

Consumer-Lending Discrimination in the FinTech Era*

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

May 2019

Abstract

Discrimination in lending can occur either in face-to-face decisions or in algorithmic scoring. We provide a workable interpretation of the courts' *legitimate-business-necessity* defense of statistical discrimination. We then estimate the extent of racial/ethnic discrimination in the largest consumer-lending market using an identification afforded by the pricing of mortgage credit risk by Fannie Mae and Freddie Mac. We find that lenders charge Latinx/African-American borrowers 7.9 and 3.6 basis points more for purchase and refinance mortgages respectively, costing them \$765M in aggregate per year in extra interest. FinTech algorithms also discriminate, but 40% less than face-to-face lenders. These results are consistent with both FinTech and non-FinTech lenders extracting monopoly rents in weaker competitive environments or profiling borrowers on low-shopping behavior. Such strategic pricing is not illegal per se, but under the law, it cannot result in discrimination. The lower levels of price discrimination by algorithms suggests that removing face-to-face interactions can reduce discrimination. Further silver linings emerge in the FinTech era: (1) Discrimination is declining; algorithmic lending may have increased competition or encouraged more shopping with the ease of platform applications. (2) We find that 0.74-1.3 million minority applications were rejected between 2009 and 2015 due to discrimination; however, FinTechs do not discriminate in loan approval.

JEL classification: G21, G28, G23, J14, K22, K23, R30

Keywords: Discrimination; FinTech; mortgages; credit scoring; algorithmic underwriting; big-data lending; platform loans; disparate impact; statistical discrimination; legitimate business necessity

*Financial support from the Fisher Center for Real Estate and Urban Economics is gratefully acknowledged. For helpful comments and suggestions, we thank Manuel Adelino, Sanjiv Das, Andreas Fuster, Andres Liberman, Manju Puri, Raghuram Raju, Amit Seru, Ansgar Walther, Justin Wolfers, and seminar participants at U.C. Berkeley, Cornell, the University of Michigan, USC, NYU Stern, U.T. Austin, Texas A&M, Michigan State University, Chicago-Booth, Johns Hopkins University, Consumer Finance Protection Bureau, the George Washington-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, the 2018 EFA meeting, the 2019 AEA meeting, the 2019 NBER conference on "Big Data: Long-Term Implications for Financial Markets and Firms," the 2019 Adam Smith Workshop in Corporate Finance, the 2019 FMA Wine Country Finance Conference, the Federal Reserve Board, and Freddie Mac.

Table 1: Summary Statistics

Panel A reports summary statistics for the pricing estimations. Data are GSE, 30-year fixed rate mortgage originations obtained from a loan-level merge of HMDA, ATTOM, McDash, and Equifax data. Loan amount, applicant income and Latinx-/African-American are from HMDA. Interest rate, LTV, and credit score are from McDash-Equifax. Top 25 Volume Lender is calculated annually from volume of loans by lender. FinTech is a platform identifier from Buchak et al (2017). Panels B and C report statistics for accept/reject analyses. Because we only have HMDA data for rejections, we proxy for LTV, credit score, debt outstanding, and debt-to-income ratio using census tract medians. For all panels, Top 25 Volume Lender is calculated annually from the volume of loans by lender. FinTech is a platform identifier from Buchak et al (2017).

Panel A: For Pricing Analysis: GSE, 30-Year Fixed Rate Mortgage Acceptances (N = 3,577,010)

	Mean	St. Deviation	Minimum	Median	Maximum
Interest Rate % (McDash)	4.50%	0.56%	2.00%	4.50%	12.50%
Loan Amount \$,000	\$234.0	\$122.6	\$30.0	\$210.0	\$729.0
Applicant Income \$,000	\$107.2	\$92.0	\$19	\$89	\$9,980
Credit Score (McDash-Equifax)	755.8	43.4	620	766	850
Loan-to-Value(McDash-Equifax)	0.744	0.165	0.300	0.774	1.300
FinTech	0.043	0.203			
Top 25 Lender	0.523	0.499			
Latinx-/African-American	0.110	0.313			
Purchase=1; Refinance=0	0.418	0.493			

Panel B: For Accept/Reject Analysis: Conventional Mortgage Acceptances (N = 6,648,413)

	Mean	St. Deviation	Minimum	Median	Maximum
Loan Amount \$,000	\$213.9	\$114.7	\$30.0	\$191.0	\$729.0
Applicant Income \$,000	\$108.3	\$103.2	\$19	\$89	\$9,999
Credit Score (census tract)	750.8	24.1	620	756	832
Loan-to-Value (census tract)	0.791	0.101	0.300	0.799	1.283
Debt Outstanding (census tract)	18,180	8,145	0	17,739	529,506
Debt-to-Income% (census tract)	32.7	3.6	1.0	32.5	63.0
FinTech	0.042				
Top 25 Lender	0.522				
Latinx-/African-American	0.119				
Purchase=1; Refinance=0	0.308				

Panel C: For Accept/Reject Analysis: Conventional Mortgage Rejections (N = 6,535,664)

	Mean	St. Deviation	Minimum	Median	Maximum
Loan Amount \$,000	\$187.3	\$101.2	\$30.0	\$166.0	\$428.0
Applicant Income \$,000	\$97.4	\$129.7	\$19	\$75	\$9,999
Credit Score (census tract)	744.2	26.9	620.0	749.0	830.0
Loan-to-Value (census tract)	0.812	0.099	0.300	0.800	1.283
Debt Outstanding (census tract)	18,322	9,290	0	17,715	513,857
Debt-to-Income% (census tract)	32.9	3.9	1.0	33.0	61.0
FinTech	0.055				
Top 25 Volume Lender	0.515				
Latinx-/African-American	0.186				
Purchase=1; Refinance=0	0.173				

Table 2: Interest Rate Discrimination

Panel A reports discrimination results using the GSE grid for identification. The dependent variable is the interest rate on originated GSE 30-year fixed-rate mortgages. Estimates for purchase mortgages are in Columns (1) and (2); estimates for refinances are in Columns (3) and (4). Columns (1) and (3) report raw differences in means, as a starting point for understanding the role of the credit risk model. Columns (2) and (4) report discrimination estimates for the full credit risk model. We regress the interest rate on the GSE grid fixed effects and the month-year effects; identifying discrimination as the estimate on a Latinx/African-American indicator variable. Standard errors are clustered at the lender level. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels. Panel B presents economic magnitude calculations, aggregating the findings in Panel A to the mortgage market outstanding as of the end of 2018. Latinx-/African-American percentage representation in the mortgage market float (17.3%) is from the Survey of Consumer Finances. The aggregate housing debt is from the Federal Reserve Bank of New York.

Panel A: Estimates

	<i>Dependent Variable: Mortgage Interest Rate</i>			
	<i>Purchase Mortgages</i>		<i>Refinance Mortgages</i>	
	(1)	(2)	(3)	(4)
Latinx-/African-American	0.000903*** [0.000102]	0.000788*** [3.11e-05]	0.000298*** [7.98e-05]	0.000356*** [2.92e-05]
Observations	1,495,021	1,495,021	2,081,989	2,081,807
R-squared	--	0.729	--	0.694
Month-Year FE	N	Y	N	Y
GSE Grid FE	N	Y	N	Y

Panel B: Economic Magnitude Calculation

A. Market Size of Housing Debt (Federal Reserve of New York) (\$M)	\$9,536,000
B. African-American/Latinx % of the Float in Mortgage Market (SCF data)	0.173
C. Extra Interest Payments per Basis Point of Discrimination (\$M)	\$164.97
Discrimination Estimates from Panel (A):	
D. Extra Interest Rate (bps): Purchase Mortgages (estimate from column 2)	7.88
E. Extra Interest Rate (bps): Refinance Mortgages (estimate from column 4)	3.56
F. Share of refinance loans in stock of float	0.75
G. Weighted average extra interest rate ($= D*(1-F) + E*F$)	4.64
Annual Aggregate Extra Interest Paid by Latinx-/African-Americans ($= C*G$) (\$M)	\$765.47

Table 3: Interest Rate Discrimination - FinTech Results

This table replicates our main credit risk specification, but only for the subsample of FinTech lenders. The columns reproduce the specification of column (2) and (4) of Table 2, regressing interest rates on the GSE grid dummy variables, month-year effects, and an indicator for whether the borrower is Latinx- or African-American. The sample is the list of FinTech platforms from Buchak et al (2017). Column (1) reports purchase-mortgage estimates, and column (2) reports refinance mortgages. Standard errors clustered at the lender level are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

<i>Sample:</i>	<i>Dependent Variable: Mortgage Interest Rate</i>	
	<i>Purchase</i>	<i>Refinance</i>
	<i>FinTech Lender</i>	<i>FinTech Lender</i>
	(1)	(2)
Latinx-/African-American	0.000531*** [4.22e-05]	0.000197** [6.20e-05]
Observations	42,318	111,912
R-squared	0.729	0.707
Year FE	Y	Y
GSE Grid FE	Y	Y

Table 4: Interest Rate Discrimination - Robustness to Lender and Geography Fixed Effects

This table mitigates the concern that our Table 2 estimates are picking up differential costs of delivering a mortgage by lender or by geography. Panels A and B report estimates for interest rate discrimination for purchase and refinance mortgages, respectively. Column (1) repeats the OLS estimate of Table 2, regressing interest rates on the GSE grid-dummy variables, month-year fixed effects, and an indicator for whether the borrower is ethnically Latinx or African-American. This is the main credit risk model. Column (2) adds county fixed effects; column (3) adds lender fixed effects; and column (4) adds lender crossed with county fixed effects. Standard errors in brackets are clustered at the lender level. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

Panel A: Purchases

	<i>Dependent Variable: Mortgage Interest Rate</i>			
	<i>Varying Fixed Effects</i>			
	(1)	(2)	(3)	(4)
Latinx-/African-American	0.000788*** [3.11e-05]	0.000695*** [2.41e-05]	0.000545*** [2.67e-05]	0.000516*** [3.26e-05]
Observations	1,495,021	1,495,005	1,493,797	1,468,357
R-squared	0.729	0.733	0.748	0.759
Month-Year FE	Y	Y	Y	Y
GSE Grid FE	Y	Y	Y	Y
County FE	N	Y	Y	Y
Lender FE	N	N	Y	Y
County x Lender FE	N	N	N	Y

Panel B: Refinances

	<i>Dependent Variable: Mortgage Interest Rate</i>			
	<i>Varying Fixed Effects</i>			
	(1)	(2)	(3)	(4)
Latinx-/African-American	0.000356*** [2.92e-05]	0.000364*** [2.80e-05]	0.000223*** [2.19e-05]	0.000202*** [1.78e-05]
Observations	2,081,807	2,081,798	2,080,699	2,052,246
R-squared	0.694	0.697	0.712	0.721
Month-Year FE	Y	Y	Y	Y
GSE Grid FE	Y	Y	Y	Y
County FE	N	Y	Y	Y
Lender FE	N	N	Y	Y
County x Lender FE	N	N	N	Y

Table 5: Interest Rate Discrimination - Robustness

This table addresses robustness concerns for interpreting discrimination in pricing results. All specifications use the formulation of the main credit risk model of Table 2, columns (2) and (4), which include GSE-grid dummies and month-year fixed effects. Panels A and B present results for purchase and refinance mortgages, respectively. Column (1) address the possibility that our loading on the Latinx/African-American variable is due to points paid by non-minority borrowers. We limit the sample to mortgages with LTVs precisely at 0.8 and borrower's whose non-mortgage debt is greater than 30% of income. These are on average likely to be borrowers fully utilizing their cash for down payment, leaving no cash to pay (positive) points. Column (2) addresses the possibility that our results are driven by minority borrowers taking negative points (a rebate) to pay closing costs. Borrowers who have LTVs between (0.70, 0.74) and (0.75, 0.795) are not at the edge of the GSE grid bucket in LTV. These borrowers could pay slightly less in down payment without incurring an interest rate increase, so that taking a negative point rebate for closing costs would be suboptimal. Column (3) restricts the analysis to only non-top-25-volume lenders, addressing robustness of our results to the concern that large lenders retain the MBS and servicing rights after securitizing through the GSEs. Small lenders do not service GSE loans. Column (4) drops borrowers who do not designate their race or ethnicity directly in HMDA. Standard errors clustered at the lender level are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels. mean (0.795 to 0.801).

Dependent Var:	Mortgage Interest Rate			
	(1)	(2)	(3)	(4)
<i>Robustness Concern:</i>	<i>Positive Points</i>	<i>Negative Points</i>	<i>Residual Risk via Servicing or MBS Holding</i>	<i>Ethnicity Designation</i>
<i>Sub-Sample:</i>	<i>0.795 < LTV < 0.801 and Other Debt/Income > 0.3</i>	<i>LTV not near Grid Cell Maximum</i>	<i>Small Lenders</i>	<i>Only use HMDA-Classified Ethnicity Observations</i>
<i>Reasoning:</i>	<i>At budget constraint</i>	<i>Borrower will not face higher interest rate for slightly larger loan to cover closing costs</i>	<i>Unlikely to service the loans or hold as MBS on balance sheet</i>	<i>Eliminate software errors in race/ethnicity classification</i>
Panel A: Purchases				
Latinx-/African-American	0.000873*** [5.27e-05]	0.000938*** [4.02e-05]	0.000865*** [3.50e-05]	0.000809*** [3.26e-05]
Observations	241,847	337,115	846,547	1,370,384
R-squared	0.749	0.725	0.729	0.729
Month-Year FE	Y	Y	Y	Y
GSE Grid FE	Y	Y	Y	Y
Panel B: Refinances				
Latinx-/African-American	0.000251*** [5.96e-05]	0.000333*** [4.28e-05]	0.000384*** [2.21e-05]	0.000384*** [3.19e-05]
Observations	102,627	380,625	859,715	1,857,191
R-squared	0.746	0.705	0.732	0.695
Month-Year FE	Y	Y	Y	Y
GSE Grid FE	Y	Y	Y	Y

Table 6: Application Rejection Discrimination

This table reports discrimination in rejection rates for mortgages in our full sample of HMDA mortgages. The dependent variable is an indicator for an application being rejected by the lender. The sample is the set of all HMDA 30-year, fixed-rate mortgage applications run through GSE underwriting. The credit-risk model controls consist of three sets of variables, with the inclusion noted beneath the estimation. The first column presents raw mean differences. Column (2) presents the full set of HMDA data publicly available, which includes the year, applicant income, and applicant loan amount, including 21 piecewise-linear functions of income and 47 of loan amount. Column (3) includes the full set of variables used in the black box of the GSE underwriter system that determines GSE acceptability of applications, which adds in census tract proxies for LTV, credit score, debt-to-income (DTI), and total debt. We include quartiles (25th, 50th and 75th percentiles) of each variable to capture within-tract dispersion. We construct these census tract variables from McDash (for LTV, total debt, and credit score) and from the GSE data (for DTI). Standard errors clustered at the lender level are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

Panel A: Purchase Mortgages

	<i>Dependent Variable: Application Rejection</i>		
	(1)	(2)	(3)
Latinx/African-American	0.141*** [0.00510]	0.109*** [0.00508]	0.0957*** [0.00448]
Observations	3,179,813	3,179,813	3,179,273
R-squared	--	0.055	0.063
Application Income & Application Loan Amount: 68 piecewise-linear splines	N	Y	Y
Year FE	N	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	N	N	Y

Panel B: Refinance Mortgages

	<i>Dependent Variable: Application Rejection</i>		
	(1)	(2)	(3)
Latinx/African-American	0.122*** [0.00867]	0.0896*** [0.00577]	0.0728*** [0.00505]
Observations	10,004,264	10,004,264	10,002,727
R-squared	--	0.049	0.055
Application Income & Application Loan Amount: 68 piecewise-linear splines	N	Y	Y
Year FE	N	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	N	N	Y

Table 7: Application Rejection Discrimination - FinTech Lenders

This table reports discrimination in rejection rates among FinTech lenders for mortgages in our full sample of HMDA mortgages. The dependent variable is an indicator for an application being rejected by the lender. The sample is the set of stand-alone FinTech platforms from Buchak et al (2017). The credit risk model controls consists of three potential sets of variables, with the inclusion noted beneath the estimation. The first column presents raw mean differences. Column (2) presents the full set of HMDA data publicly available, which includes the year, applicant income, and applicant loan amount, including 21 piecewise-linear functions of income and 47 of loan amount. Column (3) includes the full set of variables used in the black box of the GSE underwriter system that determines GSE acceptability of applications, adding in census tract proxies for LTV, credit score, debt-to-income (DTI), and total debt. We include quartiles (25th, 50th and 75th percentiles) of each variable to capture within-tract dispersion. We construct these census-tract variables from McDash (for LTV, total debt, and credit score) and from the GSE data (for DTI). Standard errors clustered at the lender level are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

Panel A: Purchase Mortgages

	<i>Dependent Variable: Application Rejection</i>		
	<i>Sample: FinTech Lenders</i>		
	(1)	(2)	(3)
Latinx/African-American	0.0527* [0.0281]	0.0411 [0.0224]	0.0328 [0.0220]
Observations	70,813	70,813	70,791
R-squared	--	0.043	0.048
Application Income & Application Loan Amount: 68 piecewise-linear splines	N	Y	Y
Year FE	N	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	N	N	Y

Panel B: Refinance Mortgages

	<i>Dependent Variable: Application Rejection</i>		
	<i>Sample: FinTech Lenders</i>		
	(1)	(2)	(3)
Latinx/African-American	0.0540** [0.0213]	0.0288* [0.0154]	0.0233 [0.0132]
Observations	337,582	337,582	337,508
R-squared	--	0.052	0.058
Application Income & Application Loan Amount: 68 piecewise-linear splines	N	Y	Y
Year FE	N	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	N	N	Y

Table 8: Application Rejection Discrimination - Lender & Location Fixed Effects

This table reports robustness tests for our analysis of discrimination in rejection rates. The dependent variable is an indicator for an application being rejected by the lender. Column (1) repeats the specification from the third column of Table 6. Column (2) adds in lender fixed effects. Column (3) includes county and lender fixed effects. Column (4) includes lender crossed with county fixed effects. Column (5) repeats the column (4) specification but limits the sample to only small lenders. Panel (A) presents purchase mortgage application results, and Panel (B) presents results for refinance applications. Standard errors clustered at the lender level are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

Panel A: Purchase Mortgages

	<i>Dependent Variable: Application Rejection</i>				
	(1)	(2)	(3)	(4)	(5) Small Lenders
<i>Sample Restriction:</i>					
Latinx/African-American	0.0957*** [0.004481]	0.0745*** [0.003421]	0.0714*** [0.003191]	0.0698*** [0.003111]	0.0657*** [0.002231]
Observations	3,179,273	3,178,331	3,178,324	3,137,844	1,838,121
R-squared	0.063	0.224	0.227	0.264	0.337
Application Income & Application Loan Amount: 68 piecewise-linear splines	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	Y	Y	Y	Y	Y
Lender FE	N	Y	Y	N	N
County FE	N	N	Y	N	N
Lender # County FE	N	N	N	Y	Y

Panel B: Refinance Mortgages

	<i>Dependent Variable: Application Rejection</i>				
	(1)	(2)	(3)	(4)	(5) Small Lenders
<i>Sample Restriction:</i>					
Latinx/African-American	0.0728*** [0.005051]	0.0617*** [0.004841]	0.0628*** [0.004771]	0.0633*** [0.004761]	0.0527*** [0.002851]
Observations	10,002,727	10,001,968	10,001,964	9,954,232	4,415,763
R-squared	0.055	0.164	0.167	0.188	0.327
Application Income & Application Loan Amount: 68 piecewise-linear splines	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
LTV, Credit Score, Debt-to-Income, and Total Debt: Quartile Variables by Census Tract	Y	Y	Y	Y	Y
Lender FE	N	Y	Y	N	N
County FE	N	N	Y	N	N
Lender # County FE	N	N	N	Y	Y

Appendix Table 1: Statistics & Estimation of Interest Rate Discrimination for Shorter Maturities

This table reports statistics and analysis on GSE-issued conforming mortgages, with maturities of less than 30 years. Panel A reports statistics akin to Table 1, panel A except for the shorter maturity sampling. Data are from a loan-level merge of HMDA, ATTOM, McDash, and Equifax data. Loan amount, applicant income and Latinx-/African-American are from HMDA. Interest rate, LTV, and credit score are from McDash-Equifax. Top 25 Volume Lender is calculated annually from volume of loans by lender. FinTech is a platform identifier from Buchak et al (2017). Panel B reports discrimination results using the GSE grid for identification. The dependent variable is the interest rate. Columns (1) and (3) report raw differences in means, as a starting point for understanding the role of the credit risk model. Columns (2) and (4) report discrimination estimates for the full credit risk model. We regress the rate on the GSE grid fixed effects and the month-year effects, identifying discrimination as the estimate on an Latinx-/African-American indicator variable. Standard errors are clustered at the lender level. ***, **, and * indicate significance at the 1%, 5%, and 10% conventional levels.

Panel A: Statistics for Accepted GSE <30 Year Fixed Rate Mortgages (N = 1,390,286)

	Mean	Standard Deviation	Minimum	Median	Maximum
Interest Rate % (McDash)	3.94%	0.64%	2.00%	3.88%	7.63%
Loan Amount \$,000	\$185.4	\$100.1	\$30.0	\$163.0	\$729.0
Applicant Income \$,000	\$111.4	\$101.5	\$19	\$91	\$9,600
Credit Score (McDash-Equifax)	758.6	45.6	620	772	850
Loan-to-Value(McDash-Equifax)	0.669	0.177	0.300	0.685	1.300
FinTech	0.050				
Top 25 Lender	0.591				
Latinx-/African-American	0.113				
Purchase=1; Refinance=0	0.101				

Panel B: Estimates of Interest Rate Discrimination for Shorter Maturity Originations

	<i>Dependent Variable: Mortgage Interest Rate</i>			
	<i>Purchase Mortgages</i>		<i>Refinance Mortgages</i>	
	(1)	(2)	(3)	(4)
Latinx-/African-American	0.00134*** [0.000249]	0.00117*** [0.000162]	0.000292** [0.000137]	0.000555*** [6.85e-05]
Observations	140,613	140,613	1,249,673	1,249,634
R-squared	--	0.674	--	0.6
Month-Year FE	N	Y	N	Y
GSE Grid FE	N	Y	N	Y