Consumer-Lending Discrimination in the Era of FinTech*

Robert Bartlett
School of Law
UC Berkeley
rbartlett@berkeley.edu

Adair Morse
Haas School of Business
UC Berkeley & NBER
adair@berkeley.edu

Richard Stanton
Haas School of Business
UC Berkeley
rhstanton@berkeley.edu

Nancy Wallace
Haas School of Business
UC Berkeley
newallace@berkeley.edu

November 2018

Abstract

Ethnic discrimination in lending can occur in face-to-face decisions or in algorithmic scoring. The GSEs' model for pricing credit risk provides us with an identified setting to estimate discrimination for FinTech and face-to-face lenders, as well as to offer a workable enforcement interpretation of U.S. fair -lending laws using the court's justification of legitimate business necessity. We find that face-to-face and FinTech lenders charge Latinx/African-American borrowers 6-9 basis points higher interest rates, consistent with the extraction of monopoly rents in weaker competitive environments and from profiling borrowers on shopping behavior. In aggregate, Latinx/African-American pay $250-$500M per year in extra mortgage interest. FinTech algorithms have not removed discrimination, but two silver linings emerge. Algorithmic lending seems to have increased competition or encouraged more shopping with the ease of applications. Also, while face-to-face lenders discriminate against minorities in application rejection, FinTechs do not.

JEL classification: G21, G28, G23, J14, K22, K23, R30
Keywords: FinTech; mortgages; credit scoring; algorithmic underwriting; big-data lending; platform loans; disparate impact

* Financial support from the Fisher Center for Real Estate and Urban Economics is gratefully acknowledged. For helpful comments and suggestions, we thank Sanjiv Das, Andreas Fuster, Andres Liberman, Raghu Rau, Amit Seru, Justin Wolfers, and seminar participants at U.C. Berkeley, Cornell, the University of Michigan, USC, NYU Stern, Texas A&M, the Georgetown-Federal Reserve Board Seminar Series on Household Finance, the Spring 2018 NBER Corporate Finance meeting, the 2018 NBER Law & Economics Summer Institute, the 2018 Annual Conference on Empirical Legal Studies, the 2017 Boston JOIM Conference, the 2017 Annual Research Conference of the Center for International Securities and Derivatives Markets, the 2018 NY Fed./NYU Stern Conference on Financial Intermediation, the 2018 American Law and Economics Association (ALEA) annual meeting, the 2018 Financial Intermediation Research Society (FIRS) annual meeting, and the 2018 EFA meeting.

I. Introduction

Minority households hold 17.3% of the $13 trillion (or $2.25 trillion) in U.S. household debt as of 2017.[1] Any discrimination with respect to payments on such a large balance is thus likely to be material, affecting both minority household well-being and lender profits. For example, consider home loans, which are the largest component of consumer debt, representing approximately $9.1 trillion of the total. Each extra basis point in mortgage interest charged due to discrimination costs minority mortgage holders approximately $100 million per year.[2] Likewise, any discrimination with respect to the accept/reject decision on loan applications is also likely to be material. In 2017 census data, homeownership by white ethnicities was 72.4%, but only 48.4% and 43.0% for Latinx and African-American households, respectively. If any of this disparity is due to discrimination, it would materially bias the well-being of minority households in the United States.

In this paper, we estimate the level of discrimination in the largest consumer-loan market, conventional conforming mortgages securitized by the Government Sponsored Enterprises (GSEs), Fannie Mae and Freddie Mac. Using GSE loans allows a novel identification strategy. The GSEs charge each loan a guarantee fee (which gets added into the offer interest rate) that depends only on where in a grid of credit score and loan-to-value (LTV) ratios (both observable) the borrower lies. In return, lenders are guaranteed against credit risk. Thus, mortgage interest rate differences between loans within a given GSE grid cell of credit score and LTV cannot reflect differential credit risk, but must instead reflect strategic pricing decisions on the part of lenders. Strategic pricing is not illegal by any means, but, under the law, it cannot induce disparate impact.[3] Lenders cannot, even inadvertently, charge a higher mark-up to protected ethnic groups. Using this novel identification, we find that discrimination is 6-9 basis points (bps) in purchase-mortgage markets and 1-3 bps in refinance mortgages. Averaging across the distribution of these products in the U.S., lending discrimination currently costs African-American and Latinx borrowers $250-$500 million in extra interest per year.

Consumer lending in the United States is changing rapidly, with loan origination becoming almost exclusively algorithmic. A case in point is the Rocket Mortgage of the platform lender Quicken, which is the largest-volume mortgage product in the U.S. as of 2018. Algorithmic loan origination is not, however, just a feature of FinTech companies. We study the 2,098 largest mortgage lenders (inclusive of all the big banks) over the 2012-2018 period, finding that as of 2018, 45% of them offer complete online or app-based

---

[1] Percent-of-debt estimates are from the 2016 Survey of Consumer Finances.
[2] The calculation assumes at origination the mean mortgage is $175,000, paying 4.0 % on a 30 year fixed contract.
[3] A disparate impact occurs under U.S. anti-discrimination law when a decision maker's practices do not expressly discriminate on a protected characteristic (e.g., race or ethnicity), but nevertheless disadvantage one or more groups without a legitimate business justification. Under disparate-impact theory, proof of discriminatory intent is not required to establish liability. As noted below, the U.S. Supreme Court has held that the disparate impact theory of liability extends to the Fair Housing Act.

mortgage contracting. Value-weighted, that percentage is much higher. Indeed, nearly all of the big banks and most small lenders now act as FinTech. This is not simply a mortgage story; one only has to look to the emergence of personal lending platforms, such as Capital One and Goldman Sachs Marcus's site, to see the broader transformation in consumer lending.

With algorithmic credit scoring, the nature of discrimination changes from being primarily concerned with human biases – racism and in-group/out-group bias – to being primarily concerned with illegitimate applications of statistical discrimination. Even if agents performing statistical discrimination have no animus against minority groups, they can induce disparate impact by their use of Big Data variables. Whether these changes induce more or less discrimination was previously unknown. We find that FinTechs and face-to-face lenders are equally discriminatory in extracting rents from minorities. It is perhaps disappointing that discrimination remains even though face-to-face interactions are removed. However, the story has two silver linings for the role of FinTechs and algorithmic decision-making. First, in the loan accept/reject decision (as opposed to pricing), FinTechs seem to do the opposite of discrimination, catering to those discriminated against by face-to-face borrowers. Second, our evidence on time patterns is consistent with discrimination declining for all lenders from 2008 to 2015 due to the advent of FinTech startups and the speed of being able to shop around at algorithm-based lenders of all sorts.

At the core of our paper is the importance – in our identification and in the legal setting – of statistical discrimination in the new era of algorithmic loan decision-making. Regulators and courts face heightened hurdles to identify which Big Data variables can give rise to a successful claim of illegal discrimination under U.S. fair lending laws (see *Inclusive Communities*; 2015).[4] For economists, the courts' struggle to untangle legitimate from illegitimate statistical discrimination is the same problem as handling omitted variables in estimating discrimination. Statistical discrimination arises as a solution to a signal extraction problem. The signal extraction setting in consumer lending emerges as follows. Economists can write down a macro-founded (life-cycle) model of repayment risk that applies to everyone.[5] The problem is that some variables in this macro-fundamental model are not observable. The goal of statistical discrimination is to reconstruct this hidden fundamental information using observable proxies.

In the law, lenders can use proxy variables that produce a disparate impact on minority applicants but only if the lender can show that these variables have a *legitimate business necessity*. According to the courts, legitimate business necessity is the act of scoring credit risk. Furthermore, according to the courts, efforts to use proxy variables that produce a disparate impact for other purposes, including lenders' earning

[4] The potential for illegitimate statistical discrimination toward protected classes of borrowers was a key aspect of Congressman Emanuel Cleaver's 2017 investigation into FinTech lending.

[5] Behavioral models may correctly profile individuals on average, but some individuals would be incorrectly profiled, which could be deemed discrimination by disparate impact under the law.

of higher profit margins, do not meet this definition. In business terms, any strategic pricing that causes disparate impact, even inadvertently, is discrimination in the eyes of the law.

Lenders have used the legitimate-business-necessity defense to argue that any variable that is correlated with default is acceptable. This definition of legitimate business necessity is necessary but not sufficient to comply with the court rulings. An example is illustrative. Surely, the high school that a person attended is an empirically relevant proxy for hidden wealth, where wealth is the endowment variable in a macro-fundamental model of repayment risk. High school, however, may be correlated with ethnicity even after orthogonalizing with respect to wealth. If so, using high school would punish, or have disparate impact on, some minority households.

Our economic mapping of these court rulings on disparate impact to legitimacy in statistical discrimination yields three punchlines: (a) Scoring or pricing loans explicitly on credit-risk macro-fundamental variables is legitimate. (b) Scoring or pricing on a Big Data variable that only correlates with ethnicity through hidden fundamental variables is legitimate. (c) Scoring or pricing on a Big Data variable that has residual correlation with ethnicity after orthogonalizing with respect to hidden fundamental credit risk variables is illegitimate.

For policymakers, these punchlines suggest that regulators might take an approach of mandating that lenders provide proof of legitimacy of Big Data proxy variables using their proprietary data on otherwise-hidden fundamental variables such as wealth. This arrangement would be akin to putting the burden of value-at-risk modeling on banks, as is done in banking regulation. We discuss this more in the conclusion.

For researchers, these punchlines imply that in the age of Big Data, econometricians require a setting in which all legitimate business necessity variables are observable in order to identify discrimination without concern for omitted-variable bias. We have been able to find just such a setting, covering a large fraction of consumer lending yet free from omitted-variable concerns. We use this setting to document the extent to which discrimination is happening in the largest consumer-loan market and to illustrate how algorithmic pricing of loans may yet result in discrimination.

It is well known that, post-crisis, the GSEs (Fannie Mae and Freddie Mac) purchase, securitize, and guarantee more than 90% of the conventional conforming mortgage market in the U.S. It is less recognized that, post-crisis, the GSE actions fully determine the price of credit risk by their role as guarantors. In particular, the GSEs produce a predetermined grid pricing that prices credit risk across loan-to-value and credit score buckets. The pricing grid need not be the optimal model for predicting default among all application variables,[6] but is nevertheless the price lenders must pay the GSE to absorb risk for the MBS market. Thus, any deviation from this grid pricing reflects lenders' competitive agenda in capturing volume or profit per mortgage. Because these non-credit risk agenda are unrelated to creditworthiness, they fail to satisfy the legitimate business necessity as determined by the courts. Thus, within the grid, any additional correlation of loan pricing with ethnicity is discrimination.

Our analysis uses a data set that includes never-before-linked information at the loan-level on income, ethnicity, loan-to-value ratios, debt-to-income ratios, all contract terms (such as coupon, loan amount, installment payment structure, amortization, maturity, loan purpose, and mortgage-origination month), and indicators for whether the lender-of-record primarily used algorithmic scoring.

We find that accepted borrowers in Latinx and African-American ethnic groups pay 16.9 basis points higher interest for home purchase mortgages. Of this, 8.3 basis points are due to differences in risk in the grid and time variation (together accounting for 3/4 of the variation in pricing), leaving 8.6 bps of discrimination. Although courts have explicitly held that credit risk is the only legitimate business necessity, as economists, we believe the spirit of these decisions may perhaps include room for lenders to differentiate loan pricing based on the fixed cost of providing a loan, by lender or by geography. We thus additionally include county and lender fixed effects. Overall, we find discrimination in these specifications of 5.6 to 8.6 bps for purchase mortgages, or approximately 11-17% of lenders' average profit per loan.[7] Interest rate discrimination is almost identical for FinTech lenders (5.3 basis points extra paid by minorities) as for the overall set of lenders (5.6 basis points). In the paper text, we address the robustness of this result to other concerns of servicing rights and to the use of points.

How discrimination happens is an important question. Although, we leave a full exploration to a separate research project, we can fix ideas. Lenders may be able to extract monopoly rents from minority borrowers because such borrowers might be prone to less shopping on average. In refinance loans, we find minorities pay just 1-3 bps in higher interest rates. This result is consistent with an interpretation that monopoly price extraction of rents is easier in transactions where the borrowers have less experience or are acting in a more urgent time frame. Additionally, because lenders may price loans to capture rents in less-competitive areas, prices might be higher in financial services deserts, which might have higher ethnic minority populations.

[6] The actuarially fair GSE guarantee fee (or G-fee) is also a central policy question in the determination of the future role of the GSEs in the U.S. mortgage markets (see Elenev, Landvoigt, and Van Nieuwerburgh, 2016; Vickery and Wright, 2013.) A standard G-fee is assessed on all mortgages as a percentage of the loan balance and is collected monthly (see Fuster, Goodman, Lucca, Madar, Molloy, and Willen, 2013).

[7] According to the Mortgage Bankers' Association, the average mortgage profit is 50 basis points (see https://www.mba.org/x73719).

In our setting of the GSE guarantee, if lenders were to discriminate in the accept/reject decision, it would imply that money is left on the table. Logic suggests that such unprofitable discrimination must reflect a human bias by loan officers. This is what we find. Face-to-face lenders reject Latinx and African-American ethnicities 4% more often. FinTech lenders, on the other hand, do not discriminate at all. If anything, they seem to have negative discrimination, catering to those discriminated by face-to-face lenders.

Our paper contributes to a small but growing literature on discrimination in lending. A large literature in labor contributes to the topic of wage discrimination, but even there, our commentary on how courts and regulators can consider Big Data use may be informative. The lending discrimination literature has lagged the wage literature primarily because of the lack of data on ethnicity or race combined with an identification strategy that handles omitted variables in scoring.

Early studies looking at the raw HMDA data found that minority loan applicants were rejected much more often than white applicants even with higher incomes, but did not control for variables not collected by HMDA, such as credit history. In a widely cited paper, Munnell, Browne, McEneaney, and Tootel (1996) combined HMDA data on loan applications in Boston in 1990 with additional borrower data collected via survey by the Federal Reserve Bank of Boston, and found that after controlling for borrower characteristics, especially credit history and loan-to-value ratio, white applicants with the same property and personal characteristics as minorities would have experienced a rejection rate of 20% compared with the minority rate of 28%.

Much of the more recent literature focuses on the pre-crisis period, usually looking at subprime lending. Ghent, Hernandez-Murillo, and Owyang (2014) examine subprime loans originated in 2005, and find that for 30-year, adjustable-rate mortgages, African-American and Latinx borrowers face interest rates 12 and 29 basis points, respectively, higher than other borrowers. Bayer, Ferreira, and Ross (2018) find that after conditioning on credit characteristics, African American and Hispanic borrowers were 103% and 78% more likely than other borrowers to be in a high-cost mortgage between 2004 and 2007. Similar results were obtained by Reid, Bocian, Li and Quercia (2017).

Cheng, Lin and Liu (2015) use data from the Survey of Consumer Finances to compare mortgage interest rates for minority and non-minority borrowers. They find that black borrowers on average pay about 29 basis points more than comparable white borrowers, with the difference larger for young borrowers with low education, subprime borrowers, and women.

Focusing on the *quality* of consumer credit services, Begley and Purnanandam (2018) study the incidence of consumer complaints about financial institutions to the CFPB. They find that the level of complaints is significantly higher in markets with lower income and educational attainment, and especially in areas with a higher share of minorities, even after controlling for income and education.

In one of the few experimental papers in this area, Hanson, Hawley, Martin, and Liu (2016) show that when potential borrowers (differing only in their name) ask for information about mortgages, loan officers are more likely to respond, and give more information, to white borrowers.

Finally, Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2018) show that the use of machine-learning techniques to evaluate credit quality may result in differential impact on loan provision to minority versus non-minority borrowers. This paper conveys important knowledge in how algorithms are utilized in mortgage markets.

There are also related results from other consumer debt markets. For example, Dobbie, Liberman, Paravisini and Pathania (2018) look at data from a high-cost lender in the UK and find significant bias against immigrant and older loan applicants when measured using long-run profits. However, they find no bias when using the (short-run) measure actually used to evaluate loan examiners, suggesting that the bias is due primarily to a misalignment of firm and examiner incentives.

The rest of the paper is organized as follows. In Section II we discuss our multifaceted data. We present our methodology for the measurement of mortgage discrimination in Section III, and provide statistics showing the role of the GSE pricing grid in practice. Our empirical results are reported in Section IV. Section V concludes and offers policy perspectives on moving forward from our findings. We also include Appendix A, where we discuss anti-discrimination regulations in U.S. mortgage lending.

II. Data & Statistics

A key obstacle in prior empirical mortgage discrimination studies has been their reliance on the Home Mortgage Disclosure Act (HMDA) data. The HMDA compliance surveys cover 90% of mortgage origination in the U.S. (see Engel and McCoy, 2011),[8] and are the only data source with loan-level information on applicant ethnicity for both successful and unsuccessful loan applications.[9] HMDA also contains applicant income and (nonstandardized) information on lender name. What HMDA lacks is information on the contracting structure of the loan (exact date, interest rate, maturity, loan-to-value ratio), on the type of loan (fixed, ARM), on the property characteristics (e.g., address), and on the applicant's credit data commonly used by the GSEs and other lenders (credit score, debt-to-income ratio, etc.).

A challenge with mortgage loan data in the U.S. has been the lack of unique loan identification number and thus the lack of a direct way to link the HMDA data and other datasets containing these missing

[8] HMDA reporting is not required for institutions with assets (of the entity and its parent corporation) that are below $10 million on the preceding December 31 (see http://www.ffiec.gov/hmda/pdf/2010guide.pdf).

[9] HMDA has missing values on ethnicity (Buchak et al., 2017). We first fill in these missing values using the ethnic-name-categorization algorithm from Kerr and Lincoln (2010) and Kerr (2008). Then we check for consistency of our results including and excluding these fixes.

data. We ameliorate this deficiency with a multi-year project of linking loan-level data across the following data providers:

- HMDA (54.8 million single family residential loans accepted and 53.6 million single family loan rejections between 2008 and 2015). HMDA data include information on borrower income, ethnicity, and lender, as well as geography of the property only at the census tract.
- ATTOM (77.2 million single family loans between 2008 through 2015). ATTOM data provide transaction and assessor information including loan performance data (i.e., prepayment and default), lender names and exact location, but very little information on mortgage contract terms other than the loan amount.
- McDash (24.2 million single family residential loans between 2008 and 2015). McDash data provide loan-level data compiled by Black Night Financial Services and includes quite comprehensive information on the mortgage terms (including interest rates and the zip code of the mortgaged property).
- Equifax (23.7 million single family residential loans that merged with McDash between 2008 and 2015). Equifax data provide information on other consumer financing balances that are held by borrowers in addition to their mortgages.

Using a machine-learning protocol, we exploit overlapping variables within these four datasets to construct a merged data set of accepted loans with performance information, contract terms, and borrower information. We describe our machine-learning merging algorithm in detail in Appendix B. A key component of the merging was the McDash-ATTOM link, which we accomplished by matching performance strings; i.e., matching loans on the flow of events reported in the property registers. The Equifax-McDash merge was done by Equifax in compliance with IRB standards, and our residual data are anonymized.

An admitted weakness in our data is in the rejection data. Although the HMDA data includes rejection/accept decisions (with ethnicity, income and loan amount) so that we can perform tests of accept/reject discrimination, we must be more cautious in interpreting the accept/reject results (as opposed to the pricing results) because none of these auxiliary datasets covers loan application rejections. Thus, while we know the precise (observable) variables that the GSEs use to determine loan acceptance/rejection in their automated underwriting systems, we must augment the rejection data with proxies for credit score, debt, and loan-to-value at the census tract level. A census tract is on average 1600 households (4000 inhabitants), designed by the Census Bureau to reflect relatively uniform economic situations. We describe the construction of these proxies in Appendix B.

To standardize our analysis, we filter the data to focus on 30-year fixed rate, single-family residential loans, securitized by the GSEs over the period 2008 through 2015. We additionally eliminate from our sample any loans made within a zip code covered by the Community Reinvestment Act of 1977 (CRA), given the potential bias these zip codes would introduce into our empirical analysis.[10] The final sample consists of 5,070,900 accepted loan applications and 3,690,313 rejected applications with application dates between 2008 and 2015.

We present summary statistics in Table 1. Among rejected and accepted loans, the median loan amount was just over $100,000, with accepted loans showing evidence of positive skew through the maximum loan size ($2.4 million) and the slightly higher mean loan amount. The mean and median interest rate on loans within our sample was approximately 4.5%. Not surprisingly, compared to the average accepted loan applicant, the average rejected loan applicant generally had lower income, a slightly lower credit score, and a slightly higher LTV. The median purpose for the loan application among accepted and rejected loans was to complete a refinancing, as opposed to a home purchase.

Table 1 also reveals summary information concerning the types of lending institutions that received the loans applications in our sample. Using the list of firms identified as FinTech in Buchak et al. (2017), we find that FinTech lenders originated approximately 4.2% of accepted loans and were responsible for 5.4% of all loan rejections in our sample. As noted previously, however, many conventional lenders now regularly utilize online application forms, and we separately identified the 2,098 largest mortgage lenders over the 2012-2018 period who made available an online application form as of the summer of 2018. Loans originated by these firms constituted over 45% of our accepted loans, and they rejected over 50% of the rejected loans in our sample. Table 1 also highlights the dominance of the largest originators in the mortgage lending industry: the top 25 originators (by origination volume in their respective loan origination year) originated over 50% of all accepted loans and processed nearly 50% of all of our rejected loan applications.[11]

Finally, among accepted loans in our sample, Table 1 shows that 13.2% were made to African-American or Hispanic borrowers, while 20.8% of rejected loan applicants were African-American or Hispanic.

III. Method: Using the GSE Lending Process

When an individual applies for a conforming mortgage (where *conforming* means the loan size falls beneath a federally-set conforming loan limit), the GSE process begins. The lender feeds application observables (the credit score, income, liquid reserves, debt-to-income ratio, loan-to-value ratio, property value, etc.), into the GSE 'black box', an automated underwriter system (Desktop Underwriter for Fannie

[10] Under the CRA Act, financial institutions are required to provide a certain level of lending to CRA districts to counter the lack of financial services in lower-income districts.

[11] We create a variable of the top 25 mortgage originators per year by matching HMDA lender names with mortgage origination statistics obtained from Inside Mortgage Finance.

Mae; Loan Prospector for Freddie Mac). The GSE black box produces the accept/reject decision based on a specified set of observables contained in the application. If the GSE accepts the loan, and the lender and borrower issue the loan, the lender sells the mortgage to the GSE. In return, the GSE compensates the lender with a cash transfer.[12] The GSE then packages the loan with a pool of similar mortgages into a mortgage-backed security (MBS), issues a default-risk guarantee on this product, and sells it to the MBS market.

For 30 days, the lender holds some credit risk, but after this time period, the lender is not exposed to any prepayment or default risk. The only risk that the lender faces is put-back risk. Put-backs can occur when the documentation on income (the IRS form), credit score (the credit history pull), loan purpose (residential vs. non-occupancy) or property value (the appraisal) is falsified or missing. After the 2008 financial crisis and the half decade thereafter, because of put-backs[13] and large fines for misrepresentation, lenders ceased no-documentation GSE loans and adjusted their policies to remove the potential for falsified documentation. The magnitudes of put-backs on post-2010 originations have become a trickle compared to the early 2000 issuances.[14] Figure 1 plots put-backs over the time horizon of the loan for issuances 2000-2010. This figure is taken from an Urban Institute publication by Goodman and Zhu (2013). The figure supports our assumption that put-back risk for post-2008 issuances had diminished as well as the lack of importance of the 30-day exposure of lenders to credit risk.

III.a. GSE Process in Pricing and Identification of Rate Discrimination

Within this GSE process, the lender has three decisions to make. The first is pricing. The mortgage interest rate that a household sees consists of three parts (also see Fuster, et al (2013)). First, all mortgages face the same market price of capital, determined by the Base Mortgage Rate, which reflects the primary market interest rate for loans to be securitized by the GSEs. In effect, the Base Mortgage Rate reflects the compensation demanded by investors in the MBS market, or the credit risk-free rate.

Second, when a GSE buys a mortgage from a lender, the GSE takes a guarantee fee (or g-fee) to cover projected borrower default and operational costs. Starting in March 2008 and adjusted a handful of times since then, this g-fee varies in an 8×9 matrix of LTVs and credit scores to reflect varying credit risk across the GSE grid. Figure 2 depicts a typical GSE grid of Fannie Mae, also called the Loan Level Price

---

[12] If the originator is a large volume lender, the lender will transfer loans to the GSE in bulk and, instead of receiving cash for the mortgages, the originator receives back an MBS with a pool of similar-characteristic mortgages produced by that lender. (Sometimes these MBS products have mortgages originated by other lenders to fill out the MBS, but one should think of this pool as primarily being the lender's own issuances.) These MBS products are equally guaranteed by the GSE, but because the lender issuer also retains servicing rights, the lender may be exposed to the extra servicing costs (i.e., additional phone calls and outreach) that happen when loans become delinquent. For this reason, we show all of our results with and without the large volume lenders.

[13] The GSEs put back \$4.2 billion of pre-crisis loans in 2010 alone (American Banker, July 14, 2016).

[14] A large number of put-backs by a lender will induce the GSEs and the MBS market to add a discount to the price for products from this lender. These overall lender adjustments can be controlled for with firm fixed effects.

Adjustments (LLPAs) (see FHFA 2000; 2010; 2011; 2012; 2013 as well as Fuster and Willen, 2010). In practice these one-time fees are commonly converted into monthly "flow" payments, which are added into the interest rate as rate pass-throughs to borrowers. In quoting rates to customers, originators utilize rate sheets that expressly incorporate both the Base Mortgage Rate (generally reflected as the "par rate") as well as LLPA adjustments (colloquially referred to as "hits" to the par rate).

The third component of pricing comes from lenders discretion in quoting rates in excess of the par rate and any LLPA "hits" and rate quoting to incorporate strategic volume positioning and monopoly rent-taking. Lenders might use monopoly-like pricing based on the competition environment of a location (e.g., in areas of collusion and in financial desert environments) or as a rent extraction strategy against borrowers shop around less.

In Figure 3, we graphically depict the importance of the LLPA "grid" for purposes of pricing loans. In panels (a) and (b) we display histograms of borrower interest rates among approved loans by treatment (treat=1 is the set of Latinx and African-American borrowers). Panel (a) shows a histogram of the raw data, revealing a wide distribution of rates for both control and treated loans, as one might expect given the length of our sample period and the large number of loans in the sample. However, when we level interest rates within the grad by subtracting out the year-grid cell mean, Panel (b) shows a dramatic reduction in the distribution of interest rates for both groups of borrowers, highlighting the central role of the LLPA grid in determining interest rates for GSE mortgages. Notably, visual inspection of Panel (b) shows greater mass in the higher rates for the ethnic treatment histogram, suggesting that (even controlling for GSE pricing grid) minority borrowers are paying higher interests.

Much like this visual exercise, our identification relies on the observation that within a cell of the GSE grid, disparities in interest rates for similar grid borrowers reflect markup strategies arising from discretionary pricing, not omitted variables in credit-risk scoring. There is no disparity in fair pricing of credit risk (which equals the market rate plus any LLPA adjustments) inside the grid because the price is dictated by the GSE.[15] In particular, for application *i* occurring in the month-year *t*, we estimate:

$$interest\ rate_i = \alpha\ LatinxAfricanAmerican_i + \mu_{GSEgrid} + \mu_{month\_year} + \varepsilon_{it}, \tag{1}$$

[15] Fundamentally, the GSEs have elected to price credit based on two observable variables (LTV and credit score) that proxy for lifecycle variables in a macro-fundamental model of repayment risk. Under U.S. fair lending law, any discrimination in loan pricing that arises from this parsimonious pricing model should be justifiable as a legitimate business necessity given that these variables proxy for repayment risk. For the same reason, however, any disparate impact in loan pricing within the GSE grid is unexplainable by this model and represents illegitimate discrimination that is unrelated to the life cycle variables the GSEs have chosen to utilize in pricing credit risk.

where the mortgage interest rate is regressed on the indicator for the applicant being a Latinx or African American, GSE grid fixed effects $\mu_{GSEgrid}$ and month-year fixed effects μ_{month_year}. Because the grid fully prices credit risk, and the month-year fixed effects pick up changes in the risk-free rate, any correlation of prices with ethnicity inside the GSE grid is thus discrimination.

Notably, our identification strategy permits us to identify not only the incidence of discrimination in loan pricing but also the legal consequence of this discrimination under U.S. fair-lending laws. This stems from the fact that, as discussed in Appendix A, courts have consistently limited the legitimate business necessity defense to a lender's use of structural variables and loan practices that seek to ascertain creditworthiness.[16] Evidence of within-grid discrimination would thus lack a legitimate business necessity defense given the GSEs' decision to price credit with respect to just two observable variables pertaining to creditworthiness. That said, the interpretation of U.S. fair lending laws continues to evolve, and we would find it unsurprising were a court to permit at some point a lender to advance as a legitimate business necessity defense the ability to recoup differential overhead costs for originating loans in different geographic locations (which may correlate with applicants' ethnicity). Thus, we also utilize an alternative specification to account for this possibility which includes lender fixed effects, μ_{lender}, and county fixed effects, μ_{county}:

$$interest\ rate_i = \alpha\ LatinxAfricanAmerican_i + \mu_{GSEgrid} + \mu_{lender} + \mu_{county} + \mu_{month_{year}} + \varepsilon_{it}, \qquad (2)$$

This specification forces the appealing interpretation on the discrimination coefficient to be the differences in average interest rate charged to a minority applicant "as compared to that offered to an ethnic majority applicant by the same lender" and "as compared to rates offered to others in that same county". Although appealing econometrically, this rigor throws out some of the variation in which we are interested. Thus, we interpret our results in a range from equation (1) to equation (2).

III.b. GSE Process in Accept/Reject Decisions and Identification of Rejection Discrimination

---

[16] See A.B. & S. Auto Service, Inc. v. South Shore Bank of Chicago, 962 F. Supp. 1056 (N.D. Ill. 1997) ("[In a disparate impact claim under the ECOA], once the plaintiff has made the prima facie case, the defendant-lender must demonstrate that any policy, procedure, or practice has a manifest relationship to the creditworthiness of the applicant…"). See also Lewis v. ACB Business Services, Inc., 135 F.3d 389, 406 (6th Cir. 1998) ("The [ECOA] was only intended to prohibit credit determinations based on 'characteristics unrelated to creditworthiness.'"); Miller v. Countrywide Bank, NA, 571 F.Supp.2d 251, 258 (D. Mass 2008) (rejecting argument that discrimination in loan terms among African American and white borrowers was justified as the result of competitive "market forces," noting that prior courts had rejected the "market forces" argument insofar that it would allow the pricing of consumer loans to be "based on subjective criteria beyond creditworthiness.")

The second and third decisions the lender makes concern the accept/reject decision. Upon receiving an approval in the GSE underwriter system, the lender can still choose to reject the application. If there remains no credit risk post-transacting, money would be left on the table. Why would a lender choose to reject a GSE-accepted applicant? (i) The lender might feel that a particular borrower reflects additional put-back risk. As we have argued, such put-back risk is so small, especially in the latter half of our sample, that even if this put-back risk were residually correlated with ethnicity (which is not established), it would not be able to explain any material differences that we find in rejection rates. Thus, this argument would amount to a biased belief affecting loan decisions. (ii) The lender might be directly racist or have other in-group biases. (iii) The lender might prefer to cater to white ethnicities for their clientele. We have heard stories proposing that some community financial institutions prefer to keep a uniform customer base to cater to in-group biases in their clientele. None of these explanations fall under *legitimate business necessity*.

The third decision that a lender makes is an ex ante one. A loan officer might deter a potential borrower from applying, and this influence might be unequal across ethnicities. This argument has two sides, however. Loan officers might discourage legitimate borrowers because of biases and racism. Conversely, loan officers may inform their in-group preferred clientele of weaknesses in an application to assist the potential applicant. If we had a perfect set of both accepted and rejected applicants' data with all observable variable used in a GSE's black box, we could recreate the GSE's black-box algorithm and eliminate this possibility from biasing our results. Since our data and identification weakness concerns this point, we interpret our results with some caution.

Our estimation of rejection discrimination for application *i* in the year *y* is:

$$rejection_i = \alpha\ LatinxAfricanAmerican_i + f(HMDA\ income_i, HMDA\ loan\ amount_i) + \mu_{year} + \varepsilon_{it}. \tag{3}$$

Rejection is an indicator for an application being rejected by the GSE underwriter system. The $f(\cdot)$ function is a non-parametric function of the original HMDA data for income and loan amount. Since we do not know the exact scoring function of lenders on these variables, we implement splines (21 splines for income and 47 splines of loan amount) rather than a linear form to capture step functions, which we understand is the standard practice of lenders. We control for year fixed effects μ_{year}, rather than month fixed effects because HMDA does not provide precise dates for the rejections. This is our baseline estimation, with the complete set of data on rejections given by HMDA. However, we can do better by including a function of variables which the GSEs disclose as entering into the GSE black box underwriter system, proxied at the census tract (1600 households) level, where the census tract *c* refers to that of the property:

$$rejection_i = \alpha\ LatinxAfricanAmerican_i + f(HMDA\ income_i, HMDA\ loan\ amount_i) + g(LTV_c, credit\ score_c) + \mu_{lender} + \mu_{county} + \mu_{year} + \varepsilon_{it}. \quad (4)$$

To capture the distribution within the census tract, we include the 25th, 50th, and 75th percentile of LTV and *credit score* in the census tract in $g(\cdot)$. As in our rates estimations, we include, in some specifications, fixed effects for lender and county to control for cost differentials in providing financial services.

IV. Results

IV.a. Interest Rate Discrimination Estimates - Main

Table 2 presents within-GSE grid estimates of interest rate discrimination for 30-year fixed-rate mortgages that are approved and originated. Because lenders' pricing strategies vary by mortgage type, we use Panel A for purchase mortgages and Panel B for refinance mortgages. Models (1) and (2) are OLS regressions with varying fixed effects as noted. The column under model (3) reports an Oaxaca (1973) decomposition, showing the overall difference in rate means (left Oaxaca column), the covariate-explained part (middle Oaxaca column), and the unexplained part - i.e., discrimination - of the overall mean difference. It is easiest perhaps to start with a discussion of the Oaxaca decomposition.

The overall mean difference in the mortgage interest rate between white ethnicities and Latinx/ African-American ethnicities is 0.169%, 16.9 basis points. Of this amount, 8.31 basis points is explained by the monthly dummy variables and the GSE grid effects, leaving 8.57 basis points of discrimination. The Oaxaca decomposition is linear; thus the Oaxaca result is the same as the fixed effects specification in model (2). Of interest in model (2) is that over eighty percent of the variation across 1.2 million mortgages is explained by the GSE grid and time. (The R-squared in column 2 is 0.806.) Although this is a [very] large R-square, it should not be surprising given the g-fee + credit risk-free rate motivation for pricing the cost of mortgages.

The interpretation of these results is that conditional on being given a loan, African-American and Latinx borrowers pay an average of 8.57 basis points more than other borrowers for their purchase mortgage. The unexplained (one minus the r-square) twenty percent of the variation reflects strategic pricing on borrowers' location (perhaps due to collusion or financial deserts) or due to borrowers' behavioral characteristics (perhaps a lack of shopping). In the context of residential mortgage loans this pricing differential is economically meaningful for both borrowers and originators. For borrowers, imposing an additional 8.57 basis points per year in interest for a 30-year fixed rate amortizing loan has a present value of approximately 1 percentage point of the loan balance assuming a benchmark annual interest rate of 3.5%. Likewise for originators, sourcing mortgages with an 8 basis premium on the coupon would represent a material increase in profits over the Mortgage Bankers' Association mean profit of 50 basis points.

For refinance mortgages (Table 2, Panel B), the effects are smaller. We find 3 basis points of discrimination, after removing the time and GSE grid effects. The disparity in panels A and B suggests that borrower sophistication and hurriedness matter. Borrowers doing refi's are, by definition, experienced and are likely not in such a hurry to re-contract compared to the average purchase mortgage borrower, who may be time constrained to bid on a house on the market.

Table 3 introduces 4,900 lender fixed effects and 981 county fixed effects, building off the specification with GSE grid and month-year fixed effects. The point of this table is two-fold. First, from a court's perspective, we want to produce estimates that are immune from arguments that our discrimination estimates in Table 2 are a function of differing costs of providing loans, by lender and by geography. The courts have not ruled that *legitimate business necessity* includes locational or lender fixed costs, but we can imagine such an argument might arise. Second, from an econometrician's point of view, including lender allows an appealing within-lender comparison as discussed in the methodology. We find that lenders indeed discriminate within their organizations. Latinx and African-Americans pay 5.6 basis points higher rates for purchase mortgages and 1 basis point higher rates for refinance mortgages.

Although it is important for the robustness of our result to show that our results hold within lender and within geography, it is likely that we are throwing away some true discrimination in these estimations. Thus, we interpret our economic magnitude as a range from 8.6 to 5.6 bps for purchase mortgages and 3 to 1 bps for refinance mortgages.

To put these magnitudes in context, we can do a back-of-the-envelope calculation of interest paid. We need some aggregation assumptions. We assume the existing float of mortgages comprises 80% refinance loans and 20% purchase loans; the total U.S. mortgage market is $9.09 trillion;[17] and the share of African-Americans and Latinx borrowers is 13.20% (in our data). In our data, we can take the mean mortgage at the mean interest rate and amortize a loan with and without the extra discrimination pricing. We find that discrimination in mortgage interest rate costs African-Americans and Latinx from $250 to $500 million extra in interest annually.

IV.b. Interest Rate Discrimination Estimates – Robustness to Servicing and Points

Two additional concerns might bias our estimates – servicing costs and points. GSE loans are special in that many lenders do not retain servicing rights in the GSE process. In particular, large lenders

[17] https://www.nytimes.com/2017/05/17/business/dealbook/household-debt-united-states.html

are much more likely to be servicers of their own mortgages.[18] Our concern of servicing is very specific about residual exposure to default, because credit risk is that which the courts deemed to be *legitimate business necessity*. Although the GSE process removes credit risk directly, recall that we have also said the GSE credit risk model in the black box underwriting may not be perfect because it does not include life cycle fundamental variables, such as total wealth. Thus, if there is omitted variation in the GSE underwriting model, if a lender such as a large bank indeed observes wealth, and if a servicing contract exposed a lender to residual default risk in servicing costs (collections and human capital making phone calls), then a large lender might rationally implement a better pricing model using fundamental variables to compensates themselves for hidden servicing cost risk. The hidden wealth variable in this legal, fundamental model is hidden to the econometrician and may be correlated with ethnicity. Thus, our results could be due to such servicing cost differentials.

This is not the case. Appendix Table 1 implements the same specifications of Table 3, but limiting the sample to the non-top 25 volume lenders for every year. The results in Appendix Table 1 are not materially different from those in Table 3. The magnitudes are somewhat smaller for purchase mortgages in the lender and county fixed effects specification and somewhat larger for refinance mortgages, balancing out to the same discrimination in the overall economy.

The second robustness issue concerns points (paying a lump sum to the lender to reduce the interest rate). We are concerned with both positive points and negative points (sometimes called rebates or yield spread premium). In the case of positive points, an interpretation consistent with our Table 2-3 results is that ethnic minorities are more fully utilizing their cash for downpayment, leaving no cash to pay positive points to reduce the rates. On average, white ethnic groups may be using points more, implying lower interest rates.

We test for this in Table 4, columns (1) to (3), which again reproduce the robust specifications from Table 3. In particular, we limit the sample to borrowers precisely at the 0.80 LTV threshold. The idea is that on average, this set is likely to contain borrowers who are scraping-up funds to just make the downpayment required to meet an LTV of 0.80, which is the LTV threshold at which borrowers are exempt from purchasing mortgage loan insurance. Figure 3 illustrates how important the LTV threshold is. The penalty for going over an LTV of 0.80 in our time period is a higher rate and a need to obtain mortgage insurance. We show that discrimination is higher for these borrowers, inconsistent with a positive points story driving our results.

Another points story consistent with our results is that Latinx and African-American borrowers may be paying negative points (incurring a higher interest rate) to get a rebate in cash to pay closing costs.

---

[18] In the volume of loans, the largest servicers are Wells Fargo, J.P. Morgan Chase, Bank of America, and Citigroup. Quicken has always serviced their own loans.

Although we cannot see the pricing in our dataset, conversations with mortgage brokers suggest that the interest rate costs of taking these yield rate spreads are high. Thus, if a borrower could pay a slightly lower downpayment, retaining some cash for closing costs, without inducing any interest rate increase, this choice would be optimal for most borrowers. (This need not be true in the very short term, but applies only on average.) The borrowers who are likely to be able to slightly decrease their downpayments while not increasing increase rates are those not facing the LTV=0.80 threshold. Thus, in columns (4) to (6) of Table 4, we consider the robustness of our results to negative points by re-estimating the Table 3 specification for the sample of lower LTV borrowers. The results in these columns are very similar to those in Table 3.

IV.c. Interest Rate Discrimination Estimates – FinTech Lenders

With robustness concerns alleviated, we can move to one of our fundamental objectives, estimating price discrimination by face-to-face lenders versus algorithmic lenders. Face-to-face lenders have additional freedom to discriminate, not just because of racism and in-group bias, but because they may use soft information differentially across borrowers. An example we like to give is that imagine you are a loan officer, and someone comes into your office saying that she must have a mortgage today to make an offer on a house she has seen. The loan officer, if maximizing for the lender, would quote this applicant with a higher rate than if the applicant mentioned that she is shopping-around. If loan officers apply this soft information, or simply quote biased rates every time for a particular minority, higher discrimination should be seen in face-to-face lenders.

Algorithmic lending requires using data from applicants on geography or other variables to induce the algorithm to quote a higher price. Face-to-face lenders may be doing this too, in their preparation of different rate sheets by branch (or, in the case of small mortgage issues, simply in having particular zone pricing.).

Table 5 shows the results for FinTech lenders. We do not have a perfect indicator as to which lenders are FinTechs, or, said more broadly, which lenders have platform loans than require no face-to-face contracting. The sample in columns (1) to (3) are the list of FinTech platforms from Buchak et al (2017). However, this sample is too restrictive. Although their sample covers 47 non-banks (6 of which are FinTech platforms, including Quicken) and 31 banks, the lenders only cover 32,393 purchase and 134,712 refinance mortgages.

The sample in columns (4) to (6) are the lenders who, as of summer 2018, have a full mortgage application process online. We manually look up each of 2,098 of our largest volume lenders. A full mortgage application means that all pulling of data from the applicant (tax forms, bank information, credit report, etc.) are online and the borrowing is provided an accept/reject decision as well as a rate. The only residual underwriting that remains for the lender concerns the house (appraisal, lien check, etc.). We found that 945 of these, 45%, including almost all of the big banks, have a fully-algorithmic mortgage. This set, however, is too inclusive, in that we could not discern when these lenders began to provide algorithmic mortgages in our time series. Thus, we are looking for consistency in these two measures of FinTech.

Looking to columns (3) and (6) in Table 5, we find that FinTech purchase mortgage discrimination ranges from 5.17 - 6.54 bps for the restrictive definition (quoting columns (1) and (3)) and 5.43 – 8.91 bps for the too inclusive FinTech definition. These numbers are very internally consistent in Table 5 and are also almost identical in economic magnitude as those in Table 3. For refinance mortgages, likewise, our Table 5 estimates are internally consistent across the two definitions of FinTech and approximate the 1-3 bps of discrimination we report from Table 3.

We conclude from this that the FinTechs do no better at discrimination than face-to-face borrowers. It must be the case that any loan-officer discrimination that is removed by not seeing faces is added back by algorithms that better predict which borrowers can be captured at higher rates.

However, the story has a couple of silver linings. The first silver lining can be seen in Figure 5, a plot a time series of discrimination by a two-year window of loan issuance. Discrimination has monotonically declined between the 2008-2009 period and the 2014-2015 period. Although we cannot assert causality, we believe that this result could be due to competition from the platforms and/or the ease of shopping-around made possible by online applications. To look for evidence consistent with this interpretation, we plotted (not displayed in the current draft) Figure 5 by FinTech or not. The pattern looks the same. Thus, the pattern seems to reflect growing competition edging out the possibility for rents. We also plotted the same pattern for FHA mortgages. The pattern is the opposite. The non-GSE market seems to be discriminating more, and the GSE market, whose processes have become easier with algorithms directly connected to the GSE underwriter systems have become more uniform in pricing.

The second silver lining for the role of FinTechs comes in looking to rejection decisions. Thus, we turn our attention to rejection results.

IV.d. Rejection Discrimination Estimates – Face-to-face and FinTech Lenders

Table 6 presents the discrimination estimated by in the accept/reject decision. The dependent variable is an application rejection indicator, where a loan is deemed accepted if the lender makes an accept offer to a borrower even if the borrower does not take up the loan. The table presents six different samplings of the data, which covers more than 8.5 million observations, starting with the full sample of data for columns (1) to (3). This is close-to the full population of applications processed as 30-year, fixed-rate, single-family home GSE mortgages, run through the GSE underwriting system, except for nonconforming loans using the system for convenience. For each sample, we present three models. The first includes the application data that HMDA provides – ethnicity, income, loan amount, and year. We include 21 splines of income and 47 splines of loan amount. The second model of each sample adds census tract-level data to proxy for the other application data that HMDA does not record – LTV and credit score. We add three percentiles (the25th, 50th, and 75th) of these variables to capture the distributional spread in the census tract. We also add the census tract average loan amount. The third model finally adds lender and county fixed effects. The difference in sample between Panels A and B is important. Because large lenders may bear residual default risk in servicing costs for delinquent borrowers (not in direct exposure to default itself), we think the sample of non-top 25 volume lenders is more robust. Thus, we focus our discussion on Panel B of Table 6, noting that the full sample results (in Panel A) are very similar.

In columns (1) to (3) of Panel B, we find results consistent with residual discrimination against Latinx and African-Americans. Focusing on column (3), lenders reject minority ethnicities 4.3% more often, after controlling for the variables (or proxies) for the variables entering the black-box GSE underwriting system and after doing within-lender and within-location analysis. Note that the R-squared is 0.328 in this model. It is well known that R-squareds are severely downward biased in Linear Probability Models (which we use because of all of the fixed effects). A useful interpretation is from Gronau (1998) that the R-squared estimates the improvement in predictive probability, for predict a reject, over a naïve model that would be correct fifty percent of the time. In this light, our model does very well in picking up much of the residual variation.

A 4.3% higher rejection rate is very material. Table 1 reports that the mean rejection rate is 42%. Thus, our results suggest that rejections increase by 10% in percentage change.

Columns (4) to (9) consider discrimination by the FinTechs. As before, we use an under-inclusive (columns (4) to (6), representing 92,817 small-lender applications) definition of algorithmic decision-making in the measure of FinTech platforms from Buchak et al (2017), and we use an over-inclusive (columns (7) to (9) , representing 1,932,323 small-lender applications) measure of the set of lenders who, as of summer 2018, have a full mortgage application online. The truth is somewhere in between, thus we average the coefficients to interpret. The average of the Latinx and African-American coefficients in Panel

B for columns (6) (a coefficient of -0.0158) and (9) (a coefficient of -0.0372) is 0.0107. FinTechs discriminate much less than face-to-face lenders. In fact, columns (5) and (6) suggest that FinTech actually negatively discriminate, perhaps instead catering to Latinx and African-American clienteles to pick up populations discriminated against by face-to-face lenders.

Discrimination in rejections is not akin to discrimination in pricing for the lenders. Firms that discriminate in rejecting qualified applicants are losing money, because of the GSE guarantee and securitization process. Thus, it is rational that FinTechs have remove the potential for racism and in-group/out-group biases that cost lenders money. What is strange is the face-to-face lenders still allow procedures that allow-for discrimination. However, we must revisit our caveat to provide an alternative interpretation.

We stated that a potential selection issue may influence our estimates. In particular, a story consistent with our results is that white ethnicity loan officers are more active in informing white applicants of weaknesses in their profile ahead of the formal application process. This alone would not bias our results except that we do not have all the variables (in their exact form rather than in a proxy form) that enter into the GSE underwriter. Our results showing a higher 4.32% rejection rate for Latinx and African-Americans is large for such a bias to explain, but we cannot rule out that at least part of this coefficient is due to selection rather than discrimination. We are content with this alternative explanation, noting that this is still a form of discrimination in that the "help" provided to white ethnicities is still welfare-improving.

This potential selection/observability bias cannot, however, affect our FinTech results, in that algorithms do not perform these "help" functions of pre-informing potential applicants of profile weaknesses, or, if they do, they do not perform these functions differentially by ethnicity. Therefore, we summarize our rejection results as follows.

Our rejection results suggest two inferences. First, our results suggest that face-to-face lenders either over-reject Latinx and African-American borrower applications, or they help ethnic majority borrowers in pre-advising of application weaknesses, both of which imply a non-neutral welfare implication for minority applicants. Second, FinTech do not discriminate much, if any, in application accept/reject decisions. This represents a second silver lining in the role of algorithms in consumer lending.

V. Conclusion

Using a unique data set of mortgage loans that includes never-before-linked information at the loan-level on income, ethnicity, loan-to-value and other contract terms, we exploit the unique structure of the GSE pricing grid to identify discrimination in mortgage loan pricing. Overall, we find that conditional on obtaining a loan, African-American and Latinx pay a higher (5.6 bps) interest rate for purchase mortgages and about a 1-3 bps higher mortgage interest rate for refinance mortgages. Given the emergence of algorithmic credit scoring, we further investigate the level of discrimination among FinTech lenders, which we find to be roughly identical (5.3 basis points extra paid by minorities in purchase mortgages) to the overall set of lenders. However, in cross-sectional analysis of rejection rates, we find that African-American and Latinx borrowers are almost 4% more likely to be rejected for a mortgage than other borrowers in the overall sample, but we find no discrimination in rejection rates among FinTech lenders suggesting these lenders may cater to borrowers discriminated against by face-to-face lenders.

We also make contributions to debates within the economic literature concerning the identification of illegitimate discriminatory use of protected characteristics in credit screening and how the economist's challenge maps to the legal concept of "disparate impact" liability under U.S. fair lending laws. Fundamentally, the GSEs have elected to price credit based on two observable variables (LTV and credit score) that proxy for lifecycle variables in a macro-fundamental model of repayment risk. Under U.S. fair lending law, any discrimination in loan pricing that arises from this parsimonious pricing model should be justifiable as a legitimate business necessity given that these variables proxy for repayment risk. However, for the same reason, any disparate impact in loan pricing that is unexplained by this model points to illegitimate discrimination in loan pricing that is unrelated to the legitimate life cycle variables the GSEs have chosen to utilize in pricing credit risk. This is precisely what we find. Having access to the GSE's model for pricing credit risk thus permits us to identify empirically the incidence of discrimination as well as to make the further assertion that any such discrimination is not justified by a legitimate business necessity under U.S. fair lending laws.[19]

This framework has important implications for regulators and courts as lenders seek to exploit Big Data to improve credit scoring in consumer lending, particularly given the Supreme Court's recent requirement that a plaintiff must establish a "robust" causal connection between a lending practice and an alleged statistical disparity in lending outcomes. As our analysis suggests, loan outcomes that depart from a lender's credit scoring model that includes variables used to proxy for legitimate lifecycle variables in a macro-fundamental model of repayment risk is sufficient to identify empirically the existence of problematic discriminatory lending practices. This conclusion follows because, under the lender's own model, any such discrimination is unjustified by the legitimate business necessity of evaluating

[19] Under the Supreme Court's *Inclusive Communities* framework, our identification strategy should provide a means to satisfy the Court's requirement that a plaintiff first establish a "robust" causal connection between a lending practice and an alleged statistical disparity before shifting the burden to the lender to establish a legitimate business necessity for the practice. While we lack information as to what specific practices are producing these discriminatory effects in GSE lending, the fact that the GSE underwriting and pricing structure relies entirely on observable life-cycle variables also permits the conclusion that, whatever practices produced these outcomes, they cannot be justified as lenders' attempts to proxy for unobservable life-cycle variables. Accordingly, the practices producing these discriminatory effects must be lacking in a legitimate business necessity, to the extent courts limit this defense to factors related to a borrower/applicant's credit risk.

creditworthiness. At the same time, courts' cabining of the legitimate business necessity to evaluating borrower's creditworthiness places the burden on lenders to justify empirically why the use of particular Big Data variables is an empirically valid proxy for a legitimate lifecycle variable *even after orthogonalizing with respect to their estimate of this variable.*[20]

Finally, our findings with regard to the absence of discrimination of rejection rates in FinTech mortgage lending suggest that, in addition to the efficiency gains of these innovations, they may also serve to make the mortgage lending markets more accessible to African-American and Latinx borrowers. While we caution that this finding is tentative given that we lack access to all of the variables used by the GSEs to approve or reject loans, this positive evaluation of FinTech lending bodes wells for the needed expansion of U.S. residential mortgage markets as they continue to recover from the 2009 crisis. On a more cautionary note, however, the discipline imposed by the GSEs' underwriting and pricing requirements may help explain why the incidence of discrimination in lending is not greater than we find within our sample of loans originated by FinTech lenders or among lenders overall. To date, this less-well-understood role of the GSEs has not been considered in GSE reform proposals, nor is it obvious how such a role could be supported within a fully privatized, conventional conforming secondary mortgage market. Likewise, outside of mortgage lending, the possibility remains that, lacking formal underwriting and pricing standards, lending algorithms might proxy for unobservable life-cycle variables by relying on observable characteristics (e.g., the name of an applicant's high school) that produce illegitimate statistical discrimination.

[20] This approach necessarily assumes a lender has taken the effort to justify the inclusion of a Big Data variable as a proxy for some lifecycle variable in a model of credit risk. Even if these variables are unobservable for a particular applicant (e.g., income growth), lending institutions can nevertheless observe how these variables are distributed in the population using historical data. For instance, a lender that chooses to use the high school of an applicant as a proxy for the applicant's income growth would have presumably made this decision due to some historical relationship between observed high school and observed income growth among a sample of other individuals for which the lender has historical data. Under our approach, the burden is on the lender to demonstrate empirically why, within this sample of data, high school attended is not correlated with ethnicity after orthogonalizing high school and observed income growth.

Appendix A: Legality of Explicit and Statistical Discrimination

Discrimination in residential real estate lending is policed primarily by two federal statutes, the Fair Housing Act of 1968 (FHA) and the Equal Credit Opportunity Act of 1974 (ECOA). The FHA, which is administered by the Department of Housing and Urban Affairs (HUD), prohibits taking any action affecting the terms of the transaction on the basis of a borrower/applicant's race, religion, national origin, gender, familial status (e.g., family size or marital status), or handicap. The ECOA builds on the FHA by expanding the set of protected classes of borrowers. Under these laws, a lender in the mortgage market may not engage in, among other things, (i) refusing to extend credit, (ii) using different standards in determining whether to extend credit, or (iii) varying the terms of credit offers (e.g., loan amount, interest rate, duration and type of loan) on the basis of one of the above-mentioned protected characteristics.

Under the FHA and the ECOA, either a wronged borrower/applicant or the relevant administrative agency has the authority to bring a civil action against a lender. As in other areas of anti-discrimination law, however, the scope of these anti-discrimination mandates has been determined largely by the legal standards that courts have created for a successful claim. Historically, the primary method by which lenders were found to have engaged in prohibited discrimination was by a showing of disparate treatment --- specifically, a showing that a lender had treated borrowers or applicants differently because of a protected characteristic. For example, in Watson v. Pathway Financial, 702 F. Supp. 186 (N.D. Ill. 1988), an African-American couple successfully sued a lender under the FHA for rejecting their mortgage application because of delinquent credit card accounts. While the lender's justification was a potentially legitimate basis for denying credit, the court found the lender had violated the FHA because the applicants produced evidence that the lender had approved at least six applications from white borrowers showing similar delinquencies.

Disparate treatment claims have also been used to prohibit traditional redlining, in which a mortgage lender refuses to make loans to entire geographic areas because of their racial composition (Gano; 2017). Thus, for economists, the two sets of variables that are explicitly illegal under disparate treatment are indicator variables of the protected category (e.g., ethnicity in our case) and geography.

In addition to disparate treatment claims, a private party or governmental agency can bring a claim of lending discrimination under the FHA or ECOA under a disparate impact theory. In contrast to disparate treatment claims, claims of disparate impact do not involve any allegation of intentional discrimination in how a lender treats applicants/borrowers but rely instead on the fact that lending practices that are facially neutral in their treatment of different groups nonetheless fall more harshly on a protected category of applicants. For instance, in a joint policy statement on the enforcement of the ECOA and the FHA, the

Department of Justice, HUD, and all federal banking regulators provided the following example as an illustration of a lending policy that could give rise to a disparate impact claim[21]:

Example. A lender's policy is to deny loan applications for single-family residences for less than $60,000. The policy has been in effect for ten years. This minimum loan amount policy is shown to disproportionately exclude potential minority applicants from consideration because of their income levels or the value of the houses in the areas in which they live.

Despite agency approval of disparate impact theory, the ability of parties to pursue disparate impact claims has been hindered for two principal reasons. First, the existence of a disparate impact, by itself, is insufficient to prove illegal discrimination. Rather, after a plaintiff demonstrates that a lending practice produces a disparate impact, the lender can defend the practice as justified by a legitimate business necessity, provided no alternative policy or practice could achieve the same goal with less discriminatory effect.

Second, neither statute expressly defines the standard for proving a disparate impact violation. In fact, it was not until the Supreme Court's 2015 decision in Texas Department of Housing and Community Affairs v. Inclusive Communities Project that the Court even approved the disparate impact framework under the FHA. While the Court interpreted the FHA to permit disparate impact claims, it also required plaintiffs to establish a "robust" causal connection between a specific practice and an alleged statistical disparity. The Court imposed this requirement to ensure that " '[r]acial imbalance …does not, without more, establish a prima facie case of disparate impact' and thus protects defendants from being held liable for racial disparities they did not create." Since the decision, however, lower courts have struggled with how plaintiffs can establish proof of causality under this standard for a disparate impact case to move forward (Fulks, 2017).

The economic literature has something to contribute to the question of specifying a standard of proof.[22] Starting in the 1970s (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977), economic research shifted discrimination discussions to the statistical theory of discrimination rather than taste-based discrimination associated with Gary Becker. The language of statistical discrimination maps well to the legal theory of disparate impact. However, like courts considering disparate impact claims, economists have struggled with positing causality in statistical discrimination estimation. Criticisms of inadequate data and omitted-variable biases in estimation have plagued this literature.[23]

---

[21] See Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18,266 (Apr. 15, 1994).

[22] See commentary by <cite>GlassmanVerna:16</cite> and Winfield v. City of N.Y., No. 15CV5236-LTS-DCF, 2016 WL 6208564, at *1 (S.D.N.Y. Oct. 24, 2016); Cty. of Cook v. HSBC N. Am. Holdings Inc., 136 F. Supp.3d 952, 955 (N.D. Ill. 2015).

[23] Cited data omissions in empirical tests of discrimination include omitted information such as the loan to value ratio at origination, the debt to income ratio, all the contracting elements of the mortgage, the property characteristics and

For example, Schafer and Ladd (1981) find some evidence of interracial pricing differences, and Black and Schweitzer (1985) find indications of differences in loan terms. Yet Sandler and Biran (1995) critique any use of statistics in legal proceedings of discrimination claims due to the prevalence of omitted-variable bias and poor identification strategies in these studies. Not surprisingly, the mortgage literature on discrimination parallels these broader patterns, primarily focusing either on the methodological difficulties in providing robust statistical evidence for discrimination (See Black, Schweitzer, and Mandell (1978); Kaye (1982); Maddala and Trost (1982); Rachlis and Yezer (1993) on the efficacy of specific types of legislation.)[24]

However, we think that more can be accomplished. Consistent with the Supreme Court's decision in Inclusive Communities, the theory behind statistical discrimination provides guidance as to how one can demonstrate a "robust" causal connection between lending practices and racial or ethnic disparities in lending outcomes. In statistical discrimination, agents (in our case lenders seeking to screen applicants) have limited information and no animus against racial groups. Statistical discrimination arises as a solution to a signal extraction problem. Agents seek to reconstruct hidden information as to the expected creditworthiness of applicants using observable proxies.

We have already established that the use of a protected variable (ethnicity) is illegal under disparate treatment. However, could other variables be legal proxies for hidden information, even if they are correlated with ethnicity? Under the legal theory of disparate impact, statistical discrimination is allowable for "legitimate business necessity." Economic theory guides us that the meaning of this phrase is that a variable legitimately appears as a structural variable that maps the ability of a household to repay a loan to their economic fundamentals. In particular, one can write down a life-cycle model in which cash flow for repayments emerge from the current borrowing position (debt), cost of borrowing (credit score), income (in levels, growth, and risk), wealth, and regular expense levels (cost of living measures).

Legally, limiting the legitimate business necessity defense to structural variables that predict creditworthiness is entirely consistent with legal precedent assessing whether the refusal to extend credit violates federal antidiscrimination law.[25]

---

exact address, the applicant's ethnicity, gender, credit history and debt burden levels, lender and regulatory characteristics, all at the loan level for successful and unsuccessful applicants.

[24] For example, the Fair Housing Act of 1968, the Equal Credit Opportunity Act of 1974, the Home Mortgage Disclosure Act of 1975, the Community Reinvestment Act of 1977, and the HMDA amendments to the Financial Institutions Reform Reregulation and Enforcement Act of 1989

[25] See A.B. & S. Auto Service, Inc. v. South Shore Bank of Chicago, 962 F. Supp. 1056 (N.D. Ill. 1997) ("[In a disparate impact claim under the ECOA], once the plaintiff has made the prima facie case, the defendant-lender must demonstrate that any policy, procedure, or practice has a manifest relationship to the creditworthiness of the applicant…"). See also Lewis v. ACB Business Services, Inc., 135 F.3d 389, 406 (6th Cir. 1998) ("The [ECOA] was only intended to prohibit credit determinations based on 'characteristics unrelated to creditworthiness.'").

Further, cases involving "reverse redlining" --- the extension of credit to a protected class on inferior terms than those offered to non-protected applicants --- also suggest that only structural variables that map to repayment risk should be permitted under the "legitimate business necessity" defense in assessing disparate impact claims in rates, conditional on the extension of credit.[26]

Thus, whether in loan rejection or in loan pricing, the use of any of the foregoing structural variables should be considered "legitimate business necessity" under disparate impact theory, even if this variable statistically loads on (punishes) a particular protected category. As a matter of economic theory and legal doctrine, this is legitimate statistical discrimination.

What if a lender cannot see one of these variables (wealth, say), but can see a variable (e.g., the name of the high school attended) that correlates with wealth? Under the theory of disparate impact, this variable should be allowable if it is only disparately impacting the pool of applicants in sorting on wealth. In other words, conditional on latent hidden wealth, high school is orthogonal to loan decisions. A slightly less stringent assumption would also be consistent: conditional on latent hidden wealth, the impact of high school on loan decisions must be the same for one ethnic group as the other (ignorability).

[26] See, e.g., Miller v. Countrywide Bank, NA, 571 F.Supp.2d 251, 258 (D. Mass 2008) (rejecting argument that discrimination in loan terms among African American and white borrowers was justified as the result of competitive "market forces," noting that prior courts had rejected the "market forces" argument insofar that it would allow the pricing of consumer loans to be "based on subjective criteria beyond creditworthiness.")

Appendix B: Algorithm for Merging Mortgage Data Sets

Since there are no unique mortgage loan identifiers in the U.S., we develop an algorithm using machine-learning techniques to match loans found in two independent datasets: the McDash dataset, which contains loan-level data compiled by Black Night Financial Services, and the ATTOM dataset, which provides detailed property transaction and ownership information in addition to a time-series history of all recorded mortgage lien events such new mortgage originations, prepayments, REO, foreclosure, short sales, and arms-length sales and loan payoffs. The algorithm relies on matching distinct loan characteristics such as origination date, loan amount, and termination and distress events to precisely match each loan with its counterpart in the other dataset using a modified k-nearest neighbor classifier (see Hastie, Tibshirani, and Friedman, 2009; James, Witten, Hastie, and Tibshirani, 2015).

Merge process for newly originated fixed-rate GSE loans

There are two steps to our merging process. The first is to serialize the ATTOM data into a record format in which each property is assigned a full event-history string for each mortgage loan and the priority and performance of these mortgage loan positions. ATTOM provides very comprehensive geographic coverage for mortgage originations and terminations at the property-level for 90 percent of property Assessor's Pin Numbers and all the mortgage and lien recording for each property all sourced from public records. The second stage of the merging process is to employ functions for k-nearest neighbor algorithms using *sklearn.neighbors* in Python to fit radial kernels using *BallTree*. The k-nearest-neighbors classifier implements learning based on the 25 nearest neighbors in the corresponding zip code within the McDash mortgage data that also records the loan contract features and a loan-level string of performance characteristics. We represent each loan in each data set with a thirteen element vector that includes: 1) the original loan balance; 2) the lien position, 3) the origination date of the loan, 4) the ending date of the loan, 5) the foreclosure date of the loan (maybe null), the prepayment date of the loan (maybe null), 6) the appraised market value of the property, 7) the loan purpose (refinance or purchase), 8) loan distress dates (may be null), 9) loan REO date (may be null), 10) loan liquidation date (may be null), 11) short sale indicator variable (may be null), 12) interest rate type (fixed or variable loans), 13) property transaction value if there is a sale. Each of these elements is assigned a category subscore between 0 and 1. Each subscore is then squared to achieve a greater penalty for matches on key elements such as the loan amount. The category subscore is then scaled by a factor which represents the category's importance to the match quality relative to other elements used in the match. Each category factor is an integer between 0 and 100, and the sum of the category factors is 100. Our scoring algorithm (*get.score* in Python's *sklearn*) takes into account the 13 different elements of each matched pair of loans to calculate a score. The score roughly corresponds to the estimated error for each match, measured in hundredths of a percent. Thus, a match score of 1689 corresponds to a 16.89% chance of an incorrect match, or an 83.11% confidence in the match. We use only matches with scores of 2000 or less. For GSE loans originated between 2008 and 2015, we obtain a 90% merge rate.

Our prior machine learning strategy is less applicable for the merge of the HMDA data to McDash data, because we have only origination data in HMDA and as well as a greatly reduced set of loan characteristics at origination including: 1) the regulator type, 2) the loan type (conventional), 3) property type (1 to 4 single family residential properties, 4) loan purpose (refinance or purchase), 5) occupancy status, 6) original loan amount, 7) MSA, state, county and census tract, 8) self-reported borrower and co-borrower ethnicity (African-American, Hispanic, Asian, Caucasian, unknown), 9) borrower/coborrower gender, 10) borrower annual income, 11) year of origination, 12) denial reason if loan application is rejected, and 13) lender name. For this merge, we instead we standardize the lenders names between Dataquick and HMDA and then merge these data sets using lender names, loan amount, lien type, and the loan purpose fields. Of the 30.6 million originations in the HMDA data sets and about 10.4 million GSE loans, we successfully merged 60% of these loans. We then merged the ATTOM to McDash merged data to the ATTOM to HMDA merged data using the crosswalk developed with the k-nearest neighbor algorithm and we obtained a final data set of 3.47 million loans that are single family fixed rate GSE loans originated between 2008 and 2012.

The Equifax-enhanced subsample of originations

To obtain a final data that includes the full spectrum of underwriting characteristics that would have been available to the lender, we again merge the HMDA/ATTOM/McDash data set of fixed rate GSE loans that were originated between 2008 and 2015 to the McDash loans that are merged to Equifax data. The Equifax-enhanced originated loan sample includes other consumer credit positions of the borrowers such as: the total sum of retail, consumer finance and bank card balances; total student loan debt, total auto loan debt (sum or auto finance and auto bank debt); age of the borrower, and Vantage 3.0 score.

The HMDA sample of rejected conventional loans

The second important class of loans in our data set includes all of the conventional conforming loans in HMDA for 1-4 family residential borrowers whose loan applications were either denied by the originator, approved but not accepted by the borrower, withdrawn by the applicant, or the loan application file was closed for incompleteness. These data include information on: 1) the regulator type, 2) the loan type (conventional), 3) property type (1 to 4 single family residential properties, 4) loan purpose (refinance or purchase), 5) occupancy status, 6) original loan amount, 7) MSA, state, county and census tract, 8) self-reported borrower and co-borrower ethnicity (African-American, Hispanic, Asian, Caucasian, unknown),

9) borrower/co-borrower gender, 10) borrower annual income, 11) year of origination, 12) denial reason if loan application is rejected, and 13) lender name. They also include information on demographic and minority representation in the census tract in which the collateral on the loan is located. These variables include: the Federal Financial Institutions Examination Council (FFIEC) tract median family income to MSA median family income as a percentage, median family income for tract in thousands of dollars, tract population in thousands, tract minority population as a percentage, tract number of owner occupied units in thousands, tract number of 1- to 4-Family units in thousands.

HMDA does not include information on the credit score or the loan-to-value ratio of the rejected loan application files. For this reason, we proxy for the loan-to-value ratio by computing the mean assessed market value of all houses in each census tract in the U.S. using the panel of assessor's data from Dataquick. For each assessment we also have the year of the assessment. We then compute the ratio of the requested loan balance to the median value of all the homes in the appropriate census tract and year as reported by Dataquick to compute an estimated loan-to-value ratio for the rejected loan application. We proxy for the applicant's unobserved credit scores by using the McDash Vantage 3.0 score to compute the median credit score for each census track reported in McDash. The median census tract credit score is applied as a proxy for the credit score of the rejected loan.

References

Abowd, J., F. Kramarz, and D. Margolis, 1999, High wage workers and high wage firms, Econometrica 67, 251–333.

Aigner, D. J., and G. Cain, 1977, Statistical theories of discrimination in the labor market, Industry and Labor Relations Review 30, 175–187.

Aldrich, S. P. B., W. R. Greenberg, and B. S. Payner, 2001, A capital markets view of mortgage servicing rights, Journal of Fixed Income 11, 37–53.

Arnold, D., W. Dobbie, and C. S. Yang, 2018, Racial bias in bail decisions, Quarterly Journal of Economics (forthcoming).

Arrow, K., 1973, Higher education as a filter, Journal of Public Economics 2, 193–216. Ayres, I., 2002, Outcome tests of racial disparities in police practices, Justice Research and Policy 4, 131–142.

Bayer, P., F. Ferreira, and S. L. Ross, 2016, The vulnerability of minority homeowners in the housing boom and bust, American Economic Review: Economic Policy 8, 1–27.

Bayer, P., F. Ferreira, and S. L. Ross, 2018, What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders, Review of Financial Studies 31, 175– 205.

Becker, G. S., 1957, The Economics of Discrimination (University of Chicago Press, Chicago).

Begley, T. A., and A. K. Purnanandam, 2017, Color and credit: Race, regulation, and the quality of financial services, Working paper, University of Michigan.

Bertrand, M., and S. Mullainathan, 2004, Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, American Economic Review 94, 991–1013.

Black, H., R. L. Schweitzer, and L. Mandell, 1978, Wage discrimination: Reduced form and structural estimates, American Economic Review, Papers and Proceedings 68, 186–191.

Black, H. A., and R. Schweitzer, 1977, A canonical analysis of mortgage lending terms: Testing for lending discrimination at a commercial bank, Urban Studies 22, 13–19.

Blinder, A. S., 1973, Wage discrimination: Reduced form and structural estimates, Journal of Human Resources 8, 436–455.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru, 2017, Fintech, regulatory arbitrage, and the rise of shadow banks, Working Paper 23288, NBER.

Card, D., A. R. Cardoso, and P. Kline, 2016, Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women, Quarterly Journal of Economics 131, 633–686.

Elenev, V., T. Landvoigt, and S. van Nieuwerburgh, 2016, Phasing out the GSEs, Journal of Monetary Economics 81, 11–132.

Engel, K. C., and P. A. McCoy, 2011, The Subprime Virus: Reckless Credit, Regulatory Failure, and Next Steps (Oxford University Press, New York).

Federal Housing Finance Administration, 2009, Fannie Mae and Freddie Mac single-family guarantee fees in 2007 and 2008.

Federal Housing Finance Administration, 2010, Fannie Mae and Freddie Mac single-family guarantee fees in 2008 and 2009.

Federal Housing Finance Administration, 2011, Fannie Mae and Freddie Mac single-family guarantee fees in 2009 and 2010.

Federal Housing Finance Administration, 2012, Fannie Mae and Freddie Mac single-family guarantee fees in 2010 and 2011.

Federal Housing Finance Administration, 2013, Fannie Mae and Freddie Mac single-family guarantee fees in 2012.

Fortin, N., T. Lemieux, and S. Firpo, 2011, Decomposition methods in economics, in O. Ashenfelter, and D. Card, eds., Handbook of Labor Economics, volume 4A, chapter 1 (Elsevier, Amsterdam).

Fulks, M., 2017, Living with the robust causality requirement for disparate impact under the Fair Housing Act, Baker Donelson.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, 2018, Predictably unequal? the effects of machine learning on credit markets, Working paper, Federal Reserve Bank of New York.

Fuster, A., L. Goodman, D. Lucca, L. Madar, L. Molloy, and P. Willen, 2013, The rising gap between primary and secondary mortgage rates, Federal Reserve Bank of New York Economic Policy Review 19.

Fuster, A., and P. Willen, 2010, $1.25 trillion is still real money: Some facts about the effects of the Federal Reserve's mortgage market investments, Public Policy Discussion Paper 10-4, Federal Reserve Bank of Boston.

Gano, A., 2017, Disparate impact and mortgage lending: A beginner's guide, University of Colorado Law Review 88, 1109–1166.

Ghent, A. C., R. Hernandez-Murillo, and M. T. Owyang, 2014, Differences in subprime loan pricing across races and neighborhoods, Regional Science and Urban Economics 48, 199–215.

Glassman, A. M., and S. Verna, 2016, Disparate impact one year after Inclusive Communities, Journal of Affordable Housing & Community Development Law 25, 12–24.

Gronau, Reuben, 1998, A useful interpretation of R-squared in binary choice models, Princeton University Working Paper.

Gruenstein Bocian, D., K. Ernst, and W. Li, 2008, Race, ethnicity and subprime home loan pricing, Journal of Urban Economics 60, 110–124.

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

Hastie, T., R. Tibshirani, and J. Friedman, 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition (Springer, New York).

Haughwout, A., C. Mayer, and J. Tracy, 2009, Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing, Staff Report 368, Federal Reserve Bank of New York.

Heckman, J., 1998, Detecting discrimination, Journal of Economic Perspectives 12, 101–116.

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

James, G., D. Witten, T. Hastie, and R. Tibshirani, 2015, An Introduction to Statistical Learning with Applications in R (Springer, New York).

Kaye, D., 1982, Statistical evidence of discrimination, Journal of the American Statistical Association 77, 773–783.

Kerr, W. R., 2008, Ethnic scientific communities and international technology diffusion, Review of Economics and Statistics 90, 518–537.

Kerr, W. R., and W. F. Lincoln, 2010, The supply side of innovation: H-1B visa reforms and U.S. ethnic invention, Journal of Labor Economics 28, 473–508.

Kline, P., 2011, Oaxaca-Blinder as a reweighting estimator, American Economic Review: Papers & Proceedings 101, 532–537.

Maddala, G., and R. P. Trost, 1982, On measuring discrimination in loan markets, Housing Finance Review 1, 693–709.

Munnell, A. H., L. Browne, J. McEneaney, and G. Tootel, 1996, Mortgage lending in Boston: Interpreting HMDA data, American Economic Review 86, 25–54.

Oaxaca, R., 1973, Male-female wage differentials in urban labor markets, International Economic Review 14, 693–709.

Oaxaca, R. L., and M. R. Ransom, 1999, Identification in detailed wage decompositions, Review of Economics and Statistics 81, 154–157.

Phelps, E. S., 1972, Money, public expenditure and labor supply, Journal of Economic Theory 5, 69–78.

Rachlis, M., and A. Yezer, 1993, Serious flaws in statistical tests for discrimination in mort- gage markets, Journal of Housing Research 4, 315–336.

Reid, C.K., D. Bocian, W. Li, and R. G. Quercia, 2017, Revisiting the subprime crisis: The dual mortgage market and mortgage defaults by race and ethnicity, Journal of Urban Affairs, 39, 469-487.

Ross, S., 2002, Paired testing and the 2000 Housing Discrimination Survey, in W. Foster, A. F. Mitchell, and S. E. Fienberg, eds., Measuring Housing Discrimination in a National Study: Report of a Workshop (National Academies Press, Washington DC).

Ross, S. L., M. A. Turner, E. Godfrey, and R. R. Smith, 2008, Mortgage lending in Chicago and Los Angeles: A paired testing study of the pre-application process, Journal of Urban Economics 63, 902–919.

Sandler, A. L., and J. Biran, 1995, The improper use of statistics in mortgage lending discrimination actions, in A. Yezer, ed., Fair Lending Analysis (American Bankers Association, Washington, DC).

Shafer, R., and H. Ladd, 1981, Discrimination in Mortgage Lending (MIT Press, Cambridge, MA).

Sloczynski, T., 2015, The Oaxaca-Blinder unexplained component as a treatment effects estimator, Oxford Bulletin of Economics and Statistics 77, 588–604.

Smith, R., and M. Delair, 1999, New evidence from lender testing: Discrimination at the pre- application stage, in M. Turner, and F. Skidmore, eds., Mortgage Lending Discrimination: A Review of Existing Evidence (Urban Institute Press, Washington, DC).

Vickery, J. I., and J. Wright, 2013, TBA trading and liquidity in the agency MBS market, Economic Policy Review 19.