Results

County Fixed Effects

Table 2 shows the results of the rate spread regression models using the aggregated HMDA data. The first three specifications can be characterized by the equation:

$$y_i = \sum_{all\ z} \alpha_z X_{i,z} + \epsilon_i \quad (1)$$

where *y* is the average rate spread in tract *i* and *X* represents the tract-level variables in the model described above. The first specification only includes the binary variable "Majority Black" which indicates that over half the applicants in a tract are black. This is included to showcase the rate gap before adjusting for borrower and loan characteristics. The second specification includes borrower and loan characteristics, for example, applicant race and loan term. The difference in the estimated coefficient for this specification and specification one can be attributed to statistical discrimination on the basis of weaker borrower characteristics and taste for riskier loan types, like manufactured housing, for black borrowers. The third specification further includes tract characteristics like the prominence of the GSEs, the market share of the largest lender, and the credit proxy variable measuring the proportion of applicants denied on the basis of credit. Lastly, the fourth specification includes county-level fixed effects. This specification will henceforth be the baseline model and can be characterized by the equation:

$$y_{i,j} = \sum_{all\ z} \alpha_z X_{i,z} + \mathrm{w_j} + \epsilon_{i,j} \quad (2)$$

where *X* and *y* are the same as equation (1) and with $\mathrm{w_j}$ representing the county fixed effects. These fixed effects capture the idiosyncratic differences across geographical regions, and their effect on rate spreads.

Table 2
Regression models of rate spread on aggregated HMDA variables
Dependent variable: Rate Spread

Variables	Race	+ Borrower	+ Tract	+ County FE
	(1)	(2)	(3)	(4)
Majority Black	0.525***	0.212***	0.175***	0.118***
	(0.01)	(0.01)	(0.01)	(0.01)
Income		-0.001***	-0.001***	-0.001***
		(0.00)	(0.00)	(0.00)
Loan amount (10,000s)		-0.008***	-0.009***	-0.009***
		(0.00)	(0.00)	(0.00)
Female		0.227***	0.170***	0.139***
		(0.01)	(0.01)	(0.01)
DTI		0.007***	0.008***	0.004***
		(0.00)	(0.00)	(0.00)
LTV		0.012***	0.012***	0.009***
		(0.00)	(0.00)	(0.00)
Origination charges (1,000s)		0.068***	0.072***	0.068***
		(0.00)	(0.00)	(0.00)
Discount points (1,000s)		-0.035***	-0.037***	-0.027***
		(0.00)	(0.00)	(0.00)
Conventional loan		-0.211***	-0.127***	-0.350***
		(0.01)	(0.01)	(0.01)
Loan term (months)		-0.003***	-0.002***	-0.002***
		(0.00)	(0.00)	(0.00)
Manufactured		1.909***	1.696***	1.617***
		(0.01)	(0.01)	(0.01)
Asian		0.092***	0.061***	-0.017
		(0.01)	(0.01)	(0.01)
Native American		0.206***	0.204***	0.227***
		(0.05)	(0.05)	(0.05)
Hispanic		0.387***	0.357***	0.343***
		(0.01)	(0.01)	(0.01)
Pacific Islander		-0.563***	-0.531***	-0.162**
		(0.08)	(0.08)	(0.08)
Fannie Mae purchased			-0.389***	-0.390***
			(0.02)	(0.02)
Freddie Mac purchased			-0.303***	-0.481***
			(0.02)	(0.02)
Ginnie Mae purchased			-0.187***	-0.278***
			(0.02)	(0.02)
Denied for credit			0.547***	0.115***
			(0.04)	(0.03)
Tract to MSA median income			-0.004***	-0.002***
			(0.00)	(0.00)
Lender market share			-0.058***	0.245***
			(0.01)	(0.01)
Constant	0.796***	0.455***	0.424***	
	(0.00)	(0.04)	(0.04)	
Observations	71,232	68,139	68,139	68,139
R^2		0.70	0.71	0.61
Adjusted R^2		0.70	0.71	0.59

The table on the previous page presents OLS estimates for 4 models of rate spread. Table 1 includes descriptive statistics for all variables seen above. Column (1) solely includes the binary variable of whether over 50% of the applicants are black. Column (2) adds borrower and loan characteristics. Column (3) additionally includes tract characteristics. Column (4) includes the previous variables and adds fixed effects on county with standard errors clustered at the county level. Standard errors are shown in parentheses.
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

As seen in Table 2, the rate gap between predominantly and non-predominantly black neighborhoods is 52.5 basis points. This constitutes a large differential in mortgage payments and interest expense for black households, constraining budgets and eroding long term wealth accumulation. Assuming the average loan size and interest rate from Table 1, an increase in rate spreads by 52.5 basis points corresponds to approximately $950 of higher interest payments a year for residents of predominately black neighborhoods.

Of this gap, approximately 60% is attributable to statistical discrimination on the weaker borrower characteristics of black neighborhoods. A further 7% is attributable to tract-level characteristics and 11% to geographical characteristics. This leaves 11.8 basis points of rate differential that this study attributes to discrimination against black neighborhoods. With an average rate spread of 82 basis points across tracts, 11.8 basis points of pricing discrimination corresponds to 14.3% higher rate spreads for predominately black neighborhoods. Notably, this rate premium is paid by all borrowers in these tracts.

Further, the coefficients of all controlling factors match the expected direction discussed in the Empirical Specification section. In line with previous research, the coefficient on female is positive, indicating communities with larger female applicant pools face higher-priced mortgages (Cheng et al. 2011). The coefficients on Hispanic and Native American variables are positive as well, confirming results from previous studies showing these populations face pricing discrimination (Cheng et al. 2015; Bayer et al. 2018; Delis & Papadopoulos 2019). Lastly, the coefficient on credit denials, the novel variable to measure credit strength across tracts, is

positive and significant. This suggests that the credit denial variable is a valid, though imperfect, means of capturing credit strength in the estimation of pricing discrimination. Together, these results suggest the model is well specified.

The estimate of pricing discrimination derived in this study for predominantly black communities is also in line with estimates of pricing discrimination found in previous literature, which have estimated pricing discrimination to range between 5 and 25 basis points for black borrowers (Ghent et al. 2014; Cheng et al. 2015). However, with the more comprehensive nature of the data used in this study, these findings can be extended to the nation as a whole in contrast to the localized interpretations required of matched datasets.

Additionally, with the exclusion of rate spread from past HMDA datasets, much of the past literature has focused on the incidence of high-cost loans in contrast to the more granular price differences found in this study (DeLoughy 2012; Bayer et al. 2018). The availability of rate spread in place of the binary incidence of a high-cost loan allows for a deeper understanding of the discrimination faced by all black applicants, not just the weaker borrowers who would be targets for high-cost loans. These results suggest it is not just the weakest black borrowers that face discrimination in the market through the higher incidence of high-cost loans, but that entire communities face higher rate spreads because of the proportion of black residents in that area.

Lender Fixed Effects

In order to account for the significant impact of shopping behavior and choice of lender shown by Bhutta and Ringo (2014) and Bayer et al. (2018), this section analyzes the effect of adding lender fixed effects to the models used in Table 2. These fixed effects account for the fact that black applicants are more likely to take out loans with higher cost lenders. This study also

adds an additional interaction term to capture the potential for a lender's impact on rate spreads to increase as their market share increases.

Specification five, which uses the same variables as specification three and adds fixed effects on lender, can be represented by the equation:

$$y_{i,s} = \sum_{all\ z} \alpha_z X_{i,z} + l_s + \epsilon_{i,s} \quad (3)$$

where y and X are the same as previous specifications, and l represents lender fixed effects. Specification six, which includes fixed effects on county and lender, can be represented by the equation:

$$y_{i,j,s} = \sum_{all\ z} \alpha_z X_{i,z} + \mathrm{l_s} + w_j + \epsilon_{i,j,s} \quad (4)$$

where $y, X, w,$ and l are the same as previous specifications. Lastly, specification seven includes the interaction of l and m, which represents the market share of the most prominent lender with the fixed effect of that lender. This allows the effects of lenders to be scaled by their penetration in a tract. This specification can be represented by the equation:

$$y_{i,j,s} = \sum_{all\ z} \alpha_z X_{i,z} + \mathrm{l_s} + w_j + m_i l_s + \epsilon_{i,j,s} \quad (5)$$

In contrast to previous literature analyzing loan-level data, including lender fixed effects has a small impact on rate spread differentials in the model. Including the interacted market share and lender fixed effects term only accounts for an additional .7 basis points in rate spreads. In practice, .7 basis points has little economic impact indicating the inclusion of these effects are inconsequential to the model.

Table 3
Regression models of rate spread on aggregated HMDA variables with fixed effects
Dependent variable: Rate Spread

Variables	County FE	Lender FE	Both FE	+ Interaction
	(4)	(5)	(6)	(7)
Majority Black	0.118***	0.157***	0.118***	0.111***
	(0.01)	(0.01)	(0.01)	(0.01)
Income	-0.001***	-0.001***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Loan amount (10,000s)	-0.009***	-0.010***	-0.009***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.139***	0.144***	0.127***	0.112***
	(0.01)	(0.01)	(0.01)	(0.01)
DTI	0.004***	0.006***	0.004***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
LTV	0.009***	0.011***	0.009***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Origination charges (1,000s)	0.068***	0.065***	0.062***	0.060***
	(0.00)	(0.00)	(0.00)	(0.00)
Discount points (1,000s)	-0.027***	-0.026***	-0.025***	-0.024***
	(0.00)	(0.00)	(0.00)	(0.00)
Conventional loan	-0.350***	-0.210***	-0.345***	-0.393***
	(0.01)	(0.01)	(0.01)	(0.01)
Loan term (months)	-0.002***	-0.002***	-0.001***	-0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Manufactured	1.617***	1.436***	1.403***	1.145***
	(0.01)	(0.02)	(0.02)	(0.02)
Asian	-0.017	-0.001	-0.038***	-0.024**
	(0.01)	(0.01)	(0.01)	(0.01)
Native American	0.227***	0.274***	0.246***	0.148***
	(0.05)	(0.04)	(0.05)	(0.05)
Hispanic	0.343***	0.402***	0.317***	0.299***
	(0.01)	(0.01)	(0.01)	(0.01)
Pacific Islander	-0.162**	-0.420***	-0.238***	-0.231***
	(0.08)	(0.08)	(0.08)	(0.08)
Fannie Mae purchased	-0.390***	-0.402***	-0.363***	-0.352***
	(0.02)	(0.02)	(0.02)	(0.02)
Freddie Mac purchased	-0.481***	-0.421***	-0.486***	-0.484***
	(0.02)	(0.02)	(0.02)	(0.02)
Ginnie Mae purchased	-0.278***	-0.260***	-0.276***	-0.293***
	(0.02)	(0.02)	(0.02)	(0.02)
Denied for credit	0.115***	0.403***	0.084**	-0.010
	(0.03)	(0.04)	(0.03)	(0.03)
Tract to MSA median income	-0.017***	-0.011***	-0.015***	-0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Lender market share	0.245***	0.200***	0.268***	0.18
	(0.01)	(0.01)	(0.01)	(0.36)
Observations	68,139	68,139	68,139	68,139
R^2	0.61	0.53	0.64	0.68
Adjusted R^2	0.59	0.52	0.62	0.64

The table on the previous page presents OLS estimates for 4 models of rate spread. Table 1 includes descriptive statistics for all variables seen above. Column (4) is the same as specification 4 in Table 2. Column (5) includes all HMDA variables included in specification 3 and adds fixed effects on the most prominent lender in each tract with standard errors clustered at the lender level. Column (6) includes the variables in Column (4) and Column (5) with fixed effects on both county and lender. Column (7) includes all variables in Column (6) but adds the interaction between market share and the fixed effects on lender. Bootstrapped standard errors are presented in Column (6) and (7). Standard errors are in parenthesis.
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Considering the identical rate spread differentials in specification four and six, it is likely that geographic fixed effects largely encapsulate the lender effects seen in previous studies. This may come as a result of aggregating the loans at the tract level. Whereas lenders vary widely at the national level, and thus for the individual applicants, at the local level only a subset of all national lenders will operate. These results suggested that while individual black applicants are more likely to use lenders that have a higher incidence of high-cost loans, black neighborhoods do not show the same aggregate shopping behavior once the geographic distribution of lenders is accounted for. Also, in contrast to this paper, Bhutta and Ringo (2014) used a 1% matched dataset of credit statistics, and Bayer et al. (2018) used a matched dataset for seven large MSAs for their analysis. Matching the HMDA data with smaller subsets limits the overall scope of the analysis and constrains the use of geographic fixed effects. The use of a national dataset allows for county-level effects to be accounted for, which may encapsulate the lender effects found by Bhutta and Ringo (2014) and Bayer et al. (2018).

Additionally, previous studies used the binary incidence of a high-cost loan in place of continuous rate spreads. It may be that certain lenders are more willing to give high-cost loans, but on average, do not charge significantly higher rate spreads than other lenders. It is also possible that lenders are now more hesitant to offer high-cost loans. As mentioned previously, a number of lawsuits have been filed in recent years, and this may have had the effect of reducing the incidence of high-cost loans, even in cases where they were not inherently discriminatory.

Limitations

Though the expanded HMDA dataset corrects for a number of the omitted variables of past releases of the data, a number of variables are still omitted and thus absent in this study. Principally these variables include foreclosure rates, prepayments, and a direct credit variable. It has long been speculated that black applicants face higher rate spreads and denial rates because of a higher tendency to default than white borrowers. The Financial Crisis appears to support this argument since minority borrowers did face higher rates of foreclosure, but this view is simplistic and does not consider predatory tactics that put minorities in high-cost loans which are more likely to default (Bayer et al. 2016; Chan et al. 2015; Mayer & Sherlund 2008). Accounting for these practices and the weaker baseline financial characteristics of these borrowers, Kau et al. (2011) does not find any significant difference in default rates for black applicants. Further, they do find a significant impact on prepayment rates, but these effects lower the probability of prepayment, which should increase the desirability of loans originated to black applicants.

Unlike most other fixed-income investments, such as US treasury bonds, investors in mortgages take on prepayment risk. This is the risk that borrowers prepay more rapidly when rates fall and prepay more slowly when rates rise. Both of these behaviors are disadvantageous to mortgage investors. The findings of Kau et al. (2011) indicate that black borrowers are less reactive to changes in interest rate, which should make these loans more attractive to lenders. In short, if race is going to be taken into account on the basis of default and prepayment risk, black borrowers should receive a rate spread discount for their expected behavior, not a rate premium. In the context of this study, this would suggest that omitting foreclosure and prepayment variables underestimates the pricing discrimination faced by black communities.

In regard to credit variables, this study does include a measure of credit approximated by the credit denial rates within a tract. However, while this does serve to add credit factors to the model, this measure of credit is likely noisy and potentially does not fully capture the full effect of credit on rate spreads. This would predominantly be an issue in tracts with significant disparity between rejected and accepted applicant pools. At the level as granular as the tract, which is typically a small geographical area with approximately 4,000 residents[1], this disparity is likely small but, in some cases, may still bias the estimated strength of credit in an area.

Further, the inclusion of county and lender fixed effects also helps account for inefficiencies in this study's measure of credit. County fixed effects encapsulate the strength of credit for residents in that region, and lender fixed effects encapsulate the variation in credit present in tracts that predominantly use that lender. This is especially true for lenders that cater to weaker borrowers. In aggregate, the use of credit denials and fixed effects likely do not fully capture the variation in credit between black and non-black neighborhoods, but together these variables likely significantly reduce the potential credit bias in the estimate for loan pricing discrimination.

[1] See: https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf