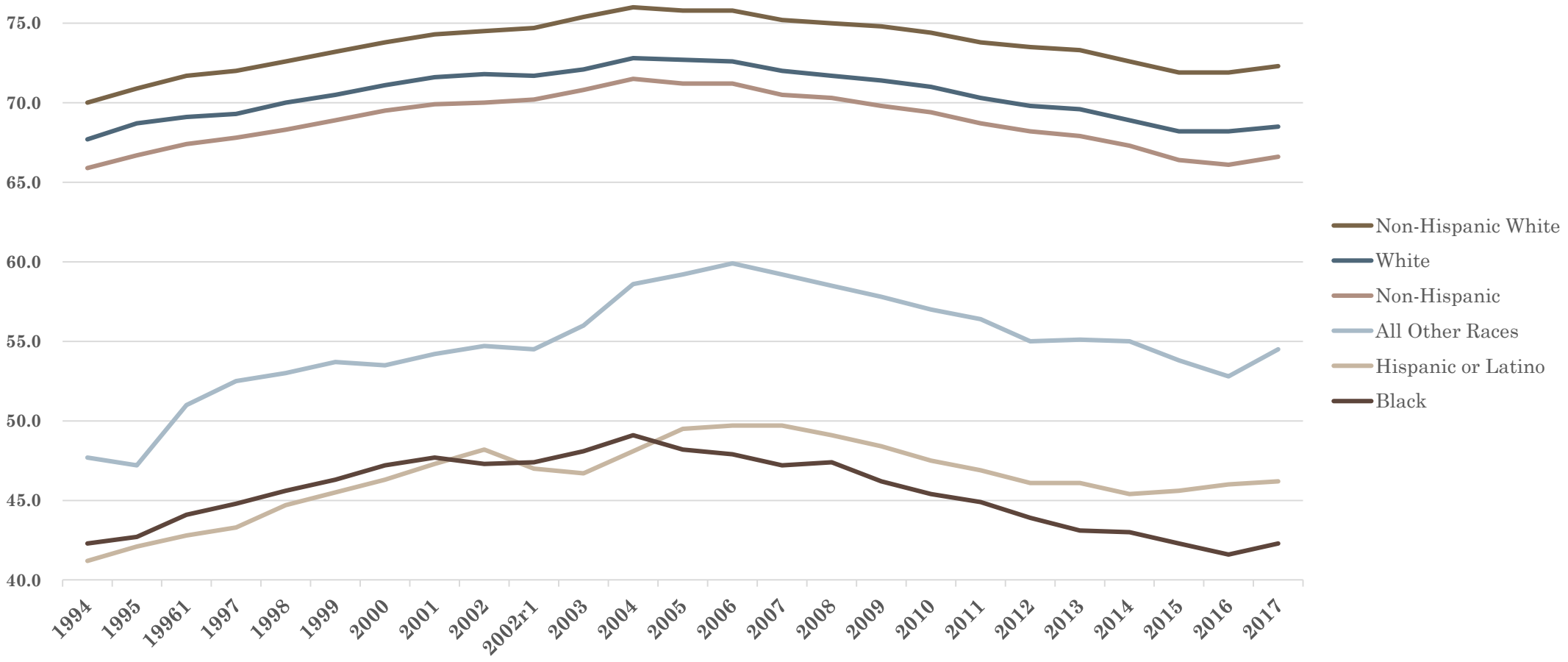
An Unequal Dream

The Gap in Homeownership for White and Black Americans

By Michael Nicholson

The Gap in Homeownership

Homeownership by Race



Data

HDMA LAR Data

- Home Mortgage Disclosure Act
  - Enacted in 1975
  - Requires financial institutions to provide mortgage data to the public
- Loan Application Register
  - Loan level data released annually
  - Includes data on race, ethnicity, income, type (conventional loan, FHA loan, VA loan, etc.), amount, action taken, reason denied (if applicable), etc.
- Dodd-Frank
  - Beginning in 2018, the required characteristics in LAR data was expanded
  - These new data points include credit score, DTI, LTV, age, fees, interest rate and expanded information on commercial and business loans
    - Credit scores are stripped from publicly released versions of the data to protect the autonomy of the borrower

Excluded Data

- The HDMA data includes applications for refinancing, commercial loans, pre-approval requests, investment properties, second home, vacation properties and other non-primary home loan applications
- I focused my analysis on primary residence mortgage applications to normalize the data and to avoid any idiosyncrasies of these other loan types
- I also intend to focus on discrimination in credit availability that impacts the homeownership rate gap
  - These other loan types provide the potential for interesting analysis but do not directly contribute to the disparity in homeownership since you are a homeowner whether you own 1 or 10 homes
- A few data points contain obvious data entry errors, e.g. a loan with a maturity of “360360”
  - I removed these points from the dataset as they arose

Results

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)

- Before adjustment, predominately black neighborhoods face rate spreads that are 52.5 basis points higher than non-black neighborhoods
  - This equates to approximately $950 a year in higher interest expense
- Borrower and tract characteristics account for ~70% of this gap
  - This holds for any order of adding these variables to the regression
- Geographic effects account for ~10%
- This leaves 11.8 basis points I predominately attribute to loan pricing discrimination
  - This corresponds to 14.3% higher rate spreads for black neighborhoods

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)

- Past literature has shown that black borrowers are much more likely to use high-cost lenders
  - This necessitates the use of lender fixed effects
- In contrast to past results, I find that county-level fixed effects largely encapsulate the lender fixed effects
  - This may result from the geographic effects capturing the geographic distribution of lenders
  - This is apparent in the identical Majority Black coefficient in specification 4 and 6
- Adding the interaction of lender fixed effects and their market share does have a small impact, but .7 basis points has little economic significance in practice

Robustness

Regression models of rate spread on aggregated HMDA variables with fixed effects

Dependent variable: Rate Spread

Variables	All Terms	30 Year	Non-30 Year
	(4)	(7)	(8)
Majority Black	0.118***	0.109***	0.151***
	(0.01)	(0.01)	(0.03)
Observations	68,139	67,792	30,143
R^2	0.61	0.58	.51
Adjusted R^2	0.59	0.55	.47

- To compare my results to past studies utilizing only 30-year-loans, I reran my models using only data from 30 and non-30 year loans
- Overall, the estimated loan pricing discrimination differential for 30 year loans and all loans is small
  - However, it does appear only considering 30-year loans slightly underestimates pricing discrimination
- More notably, the differential between solely 30 and non-30 year loans is much larger at ~4 basis points
- In of itself, this may not indicate greater rates of pricing discrimination in non-30 year loans since some features like adjusted-rate are not present in the model
  - This does, however, showcase yet another area for which black households face higher loan pricing
  - This is especially problematic if black borrowers are steered to these loan types

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

Variables	Majority Black	Quarter Black	Proportion Black
	(4)	(9)	(10)
Majority Black	0.118***		
	(0.01)		
Quarter Black		0.080***	
		(0.00)	
Proportion Black			0.224***
			(0.01)
Observations	68,139	68,139	68,139
R^2	0.61	0.61	0.61
Adjusted R^2	0.59	0.59	0.59

- The Majority Black variable is heavily skewed with only 5% of tracts meeting the threshold of having over 50% black applicants
- To check the robustness of using this as an explanatory variable, I reran the model using a binary variable for if over 25% of the applicants in a tract were black and a continuous black population variable
- The results of these models are in line with economic intuition, with the Quarter Black tracts showing a smaller but significant estimate of pricing discrimination and the continuous variable also showing a significant estimate for pricing discrimination
  - Note: the Proportion Black coefficient can be interpreted as a 2.24 increase in expected rate spreads for a 10 percentage point increase in black applicants

Next Steps

- Account for potential tract-level confounding factors
  - I am limited in the available data to append since most datasets are not at the granularity of the tract level
    - Most census data is not released beyond the county or MSA level
    - Fed data is only available to the county level
    - FBI data is released at the department level
  - At present, I have set up an API with the 2017 American Household Survey administered by the Census Bureau
    - This dataset has tract-level aggregate counts for unemployment, uninsured, government subsidy recipients, education attainment and vacant housing
    - This variables account for or are highly correlated with potential omitted variables in the model
      - For example, crime stats are not available at the tract level but correlates with both vacancy rates, unemployment and government subsidies