 period (2008–2013), we find a nearly three times higher probability (0.66%) that white applicants are granted a loan compared with African-American applicants *ceteris paribus* (Column III).

The role of the applicant's ethnicity seems also to be significantly more important after the subprime crisis. Specifically, during the post-2007 period, Hispanic applicants have a 0.54% lower probability of receiving a loan than non-Hispanics (almost four times higher than the 0.14% of the pre-crisis period). The estimated coefficient for the *gender* variable is quite low in the pre-crisis period, when female applicants had a slightly higher probability (0.07%) to receive a loan than male applicants. This probability diminishes during the post-2007 period (0.01%), and it is statistically insignificant at any conventional level.

We also experiment with estimating Eq. (5) by year, and Fig. 1 reports the coefficient estimates in graphical form. Apparently, the annual estimates of *race*, *gender*, and *ethnicity* are in line with the preceding suggestion that loan origination discrimination levels are particularly low before the subprime crisis and that a slight increase occurred after 2007 in the absolute values of the estimates of *race* and *ethnicity*. Still, we cannot suggest that extensive discrimination takes place against African-American or Hispanic applicants because the maximum (absolute) value on both race and ethnicity is about 0.6%.

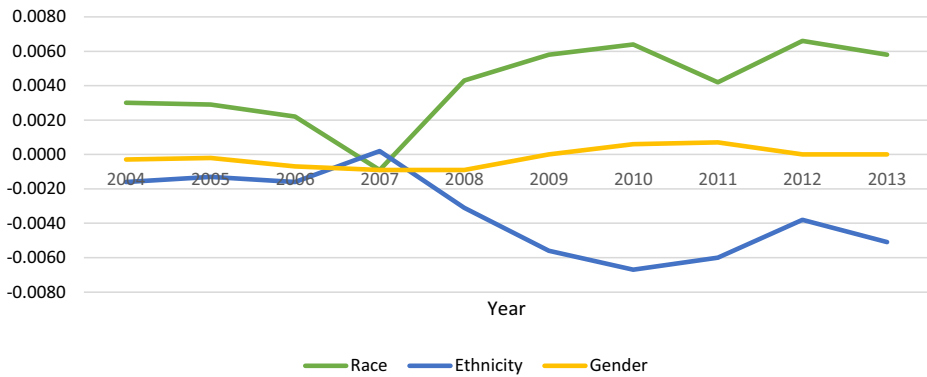
The regressions of Table 3 replicate those of Table 2 using Erickson et al.'s (2014) errors-in-variables model for estimating Eq. (5). Following the proposed procedure in Erickson et al., we first use the within transformation to account for the census tract and year fixed effects and then use the transformed variables in Eq. (5), assuming that both the transformed  $u_i^h$  and  $y$  variables are measured with error.

The results show that, if anything, the coefficient estimates on *race*, *ethnicity*, and *gender* are a bit smaller than the equivalent in Table 2. Thus, the role of measurement error seems to be minimal and, if anything, producing an upward bias in the results. We also find the same pattern concerning the role of the crisis, with the estimates in the post-2007 period being somewhat larger than those of the pre-crisis period. Even in the post-2007 period, however, the estimates' economic significance is fairly low, indicating for example that white applicants are 0.49% more likely to have a loan approved compared with African-American applicants, *ceteris paribus*.

In Table 4, we include bank fixed effects in Eq. (5).<sup>14</sup> Thus, these specifications control for each bank's business model and consider within-bank discrimination of mortgage loan applicants. The results are almost the same as those of Table 2.

Further, in Table 5, we examine whether our findings differ by supervisory agency. We split our sample into three groups based on the supervisor being the OCC, the FRS, or the FDIC. We find some evidence that, at least in the full sample period, discriminatory practices are lowest in the OCC-supervised banks and highest in the FDIC-supervised banks. This result is somewhat expected because in general, the FDIC supervises the relatively small local banks. Nonetheless, the differences and magnitude of the coefficients show a very low level of discriminatory practices irrespective of the supervisor.

<sup>14</sup> As we showed earlier, including or excluding income in the first stage of the model does not significantly affect our results. From this point onward, we retain the specification with income included in the first stage of the model.



**Fig. 1** Discrimination in loan origination by year

So far, we have assumed that nonlinear effects between the different forms of discrimination in loan origination are not present. To relax this assumption, we introduce in Eq. (5) a number of interaction terms between (i) race and gender, (ii) race and ethnicity, and (iii) gender and ethnicity. We further refine our results by including the triple interaction between race, gender, and ethnicity. In Table 6 we report our findings, with the majority of the interaction variables being statistically insignificant. Specifically, the coefficient estimates on the interaction term between race and ethnicity and the triple interaction term among race, gender, and

**Table 3** Discrimination in loan origination: Error-in-variables model. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (5), given estimates from Eqs. (2)–(4). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year and census-tract fixed effects. Equation (5) is estimated using the Erickson–Whited (2002) errors-in-variables model. Columns I and IV report the results for the period 2004–2013, Columns II and V the results for the pre-crisis period (2004–2007), and Columns III and VI for the crisis and post-crisis period (2008–2013). Columns I–III report the results when income is included in the first stage of the model and Columns IV–VI when income is not included in the first-stage regressions

	Income is included in the first stage of the model			Income is not included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013	IV 2004–2013	V 2004–2007	VI 2008–2013
Race	0.0019*** (22.42)	0.0004*** (3.97)	0.0049*** (24.20)	0.0019*** (22.39)	0.0004*** (3.93)	0.0049*** (24.21)
Income	0.0441*** (1244.42)	0.0414*** (937.65)	0.0471*** (821.19)	0.0147*** (414.67)	0.0139*** (314.39)	0.0157*** (273.50)
Gender	-0.0006*** (-13.75)	-0.0008*** (-14.72)	-0.0001 (-1.42)	-0.0006*** (-13.76)	-0.0008*** (-14.74)	-0.0001 (-1.42)
Ethnicity	-0.0013*** (-17.14)	-0.0001 (-1.34)	-0.0042*** (-26.60)	-0.0013*** (-17.11)	-0.0001 (-1.31)	-0.0042*** (-26.61)
First-stage residual differential $u_i^h$	-0.2004*** (-16,742.7)	-0.2019*** (-14,070.5)	-0.1959*** (-9458.9)	-0.2004*** (-16,757.2)	-0.2019*** (-14,080.0)	-0.1959*** (-9471.67)
Observations	6,452,279	4,788,873	1,658,207	6,452,279	4,788,873	1,658,207
R-squared	0.9782	0.9771	0.9823	0.9782	0.9771	0.9824
Census Tract effects	YES	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively



**Table 4** Discrimination in loan origination: Including bank fixed effects. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (5), given estimates from Eqs. (2)–(4). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year, bank, and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	0.0035*** (24.43)	0.0020*** (12.88)	0.0061*** (16.32)
Income	0.0442*** (813.58)	0.0414*** (632.60)	0.0471*** (481.25)
Gender	−0.0007*** (−11.08)	−0.0008*** (−11.95)	−0.0001 (−1.17)
Ethnicity	−0.0026*** (−21.71)	−0.0015*** (−11.06)	−0.0057*** (−20.46)
First-stage residual differential $u_i^h$	−0.1971*** (−5839.5)	−0.1984*** (−4948.9)	−0.1927*** (−3107.5)
Observations	6,452,272	4,788,866	1,658,199
Adjusted R-squared	0.9665	0.9655	0.9710
Census-tract effects	YES	YES	YES
Bank effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

ethnicity are statistically insignificant, while the interaction term between race and gender is positive and significant only for the post-crisis period. The interaction term between the gender and the ethnicity is positive and significant at the 10% level indicating that the probability of male applicants to be granted a loan increases when the applicants belong to a Hispanic minority group. Even in this case, however, the effect is economically very small.

An important sensitivity test is to use mortgage applications only in zip codes with mortgage-loan misrepresentation values at the lowest 5% centile of the distribution of the misrepresentation measure compiled by Piskorski et al. (2015). Specifically, we adopt the following procedure: (1) We match the census tracts in the HMDA annual files with zip codes using the Missouri Census Data bridge for the 2000 census.<sup>15</sup> For a one-to-one match between census tracts and zip codes, we rely on population weights. We use only zip codes for which census tracts cover more than 85% of a zip code's population. (2) We match the zip codes with the misrepresentation measure compiled by Piskorski et al. (2015). Although these data refer to the 2005–2007 period and concern misrepresentations with regard to owner-occupied property or deliberately omitted information on a second lien, Mian and Sufi (2015) show that this type of misrepresentation is also strongly related to fraudulent income overstatement. Moreover, Mian and Sufi (2015) show that mortgage fraud is a fixed characteristic of such zip codes (does not have a significant time variation) and thus could be used for the full sample period. (3) For our sensitivity analysis, we keep only those applications in HMDA annual files for which the misrepresentation measure is at the low 10% centile of its distribution. The results, reported in

<sup>15</sup> Available online in <http://mcdc2.missouri.edu/websas/geoconr2k.html>.

**Table 5** Discrimination in loan origination: Evidence by the three main supervisory authorities. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (5), given estimates from Eqs. (2)–(4). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year and census-tract fixed effects. Columns I, IV, and VII report results for the period 2004–2013, Columns II, V, and VIII for the pre-crisis period (2004–2007), and Columns III, VI, and IX for the crisis and post-crisis period (2008–2013). Columns I–III report results for banks supervised by the OCC, Columns IV–VI for the banks supervised by the Fed, and Columns VII–IX for the banks supervised by the FDIC

	Office of the Comptroller of the Currency (OCC)			Federal Reserve System (FRS)			Federal Deposit Insurance Corporation (FDIC)		
	I	II	III	IV	V	VI	VII	VIII	IX
	2004–2013	2004–2007	2008–2013	2004–2013	2004–2007	2008–2013	2004–2013	2004–2007	2008–2013
Race	0.0025*** (11.95)	0.0007*** (3.12)	0.0050*** (9.37)	0.0037*** (16.50)	0.0032*** (13.51)	0.0040*** (4.89)	0.0065*** (12.64)	0.0056*** (8.60)	0.0062*** (7.49)
Income	0.0459*** (587.43)	0.0461*** (505.11)	0.0405*** (266.17)	0.0332*** (324.25)	0.0320*** (278.59)	0.0333*** (153.53)	0.0560*** (476.9753)	0.0515*** (314.079)	0.0607*** (357.88)
Gender	−0.0006*** (−6.68)	−0.0006*** (−6.70)	−0.0002 (−1.40)	−0.0009*** (−7.97)	−0.0009*** (−7.85)	−0.0004 (−1.44)	0.0003*** (2.21)	0.0002 (1.07)	0.0003*** (2.29)
Ethnicity	−0.0032*** (−19.48)	−0.0016*** (−8.64)	−0.0062*** (−17.55)	−0.0022*** (−11.56)	−0.0019*** (−9.46)	−0.0027*** (−3.79)	−0.0046*** (−9.94)	−0.0039*** (−6.31)	−0.0034*** (−7.59)
First-stage residual differential $u_i^h$	−0.1942*** (−3900.0)	−0.1970*** (−3103.9)	−0.1881*** (−2403.2)	−0.1984*** (−3872.8)	−0.1988*** (−3689.3)	−0.1942*** (−1174.6)	−0.2044*** (−1959.0)	−0.2063*** (−1350.2)	−0.2024*** (−1417.2)
Observations	3,054,842	2,230,499	817,438	2,281,753	1,990,686	282,669	1,011,918	534,296	470,669
Adjusted R-squared	0.9669	0.9655	0.9714	0.9664	0.9658	0.9758	0.9677	0.9653	0.9707
Census Tract effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES	YES	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

**Table 6** Discrimination in loan origination: Interaction terms between the discrimination variables. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (5), given estimates from Eqs. (2)–(4). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year, bank, and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	0.0041*** (20.34)	0.0025*** (11.48)	0.0075*** (13.74)
Income	0.0442*** (812.41)	0.0413*** (630.25)	0.0471*** (479.99)
Gender	−0.0005** (−2.05)	−0.0007*** (−2.76)	0.0012* (1.79)
Ethnicity	−0.0029** (−2.15)	−0.0018 (−1.25)	−0.0086** (−2.41)
Race*Gender	−0.0003 (−1.14)	−0.0002 (−0.79)	−0.0016** (−2.30)
Race*Ethnicity	−0.0010 (−0.69)	−0.0009 (−0.57)	0.0019 (0.53)
Gender*Ethnicity	0.0034* (1.90)	0.0033* (1.73)	0.0021 (0.43)
Race*Gender*Ethnicity	−0.0015 (−0.84)	−0.0015 (−0.76)	−0.0002 (−0.03)
First-stage residual differential $u_i^h$	−0.1974*** (−5910.3)	−0.1988*** (−5029.0)	−0.1929*** (−3140.5)
Observations	6,452,279	4,788,873	1,658,207
Adjusted R-squared	0.9661	0.9651	0.9705
Census-tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

Table 14 in the Appendix, show that like the error-in-variables model, income misreporting does not affect our results.

A final sensitivity test involves orthogonalizing the residuals in Eqs. (2) or (3) relative to  $d_r$ ,  $d_g$  and  $d_e$  before they are plugged into Eq. (5). Even though the pairwise correlation coefficients never exceed 0.0645, this exercise aims to further reduce potential endogeneity stemming from such correlation. The results, reported in Table 15 of the Appendix, are very close to our baseline results in Table 2.

## 4.2 Discrimination in loan pricing

In Table 7, we report our baseline results from estimating Eq. (6). For expositional brevity, we report only the results from specifications that include income in the first stage of the model (again, this decision does not affect the results). Column I reports the results for the full sample and shows that white applicants pay a lower spread on the originated loans by 0.37 basis points.<sup>16</sup> This finding holds despite the inclusion of  $u_i^S$  in Eq. (6) and its very high statistical and economic significance.

<sup>16</sup> For the mean applicant, the spread is 5.07 basis points, implying that white applicants pay approximately a 7.3% lower spread.

**Table 7** Discrimination in loan pricing: Baseline results. The table reports coefficient estimates and t-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). The dependent variable is *Spread*, and all variables are defined in Table 1. Equations (3) and (6) include year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	−0.3740*** (−57.17)	−0.3880*** (−57.70)	−0.0418 (−1.44)
Income	−0.0988*** (−24.29)	−0.0867*** (−18.34)	−0.1118*** (−13.60)
Gender	−0.0529*** (−12.96)	−0.0643*** (−14.92)	0.0629*** (5.20)
Ethnicity	0.1052*** (18.03)	0.1116*** (18.40)	0.0471** (2.10)
Residual of Eq. (3) $u_i^s$	−0.8729*** (−366.70)	−0.9139*** (−365.31)	−0.4760*** (−61.65)
Observations	594,302	522,028	66,328
Adjusted R-squared	0.3951	0.3187	0.4925
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

Interestingly, this difference comes from the pre-crisis period. Column II of Table 7 shows that during the pre-crisis period, the average white applicant paid a lower spread on the originated loans by 0.39 basis points, whereas in the post-2007 period, the relevant coefficient is statistically insignificant. This finding is intuitive conditional on the results discussed in Section 4.1. In the pre-crisis period, deemed a low-risk period, banks were willing to originate the loans but required somewhat higher spreads from minority applicants. In the post-2007 period, deemed a very risky period, the probability is slightly higher that banks will discriminate by not granting the loan.

Concerning the effect of *ethnicity*, the results are economically weaker and show that Hispanic applicants obtain mortgage loans with approximately 0.11 basis points higher spread. Similar to *race*, the effect is considerably higher in the pre-crisis period. Finally, the effect of *gender* is even less potent, with male applicants paying 0.05 lower spreads in the whole sample. When we differentiate between the pre-2007 and post-2007 periods, we find that in the risky post-2007 period, women are given lower spreads by approximately 0.06 basis points. Even though this effect is economically small, we can attribute such bank behavior to the general perception that women are more risk averse in their financial decisions and that banks price in this risk aversion during periods of high uncertainty.

In Table 8 we replicate the regressions of Table 7 using the errors-in-variables model. We work under the assumption that both the residual of Eq. (3),  $u_i^h$ , and applicants' income are measured with error. Using this model is potentially more important in the loan pricing equation than in the loan origination equation, because the former does not eliminate bias through Eq. (4). Still, changes in the results are not significant, implying low measurement error in  $u_i^h$  and income.

**Table 8** Discrimination in loan pricing: Error-in-variables model. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). The dependent variable is *Spread*, and all variables are defined in Table 1. Equations (3) and (6) include year and census-tract fixed effects. Equation (6) is estimated using the Erickson–Whited (2002) errors-in-variables model. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	−0.3776*** (−66.73)	−0.3917*** (−66.02)	−0.0440* (−1.88)
Income	−0.1076*** (−30.3822)	−0.0958*** (−23.50)	−0.1152*** (−17.84)
Gender	−0.0525*** (−14.5141)	−0.0639*** (−16.51)	0.0632*** (6.41)
Ethnicity	0.1072*** (20.2407)	0.1137*** (20.26)	0.0480*** (2.69)
Residual of Eq. (3) $u_i^s$	−0.9196*** (−393.64)	−0.9627*** (−383.50)	−0.5015*** (−81.24)
Observations	594,302	522,028	66,328
Adjusted R-squared	0.2715	0.2249	0.2925
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

In Table 9, we consider including bank fixed effects in Eq. (6). Interestingly, the coefficient estimates on *race*, *gender*, and *ethnicity* decrease somewhat. For example, white applicants obtain loans with a lower spread by 0.27 basis points, whereas the equivalent spread in the model without bank fixed effects is 0.37 basis points. This finding is expected because in the Table 9 regressions we consider only within-bank discrimination, whereas in previous tables we also consider across-bank discrimination. The pattern in the results regarding the effect of the subprime crisis is very similar to the one in Table 7.

We also repeat our analysis on the three main regulatory agencies. The findings, reported in Table 10 indicate that our baseline results from Eq. (6) are identical with the ones obtained from banks supervised by the FRS. This finding is expected given that approximately 65% of the higher-priced mortgage loans in our sample are originated by the above banking institutions. Interestingly, discriminatory behavior is the lowest in banks that are under the jurisdiction of the FDIC. The average white applicant in these institutions pays approximately a 0.18 basis points lower spread on the originated loans which is a substantially lower difference compared with the 0.34 and 0.25 basis points lower spread that white individuals pay in FRS-supervised and OCC-supervised banks, respectively.

Symmetrically to the relevant tests for Eq. (5), we introduce in Eq. (6) interaction terms to examine possible nonlinear effects in the three forms of discrimination. We report our findings in Table 11. We note that there are some nonlinear effects originating in the pre-crisis period. Specifically, the 0.36 points lower spread paid by the average white applicant is by 0.25 points higher for white Hispanic applicants in the period 2004–2007. Further, the interaction term between *gender* and *ethnicity* shows that white male applicants pay an extra 0.07 points lower spread compared to white female applicants in the period 2004–2007. These nonlinear effects, even though economically small, are in line with the results of some discrimination in loan

**Table 9** Discrimination in loan pricing: Including bank fixed effects. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). The dependent variable is *Spread*, and all variables are defined in Table 1. Equations (3) and (6) include year, bank, and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	−0.2671*** (−45.04)	−0.2756*** (−45.27)	−0.0839*** (−3.07)
Income	−0.1225*** (−32.85)	−0.1116*** (−25.95)	−0.1165*** (−14.62)
Gender	−0.0245*** (−6.60)	−0.0314*** (−8.04)	0.0452*** (3.96)
Ethnicity	0.0454*** (8.28)	0.0491*** (8.62)	0.0310 (1.47)
Residual of Eq. (3) $u_i^s$	−0.8323*** (−361.73)	−0.8702*** (−362.69)	−0.4393*** (−51.99)
Observations	594,247	521,749	65,989
Adjusted R-squared	0.5039	0.4471	0.5807
Census-tract effects	YES	YES	YES
Bank effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

pricing, especially in the pre-crisis period. The coefficient on the triple interaction term between race, gender, and ethnicity is statistically insignificant.

Finally, we check whether our results change when using a truncated regression and maximum likelihood. The estimator is preferred when the depended variable is censored, as is in the case with *Spread*, which has a lower limit equal to 1.5 basis points. We report the relevant results in Table 16 of the Appendix. The results are essentially the same with those of Table 7; again, we attribute this similarity to our very large data set with dense observations.

## 5 Conclusions and policy implications

This paper revisits the issue of mortgage lending discrimination in the United States and proposes a method that lowers the omitted-variable bias stemming from the use of the HMDA database. We study discrimination based on race, ethnicity, and gender, using loan applications from 2004 through 2013, which includes the subprime crisis. Further, we identify both discrimination in loan origination (i.e., the bank's decision to grant or deny the loan for reasons related to race, ethnicity, and gender) and in loan pricing (i.e., the bank's decision to grant different loan spreads for reasons related to race, ethnicity, and gender).

Our results show that the effect of race and ethnicity in a bank's decision to grant a mortgage loan is significantly lower compared with the corresponding results from previous studies on mortgage lending discrimination. Specifically, according to our preferred specification, white applicants have only a 0.39% higher probability to be granted a mortgage loan than African Americans, while non-Hispanic applicants have a 0.25% higher probability than

**Table 10** Discrimination in loan pricing: Evidence by the three main supervisory authorities. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). The dependent variable is *Spread<sub>it</sub>*, and all variables are defined in Table 1. Equations (3) and (6) include year and census-tract fixed effects. Columns I, IV, and VII report results for the period 2004–2013, Columns II, V, and VIII for the pre-crisis period (2004–2007), and Columns III, VI, and IX for the crisis and post-crisis period (2008–2013). Columns I–III report results for banks supervised by the OCC, Columns IV–VI for banks supervised by the Fed, and Columns VII–IX for banks supervised by the FDIC

	Office of the Comptroller of the Currency (OCC)			Federal Reserve System (FRS)			Federal Deposit Insurance Corporation (FDIC)		
	I	II	III	IV	V	VI	VII	VIII	IX
	2004–2013	2004–2007	2008–2013	2004–2013	2004–2007	2008–2013	2004–2013	2004–2007	2008–2013
Race	-0.2471*** (-16.71)	-0.2663*** (-16.96)	-0.0181 (-0.35)	-0.3444*** (-42.61)	-0.3466*** (-42.65)	-0.0671 (-0.68)	-0.1814*** (-8.02)	-0.2020*** (-7.88)	-0.1139*** (-2.33)
Income	-0.1076*** (-12.53)	-0.1240*** (-12.27)	-0.0801*** (-4.43)	-0.0473*** (-7.62)	-0.0357*** (-5.56)	-0.0364 (-1.36)	-0.1553*** (-18.45)	-0.1098*** (-9.05)	-0.1723*** (-14.18)
Gender	-0.0250*** (-2.71)	-0.0377*** (-3.72)	0.0442* (1.84)	-0.0522*** (-9.93)	-0.0548*** (-10.32)	0.0619 (1.60)	-0.0022 (-0.20)	-0.0275*** (-2.06)	0.0392** (2.20)
Ethnicity	-0.0457*** (-3.26)	-0.0471*** (-2.99)	-0.0142 (-0.42)	0.0911*** (12.42)	0.0904*** (12.26)	0.0180 (0.22)	0.0743*** (4.21)	0.0811*** (4.18)	0.0717* (1.75)
Residual of Eq. (3) $u_i^d$	-0.6273*** (-106.58)	-0.6328*** (-97.34)	-0.6079*** (-39.20)	-0.9864*** (-337.48)	-0.9939*** (-338.33)	-0.4026*** (-13.96)	-0.5651*** (-84.69)	-0.6784*** (-86.77)	-0.3634*** (-29.64)
Observations	105,329	86,517	16,275	361,879	353,936	6107	74,903	42,741	30,574
Adjusted R-squared	0.3730	0.2767	0.5172	0.3706	0.3498	0.5423	0.5411	0.4623	0.5528
Census Tract effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES	YES	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively



**Table 11** Discrimination in loan pricing: Interaction terms between the discrimination variables. The table reports coefficient estimates and t-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). The dependent variable is *Spread*, and all variables are defined in Table 1. Equations (3) and (6) include year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	–0.3541*** (–39.49)	–0.3573*** (–38.69)	–0.0523 (–1.19)
Income	–0.1008*** (–24.72)	–0.0866*** (–18.30)	–0.1117*** (–13.58)
Gender	–0.0048 (–0.52)	–0.0049 (–0.53)	0.0605 (1.22)
Ethnicity	–0.1274** (–2.48)	–0.1232** (–2.35)	–0.3471 (–1.57)
Race*Gender	–0.0529*** (–4.96)	–0.0724*** (–6.56)	0.0043 (0.08)
Race*Ethnicity	0.2573*** (4.94)	0.2508*** (4.73)	0.4085* (1.83)
Gender*Ethnicity	–0.1119 (–1.62)	–0.1230* (–1.75)	0.0934 (0.28)
Race*Gender*Ethnicity	0.0880 (1.26)	0.1125 (1.59)	–0.1068 (–0.32)
Residual of Eq. (3) $u_i^s$	–0.8729*** (–366.7)	–0.9138*** (–365.3)	–0.4761*** (–61.65)
Observations	594,302	522,028	66,328
Adjusted R-squared	0.3952	0.3189	0.4925
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

Hispanic applicants. For gender, the coefficient estimates are statistically insignificant. In contrast, when we use the main methodology of previous studies, we find estimates close to or higher than these studies (about 12% for race and 8.5% for ethnicity). We thus believe that our empirical strategy, even though not panacea, captures a great deal of the omitted-variable bias and allows for more-accurate estimation of discriminatory practices across the whole spectrum of insured U.S. banking institutions.

Our analysis does not suggest by any means that lending institutions do not engage in discriminatory practices. We simply posit that the average estimates that 8 or more out of 100 applicants are denied mortgage loans based on their race are inaccurate. In contrast, our methodology provides an instrument for (1) academics to identify discriminatory practices using basic data across the entire United States, (2) regulators to identify discrimination beyond doubt and punish it with enforcement actions, and (3) policymakers to initiate new laws and regulations that prevent discrimination.

With respect to loan spreads, we find that white and non-Hispanic applicants are charged 0.37 and 0.11 basis points lower spreads on the originated loans respectively, than African-American and Hispanic applicants. We also show that these results come almost entirely from the pre-crisis period. Combined with our finding that racial discrimination in loan origination increases from 0.23% in the pre-crisis period to 0.66% during the crisis and post-2007 period,

our finding on discrimination in loan pricing suggests that after the crisis erupted, discrimination simply shifted from loan pricing to loan (non)-origination. This finding itself provides evidence that discriminatory practices, even though relatively isolated events, do take place and require caution. Further, our finding reminds us of the old and influential economics and sociology literatures suggesting that crises mostly hurt minorities and steps are needed to protect these minorities during economic downturns.

## Appendix

The tables in this appendix are for online use only. They include the following information:

- Table 12 reports the results from the estimation of Eq. (5) without the previous estimation of Eqs. (2)–(4) and the inclusion of the component  $e$  among the regressors in Eq. (5).
- Table 13 reports the results from an analysis similar to Munnell et al. (1996). See also footnote 12 of the main text.
- Table 14 reports the results from the analysis on income misreporting, which includes only zip codes with mortgage-loan misrepresentation values at the lowest 5% centile of the distribution of the misrepresentation measure compiled by Piskorski et al. (2015).
- Table 15 reports the results after orthoganilizing the three discrimination dummy variables and  $u_i^D$  and  $u_i^S$ .
- Table 16 compares the results from the logit model without fixed effects (marginal effects) with the equivalent OLS results (again without fixed effects).

**Table 12** Results from the estimation of Eq. (5) without a first stage. The table reports coefficient estimates and  $t$ -statistics (in parentheses) from the estimation of Eq. (5). The dependent variable is *Action type*, and all variables are defined in Table 1. Equation (5) includes year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the post-crisis period (2008–2013)

	I 2004–2013	II 2004–2007	III 2008–2013
Race	0.1282*** (172.93)	0.1263*** (157.36)	0.1194*** (59.57)
Income	0.0395*** (153.38)	0.0366*** (118.49)	0.0443*** (91.64)
Gender	0.0028*** (8.93)	0.0045*** (12.53)	−0.0020*** (−3.21)
Ethnicity	−0.0850*** (−138.34)	−0.0838*** (−124.28)	−0.0786*** (−51.10)
Observations	6,452,279	4,788,873	1,658,207
Adjusted R-squared	0.0842	0.0838	0.0985
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

**Table 13** Comparison of our baseline results with those of Munnell et al. (1996) for the year 1990. The table reports coefficient estimates and  $t$ -statistics (in parentheses) from the estimation of Eq. (5) using data for the year 1990 for Boston (Suffolk County), as in Munnell et al. (1996). We identify 1200 black and Hispanic applicants and choose a random sample of white applicants to produce roughly equal numbers of white and minority denials (10% rejection rate and 3300 applications by whites). Using that sample, we first run a basic probit model (Column I). Then, we repeat the analysis using our method (Column II). The dependent variable is *Action type*, and all variables are defined in Table 1. Equation (5) includes census-tract fixed effects

	I Munnell et al.	II Our method
Race	0.1142*** (7.79)	0.0096*** (3.36)
Income	0.0180 (1.63)	0.0366*** (16.77)
Gender	0.0078 (0.83)	0.0016 (0.96)
First-stage residual differential $u_i^h$		-0.1824*** -142.66
Observations	4500	4500
Adjusted R-squared	0.0709	0.9724
Census Tract effects	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

**Table 14** Evidence from zip codes with low mortgage-loan misrepresentation values. The table reports coefficient estimates and  $t$ -statistics (in parentheses) from the estimation of Equation (5), given estimates from Eqs. (2) through (4). The sample includes only zip codes with mortgage-loan misrepresentation values at the lowest 5% centile of the distribution of the misrepresentation measure compiled by Piskorski et al. (2015). The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3), and (5) include year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	0.0042*** (25.32)	0.0023*** (13.15)	0.0060*** (16.10)
Income	0.0482*** (720.16)	0.0445*** (627.12)	0.0493*** (462.60)
Gender	-0.0002 (-1.18)	-0.0002 (-1.35)	-0.0001 (-0.97)
Ethnicity	-0.0029*** (-23.12)	-0.0019*** (-14.25)	-0.0064*** (-38.91)
First-stage residual differential $u_i^h$	-0.1635*** (-4624.5)	-0.1688*** (-4055.7)	-0.1627*** (-2562.3)
Observations	322,614	239,443	82,910
Adjusted R-squared	0.9485	0.9426	0.9508
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

**Table 15** Results after orthogonalization of discrimination dummies and errors from Eqs. (2) and (3). The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Equation (5), given estimates from Eqs. (2) through (4). Prior to the estimation, we orthogonalize *Race*, *Gender*, *Ethnicity* relative to  $u_i^D$ . The dependent variable is *Action type*, and all variables are defined in Table 1. Equations (2), (3) and (5) include year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the post-crisis period (2008–2013)

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	0.0033*** (51.86)	0.0033*** (44.96)	0.0025*** (21.63)
Income	0.0440*** (503.15)	0.0416*** (386.47)	0.0447*** (295.70)
Gender	−0.0005*** (−9.84)	−0.0004*** (−7.90)	−0.0006*** (−7.55)
Ethnicity	−0.0025*** (−40.46)	−0.0025*** (−34.31)	−0.0022*** (−19.48)
First-stage residual differential $u_i^h$	−0.1832*** (−4951.1)	−0.1833*** (−4225.6)	−0.1827*** (−2606.1)
Observations	6,452,279	4,788,873	1,658,207
Adjusted R-squared	0.9000	0.8929	0.9240
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

**Table 16** Discrimination in loan pricing: Results from a truncated regression model. The table reports coefficient estimates and *t*-statistics (in parentheses) from the estimation of Eq. (6), given estimates from Eq. (3). Estimation method is the truncated regression model. The dependent variable is *Spread*, and all variables are defined in Table 1. Equations (3) and (6) include year and census-tract fixed effects. Column I reports the results for the period 2004–2013, Column II the results for the pre-crisis period (2004–2007), and Column III for the crisis and post-crisis period (2008–2013). In all regressions, income is included in the first stage of the model

	Income is included in the first stage of the model		
	I 2004–2013	II 2004–2007	III 2008–2013
Race	−0.3520*** (−53.21)	−0.3914*** (−58.11)	−0.0501 (−1.52)
Income	−0.0897*** (−22.14)	−0.0941*** (−20.50)	−0.1020*** (−12.43)
Gender	−0.0571*** (−13.51)	−0.0689*** (−15.21)	0.0611*** (4.91)
Ethnicity	0.0907*** (16.81)	0.1031*** (17.44)	0.0460** (2.05)
Residual of Eq. (3) $u_i^s$	−0.8633*** (−361.22)	−0.9050*** (−358.75)	−0.4730*** (−60.28)
Observations	594,155	520,930	66,276
Census Tract effects	YES	YES	YES
Year effects	YES	YES	YES

\*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively

## References

- Avery RB, Canner GB, Cook RE (2005) New information reported under HMDA and its application in fair lending enforcement. *Fed Reserv Bull* 91:344–394
- Avery RB, Bhutta N, Brevort KP, Canner GB, Gibbs CN (2010) The 2008 HMDA data: the mortgage market during a turbulent year. *Fed Reserv Bull* 96:169–211
- Bayer P, Ferreira F, Ross SL (2014). Race, ethnicity and high-cost mortgage lending. NBER working paper No. 20762
- Bocian DG, Ernst KS, Li W (2008) Race, ethnicity and subprime home loan pricing. *J Econ Bus* 60:110–124
- Cheng P, Lin Z, Liu Y (2011) Do women pay more for mortgages? *J Real Estate Financ Econ* 43:423–440
- Cheng P, Lin Z, Liu Y (2014) Racial discrepancy in mortgage interest rates. *J Real Estate Financ Econ* 51:101–120
- Courchane MJ (2007) The pricing of home mortgage loans to minority borrowers: how much of the APR differential can we explain? *Journal of Real Estate Research* 29:399–440
- Day TE, Liebowitz SJ (1998) Mortgage lending to minorities: Where's the bias? *Econ Inq* 36:3–28
- DeLoughy ST (2012) Risk versus demographics in subprime mortgage lending: evidence from three Connecticut cities. *J Real Estate Financ Econ* 45:569–587
- Erickson T, Jiang CH, Whited TM (2014) Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *J Econ* 183:211–221
- European Union, Agency for Fundamental Rights (FRA) (2010) Protecting fundamental rights during the economic crisis, working paper (December). Available at: [https://fra.europa.eu/sites/default/files/fra\\_uploads/1423-FRA-Working-paper-FR-during-crisis-Dec10\\_EN.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/1423-FRA-Working-paper-FR-during-crisis-Dec10_EN.pdf)
- Faber JW (2013) Racial dynamics of subprime mortgage lending at the peak. *Housing Policy Debate* 23:328–349
- Ferguson M, Peters S (1995) What constitutes evidence of discrimination in mortgage lending? *J Financ* 50:739–748
- Financial Crimes Enforcement Network (2006) Mortgage loan fraud: An industry assessment based upon suspicious activity report analysis. Available at: [http://www.fincen.gov/news\\_room/tp/reports/pdf/MortgageLoanFraud.pdf](http://www.fincen.gov/news_room/tp/reports/pdf/MortgageLoanFraud.pdf)
- Ghent AC, Hernández-Murillo R, Owyang MT (2014) Differences in subprime loan pricing across races and neighborhoods. *Reg Sci Urban Econ* 48:199–215
- Harrison GW (1998) Mortgage lending in Boston: a reconsideration of the evidence. *Econ Inq* 36:29–38
- Haughwout A, Mayer C, Tracy J, Jaffee DM, Piskorski T (2009) Subprime mortgage pricing: the impact of race, ethnicity, and gender on the cost of borrowing. *Brookings-Wharton Papers on Urban Affairs*, 33–63
- Holmes A, Horvitz P (1994) Mortgage redlining: Race, risk, and demand. *J Financ* 49:81–99
- Horne DK (1997) Mortgage lending, race and model specification. *J Financ Serv Res* 11:43–68
- Hoynes H, Miller DL, Schaller J (2012) Who suffers during recessions? *J Econ Perspect* 26:27–47
- Hubbard RG, Palia D, Yu W (2012) Analysis of discrimination in prime and subprime mortgage markets. Available at SSRN: <http://ssrn.com/abstract=1975789>
- International Labour Office (2011) Equality at work: the continuing challenge. Global report under the follow-up to the ILO declaration on fundamental principles and rights at work. International labour conference, 100th session, Geneva
- LaCour-Little M (1999) Discrimination in mortgage lending: a critical review of the literature. *J Real Estate Lit* 7:15–50
- Ladd H (1998) Evidence of discrimination in mortgage lending. *J Econ Perspect* 12:41–62
- Mian AR, Sufi A (2015) Fraudulent Income Overstatement on Mortgage Applications during the Credit Expansion of 2002 to 2005. NBER Working Paper No. 20947
- Munnell AH, Tootell GM, Browne LE, McEneaney J (1996) Mortgage lending in Boston: Interpreting HMDA data. *Am Econ Rev* 86:25–53
- National Association for the Advancement of Colored People (NAACP) (2007) The NAACP filed a historic lawsuit against mortgage lenders alleging racial discrimination. Available at: <http://www.naACP.org>
- Ondrich J, Ross S, Yinger J (2003) Now you see it, now you don't: why do real estate agents withhold available houses from black customers? *Rev Econ Stat* 85:854–873
- Pager D, Shepherd H (2008) The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annu Rev Sociol* 34:181–209
- Pew Research Centre (2011) Wealth gaps rise to record highs between Whites, Blacks and Hispanics. Social and Demographic Trends. Available at: [http://www.pewsocialtrends.org/files/2011/07/SDT-Wealth-Report\\_7-26-11\\_FINAL.pdf](http://www.pewsocialtrends.org/files/2011/07/SDT-Wealth-Report_7-26-11_FINAL.pdf)
- Piskorski T, Seru A, Witkin J (2015) Asset quality misrepresentation by financial intermediaries: evidence from the RMBS market. *J Financ* 70:2635–2678
- Quillian L (2006) New approaches to understanding racial prejudice and discrimination. *Annu Rev Sociol* 32:299–328
- Reskin B (2012) The race discrimination system. *Annu Rev Sociol* 38:17–35
- Ross SL (2005) The continuing practice and impact of discrimination. University of Connecticut. Department of Economics Working Paper Series 19

- Ross SL, Yinger J (2002) *The color of credit: mortgage lending discrimination, research methodology and fair lending enforcement*. MIT Press, Cambridge
- Ross SL, Yinger J (2006) Uncovering discrimination: a comparison of the methods used by scholars and civil rights enforcement officials. *American Law and Economics Review* 8:562–614
- Rugh JS, Massey DS (2010) Racial segregation and the American foreclosure crisis. *Am Sociol Rev* 75:629–651
- Sen M (2012) Quantifying discrimination: Exploring the role of race and gender and the awarding of subprime mortgage loans. Available at SSRN: <http://ssrn.com/abstract=1593183>
- Turner MA, Skidmore F (1999) *Mortgage lending discrimination: a review of existing evidence*. Urban Institute Press, Washington, DC
- United Nations Educational, Scientific and cultural organization (UNESCO) (2009). The impact of the economic crisis on discrimination and xenophobia. Global Migration Group (GMG). Available at: [www.globalmigrationgroup.org/sites/default/files/uploads/documents/UNESCO\\_Fact-sheet\\_final.pdf](http://www.globalmigrationgroup.org/sites/default/files/uploads/documents/UNESCO_Fact-sheet_final.pdf)
- Wheeler CH, Olson LM (2015) Racial differences in mortgage denials over the housing cycle: Evidence from US metropolitan areas. *J Hous Econ* 30:33–49
- Williams R, Nesiba R, McConnell ED (2005) The changing face of inequality in home mortgage lending. *Soc Probl* 52:181–208
- Wooldridge JM (2009) *Introductory econometrics: a modern approach*. South-Western Cengage Learning, Mason
- Yinger J (1986) Measuring racial discrimination with fair housing audits: caught in the act. *Am Econ Rev* 76: 881–893