

Do Lenders Still Discriminate? A Robust Approach for Assessing Differences in Menus

Paul Willen¹ David Hao Zhang²

¹Federal Reserve Bank of Boston

²Harvard University

System Applied Micro Meeting
On Zoom, June 12, 2020

Not an official position of the Federal Reserve Bank of Boston or the Federal Reserve System.

©2020 by Paul S. Willen and David Hao Zhang.

Introduction



- Many studies have shown that black borrowers pay higher interest rates on their mortgages compared to observationally similar white borrowers.

Introduction



- Many studies have shown that black borrowers pay higher interest rates on their mortgages compared to observationally similar white borrowers.
- Disagreement about role of points
 - In the US, paying more in discount points reduces the interest rate.

Introduction



- Many studies have shown that black borrowers pay higher interest rates on their mortgages compared to observationally similar white borrowers.
- Disagreement about role of points
 - In the US, paying more in discount points reduces the interest rate.
- Did lenders offer black borrowers worse menus in terms of rates and discount points than observationally similar white borrowers? How can we tell if lenders **discriminated in menus**, given data on outcomes?

Background: the rate-point tradeoff



FNMA (GOLD): Conforming & High Balance Fixed Rate

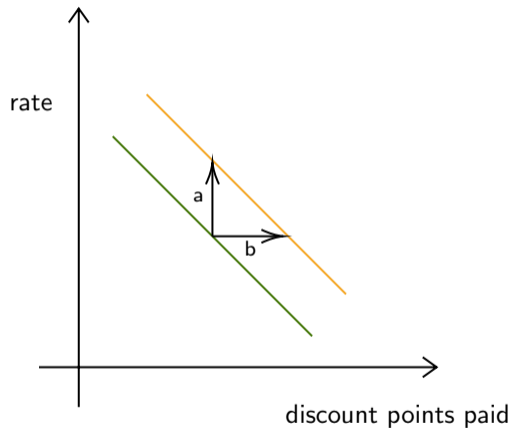
MORRIS Plan 113	30 Year Fixed Rate Conf		
Rate	15 Day	30 Day	45 Day
2.875	(1.643)	(1.353)	(1.353)
3.000	(1.916)	(1.626)	(1.636)
3.125	(2.105)	(1.815)	(1.815)
3.250	(2.751)	(2.461)	(2.461)
3.375	(3.185)	(2.895)	(2.895)
3.500	(3.387)	(3.097)	(3.097)
3.625	(3.077)	(2.787)	(2.787)
3.750	(3.591)	(3.301)	(3.301)
3.875	(3.897)	(3.607)	(3.607)
4.000	(4.006)	(3.716)	(3.716)

Background: the rate-point tradeoff



FNMA (GOLD): Conforming & High Balance Fixed Rate

MORRIS Plan 113	30 Year Fixed Rate Conf		
Rate	15 Day	30 Day	45 Day
2.875	(1.643)	(1.353)	(1.353)
3.000	(1.916)	(1.626)	(1.636)
3.125	(2.105)	(1.815)	(1.815)
3.250	(2.751)	(2.461)	(2.461)
3.375	(3.185)	(2.895)	(2.895)
3.500	(3.387)	(3.097)	(3.097)
3.625	(3.077)	(2.787)	(2.787)
3.750	(3.591)	(3.301)	(3.301)
3.875	(3.897)	(3.607)	(3.607)
4.000	(4.006)	(3.716)	(3.716)



Outline

- We show that, perhaps surprisingly, the task of assessing discrimination in menus is non-trivial. Intuitively appealing heuristics used in the literature such as:
 1. “Controlling” for points and compare rate, or the other way around,
 2. Adjusting for points using a range of possible rate-point trade-offs,
 3. Comparing means without controlling (looking at whether black consumers pay more in both rates and points).

Can lead to false positives/negatives and even contradictory results.

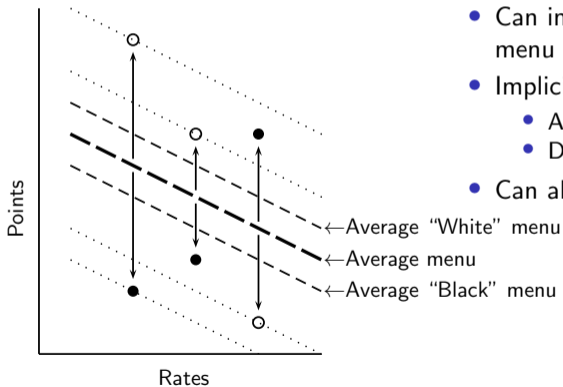
- To address this problem, we propose (i) a new, robust, statistical test of equality in menus and (ii) a new difference in menus (DIM) measure to compare two distributions of menus.
- Preliminary results using the 2018 Home Mortgage Disclosure Act (HMDA) data, show that black borrowers were offered different menus in terms of rates and points by lenders in the mortgage market.

Standard Regression Approach

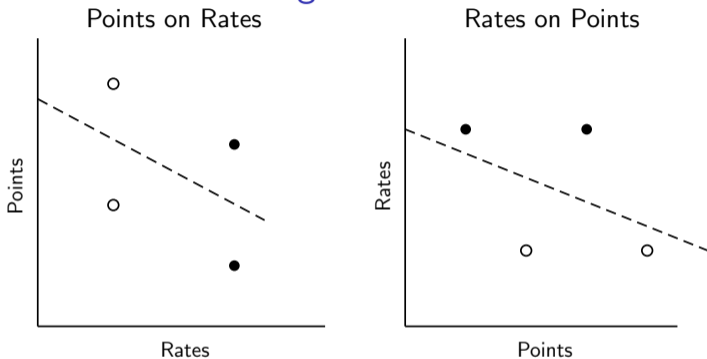
- Regression

$$\text{Points}_i = \alpha + \beta \text{Rate}_i + \varepsilon_i$$

- Can interpret regression residuals in terms of menu intercepts
- Implicit Assumption
 - All menus share the same slope
 - Different intercepts
- Can also regress Rates on Points.

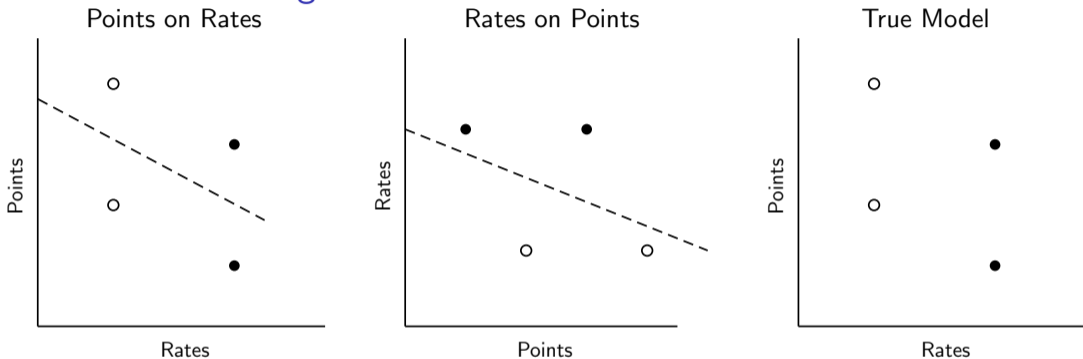


Regressions can contradict each other



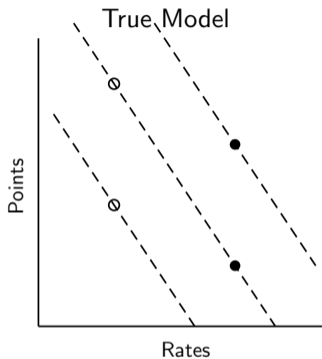
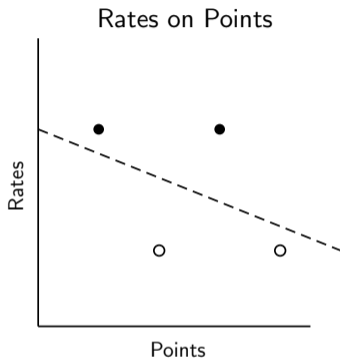
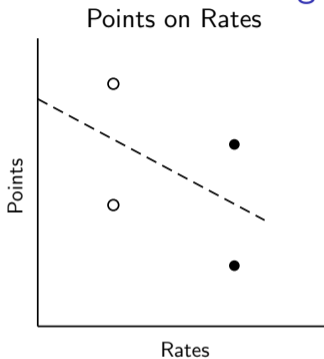
- Points on rates shows no discrimination [Bhutta and Hizmo, 2019]
- With the same data, Rates on points shows discrimination.
- “Reverse Regression” problem [Goldberger, 1984]

Regressions can contradict each other



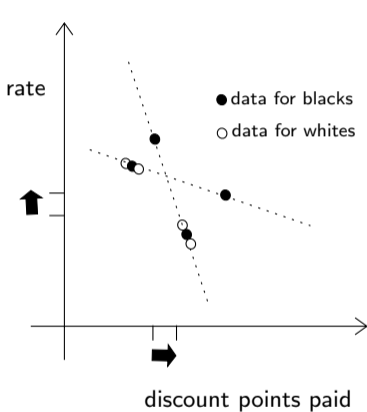
- Points on rates shows no discrimination [Bhutta and Hizmo, 2019]
- With the same data, Rates on points shows discrimination.
- “Reverse Regression” problem [Goldberger, 1984]
- True model
 - Neither regressions gets the true model.
 - What is the problem here?

Regressions can contradict each other

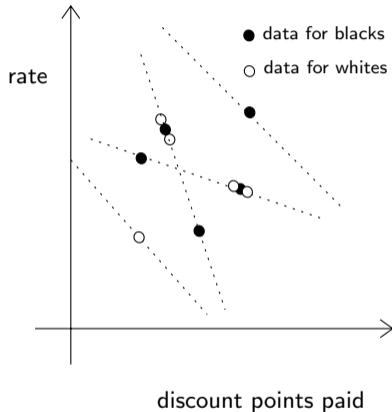


- Points on rates shows no discrimination [Bhutta and Hizmo, 2019]
- With the same data, Rates on points shows discrimination.
- “Reverse Regression” problem [Goldberger, 1984]
- True model
 - Neither regressions gets the true model.
 - What is the problem here?

Even comparing means can be misleading, when slopes can differ



(a) **False positive**, black consumers pay more in both rate and points

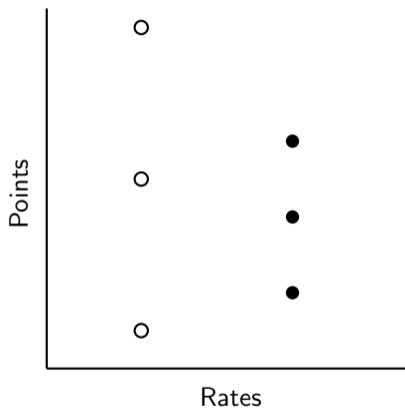


(b) **False negative**, black consumers pay the same average rate and points

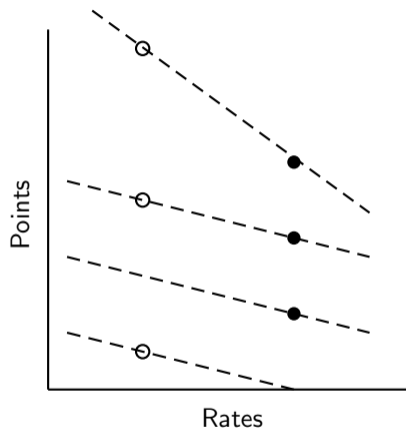
- This is the serious problem: adjusting for a range of slopes as in Bartlett et al. [2019] can also be problematic.

Discrimination

- Suppose that we get the following data

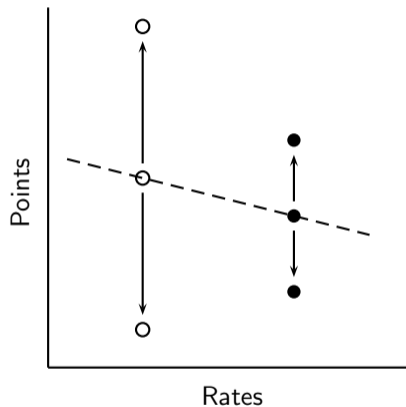


Discrimination



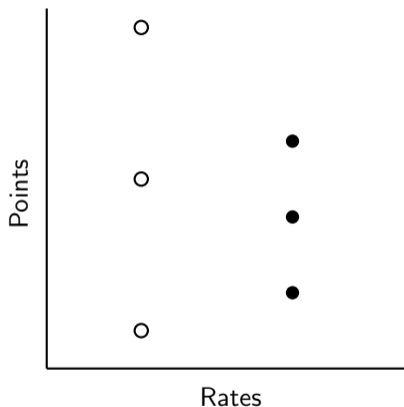
- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers

Discrimination



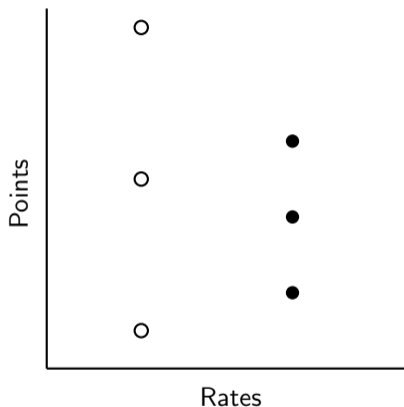
- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination

Discrimination



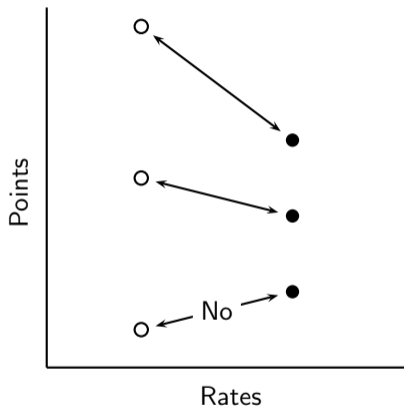
- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination
- Definition of discrimination:
Discrimination occurs two if different groups of people receive different distributions of menus.

Discrimination



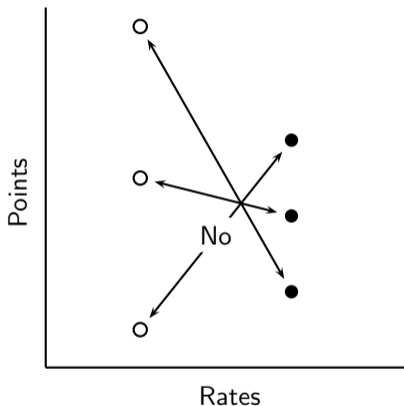
- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination
- Definition of discrimination:
Discrimination occurs two if different groups of people receive different distributions of menus.
- (Ignoring statistical error) If there are N borrowers of each type 1 and type 2, and we cannot find N downward-sloping menus such that we can match all type 1 borrowers with a type 2 borrower, then there discrimination exists.

Discrimination



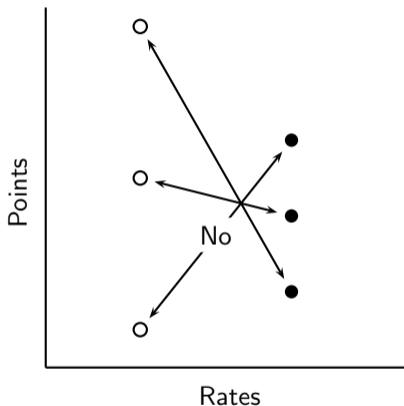
- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination
- Definition of discrimination:
Discrimination occurs two if different groups of people receive different distributions of menus.
- (Ignoring statistical error) If there are N borrowers of each type 1 and type 2, and we cannot find N downward-sloping menus such that we can match all type 1 borrowers with a type 2 borrower, then there discrimination exists.

Discrimination



- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination
- Definition of discrimination:
Discrimination occurs two if different groups of people receive different distributions of menus.
- (Ignoring statistical error) If there are N borrowers of each type 1 and type 2, and we cannot find N downward-sloping menus such that we can match all type 1 borrowers with a type 2 borrower, then there discrimination exists.

Discrimination



- Suppose that we get the following data
- True model: 4 menus
 - Discrimination: 1 white borrower gets a better menu than 1 black borrowers
- Regression finds no evidence of discrimination
- Definition of discrimination:
Discrimination occurs two if different groups of people receive different distributions of menus.
- (Ignoring statistical error) If there are N borrowers of each type 1 and type 2, and we cannot find N downward-sloping menus such that we can match all type 1 borrowers with a type 2 borrower, then there discrimination exists.
- If we find discrimination, can ask whether borrowers would like to switch to another group's menus.

Model

- For each consumer i , we observe their choice x_i from a menu m such that $x_i \in m$.
- Assumption 1: $x_i \in \mathbb{X}$ is finite, so a distribution of menus $m \sim M$ can be defined.
- Lenders offer a distribution of menus $m_1 \sim M_1$ for blacks and $m_2 \sim M_2$ for whites.
 - Is there equality in menus, such that $M_1 = M_2$?
 - If not, can we say that one set of menus is “better” than another set?
- For testing for equality of menus $M_1 = M_2$, Assumption 2: There exists some ex ante restrictions on whether consumers with choices x_1, x_2 could have faced the same menus.
 - Restriction from preferences: e.g. paying more in both rates and points is dominated.
 - Restriction from industry background: e.g. each discount point paid is worth between 1/8 to 1/4 in interest rate [Bartlett et al., 2019]. We use this, with robustness checks using 1/16 to 1/2 as the range.

Restrictions on whether consumers could have faced the same menus, from preferences

Fairly general restrictions from the assumption that {higher rate, paying more points} is dominated.

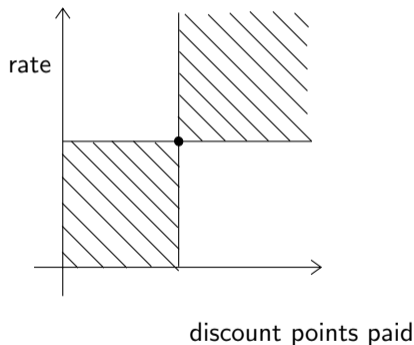
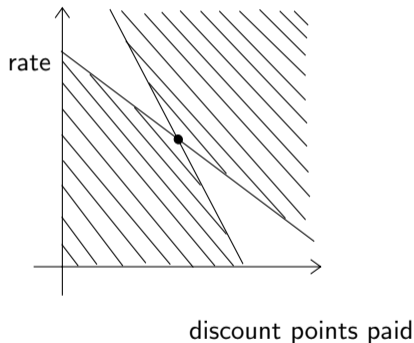


Figure: Restriction on what cannot be chosen from the same menus, based on dominance.

Restrictions on whether consumers could have faced the same menus, from industry background

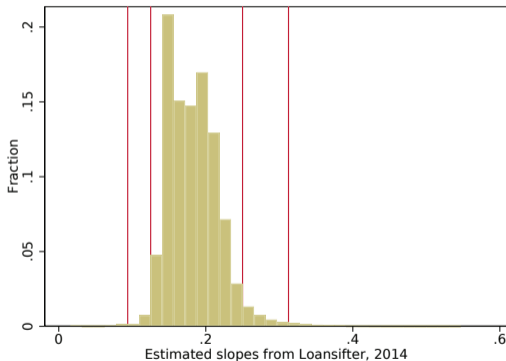
Can improve the restricted area if you know more about the setting:



Here, we use Bartlett et al. [2019]’s “rule of thumb” restriction that each point paid is worth $1/8$ to $1/4$ of rate, and $3/32$ to $5/16$ for robustness.

Menu restrictions, evidence

Looking at ratesheets, Bartlett et al. [2019]’s “rule of thumb” includes $\sim 95\%$ of all menus in Loansifter. Our robustness expansion includes $\sim 99\%$:



Source: 2014 LoanSifter data interpolated as in Fuster, Lo, and Willen [2019], slopes estimated by taking $(\text{Rate}_{103} - \text{Rate}_{99})/4$.

Sharp identification of equality in menus

Let $\phi(x_1, x_2)$ be an indicator function for whether the choices x_1, x_2 could have come from the same menus. Then, formalizing the “connecting each white choice to each black choice” intuition:

Theorem

A density of choices $f_1(x), f_2(x)$ can be generated from the same underlying distribution of menus $M_1 = M_2$ if and only if there exists a coupling $\pi(x_1, x_2) : \mathbb{X} \times \mathbb{X} \rightarrow [0, 1]$ with marginal distributions $E_{x_2} \pi(x_1, x_2) = f_1(x_1), E_{x_1} \pi(x_1, x_2) = f_2(x_2)$ s.t.:

$$E_{\pi(x_1, x_2)} \phi(x_1, x_2) = 1 \quad (1)$$

Here, the coupling $\pi(x_1, x_2)$ is the matching function.

Test statistic for equality in menus

To derive a workable test statistic, re-write Theorem 1 using an adaptation of Hall [1935]'s marriage theorem:

Corollary

A density of choices $f_1(x)$, $f_2(x)$ can be generated from the same underlying distribution of menus $M_1 = M_2$ if and only if and only if $\forall s : \mathbb{X} \rightarrow [0, 1]$:

$$\sum_{x_1} f_1(x_1)s(x_1) - \sum_{x_2} f_2(x_2) \max_{x_1} [\phi(x_1, x_2)s(x_1)] \leq 0 \quad (2)$$

This implies a test statistic:

$$\hat{T}_{n_1, n_2} = \max_{s: \mathbb{X} \rightarrow [0, 1]} \sum_{x_1} \hat{f}_1(x_1)s(x_1) - \sum_{x_2} \hat{f}_2(x_2) \max_{x_1} [\phi(x_1, x_2)s(x_1)] \quad (3)$$

Where the extent to which $\hat{T}_{n_1, n_2} > 0$ is incompatible with equality in menus.

Test statistic computation

Computationally, the test statistic \hat{T}_{n_1, n_2} can be efficiently computed via a linear program that searches over vectors of s, q :

$$\hat{T}_{n_1, n_2} = \max_{s, q} \sum_{x_1} \hat{f}_1(x_1) s(x_1) - \sum_{x_2} \hat{f}_2(x_2) q(x_2) \quad (4)$$

$$\text{s.t} \quad (5)$$

$$q(x_2) \geq \phi(x_1, x_2) s(x_1), \forall x_1 \quad (6)$$

$$0 \leq s(x_1) \leq 1 \quad (7)$$

Asymptotic distribution for test statistic

Using the numerical bootstrap of Hong and Li [2020]:

- Possible because the mapping

$T = \max_{s: \mathbb{X} \rightarrow [0,1]} \sum_{x_1} f_1(x_1)s(x_1) - \sum_{x_2} f_2(x_2) \max_{x_1} [\phi(x_1, x_2)s(x_1)]$, is directionally differentiable.

- Note that it is not fully differentiable: conventional bootstrap will fail.

$$\hat{J}_{n_1, n_2} = \frac{1}{\epsilon_n} (T(\hat{f}_1 + \epsilon_n h_1, \hat{f}_2 + \epsilon_n h_2) - T(\hat{f}_1, \hat{f}_2)) \quad (8)$$

$$c_\alpha = \inf \{c : P(\hat{J}_{n_1, n_2} \leq c) \geq \alpha\} \quad (9)$$

$$\text{95th percent CI: } [\hat{T}_{n_1, n_2} - \hat{c}_{0.975}, \hat{T}_{n_1, n_2} - \hat{c}_{0.025}] \quad (10)$$

$$\text{p-value: } P(\hat{J}_{n_1, n_2} \geq \hat{T}_{n_1, n_2}) \quad (11)$$

Where $\epsilon_n \rightarrow 0$ and $\sqrt{n}\epsilon_n \rightarrow \infty$, and h_1, h_2 are the errors in the multinomial f_1, f_2 the distribution of which can be simulated. Following Hong and Li [2020], can use double bootstrap to find a suitable rate for ϵ_n (for now, $\epsilon_n = n^{-1/4}$, where $n = n_1 + n_2$).

Data

- 2018 Home Mortgage Disclosure Act (HMDA) data, with has rates, points, lender, LTV, DTI, and race.
 - Important missing piece: FICO score.
 - We are working on getting confidential HMDA to address this.
- For new purchase mortgages, we build a sample based on exact matches of {lender, county, LTV category, DTI category, loan amount category, loan type} between blacks and whites, and conduct our analysis on this sample.
 - We also de-mean discount points by {lender, county, program}, since points may mean different things at different lenders with different programs.

Test of Inequality in Menus, Results

Table: Testing for equality in menus between black and white mortgage borrowers.

	Conforming	Jumbo	FHA	VA
Test statistic (\hat{T}_{n_1, n_2})	.081***	.068***	.063***	.055***
95% CI	[.075, .087]	[.023, .108]	[.054, .067]	[.044, .067]
p-value	<.002	<.002	<.002	<.002

Assuming each point paid is worth between 1/8 to 1/4 in rate, following Bartlett et al. [2019]. Confidence interval and p-value through 500 numerical bootstraps.

Test of Inequality in Menus, Robustness

Table: Testing for equality in menus between black and white mortgage borrowers, robustness.

	Conforming	Jumbo	FHA	VA
Test statistic (\hat{T}_{n_1, n_2})	.067***	.055***	.048***	.041***
95% CI	[.060, .074]	[.008, .086]	[.041, .055]	[.030, .051]
p-value	<.002	.002	<.002	<.002

Assuming each point paid is worth between 3/32 to 5/16 in rate. Confidence interval and p-value through 500 numerical bootstraps.

Measure of Differences in Menus

Thought experiment: **if black and white consumers switched menus, how much more on average would black consumers be willing to pay, in terms of interest rates, so as to remain indifferent as before.**

- We compute a lower bound for this, which we call a difference in menus $\underline{DIM}_{1 \rightarrow 2}$ metric.
- If consumers have *constant marginal utility over interest rates* and welfare weights are uniformly equal to 1, equivalent to a lower bound for the welfare change when black consumers switch to white menus:

$$\Delta W_{\mathcal{I}_1, 1 \rightarrow 2} = \sum_{i \in \mathcal{I}_1} \pi_{1 \rightarrow 2}(i, j) (u_i(m_j) - u_i(m_i)) \quad (12)$$

$$\geq \underline{DIM}_{1 \rightarrow 2} \quad (13)$$

- Could do with points also, but the lower bound is less informative there due to much more variance in rate than in points. Most of the literature also focuses on rate. . .

Measure of Difference in Menus, Assumptions

Two more assumptions to formalize this:

- Assumption 3: Paying more in both rates and points is dominated in terms of preferences.
 - If constant marginal utility in rate, direct mapping to change in within-group welfare under a switch of menus.
 - If not, still a lower bound for the average “willingness-to-pay” in terms of rate.
- Assumption 4: Menus are complete in at least one dimension.
 - In the sense that either all points are available, or all interest rates are available in all menus.

Importantly, we leave the assignment rule $\pi(i, j)$ for switching menus arbitrary: thus, we make no assumptions about whether menus are randomly assigned or not in our setting.

Measure of Differences in Menus, Illustration

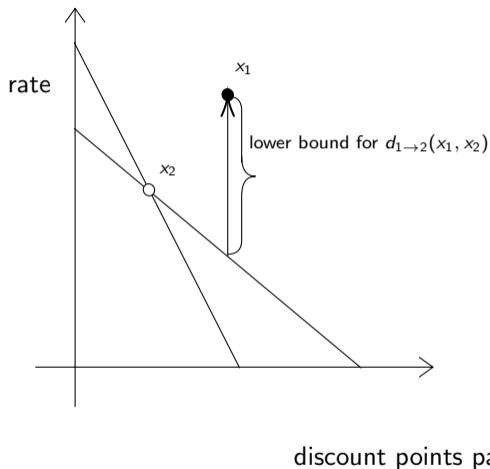


Figure: Lower bound for how much higher interest rate consumer 2 is willing to pay in order to remain indifferent with consumer 1's menu.

Measure of Differences in Menus, Computation

Suppose black consumers switch to white menus, under an arbitrary assignment rule. What is the mean increase in interest rate they are willing to pay to remain indifferent?

Computed by searching through all ways of assigning black consumers white menus through a “matching function” $\pi(x_1, x_2) : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$:

$$\underline{D\hat{I}M}_{1 \rightarrow 2} = \min_{\pi} E_{\pi} \underline{d}_{1 \rightarrow 2}(x_1, x_2) \quad (14)$$

$$\text{s. t.} \quad (15)$$

$$E_{x_2} \pi(x_1, x_2) = \hat{f}_1(x_1) \quad (16)$$

$$E_{x_1} \pi(x_1, x_2) = \hat{f}_2(x_2) \quad (17)$$

$$\pi(x_1, x_2) \geq 0 \quad (18)$$

Measure of Differences in Menus, Results

Table: A lower bound for the average interest rate increase (bps) needed for consumers to remain indifferent after switching to another group's menus

	Conforming	Jumbo	FHA	VA
Black to white ($\underline{\hat{D}IM}_{1 \rightarrow 2}$)	6.46	5.81	5.26	3.81
White to black ($\underline{\hat{D}IM}_{2 \rightarrow 1}$)	-14.21	-11.46	-13.45	-11.66

Assuming each point paid is worth between 1/8 to 1/4 in rate, following Bartlett et al. [2019].
(Standard errors to come - probably have to be through subsampling)

Measure of Differences in Menus, Robustness

Table: A lower bound for the average interest rate increase (bps) needed for consumers to remain indifferent after switching to another group's menus

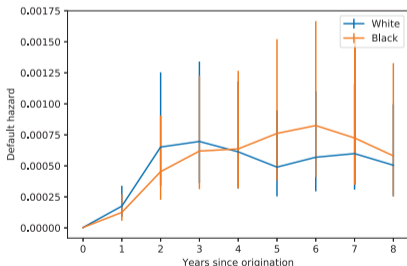
	Conforming	Jumbo	FHA	VA
Black to white ($\underline{\hat{D}IM}_{1 \rightarrow 2}$)	3.62	3.71	2.19	.88
White to black ($\underline{\hat{D}IM}_{2 \rightarrow 1}$)	-17.18	-13.62	-16.52	-14.43

Assuming each point paid is worth between 3/32 to 5/16 in rate.
(Standard errors to come - probably have to be through subsampling)

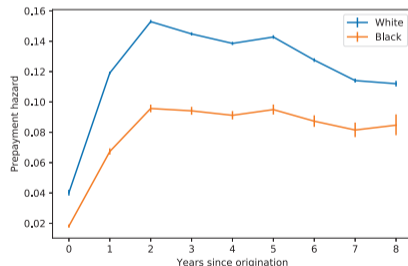
Defaults and Prepayments by Race

Does defaults explain the difference? First, it's not clear why lenders should care for mortgages sold to the GSEs, but even if they do, annual default rates don't vary much by race whereas prepayment rates vary a lot:

Figure: Defaults and Prepayments by Race



(a) Logit estimates of default probabilities by race



(b) Logit estimates of prepayment probabilities by race

Concluding Thoughts

- Fundamentally, this is about **equality of outcomes** (rates and points actually observed) vs **equality of opportunity** (menus).
- We can typically assess equality of outcomes, and most of the discrimination literature focuses on it, but equality of opportunity is more elusive.
 - Opportunity is typically unobserved, and we showed that heuristic statistics based on outcomes do not say much about opportunity, despite their intuitive appeal.
- We propose one new approach for detecting inequality in opportunity, which we call “discrimination in menus”, and showed that black borrowers are still discriminated against in the mortgage market.

References I

- Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2019. “Consumer-Lending Discrimination in the FinTech Era.” *Working Paper*.
- Bhutta, Neil, and Aurel Hizmo. 2019. “Do Minorities Pay More for Mortgages?” *Working Paper*.
- Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen. 2019. “The Time-Varying Price of Financial Intermediation in the Mortgage Market.” *Working Paper*.
- Goldberger, Arthur S. 1984. “Reverse Regression and Salary Discrimination.” *The Journal of Human Resources* 19(3): 293–318. ISSN 0022166X.
URL <http://www.jstor.org/stable/145875>.
- Hall, P. 1935. “On Representatives of Subsets.” *Journal of the London Mathematical Society* s1-10(1): 26–30. doi: 10.1112/jlms/s1-10.37.26.
URL <https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s1-10.37.26>.

References II

Hong, Han, and Jessie Li. 2020. “The numerical bootstrap.” *Ann. Statist.* 48(1): 397–412.
doi: 10.1214/19-AOS1812.
URL <https://doi.org/10.1214/19-AOS1812>.