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

Paul Willen

Federal Reserve Bank of Boston

paul.willen@bos.frb.org

David Hao Zhang[*]

Harvard Business School

dzhang@hbs.edu

September 23, 2020

Click for most recent version

Abstract

In the US, borrowers can choose to get lower interest rates on their mortgages by paying more discount points upfront. It is often observed that minority borrowers pay the same number of points as white borrowers controlling for the interest rate, but pay higher interest rates than white borrowers controlling for points. How can researchers tell if lenders charge minority borrowers more than observationally similar white borrowers? We show that the answer to this question involves a "menu problem," and intuitively appealing metrics of lender discrimination that are widely used in the literature can lead to false positives, false negatives, and contradictory results due to the existence of unobserved hetereogeneity in menus. This menu problem extends well beyond the mortgage market, but is often under-appreciated. To address this problem, we define (i) a new, robust, test statistic for equality in menus and (ii) a new difference in menus (DIM) metric for assessing whether one group of consumers would like to switch to another groups' menus, both based on pairwise dominance relationships in the data. We show how these metrics can be computed using methods from optimal transport, and devise a procedure for generating uniformly valid critical values in this class of problems based on directional differentiation. We implement our metrics on a new dataset matching 2018-2019 Home Mortgage Disclosure Act (HMDA) data to OptimalBlue rate locks. We find that Black and Hispanic borrowers were offered worse menus in terms of rates and points by lenders for conforming mortgages compared to Non-Hispanic White borrowers, and that the difference we detect can be largely explained by lender pricing differences across neighborhoods.

1 Introduction

The assessment of whether lenders overcharge minority borrowers in mortgage markets is an important topic both in terms of academic research and with regards to its policy relevance.[1] Many studies have found that minority consumers pay higher interest rates than observationally similar white consumers in the mortgage market.[2] While this can be interpreted as evidence that lenders systematically discriminated against minority borrowers by offering them worse terms on their mortgages, another potential explanation is that minority consumers were simply more constrained in their choices of discount points. In the US, borrowers can receive a lower rate on their mortgage in exchange for paying more discount points upfront. This discount points explanation still suggests that minority consumers may be disadvantaged in a broader sense, but has very different policy implications than one in which the lenders themselves are systematically offering minority consumers worse menus compared to white borrowers. Given data on rates and points, we study whether lenders offered minority consumers worse menus of rates and discount point options compared to observationally similar white consumers.

We show that the empirical assessment of whether lenders offered minority consumers worse menus of rates and points involves a non-trivial "menu problem," where the researcher seeks to assess differences in the distribution of menus given data on choices. Perhaps counter-intuitively, appealing heuristic approaches to analysis used in the literature, such as (1) controlling for rates/points (e.g. comparing the points paid by minority and white consumers getting the same rate) and (2) comparing means (e.g. checking if minority consumers on average pay both a higher interest rate and more discount points, or, more generally, adjusting by a rate-point trade-off from external sources) can lead to false positives, false negatives, and even contradictory assessments of lender discrimination in menus. Thus, the task of assessing whether one group faced a different distribution of menus than another group, or of testing for the existence of an "inequality in opportunity" more broadly speaking, is surprisingly complicated, though under-appreciated by the literature which tends to

---

[1]Since the financial crisis, the Department of Justice obtained settlements of well over $500 million for overcharging Black and Hispanic borrowers in violation of the Fair Lending Act, as explained in Bhutta and Hizmo (Forthcoming), including $335 million for Bank of America (on behalf of Countrywide), $175 million for Wells Fargo, and $55 million for JP Morgan Chase. Recently, on Jun. 12, 2019, Sen. Elizabeth Warren wrote on twitter: "For generations, lenders have given African American & Latino families fewer loans at worse terms than similar white borrowers. Tech alone won't fix the problem. A new analysis found that discrimination is hardwired into lending algorithms. I want answers." `https://twitter.com/senwarren/status/1138909674781237253`.

[2]See e.g. Black and Schweitzer (1985), Boehm, Thistle, and Schlottmann (2006), Bocian, Ernst, and Li (2008), Ghent, Hernández-Murillo, and Owyang (2014), Cheng, Lin, and Liu (2015), Bartlett et al. (2019). Relatedly, Munnell et al. (1996) and Tootell (1996) finds that black consumers are more likely to be rejected for mortgages.

focus on simple comparisons of outcomes.

The menu problem is pervasive in the mortgage discrimination literature. As we discussed, there are two main methods by which the literature has assessed discrimination in mortgage pricing. First, Courchane and Nickerson (1997), Black, Boehm, and DeGennaro (2003), and Bhutta and Hizmo (Forthcoming) looks at whether black borrowers paid more in points conditional on rate, for which Courchane and Nickerson (1997) and Black, Boehm, and DeGennaro (2003) found a differential while Bhutta and Hizmo (Forthcoming) did not in a sample of FHA mortgages. Second, Woodward (2008), Woodward and Hall (2012), Bartlett et al. (2019) compares the interest rate of minority and White borrowers after adjusting for points using a known range of rate-point trade-offs, and found that minorities pay more for mortgages. We show that both of these heuristic approaches to assessing discrimination in menus can lead to false positives, false negatives, and contradictory results. Furthermore, we find in our sample that the both the difference in methodology and the difference in samples are both needed to explain the distinct conclusions of Bartlett et al. (2019) and Bhutta and Hizmo (Forthcoming). Our robust solution to the menu problem would allow researchers to assess discrimination in mortgage markets in a more consistent and theoretically sound way.

The menu problem also extends well beyond the mortgage discrimination setting. For example, when workers make decisions that trades off wages and hours worked, one may wish to assess whether the fact that men work longer hours than women can explain the gender differences in wage-hour outcomes as in the model of Goldin (2014). Our results in Section 2 imply that popular measures of gender inequality such as the the gender pay gap conditional on hours, or even the gender pay gap after adjusting for all relevant average compensating differentials, is not necessarily informative about whether the data on outcomes can be explained by heterogenous preferences over hours worked. Generally speaking, the menu problem is relevant whenever we wish to assess equality in opportunity given data on outcomes while allowing for heterogeneous preferences across groups.[3] Therefore, the problem that we point out and the solution we propose may be of broader interest.

In this paper, we propose (i) a new test statistic for detecting differences in the distribution of menus offered to two groups and (ii) a new lower bound measure for assessing differences in menus (DIM) for the extent to which one group of consumers would like to switch to another group's menus in Section 3. Both metrics are based on pairwise dominance relationships in the data (i.e. getting a lower rate and paying fewer points dominates that with higher rates and higher points) which can be supplemented by by industry knowledge.

---

[3]The problem of unobserved menus also appears in demand estimation with unobserved choice sets, as surveyed in Crawford, Griffith, and Iaria (2019). However, the problem is different in that here we are not concerned with demand estimation, and are instead focused on assessing whether two consumer groups could have faced the same distribution of menus based on data on their choices.

Based on these pairwise relationships, we ask the question of whether the data *can* be rationalized by a model of equality in menus but heterogeneity in preferences, and if not we compute an average difference in menus perceived by one group of consumers when switching to another group's menus.

As a technical contribution, we derive a new approach to uniformly valid inference in optimal transport problems in Section 4 which we implement for our metrics. Our metrics can be written as solutions to optimal transport problems, which are computationally well-understood and can be efficiently computed (Galichon, 2016). However, conventional approaches to inference such as bootstrapping fails in this class of problems because the objective function can be non-differentiable (Fang and Santos, 2018). To do inference, we take the directional derivatives of the optimal transport objective with respect to its marginal distributions and apply the results of Fang and Santos (2018) and Shapiro (1991), which we then combine with a Bonferroni correction following Romano, Shaikh, and Wolf (2014) and McCloskey (2017) to address estimation error in the marginal distributions. We show that this approach leads to uniformly valid size control for hypothesis testing in optimal transport problems, and illustrate its efficacy in a Monte Carlo simulation. Our approach to inference may be useful for other researchers that may wish to conduct inference on the value of optimal transport problems, some of which are covered in Galichon (2016).

Empirically, in Section 5 we use our metrics to assess racial discrimination in the 2018-2019 Home Mortgage Disclosure Act (HMDA) data matched to Optimal Blue rate locks. We show that we can detect inequality in menus offered by the same lender and county and within narrow covariate groups for conforming mortgages for both Black and Hispanic borrowers. Furthermore, we show that the on average Black borrower getting conforming mortgages would be willing to increase their interest rate by at least 2.0 basis points in order to switch to the menus of non-Hispanic White borrowers. Similarly, Hispanic borrowers are on average willing to pay 1.5 basis points in interest rate in order to switch menus with non-Hispanic White borrowers. Our finding that racial differences in lender pricing remains relevant for conforming mortgages is consistent with Bartlett et al. (2019), although the amount of interest rate discrimination we detect is smaller in magnitude. On the other hand, we do not detect interest rate discrimination in FHA mortgages, which is consistent with Bhutta and Hizmo (Forthcoming). Finally, we detect large differences in menus across neighborhoods with varying levels of minority populations, and do not detect differences in menus by race after controlling for the racial distribution of neighborhoods, suggesting that differential pricing by neighborhood largely explains the differences in menus we detect.[4]

---

[4]The literature on pricing differences by neighborhood is mixed, with Nothaft and Perry (2002) and Haughwout et al. (2009) finding that neighborhood does not explain much in terms of differential interest

One may question whether assessing the differences in the menus faced by minority and white borrowers who are observationally similar within the same lender and county is a relevant definition of "discrimination," since there may be unobservables and neighborhood characteristics that affects risk which lenders may price on (Heckman, 1998). We have two reasons for focusing on comparing such observationally similar borrowers. First, as explained in Bartlett et al. (2019) for conforming mortgages the use of race or variables that are strongly correlated with race such as the racial composition of neighborhood, which the government-sponsored enterprises (GSEs) do not price on when mortgages are sold to them, is likely illegal, particularly for our sample of lenders who are unlikely to hold mortgages in portfolio. Therefore, our assessment of equality in menus conditional on our observables is of policy relevance even though the nomenclature of "discrimination" can be debated. Second, as shown in the recent studies of Kau, Fang, and Munneke (2019) and Gerardi, Willen, and Zhang (2020), minority borrowers have much lower prepayment rates and comparatively similar default rates compared to white borrowers conditional on observables, making loans from minority borrowers likely more valuable than those of white borrowers from a cash-flow perspective. Therefore, statistical discrimination based on expected racial differences in loan performance would not necessarily be able to explain the adverse pricing of loans for minorities even if it were legal.

The rest of this paper is structured as follows. Section 2 explains the motivation of our paper by exploring why heuristic approaches to analyzing discrimination in menus may be misleading. It also provides intuition for our approach. Section 3 formally defines our metrics for assessing discrimination in menus. Section 4 describes a methodology for conducting inference on our metrics. Section 5 shows our data and empirical results. Section 6 concludes.

2 Motivation and Intuition

In this section we discuss why intuitively appealing approaches for assessing discrimination in menus may be misleading, and provide intuition for our test of inequality in menus. By way of background, there are two dimensions of pricing for mortgages in the US, an upfront fee/discount points and the interest rate, where each point is worth 1% of the loan amount. Consumers can have the option of picking a particular rate and discount point combination that best suits their preferences and financial constraints, the choices of which

---

rates, while more recently Ghent, Hernández-Murillo, and Owyang (2014) finds large differences in mortgage interest interest rates by neighborhood. The authors of Bartlett et al. (2019) suggested that differential pricing by neighborhoods may be an important mechanism through which lenders overcharge minority borrowers in the age of FinTech: https://www.courant.com/business/hc-ls-bias-online-minority-borrowers-20181125-story.html.

from an example ratesheet we plot in Figure 1. In particular, borrowers can "pay discount points" to reduce their interest rate, or receive money from the lender to help cover their closing costs by "getting lender credit/negative points." The sense in which we think about lender discrimination in menus, then, is for minority borrowers to receive a worse rate-point schedule than white borrowers.
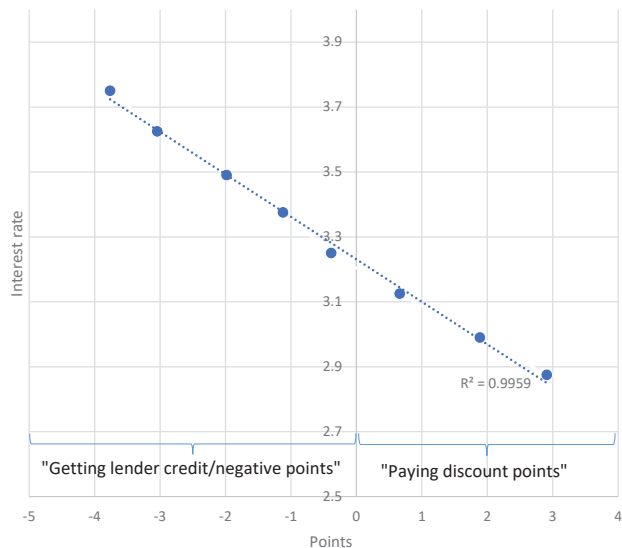

Figure 1: An example set of menu items from a lender ratesheet.

One natural heuristic approach to assessing whether lenders offered black and white borrowers different menus is to "control" for one dimension of the menu in a regression. That is, to look at whether minority borrowers who received the same interest rate as white borrowers paid more points, conditional on them being observationally similar. This was the approach used in Courchane and Nickerson (1997), Black, Boehm, and DeGennaro (2003), and Bhutta and Hizmo (Forthcoming). Relatedly, one may also look at whether minority borrowers who paid the same number of points as white borrowers received higher interest rates. However, in practice when there is unobserved heterogeneity in the menus being offered, such an approach can lead to false positives and negatives, which we illustrate in Figure 2. In that figure, we represent example data from minority and white borrowers using black and white dots, respectively, and menus by dotted lines. The researcher wishes to evaluate whether the minority and white borrowers were offered the same distribution of menus. In the left panel, Figure 2a shows a false positive situation in which minority borrowers paid more in rate controlling for points, and more points controlling for rate, even though lenders offered both minority and white consumers the same distribution of menus. The only difference driving this result is that minority consumers chose to pay fewer points on every menu. In the right panel in Figure 2b, we illustrate a false negative situation in which

minority consumers paid the same rate conditional on points but faced a worse distribution of menus as the bottom menu (the most advantageous menu) were only offered to white borrowers while the second to bottom menu were only offered to minority borrowers.
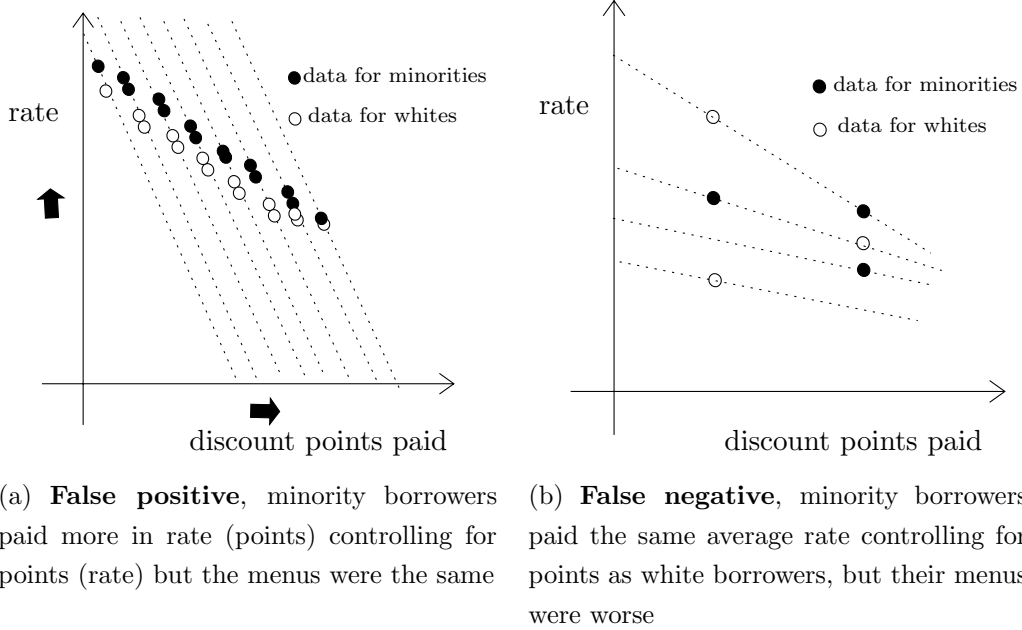

(a) **False positive**, minority borrowers paid more in rate (points) controlling for points (rate) but the menus were the same

(b) **False negative**, minority borrowers paid the same average rate controlling for points as white borrowers, but their menus were worse

Figure 2: False positives and false negatives from controlling for one direction of the menu

The core problem with the "controlling" for one dimension of the menu approach of assessing discrimination in menus is that it implicitly assumes that (i) all menus share the same shape, with unobserved heterogeneity in menus being due to an additive error term, and (ii) that this shape can be consistently estimated in the regression. In the situations in Figure 2a where consumers react the underlying heterogeneity in menus by changing their choices, the slopes of the menus suggested by the data are not the same as the true menu slopes, thus breaking assumption (ii) in a classic case of omitted variables bias.

Another way to see that assumption (ii) is almost always a problem is that, mechanically, regressing y on x yields a different slope than x on y, which implies regression estimates that could lead to different (and sometimes contradictory) estimates of discrimination depending on which menu dimension the researcher chooses to control for. For example, the situation in Figure 3 shows that it is possible for a regression of points on rate to show no discrimination against minorities while for an regression of rate on points to show discrimination with the same example data. This is not particular to the linear regression case, and as we show in Appendix Figure A.1 can appear with general conditional expectations. The possibility for contradiction can be viewed as a variation of the reverse regression problem

7

of Goldberger (1984). However, the menu problem goes far beyond the reverse regression problem, since even when the regressions/unconditional means are consistent, as in the case of Figures 2 and 4, the estimates can still be misleading.

(a) Points on rates shows no discrimination (b) With the same data, rates on points (flipped axes) shows discrimination
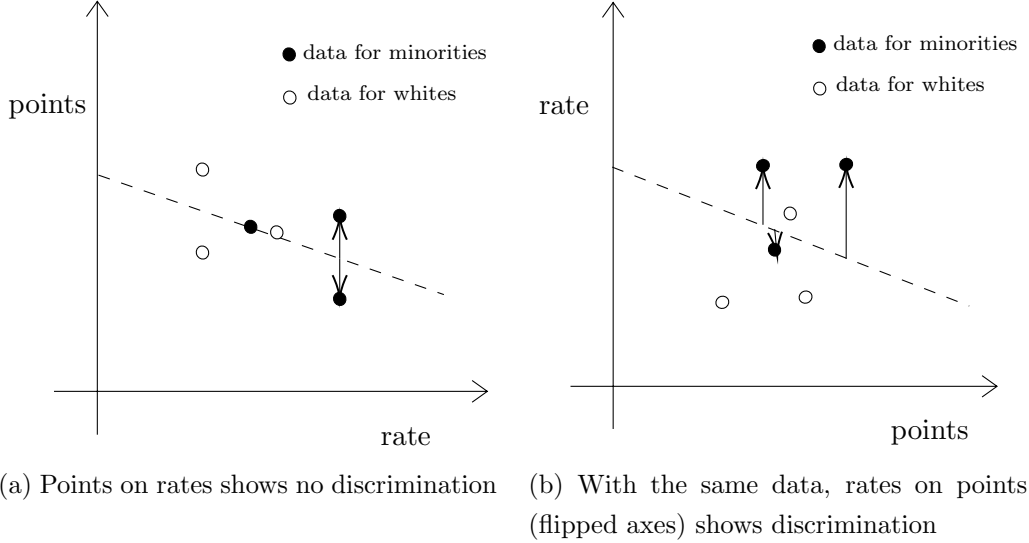
Figure 3: How the choice of which menu dimension to control for can lead to opposite findings of discrimination

A second intuitively appealing and seemingly surefire approach for assessing discrimination is to compare means, thus avoiding the problem of estimating the slope of the menu as in assumption (ii) of the first approach altogether. In other words, the researcher may wish to check if minority consumers paid more on average in both rates and discount points, such that they are disadvantaged in both dimensions, compared to observationally similar white borrowers. A variation of this approach is to take a pre-defined range of rate-point trade-offs as the slope estimated from external sources, which was done in Woodward (2008), Woodward and Hall (2012), and Bartlett et al. (2019). While this avoids the problem that regressions may inconsistently estimate menu slopes, it can still lead to false positives and false negatives when slopes are not constant across menus, thus breaking the assumption (i) of the regression approach. When slopes differ across lenders, Figure 4a illustrates how a false positive in which minority consumers pay more on average in terms of both rates and points but faced the same distribution of menus as white borrowers can occur, and Figure 4b illustrates a false negative possibility in which minority borrowers paid the same average rates and points as white borrowers but did face worse menus. This is a realistic problem because we know from ratesheet data that an unobserved heterogeneity in slopes does exist in the mortgage setting as the rate-point trade-off do vary substantially (Figure 8), perhaps due to cost and market power considerations.

8

(a) **False positive**, minority consumers paid more on average in both rates and points but their menus are the same

(b) **False negative**, minority consumers pay the same average rate and points as white borrowers, but their menus were worse
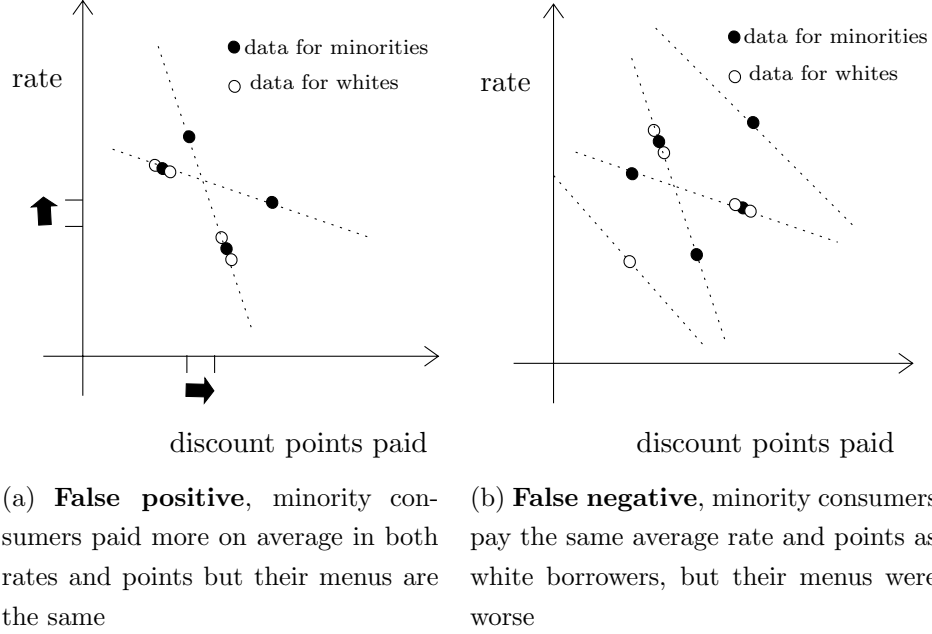
Figure 4: False positives and false negatives from checking if minority consumers paid more on average in both rates and points

More specifically, the mechanism for when comparing means would lead to false positives, as in the case of Figure 4a is that, in the example, minority borrowers react less to differences in the slopes of the rate-point menus compared to white borrowers. This can be a realistic possibility, because there are two directions in which constraints can drive borrower choices of mortgage points. First, if borrowers are cash constrained when getting the loan, they may prefer to pay fewer points (or, get more lender credit/negative points) and get a low closing cost mortgage regardless of what rate-point tradeoffs lenders offer. Second, if borrowers are debt-to-income (DTI) constrained, they may need to pay more points to buy down the rate so as to be able to borrow more.[5] Therefore, the simple observation that minority borrowers on average pay more than white borrowers may reflect the fact that minority borrowers are more constrained in their choices, which is a form of "disadvantage" that is not necessarily due to lenders discriminating against them by offering them different menus.

Furthermore, we speculate that the heterogeneity in menus may even be stronger in the wage-hour choice within labor markets than our mortgage setting, since workers can be quite diverse in terms of their choice sets. This makes the menu problem salient in those contexts and suggests that existing analyses of the gender pay gap using both heuristic (e.g. the gender pay gap conditional on hours worked) and structural approaches that rule out heterogeneity in the slope of the trade-off (e.g. Bell (2019)), though informative about inequality in

---

[5]See e.g.: `https://www.thetruthaboutmortgage.com/dti-debt-to-income-ratio/`.

outcomes, may not be informative about inequality of opportunity in labor markets.[6] While heterogeneity in the slopes of trade-offs is more difficult to deal with compared to restricting the heterogeneity across menus to an additive "productivity" term, Figure 4 shows that the consequences of assuming away this possibility can be quite severe, particularly in settings where it is plausible that one group of agents may respond less to differences in slope than another group.

As we explained, both Heuristic 1 ("controlling" for a dimension of the menu) and Heuristic 2 (comparing means) can lead to false positives and false negatives as assessments of discrimination in menus. As a way to resolve this problem, we define alternative metrics for assessing discrimination in menus based on whether the data *can* be rationalized by a model in which all groups of borrowers faced the same distribution of menus. Ignoring for now sampling error to build intuition, the common thread in the false positive situations of Figure 2a and 4a is that there exists a common distribution of menus underlying both minority and white choices, in the sense of there being a possible one-to-one match between minority and white borrowers where within each match both the borrowers' choices could have come from the same menu. This is the criteria we use for assessing equality in menus.

Our approach can also detect discrimination in menus where the heuristic approaches lead to false negatives. In particular, we illustrate in Figure 5 how the situation in which a regression of rates on race controlling for points would lead to a false negative would fail the one-to-one matching condition under the assumption that borrowers with pairwise strictly dominated choices (ie. paying more in terms of both rates and points) could not have shared menus with one another. In other words, there is no way to construct a common distribution of menus for minority and white borrowers that explains the data. Analogously, the false negative example from comparing means in Figure 4b also fails our criterion.

---

[6]More specifically, Bell (2019) suggested a new methodology for estimating and adjusting for average compensating differentials to deal with the wage-amenities menu problem, akin to the second heuristic we discussed for the mortgage context but with a novel estimation strategy for the wage-amenity trade-off. We show that the approach of adjusting by an average compensating differential can still be problematic when there exists heterogeneity in the slope of the wage-amenity trade-off.
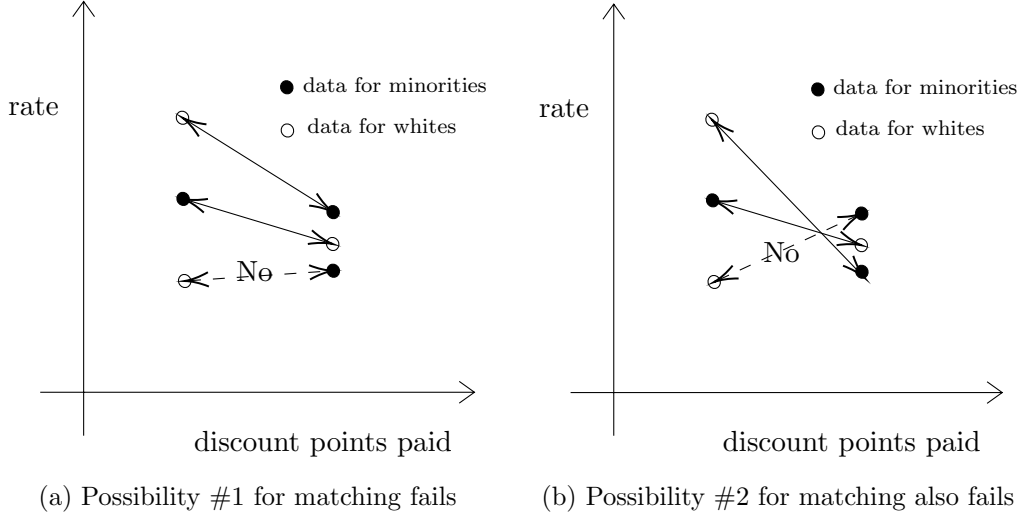
(a) Possibility #1 for matching fails   (b) Possibility #2 for matching also fails

Figure 5: How data from Figure 2b fails a "perfect matching" condition

Since our metrics are based whether a set of preferences can rationalize the data under equality in menus, they are *robust* to the false positives that is problematic for the heuristic approaches to assessment. Furthermore, in some situations, such as those in Figures 2b and 4b, our metrics can detect discrimination when heuristic approaches fail to do so. Nevertheless, a drawback of our approach is that it stills leave some possibility for false negatives, because the mere existence of a set of preferences that explains the data under equality in menus does not mean that it is the true set of preferences. This is the weakness compared to experimental that may allow the researcher to directly observe menus, but the advantage of our approach is that it requires only data on outcomes and few assumptions on the data generating process. Furthermore, we show empirically in Section 5 that our approach is useful in the mortgage setting as we can strongly reject equality in menus for conforming mortgages.

3   Robust metrics for assessing discrimination in menus

In this section we define our robust metrics for assessing discrimination in menus, building on the intuition from Section 2 that equality in menus should imply the existence of a one-to-one "match" between minority and white consumers who could have faced the same menus in the data. We keep our model fairly simple, where menus are simply treated as a collection of items, which we define more formally in Section 3.1. We then define a direct test metric for equality in menus Section 3.2, a more welfare relevant differences in menus metric for whether one group of consumers would like to switch to another distribution of menus in

Section 3.3. We leave inference on these metrics to Section 4.

3.1 Model

A menu item has values over k dimensions of attributes,[7] which we encode by $x \in \mathbb{X} \subset \mathbb{R}^k$. A menu $\mathbf{m} \subseteq \mathbb{X}$ is a collection of such menu items which are presented to the borrowers. When presented with a menu $\mathbf{m}$, we observe borrower i making a choice which maximizes their utility over menu items $u_i(x)$. That is, we observe choices x_i where:

$$x_i \in \arg\max_x \{u_i(x) : x \in \mathbf{m}\} \tag{1}$$

To keep the distribution of menus Lebesgue measurable and to implement the inference procedure of Section 4, we make the simplifying assumption that the set of items available to choose from is finite:

Assumption 1. *(Finiteness) The set of possible menu items, $\mathbb{X}$, is finite.*

In empirical contexts where choices are continuous but can be binned, Assumption 1 is approximately true. Under Assumption 1, we can consider a probability distribution over possible menus $\mathbf{m} \sim \mathbf{M}$. This setup is fairly general and follows from consumers having standard (ie. complete and transitive) preferences over menu items.[8]

3.2 A robust test for inequality in menus

Suppose borrowers who are observationally similar in groups 1 and 2 faces distributions of menus, $\mathbf{M}_1, \mathbf{M}_2$, respectively, such that $\mathbf{m}_1 \sim \mathbf{M}_1$ for borrowers in group 1 and $\mathbf{m}_2 \sim \mathbf{M}_2$ for borrowers in group 2. The researcher wishes to compare the distribution menus across two groups of borrowers. More specifically, testing for equality of menus can be written as testing the null hypothesis that the distribution of menus being offered to both groups are equal. That is:

$$\mathcal{H}_0 : \mathbf{M}_1 = \mathbf{M}_2 \tag{2}$$

$$\mathcal{H}_1 : \neg\mathcal{H}_0 \tag{3}$$

---

[7]In our context, the two dimensions of a menu item in the mortgage context are rates and discount points.

[8]The assumption that consumers maximize utility over menu items does rule out more behavioral representations of preferences over menus such as Gul and Pesendorfer (2001), in which the existence of some menu items may "tempt" consumers to change their rankings of other menu items. In that case, our statistical test of inequality in menus would still be valid, but the interpretation of our more welfare relevant differences in menus metric would be nuanced.

To go from data on choices to statements about menus, we place restrictions on what choices could have been plausibly made from the same menus in terms of borrower preferences. For mortgages, it is plausible to assume that paying more in both interest rates and discount points is a dominated choice (and indeed would not be offered as a choice by the loan originator), which is the intuition we used in Figure 5 to reject equality in menus in that situation. We formalize this as Assumption 2:

Assumption 2. *(Dominance Restriction on Preferences) Paying more in rates and points is dominated. More formally, let $x_1 = [r_1, y_1], x_2 = [r_2, y_2]$ where r_1, r_2 represents rates and y_1, y_2 represents points. Then, if $r_1 > r_2, y_1 \geq y_2$ or $r_1 \geq r_2, y_1 > y_2$ then $u_i(x_1) < u_i(x_2), \forall i$.*

We illustrate in Figure 6 the restrictions on what observed choices may come from on the same menu under Assumption 2. The observed choice of the borrower, shown as the black dot, implies that they did not have the lower-left dashed quadrant available on their menu since otherwise they would have chosen it. Similarly, any choice on the upper-right quadrant could not have been from the same menu as the choice indicated, as that agent would have an incentive to switch to that choice. Note that while Assumption 2 is defined in the form of preferences, it could have also been defined in terms of menus which would have led to the same restrictions. Defining it in terms of preferences, though, is because it will allow us to define a welfare-relevant metric for assessing differences in menus later on.
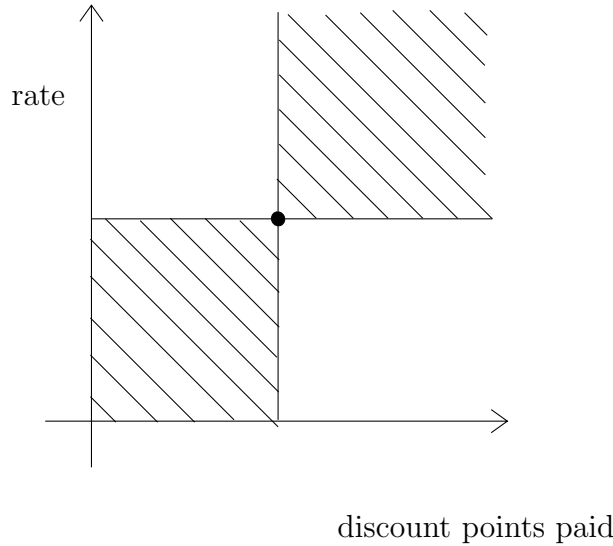

Figure 6: Restriction on what cannot lie on the same menus from dominance.

While Assumption 2 is sufficient for rejecting equality in menus in the example situation of Figure 5, in our empirical application it is too weak of a restriction to be informative. For our empirical analyses, we further adopt the industry rule of thumb of Bartlett et al. (2019)

that each point paid reduces the interest rate on a mortgage by between 1/8 to 1/4 for conforming mortgages, with an expanded range for FHA mortgages. This is an assumption about menus rather than about preferences, which we formalize as Assumption 3:

Assumption 3. *(Restriction on Menus) In menus, each point paid reduces the rate by between* $[a, b]$. *More formally,* $x_1 = [r_1, y_1]$ *and* $x_2 = [r_2, y_2]$ *can only lie on the same menu* $\{x_1, x_2\} \subseteq \mathbf{m}$ *only if:*

$$a \leq \frac{r_2 - r_1}{y_2 - y_1} \leq b \tag{4}$$

with $x_1 = [r_1, y_1], x_2 = [r_2, y_2]$, r_1, r_2 *representing rates and* y_1, y_2 *representing points.*

We illustrate in Figure 7 the effect of defining a menu set based on Assumption 3. As Figure 7 indicates, range of possible choices that could have come from the same menu as the consumer with the choice illustrated by the black dot is more restricted under this assumption, compared with using only dominance relationships in terms of preferences as in Assumption 2. Thus, this improves our ability to detect discrimination in menus. Nevertheless, as we mentioned earlier, Assumption 2 is still needed to make welfare comparisons of menu distributions in Section 3.3.
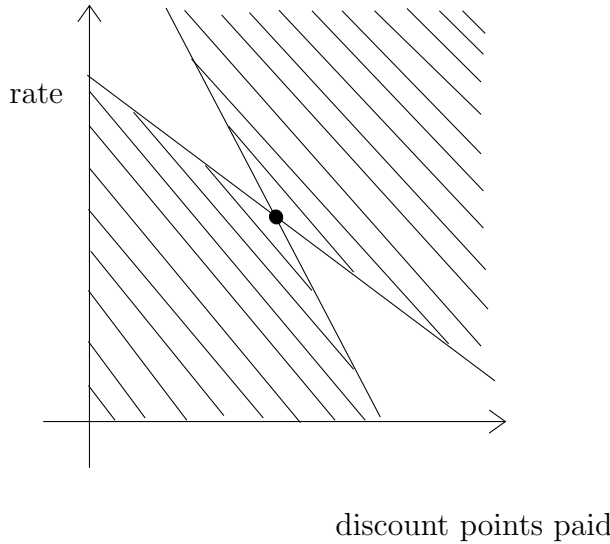

Figure 7: Restriction on what cannot lie on the same menus from Bartlett et al. (2019)'s "rule of thumb".

Empirically, we find that Bartlett et al. (2019)'s industry rule of thumb which motivated Assumption 3 covers a vast majority of menus based on data from a sample of lender

ratesheets which enumerates the rate and discount point menus.[9] Using the LoanSifter data from Fuster, Lo, and Willen (2019) in 2014, we estimate slopes of menus within their sample of 30 year conforming, purchase mortgages across seven different MSAs (Chicago, Houston, Los Angeles, Miami, New York City, Seattle, and San Francisco) and a range of loan amounts, FICO scores, and LTVs. The sample construction is discussed in more detail in Fuster, Lo, and Willen (2019). We estimate the slopes of the rate-point trade-off by taking the difference in the interpolated rate from 0 points to 2 points and dividing by 2. In this sample, the Bartlett et al. (2019)'s rule that each point paid is worth between 1/8 to 1/4 of a point covers 94.4% of all ratesheet observations for conforming mortgages. For FHA mortgages, we use an expanded rule that each point is worth between 1/32 to 1/4 in rate, which covers 97.6% of all observations. We illustrate this in Figure 8.

(a) Conforming mortgages, Bartlett et al. (2019) restriction in red

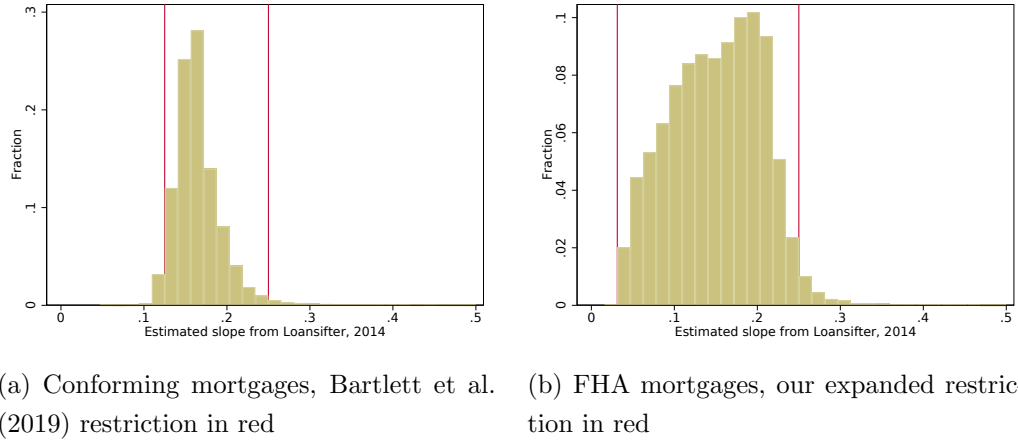(b) FHA mortgages, our expanded restriction in red

Figure 8: Ratesheet evidence for our menu slopes Assumption 3.

Note: these figures are constructed using 2014 LoanSifter rate sheet data from Fuster, Lo, and Willen (2019) for conforming and FHA mortgages. The slope of the rate-point tradeoff is estimated by taking the interpolated rate from 0 to 2 points and dividing by 2. The red lines in Figure 8a represent 1/8 and 1/4, and the red lines in Figure 8b represent 1/32 and 1/4.

There is substantial heterogeneity between the menu slopes across lender-weeks which we see in Figure 8, perhaps reflecting market power or lender and time specific costs. The existence of this heterogeneity interacted with possible differences in preferences between the two groups makes the "comparing means" heuristic and its variations prone to false positives and negatives.

Under Assumption 3, we can define an indicator function for whether choices x_1, x_2 could

---

[9]It is possible for the algorithm that generates the ratesheet to be discriminatory if they rely on problematic variables such as neighborhoods that correlate highly with race, and it is also possible for lenders to give special discounts on top of ratesheets.

have come from the same menu:

$$\phi(x_1 = [r_1, y_1], x_2 = [r_2, y_2]) = \begin{cases} 1 \text{ , if } a \leq \frac{r_2 - r_1}{y_2 - y_1} \leq b \\ 0 \text{ , otherwise.} \end{cases} \tag{5}$$

After defining this function, we have the following identification result for when vectors of choice probabilities $\mathbf{p}_1 = [p_1(x_1), p_1(x_2), \ldots], \mathbf{p}_2 = [p_2(x_1), p_2(x_2), \ldots]$ from observationally similar groups of borrowers can be rationalized under the null hypothesis of equality in the distribution of menus $\mathcal{H}_0 : M_1 = M_2$:

Theorem 1. *Under Assumptions 1 and 3, choice probabilities $\mathbf{p}_1, \mathbf{p}_2$ can be generated from the same underlying distribution of menus $\mathbf{M}_1 = \mathbf{M}_2$ if and only if there exists a coupling with probability mass function $\boldsymbol{\pi} : \mathbb{X} \times \mathbb{X} \to [0, 1]$ with implied marginal densities $\sum_{x_2} \boldsymbol{\pi} = \mathbf{p}_1, \sum_{x_1} \boldsymbol{\pi} = \mathbf{p}_2$ such that:*

$$T \equiv 1 - E_{\boldsymbol{\pi}} \phi(x_1, x_2) = 0 \tag{6}$$

Proof. See Appendix A.1.1. □

Theorem 1 formalizes the intuition from Section 2 that equality in menus should imply a way to "match" observations to one another such that each pair can come from the same set of menus. In particular, $\boldsymbol{\pi}$ serves as a coupling or "matching function" where each of its entries $\pi(x_1, x_2)$ represents the extent to which choice probabilities $p_1(x_1)$ and $p_2(x_2)$ came from the same menu, and the requirement of Equation (6) is that this coupling should lie entirely in the area where $\phi(x_1, x_2) = 1$. Based on Theorem 1, we are ready to define a test statistic that looks at by how much is this matching deficient:

Definition 1. *Our test statistic for equality in menus, $\hat{T}$, given sample choice probabilities $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2$, is:*

$$\hat{T} = \min_{\boldsymbol{\pi}} 1 - E_{\boldsymbol{\pi}} \phi(x_1, x_2), s.t. \sum_{x_2} \boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1} \boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0 \tag{7}$$

Where the extent to which $\hat{T} > 0$ indicates a failure of the perfect matching condition in Theorem 1 in sample, which is evidence against equality in menus. The statistic $\hat{T}$ in Equation (7) is a finite dimensional optimal transport objective that can be efficiently computed using a linear program. Galichon (2016) reviews optimal transport methods and its application to economics. To deal with sampling error inherent in $\hat{T}$, we discuss in Section 4 on how critical values for $\hat{T}$ from Equation (7), as an optimal transport objective, can be consistently simulated.

3.3 Metric for assessing differences in menus

The direct test of inequality in the distribution of menus in Section 3.2, while indicative of discrimination in menus, has the drawback that it may not be the object of interest for researchers. A statistical rejection of equality in menus may by itself carry little information about welfare, since the fact that the distribution of menus presented to one group is different in some aspect than the distribution of menus presented to another group may not be of welfare consequence, a problem which Abadie (Forthcoming) discusses in more detail. Rather, for the purposes of comparing menus across two distributions we want a metric for assessing whether menus from one group is meaningfully "better" than that of another group. For this purpose, we ask the question: *if black consumers were instead assigned white menus, how much better off would they be?*

Conceptually, we consider the object of interest to be the change in welfare when black consumers were instead assigned white menus, under an assignment rule $\pi(i, j)$ that map each consumer $i \in \mathcal{I}_1$ from group 1 to the menu of consumer $j \in \mathcal{I}_2$ from group 2. Giving all consumers the same welfare weight, this objective can be represented by Equation (8):

$$\Delta W_{\mathcal{I}_1 \to \mathcal{I}_2, \boldsymbol{\pi}} = \sum_{i \in \mathcal{I}_1, j \in \mathcal{I}_2} \pi(i, j)(u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i)) \tag{8}$$

To get at $\Delta W_{\mathcal{I}_1 \to \mathcal{I}_2, \boldsymbol{\pi}}$, we proxy for the utility difference $u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i)$ through a metric $d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j)$ which measures by how much would consumer i be willing to increase the interest rates on their loan in order to switch from menu $\mathbf{m}_i$ to menu $\mathbf{m}_j$. If consumers have constant marginal utility over interest rates such that utility can be represented as $u_i(x = [r, y]) = r + f(y)$, then this metric is directly proportional to the utility change for consumer i after switching to menu j:

$$u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i) = d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j) \equiv \sup\{\delta \in \mathbb{R} : u_i(\{x + \delta e^r, x \in \mathbf{m}_j\}) \geq u_i(\mathbf{m}_i)\} \tag{9}$$

Where e^r represents a basis vector that is equal to 1 at the location indexing interest rates. Since $d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j)$ measures by how much would consumer i be willing to increase the interest rates to switch from $\mathbf{m}_i$ to $\mathbf{m}_j$, even if consumers do not have constant marginal utility over interest rates it is still meaningful as a "willingness to pay" metric. In the rest of this section we will show how we can compute an informative lower bound for this metric given the data, $\underline{d}_{i \to j}(x_i, x_j) \leq d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j), x_i \in \mathbf{m}_i, x_j \in \mathbf{m}_j$, which then leads our differences in menus measure.[10]

---

[10]We note that while we define our $\underline{DIM}_{1 \to 2}$ over "willingness to pay" in terms of interest rates, we could have also defined it using points, but due to the possible existence of cash on hand constraints which are

To define our lower bound, we make an additional assumption that menus are complete in points, such that all choices of points are available to borrowers, which we formalize as Assumption 4. This is an approximation since lenders may limit the choices of points to certain decimals (e.g. 0.134, 0.266, ...) rather than literally the full range, but the implications of such small gaps in menus is likely small. Another complication is that there may be information constraints on the part of borrowers such that they do not "see" their full choice set (ie. some borrowers may not know that they can pay/receive points), but as long as these information constraints are held constant in the counterfactual where they switch to another group's menus, our lower bound metric would remain valid and our results would not be impacted.

Assumption 4. *(Completeness) The menus are complete in discount points. More specifically, $\forall \mathbf{m}, \forall y', \exists x = [r, y'] \in \mathbf{m}$.*

The effect of this Assumption 4 is illustrated in Figure 9. Under the assumption that the mortgage menus are complete in discount points, we can meaningfully say that the minority borrower whose choice is represented by the black dot would have preferred the menu of the white borrower whose choice is represented by the white dot, because there exists a level of discount points such that all possible choices that are in white borrower' menu dominates the minority borrower's choice. Otherwise, the minority borrower might not have preferred the white borrower's menu because the white borrower's menu could have been a singleton that the minority borrower dislikes. Therefore, adding in the approximately true assumption of completeness in points sharpens the comparison of menus.

---

likely binding for many consumers constant marginal utility is very unlikely to hold for points which makes for a weaker welfare interpretation.
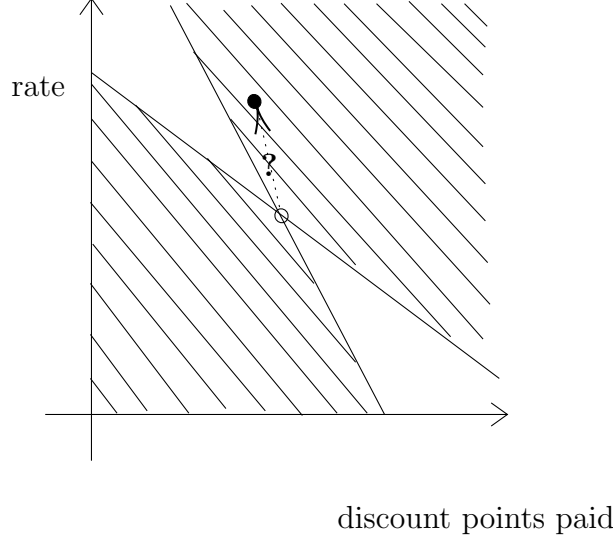
Figure 9: Impact of assuming that menus are complete in either rates or points

We illustrate in Figure 10 how we can construct a lower bound for the willingness to pay in terms of interest rates $\underline{d}_{i \to j}(x_i, x_j)$ under Assumption 4. There, a consumer 1 who made a choice x_1 from an unobserved menu $\mathbf{m}_1$ has made a choice that is dominated (in terms of paying a higher rate at the same level of points) by any possible the menu of consumer 2 who made a choice x_2 from a menu $\mathbf{m}_2$. This implies that, by revealed preference for consumer 1, the menu that consumer 2 faced is better than consumer 1's menu, or $\mathbf{m}_2 \succ_1 \mathbf{m}_1$. For consumer 1 to possibly become indifferent between $\mathbf{m}_1$ and $\mathbf{m}_2$, $\mathbf{m}_2$ needs to be shifted up by at least the amount indicated in the figure in the dimension of interest rates. In other words, a lower (sharp) lower bound for $d_{1 \to 2}(\mathbf{m}_1, \mathbf{m}_2)$ is $\underline{d}_{1 \to 2}(x_1, x_2)$ in the sense that consumer 1 is willing to pay at least $\underline{d}_{1 \to 2}(x_1, x_2)$ more in interest rate in order to get consumer 2's menu.
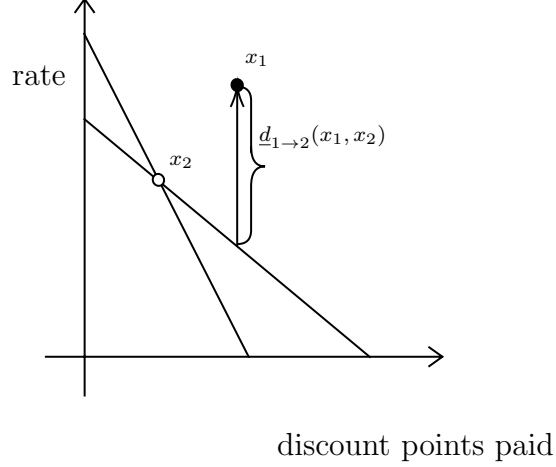
19

Figure 10: Lower bound for consumer 2's willingness to pay a higher interest rate in order to get consumer 1's menu.

Formalizing the intuition from Figure 10, we define our lower bound for the willingness of consumer 1 to switch to consumer 2's menu under Assumptions 1 to 4 as follows:

$$\underline{d}_{1\to2}(x_1, x_2) \equiv r_1 - r_2 + a\max(y_1 - y_2, 0) + b\min(y_1 - y_2, 0) \leq d_{1\to2}(x_1, x_2) \qquad (10)$$

Where in the first line j indexes points, and in the second line we take take it to the mortgage setting and let $x_1 = [r_1, y_1], x_2 = [r_2, y_2]$ where r_1, r_2 are rates and y_1, y_2 are points. The lower bound for how much consumer 1 would be willing to pay to switch to the menus of consumer 2, $\underline{d}_{1\to2}(x_1, x_2)$, then allows us to define our differences in menus (DIM) metric:

Theorem 2. *Under Assumptions 1 to 4, choice probabilities* $\mathbf{p}_1, \mathbf{p}_2$ *implies a DIM metric:*

$$\underline{DIM}_{1\to2} = \min_{\boldsymbol{\pi}} E_{\boldsymbol{\pi}}\underline{d}_{1\to2}(x_1, x_2), s.t. \sum_{x_2} \boldsymbol{\pi} = \mathbf{p}_1, \sum_{x_1} \boldsymbol{\pi} = \mathbf{p}_2, \boldsymbol{\pi} \geq 0 \qquad (11)$$

where $\underline{DIM}_{1\to2}$ *serves as a lower bound for the average willingness to pay in terms of interest rates for borrowers in group 1 to switch menus with borrowers in group 2. Furthermore, if borrowers have constant marginal utility in interest rate, then:*

$$\underline{DIM}_{1\to2} \leq \Delta W_{\mathcal{I}_1\to\mathcal{I}_2, \boldsymbol{\pi}} \qquad (12)$$

Proof. Follows from the definitions in Equation (9) and Equation (10). $\qquad\square$

Theorem 2 shows that, when all consumers have the same constant marginal utility over interest rates (normalized to 1), our DIM metric is as a lower bound for the change in welfare for when consumers in group 1 are instead assigned menus from group 2 in an arbitrary way

$\Delta W_{\mathcal{I}_1, 1 \to 2, \pi_{1 \to 2}}$. If instead consumers do not have constant marginal utility over interest rates, then the $\underline{DIM}_{1 \to 2}$ metric could still be interpreted as the average increase in interest rates consumers in group 1 would be willing to pay in order to switch to menus from group 2. Furthermore, by Theorem 1, equality in menus would imply that $\underline{DIM}_{1 \to 2} \leq 0$, so a finding that $\underline{DIM}_{1 \to 2} > 0$ is also rejection of equality in menus in a welfare relevant way.

The sample analogue of the DIM metric follows immediately from Definition 2.

Definition 2. *Our empirical differences in menus metric, $\underline{D\hat{I}M}_{1 \to 2}$, given choice probabilities $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2$, is:*

$$\underline{D\hat{I}M}_{1 \to 2} = \min_{\boldsymbol{\pi}} E_{\boldsymbol{\pi}} \underline{d}_{1 \to 2}(x_1, x_2), s.t. \sum_{x_2} \boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1} \boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0 \tag{13}$$

In terms of inference, the sample DIM metric in Definition 2 is also the value of an finite dimensional optimal optimal transport problem. We discuss inference in the following Section 4.

4 Inference for Optimal Transport

In this section we devise a new procedure for conducting hypothesis testing on the values of optimal transport problems, which includes our metrics derived in Section 3. This is needed because the objective values of our model can be non-differentiable and therefore simple bootstrap methods are not consistent (Fang and Santos, 2018). More specifically, we consider hypothesis testing on the value φ of a finite dimensional optimal transport problem with cost function $\phi(x_1, x_2), x_1, x_2 \in \mathbb{X}$ and marginal distributions $\mathbf{p}_1, \mathbf{p}_2$:

$$\hat{\varphi}(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) = \min_{\boldsymbol{\pi}} E_{\boldsymbol{\pi}} \phi \text{ s.t. } \sum_{x_2} \boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1} \boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0 \tag{14}$$

Where the hypothesis is in the form of the value of the optimal transport $\varphi(\mathbf{p}_1, \mathbf{p}_2)$ as a function of the true marginal distributions $\mathbf{p}_1, \mathbf{p}_2$ are less than or equal to some value φ_0:

$$\mathcal{H}_0 : \varphi(\mathbf{p}_1, \mathbf{p}_2) \leq \varphi_0 \tag{15}$$

$$\mathcal{H}_a : \varphi(\mathbf{p}_1, \mathbf{p}_2) > \varphi_0 \tag{16}$$

The form of the null hypothesis in Equation (15) is especially relevant to us because both our test of equality in menus (that is, whether $T \leq 0$) and our lower bound DIM metric (that is, whether $DIM \geq DIM_0$) can be expressed in terms of it. We provide a methodology to conduct this hypothesis test by looking at the asymptotic distribution of

$\hat{\varphi}$ and finding a critical value to compare the observed $\hat{\varphi}$ to under $\mathcal{H}_0$. This can then be inverted into a (one-sided) confidence interval for the true value of φ. As an overview, we combine the directional derivatives approach of Fang and Santos (2018) and Shapiro (1990) with a size correction of Romano, Shaikh, and Wolf (2014) and McCloskey (2017) which allows us to conduct hypothesis testing for optimal transport with uniform size control. In doing so, we prove the directional differentiability for optimal transport problems on finite domains, and show how the general approach can be implemented as a Linear Program with Complementarity Constraints (LPCC).

We make use of the Hadamard directional differentiation definition in Fang and Santos (2018) adapted to the optimal transport setting. Here, the value of an optimal transport represents a map $\varphi : \mathbb{D}_\varphi \to \mathbb{R}$, where $\mathbb{D}_\varphi = \mathcal{P}_\mathbb{X} \times \mathcal{P}_\mathbb{X}$, $\mathcal{P}_\mathbb{X}$ is the set of probability measures on $\mathbb{X}$. Let $\mathbb{D}_0 = \{P_1 - P_2 : P_1 \in \mathcal{P}_\mathbb{X}, P_2 \in \mathcal{P}_\mathbb{X}\}$ be the set of possible differences in probability measures, and $\theta = \{p_1, p_2\}$ be the marginal distributions, then:

Definition 3. *A map $\varphi : \mathbb{D}_\varphi \to \mathbb{R}$ is said to to be Hadamard directionally differentiable at $\theta \in \mathbb{D}_\varphi$ tangentially to the set $\mathbb{D}_0$, if there is a continuous linear map $\varphi'_\theta : \mathbb{D}_0 \to \mathbb{R}$ such that:*

$$\lim_{n \to \infty} \left\| \frac{\varphi(\theta + t_n \mathbf{h}_n) - \varphi(\theta)}{t_n} - \varphi'_\theta(\mathbf{h}) \right\| = 0 \tag{17}$$

for all sequences $\{\mathbf{h}_n\} \subset \mathbb{D}_0$ and $\{t_n\} \subset \mathbb{R}^+$ such that $t_n \to^+ 0, \mathbf{h}_n \to \mathbf{h} \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + t_n \mathbf{h}_n \in \mathbb{D}_\varphi$ for all n.

The main difference between the Hadamard directional differentiability in Definition 3 from Fang and Santos (2018) and the typical notion of differentiability is that t_n is restricted to be positive. That is, loosely speaking we are only considering the derivative "in the direction $\mathbf{h}$" for each $\mathbf{h}$.

We show in Theorem 3 that the value of all Monge-Kantorovich optimal transport problems with bounded cost functions on finite spaces is Hadamard directionally differentiable in the sense of Definition 3. In particular, Theorem 3 is a generalization of Sommerfeld and Munk (2018) which shows that the Wasserstein metric on finite spaces, which is the value of an optimal transport problem with the cost function restricted to distance metrics, is directionally differentiable.[11]

Theorem 3. *The value φ of a optimal transport problem with cost function $\phi(x_1, x_2), x_1, x_2 \in \mathbb{X}$, where $M = \sup |\phi| < \infty$ and $\dim(\mathbb{X}) < \infty$, is Hadamard directionally differentiable, with*

---

[11]It is also related to Tameling, Sommerfeld, and Munk (2019) who prove that the Wasserstein distance on countable metric spaces is directionally differentiable. Our Theorem 3 can be similarly extended to countable metric spaces under the assumption that the cost function ϕ is continuous.

derivative equal to:

$$\varphi'_{\mathbf{p}_1,\mathbf{p}_2}(\mathbf{h}_1,\mathbf{h}_2) = \max_{\mathbf{u},\mathbf{v}\in\Psi^*(\mathbf{p}_1,\mathbf{p}_2)} \mathbf{h}_1^T\mathbf{u} + \mathbf{h}_2^T\mathbf{v} \tag{18}$$

where $\Psi^*(\mathbf{p}_1,\mathbf{p}_2) = \{\mathbf{u},\mathbf{v} : \mathbf{p}_1^T\mathbf{u} + \mathbf{p}_2^T\mathbf{v} = \varphi(\mathbf{p}_1,\mathbf{p}_2), u(x_1) + v(x_2) \leq \phi(x_1,x_2)\forall x_1,x_2\}$ *is the set of dual solutions to the linear programming problem, for all* $\{\mathbf{p}_1,\mathbf{p}_2\} \in \mathbb{D}_\varphi$, *tangentially to the set* $\mathbb{D}_0$.

Proof. See Appendix A.1.2. □

Since $\mathbb{D}_0$ is closed and under i.i.d. sampling $\{\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2\}-\{\mathbf{p}_1,\mathbf{p}_2\}$ approaches a multivariate normal distribution, Theorem 2.1 of Fang and Santos (2018) immediately implies that:

$$r_n[\varphi(\{\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2\}) - \varphi(\{\mathbf{p}_1,\mathbf{p}_2\})] = \varphi'_{\{\mathbf{p}_1,\mathbf{p}_2\}}(r_n[\{\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2\} - \{\mathbf{p}_1,\mathbf{p}_2\}]) + o_p(1) \tag{19}$$

Such that the asymptotic distribution of $\varphi(\{\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2\})$ is obtained via the directional Delta method. The remaining challenge in applying in Theorem 3 for inference is that the true $\mathbf{p}_1,\mathbf{p}_2$ in Equation 19 is not known, and thus the directional derivative must be estimated. We deal with this problem using the logic of Romano, Shaikh, and Wolf (2014) and McCloskey (2017). Suppose we obtain a confidence band for $[\mathbf{p}_1,\mathbf{p}_2]$ at level β such that:

$$\limsup_{n\to\infty} \Pr([\mathbf{p}_1,\mathbf{p}_2] \in \hat{\mathcal{P}}_\beta) \geq 1 - \beta \tag{20}$$

Then, the directional derivative within this confidence band can be computed as:

$$\hat{\varphi}'_\beta(\mathbf{h}_1,\mathbf{h}_2) = \max_{\mathbf{u},\mathbf{v}\in\Psi(\mathbf{p}_1,\mathbf{p}_2):[\mathbf{p}_1,\mathbf{p}_2]\in\hat{\mathcal{P}}_\beta} \mathbf{h}_1^T\mathbf{u} + \mathbf{h}_2^T\mathbf{v} \tag{21}$$

Where $\Psi = \{\mathbf{u},\mathbf{v} : \mathbf{p}_1^T\mathbf{u} + \mathbf{p}_2^T\mathbf{v} \leq \varphi_0, u(x_1) + v(x_2) \leq \phi(x_1,x_2)\forall x_1,x_2\}$ are the set of dual solutions under the null hypothesis $\mathcal{H}_0 : \varphi \leq \varphi_0$.

Next, we can draw h_1, h_2 from a multivariate Normal distribution with the asymptotic covariance matrix of $\hat{p}_1, \hat{p}_2$, we can compute a critical value at level $1 - \alpha + \beta$:

$$\hat{c}^a_{1-\alpha+\beta} = \inf\{c \in \mathbb{R} : \Pr(\hat{\varphi}' \leq c) \geq 1 - \alpha + \beta\} \tag{22}$$

In what follows, we will prove uniform coverage for a test based on checking whether the observed value $\hat{\varphi}$ is greater than the critical value $\hat{c}^a_{1-\alpha+\beta}$ or not, and suggest a computationally tractable version of it as a linear program with complementarity constraints

(LPCC).[12]

Corollary 1. *Suppose we have uniform confidence bands for $[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta$ which provide uniform coverage as in Equation (20), then under $\mathcal{H}_0 : \varphi \leq \varphi_0$:*

$$\limsup_{n \to \infty} \sup_{[\mathbf{p}_1, \mathbf{p}_2] \in \mathcal{P}_0} \Pr(\hat{\varphi} \geq \hat{c}^a_{n, 1-\alpha+\beta}) \leq \alpha \qquad (23)$$

where $\hat{c}^a_{n, 1-\alpha+\beta}$ is computed as in Equation (22).

Proof. See Appendix A.1.3. □

Corollary 1 implies that uniformly valid hypothesis testing for the value of φ can be conducted by first computing a set of uniform confidence bands $[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta$, and then maximizing over all directional derivatives within these bands as in Equation (21). Computationally, directly maximizing over the directional derivative defined in Equation (21) is difficult because it involves optimizing over a non-linear dual value constraint $\mathbf{p}_1^T \mathbf{u} + \mathbf{p}_2^T \mathbf{v} \leq \varphi_0$. To deal with this, we replace it with a complementary slackness condition $\boldsymbol{\pi}^T \mathbf{s} = 0, \boldsymbol{\pi} \geq 0, \mathbf{s} \geq 0$ where $u(x_1) + v(x_2) + s(x_1, x_2) = \phi(\mathbf{p}_1, \mathbf{p}_2)$ which implies that the elements of $\boldsymbol{\pi}$ and $\mathbf{s}$ cannot be positive simultaneously. Following the operations research shorthand, we represent this constraint by $\boldsymbol{\pi} \leq 0 \perp \boldsymbol{s} \geq 0$. The derivation of complementary slackness conditions like this can be found in standard texts on optimal transport/linear programming, and in particular Hsieh, Shi, and Shum (2017) uses a similar set of conditions for their projection method. Based on this equivalency, the problem of finding critical values for the null hypothesis $\mathcal{H}_0 : \varphi \leq \varphi_0$ versus the alternative $\mathcal{H}_a : \varphi > \varphi_0$ be written as the following LPCC:

$$\hat{\varphi}'_\beta(\mathbf{h}_1, \mathbf{h}_2) = \max_{\mathbf{u}, \mathbf{v}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{s}, \boldsymbol{\pi}} \mathbf{h}_1^T \mathbf{u} + \mathbf{h}_2^T \mathbf{v} \qquad (24)$$

$$\sum_{x_2} \boldsymbol{\pi} = \mathbf{p}_1 \qquad (25)$$

$$\sum_{x_1} \boldsymbol{\pi} = \mathbf{p}_2 \qquad (26)$$

$$E_{\boldsymbol{\pi}} \phi \leq \varphi_0 \qquad (27)$$

$$u(x_1) + v(x_2) + s(x_1, x_2) = \phi(\mathbf{p}_1, \mathbf{p}_2) \qquad (28)$$

$$[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta \qquad (29)$$

$$\boldsymbol{\pi}, \boldsymbol{s} \geq 0 \qquad (30)$$

$$\boldsymbol{\pi} \leq 0 \perp \boldsymbol{s} \geq 0 \qquad (31)$$

---

[12]As explained in Hsieh, Shi, and Shum (2017), LPCCs are well-understood computationally and are implemented in software such as Knitro: https://www.artelys.com/docs/knitro/2_userGuide/complementarity.html.

To test our approach, we conduct a simulation with two possibilities for points $\{0, 1\}$ and five possibilities for rate $\{3, 3.25, 3.5, 3.75, 4\}$, where each point is worth between $1/8$ and $1/4$ in rate. Furthermore, black and white borrowers choose each rate-point options with probability $\frac{1}{10}$ such that the null discrimination of no discrimination in menus is satisfied. We compute $\hat{\mathcal{P}}_\beta$ using the plug-in sup-t band of Montiel Olea and Plagborg-Møller (2019), let $\beta = \frac{1}{10}\alpha$, and show the simulated probability that we reject equality in menus $\mathcal{H}_0 : \varphi_0 = 0$ at the 1%, 2.5%, 5%, and 10% levels in Table 1. As Table 1 shows, our approach has approximately correct size across a wide range of sample sizes and significance levels.

Table 1: Control of our size-corrected directional derivative approaches to inference

	Significance level			
Sample size	1%	2.5%	5%	10%
$n_1 = n_2 = 500$	1.2	2.6	4.4	10.3
$n_1 = n_2 = 1000$	0.5	2.2	4.7	9.6
$n_1 = n_2 = 5000$	0.8	2.3	4.8	9.8
$n_1 = n_2 = 10000$	1.0	3.4	5.8	10.8
$n_1 = n_2 = 50000$	0.9	2.8	5.0	9.8

Note: Computed via 2000 sample draws and 500 draws of $\mathbf{h}_1, \mathbf{h}_2$ from the estimated asymptotic multivariate Normal distribution for $\mathbf{p}_1, \mathbf{p}_2$ within each sample draw.

Compared to existing methodology that could be applied to the optimal transport context, the advantage of our procedure is that it achieves uniform coverage without being overly conservative. In particular, Hsieh, Shi, and Shum (2017) has a projection method for parameter inference in mathematical programming problems, which is a broader set of problems than optimal transport, but their approach is conservative.[13] Another approach that is theoretically valid in this setting is the m-out-of-n subsampling of Politis and Romano (1994), but in addition to requiring the researcher to choose a subsample size m it can require large samples for convergence.

We look at the control of existing approaches at the 5% significance level under our simulation setting in Appendix Table A.1. In that table, HSS (2017) refers to to the projection method of Hsieh, Shi, and Shum (2017), and m out of n subsampling approach refers to the method of Politis and Romano (1994), and in the final column the size-corrected direc-

---

[13]In our empirical context, this conservativeness tend to make the confidence intervals uninformative. In addition, we note that our approach cannot be readily extended to all mathematical programming problems covered in Hsieh, Shi, and Shum (2017) because not all mathematical programming programs are directionally differentiable.

tional derivatives approach is replicated from Table 1 for comparison. Table A.1 shows that Hsieh, Shi, and Shum (2017) rejects the null with probability far less than 5%, implying that it is conservative. On the other hand, the m out of n subsampling appears to be slow at converging to a rejection probability of 5% for all values of m we tried, and appears to have larger than desired rejection probability for the sample sizes that are relevant to our empirical setting.

In summary, in this section we devised a new procedure for hypothesis testing in optimal transport that is uniformly valid and not overly conservative for our empirical context. Our empirical analysis in the following Section 5 shows that we are able to strongly reject equality in menus for conforming mortgages using our methodology.

5   Empirical Estimates of Mortgage Discrimination

5.1   Data

We apply our methodology to a new dataset constructed via matching the 2018-2019 public Home Mortgage Disclosure Act (HMDA) data to OptimalBlue rate locks. The public HDMA data contains information on borrower race and ethnicity, and we take the borrowers with a HMDA derived race of "Black or African American" as our sample of Black borrowers, borrowers with a derived ethnicity of "Hispanic or Latino" as our sample sample of Hispanic borrowers, and borrowers with a HMDA derived race of "White" along with a HMDA derived ethnicity of "Not Hispanic or Latino" as our sample of Non-Hispanic White borrowers.

A limitation of the public HMDA data is that contains no information on credit scores. To deal with this problem, we create a new dataset by matching the 2018-2019 HMDA data to rate locks from the OptimalBlue platform. OptimalBlue is used for rate locking for about 30% of mortgages originated in the US, and has been used in the literature for analyses of price dispersion in mortgage markets.[14] We use a two-step matching process to match OptimalBlue and HMDA. In the first step, we exactly match on the interest rate of the loan, the loan amount, loan purpose, loan term, zip code/census tract, year, and allow a loan to value (LTV) difference of no more than 5%. Then in the second step, we find a correspondence between the lender ID in the OptimalBlue dataset and the lender identifier in the HMDA dataset, and use it to further restrict the matches. The resulting set of matches is of high quality: for the conforming and FHA new purchase mortgages we focus on, we were able to match 63% of locks to at least one loan in HMDA, and uniquely match 59% of locks to exactly one loan in HMDA. Not all locks turn into loans since borrowers can back

---

[14]See e.g. Bhutta, Fuster, and Hizmo (2019) and Bhutta and Hizmo (Forthcoming).

out, and our match rates are comparable to a 66% "lock pull-through rate," which is the rate at which locks turn into originated loans, that we understand based on industry sources.

Starting from the uniquely matched HMDA-OptimalBlue matched dataset, we further restrict our analysis to standard 30 year, new purchase fixed rate first lien mortgages on owner-occupied site built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortizing features. Table 2 compares the HMDA data, which includes the complete set of mortgages originated in the US with such characteristics, to our matched sample. We find that our matched sample has very similar average loan sizes, LTV, rates, points, and the percent composition of Black and Hispanic borrowers compared to the HMDA data, as can be seen from Table 2. This shows that lenders in the OptimalBlue sample may behave similarly to lenders not using the OptimalBlue platform. One known caveat to this is that lenders using the OptimalBlue platform tend to be smaller lenders, with the larger lenders being more likely to have their own platform for rate locking. Since these smaller lenders are not likely to keep any loans in portfolio, any use of signals not used by the GSEs to price mortgages is likely not allowed by law (Bartlett et al., 2019).

Table 2: Comparison of means between the 2018-2019 HMDA data and our HMDA-OptimalBlue matched data

	Loan Size	LTV	Rate	Points	% Black	% Hispanic	N
Panel A: Conforming mortgages							
HMDA data	$261,566	84.4	4.50	0.08	4.5	9.0	*3,730,152*
Matched sample	$258,205	83.5	4.59	0.12	3.9	8.6	*817,588*
Panel B: FHA mortgages							
HMDA data	$215,144	96.1	4.57	0.07	14.1	19.7	*1,437,088*
Matched sample	$220,031	95.7	4.63	0.13	13.7	18.8	*360,202*

Note: This table compares the 2018-2019 HMDA data and our HMDA-OptimalBlue matched sample for 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4, and outliers for rate below 2 and above 10.25, were excluded.

To control for the impact of lender and borrower characteristics within each loan program, we exactly match observations from black and white borrowers without replacement on groups of covariates, taking a random observation when multiple white borrowers can be matched to a black borrower. This creates a sample containing equal numbers of black and white borrowers that are exactly matched on their covariates. If lenders offered black and white borrowers the same distribution of menus conditional on covariates, our covariate-matched sample of black and white borrowers should then have faced the same distribution of menus. The covariates that we used to match are lender, county, month of lock, and 8

categories of FICO scores and 9 categories of LTVs as defined in the GSE Loan-Level Price Adjustment (LLPA) matrix. Thus, we control for the effects of the interactions of all these covariates in our assessment of equality in menus. This set of controls is similar to what is used in Bartlett et al. (2019). Furthermore, we compute the rate spread to the Freddie Mac Weekly Survey rate during the week of the rate lock, following Bhutta and Hizmo (Forthcoming).

Summary statistics for our matched sample of conforming, purchase mortgages is in Table 3. In particular, we find that black borrowers paid a slightly higher rate than white borrowers, by 4.7 basis points. They also paid more points, by 2.3 basis points. Therefore, black borrowers' outcomes are on average worse than white borrowers' in both dimensions, but as we showed in Figure 4 this can still be explained by differences in preferences rather than differences in the distribution of underlying menus. For our empirical analysis, we de-mean within each lender-county-month covariate group and round rates to the nearest eighths and points to the nearest 1/2, and we show in Table 3 rows this does not really affect the mean differences in rates or points.[15]

As shown in Table 3, Black borrowers paid 4.8 basis points more in interest rate and 2.5 basis points more in points for conforming mortgages. On the other hand, they only paid 2.9 basis points more in interest rate and -3.2 basis points less in points for FHA mortgages. While this comparison of means cannot be interpreted as evidence for or against discrimination in menus (as we noted in Section 2), it does show that the distribution of data underlying conforming and FHA mortgages are different. Nevertheless, we will show in the next Section 5.2 that this difference in the distribution of data between conforming and FHA mortgages is however not sufficient by itself to reconcile the results of Bartlett et al. (2019) and Bhutta and Hizmo (Forthcoming), the former of which found evidence of discrimination in conforming mortgages while the latter did not in FHA mortgages. The choice of heuristic used also contributes to the discrepancy in results.

---

[15]Both the HMDA data and the OptimalBlue data contains information about discount points paid which sometimes disagree with one another. While the literature has used both data sources, we focus on the HMDA information because it appears to have less measurement error. Our results using the OptimalBlue data on points are qualitatively similar. More specifically, we find in Appendix Table A.2 that regression of the HMDA information on points on origination charges and total loan costs has a much stronger R^2, with a coefficient closer to 1, compared to the OptimalBlue information on points. Furthermore, in a regression controlling for the HMDA points the effect of OptimalBlue points has minimal additional explanatory power for origination charges and total loan costs. We interpret this as suggestive evidence for there being more measurement error in the OptimalBlue definition of points.

Table 3: Summary statistics on the covariate matched sample

Panel A: Black and Non-Hispanic White Covariate Matched Sample

	Conforming			FHA		
	Black	White	*Difference*	Black	White	*Difference*
Raw						
Rate Spread (bps)	36.2	31.4	*4.8*	48.3	45.4	*2.9*
Points (bps)	12.6	10.1	*2.5*	10.6	13.6	*-3.0*
De-meaned & rounded						
Rate Spread (bps)	3.1	-1.7	*4.8*	2.2	-0.7	*2.9*
Points	1.5	-0.9	*2.4*	-1.6	1.6	*-3.2*
Sample size	6,398	6,398		4,711	4,711	

Panel B: Hispanic and Non-Hispanic White Covariate Matched Sample

	Conforming			FHA		
	Hispanic	White	*Difference*	Hispanic	White	*Difference*
Raw						
Rate Spread (bps)	38.3	34.5	*3.8*	49.3	45.4	*3.9*
Points (bps)	15.6	11.8	*3.8*	13.5	13.3	*-0.2*
De-meaned & rounded						
Rate Spread (bps)	2.5	-1.2	*3.7*	2.6	-1.2	*3.8*
Points	2.3	-1.3	*3.6*	-0.1	0.2	*-0.3*
Sample size	14,758	14,758		6,156	6,156	

Note: This table lists the summary statistics for our Black and White as well as Hispanic and Non-Hispanic White lender-county-month and covariate matched sample of 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were dropped, and all rate spreads were between -1.25 and 3. De-meaned & rounded rates and points were demeaned by lender-county-month and covariate group and then rounded into the nearest eighths for rate and 1/2 for points.

5.2 Results from heuristic analyses

We show the results from the heuristic approaches we discussed our sample in Table 4. In Table 4, Heuristic 1 in columns (1)-(2) refers to the approach of comparing points paid controlling for rate which was used in Courchane and Nickerson (1997), Black, Boehm, and DeGennaro (2003), and Bhutta and Hizmo (Forthcoming), and Heuristic 2 in columns (3)-(8) refers to the comparing means after adjusting for points using external values for the rate-point trade-off which was the general approach used in Woodward (2008), Woodward and Hall (2012), and Bartlett et al. (2019). The trade-offs used for Heuristic 2, in columns (3), (4), (5), and (6), respectively, were 1/8 and 1/4 for conforming and 1/32 and 1/4 for FHA mortgages. Finally, Columns (7)-(8) shows results from an alternative form of Heuristic 1, where we control for points and compare rate instead.

Table 4: Assessments of lender discrimination using two heuristic approaches in the Black and Non-Hispanic White matched sample

| | Heuristic 1 | | Heuristic 2 | | | | Alternate Heuristic 1 | |
	Conforming	FHA	Conforming		FHA		Conforming	FHA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	points	points	rate_1/8	rate_1/4	rate_1/32	rate_1/4	rate	rate
black	4.059***	-0.732	2.977***	3.324***	2.650***	2.223***	2.728***	2.612***
	(0.989)	(1.321)	(0.325)	(0.373)	(0.434)	(0.508)	(0.320)	(0.436)
Rate Decile FE	Yes	Yes	No	No	No	No	No	No
Points Decile FE	Yes	Yes	No	No	No	No	Yes	Yes
N	12271	9200	12271	12271	9200	9200	12271	9200

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table conducts heuristic analyses of data for our Black and Non-Hispanic lender-county-month and covariate-matched sample of 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded, and rate spreads below -55 basis points and above 90 basis points were excluded.

Columns (1) and(2) of Table 4 shows that approach taken in Heuristic 1 would detect lender overcharge in conforming mortgages but not FHA mortgages: with black borrowers paying 4.1 basis points more in discount points for conforming mortgages but a statistically insignificant -0.7 basis points *fewer* discount points for FHA mortgages. Thus, Column (2) replicates Bhutta and Hizmo (Forthcoming)'s result that lenders appear not to overcharge Black borrowers in terms points after controlling for rate in FHA mortgages. However, Columns(3)-(6) shows that Heuristic 2 would consistently detect discrimination for both conforming and FHA mortgages, with magnitudes of 2.9-3.3 basis points for conforming

mortgages and 2.4-2.9 basis points for FHA mortgages. In particular, discrimination is detected in FHA mortgages regardless of the precise trade-off used (whether it was 1/32 or 1/4). Thus, our result using Heuristic 2 is consistent with Bartlett et al. (2019) which focused on conforming mortgages. Therefore, both differences in the type of mortgages considered (ie. conforming vs FHA) and differences in the choice of heuristic are needed to reconcile the recent findings of Bhutta and Hizmo (Forthcoming) and Bartlett et al. (2019) in our sample.

Furthermore, comparing Column (2) to Column (8) in Table 4 shows that the choice of which menu dimension to control for in implementing Heuristic 1 (ie. whether we control for rate and compare points, or the other way around) can lead to contradictory results for the FHA sample. In particular, using an alternative specification for Heuristic 1 where we control for the decile of points instead of rate, we do find that Black borrowers pay a 2.6 basis points higher rate after controlling for points, even though they do not significantly differ in points paid after controlling for rate. The results for the Hispanic and Non-Hispanic White matched sample where shown in Appendix Tables A.3, with similar qualitative results.

Note that we do not interpret results from Table 4 as evidence for or against the hypothesis that lenders overcharged Black borrowers by offering them worse menus of rates and points: both heuristics approaches to analysis can lead to false positives and false negatives, as we showed in Section 2, so the results from such approaches are difficult to interpret. We present results from our analysis in the next Section 5.3.

5.3 Analysis of differences in menus using our metrics

In this section we present our assessment of whether lenders offered minority borrowers worse menus of rates and points. Table 5 presents results from testing for equality in menus between black and white borrowers using our Definition 1. We find in columns (1) and (3) that our test statistic for inequality is positive and highly significant for both Black and Hispanic borrowers for conforming mortgages. More specifically, for conforming mortgages in the Black vs Non-Hispanic White matched sample, our test statistic in column (1) is $\hat{T} = 2.77$, which indicates that 2.77% of Black borrowers' choices in the data could not have been matched to that of Non-Hispanic White borrowers' that could have been on the same menu. The p-value for this test statistic is less than < 0.001, indicating that the probability that this came out of random chance is less than 0.1%. For Hispanic borrowers, our test statistic is $\hat{T} = 1.53$ in column (3), with a p-value less than < 0.001. For FHA mortgages, on the other hand, we are unable to reject equality in menus between Black and Non-Hispanic White borrowers, with a test statistic of $\hat{T} = 0$ in column (2), but are able to for Hispanic and Non-Hispanic White borrowers with a test statistic of $\hat{T} = 2.62$ in column (4).

Table 5: Results from our test of equality in menus ($\hat{T}$).

| | Black vs Non-Hispanic White | | Hispanic vs Non-Hispanic White | |
| | (1) | (2) | (3) | (4) |
	Conforming	FHA	Conforming	FHA
Test statistic ($\hat{T}$)	2.77^{***}	0.00	1.53^{***}	2.62^{***}
95% CI	$[2.10, \infty)$	$[0.00, \infty)$	$[0.89, \infty)$	$[1.16, \infty)$
p-value	<0.001	1.000	<0.001	<0.001
N	12,796	9,422	29,516	12,312

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: this table shows results from our test for equality in menus based on our Definition 1, with $\hat{T}$ in units of percentage points. We use the Black and Non-Hispanic White and Hispanic and Non-Hispanic White lender-county-month and covariate-matched samples of 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded. P-values were computed using 2,000 draws from the asymptotic Normal distribution implied by $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2$ using our procedure in Section 4, and confidence intervals are through inversion of the hypothesis test.

Table 6 presents our results for our differences in menus metric as in Definition 2. Column (1) shows that for conforming mortgages Black borrowers on average are willing to pay at least $\underline{\hat{DIM}}_{1 \to 2} = 2.03$ basis points more in interest rates in order to get the Non-Hispanic White borrowers' menus for conforming mortgages. This again rejects equality in menus, and indicates that the distribution of menus faced by black borrowers is worse than those faced by white borrowers. Similarly, Column (3) shows that Hispanic borrowers are willing to pay at least $\underline{\hat{DIM}}_{1 \to 2} = 1.52$ basis points more in order to get the Non-Hispanic White borrowers' menus. Our lower bound for how much more in interest rates Non-Hispanic White borrowers would be willing to pay to switch to minority menus, on the other hand, is consistently negative. While these magnitudes are small, as explained in Bartlett et al. (2019) even a small difference in interest rate at origination leads to a large differences in payments over the lifetime of the mortgage.

Table 6: Results for our lower bound for the average interest rate increase (bps) needed for consumers to remain indifferent after switching to another group's menus ($\underline{D\hat{I}M}$)

	Black vs Non-Hispanic White		Hispanic vs Non-Hispanic White	
	Conforming	FHA	Conforming	FHA
	(1)	(2)	(3)	(4)
Minority to white ($\underline{D\hat{I}M}_{1\to2}$)	2.03***	-2.44	1.52***	-0.91
95% CI	[1.45, ∞)	[-3.23, ∞)	[1.90, ∞)	[-1.59, ∞)
White to minority ($\underline{D\hat{I}M}_{2\to1}$)	-6.89	-7.22	-6.32	-8.59
95% CI	[-7.50, ∞)	[-7.98, ∞)	[-6.72, ∞)	[-9.27, ∞)
N	12,796	9,422	29,516	12,312

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: this table shows results for our metric for differences in menus (DIM) based on our Definition 2, with $\underline{D\hat{I}M}$ in units of basis points. We use the Black and Non-Hispanic White and Hispanic and Non-Hispanic White lender-county-month and covariate group matched samples of 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded. P-values were computed using 2,000 draws from the asymptotic Normal distribution implied by $\hat{p}_1, \hat{p}_2$ using our procedure in Section 4, and confidence intervals are through inversion of the hypothesis test.

In summary, we find that lenders offered Black and Hispanic borrowers a different distribution of menus than Non-Hispanic White borrowers for conforming mortgages, with our test statistic in columns (1) and(3) of Table 5 strongly rejecting equality in menus and our differences in menus metric in columns (1) and (3) of Table 6 suggesting that Black and Hispanic borrowers would be on average willing to pay to switch to Non-Hispanic White borrowers' menus. On the other hand, the evidence from FHA mortgages is more mixed: column (2) of Table 5 shows that we cannot reject equality in menus between Black and Non-Hispanic White borrowers for FHA mortgages, and while column (4) of Table 5 shows we are able to do so for Hispanic and Non-Hispanic White borrowers we are unable to reject a zero DIM metric in column (4) of Table 5 in terms of the average increase in rate Hispanic borrowers would be willing to pay in order to receive the menus of Non-Hispanic White borrowers.

5.4   Do neighborhoods explain the differences in menus we detect?

In Section 5.3 we showed that lenders offered Black and Hispanic borrowers a less advantageous distribution of menus compared to observationally similar Non-Hispanic White borrowers for conforming mortgages, but were silent as to why this differential pricing appears. One possibility is that there are unobserved risk-related characteristics correlated with race that lenders price on. However, Kau, Fang, and Munneke (2019) and Gerardi, Willen, and Zhang (2020) shows that on originated mortgages compared to observationally similar white

borrowers, black borrowers has similar default rates and far lower prepayment rates, making pools of black borrowers' loans more valuable than pools of white borrowers' loans. Therefore, the evidence on loan performance does not support the hypothesis of risk-based adverse pricing for minority borrowers.

An alternative explanation is that lenders different menus for borrowers that buy homes in neighborhoods with a larger minority population, perhaps as a response to differences in the intensity of competition by neighborhood.[16] This can be done either through generating different ratesheets for different neighborhoods or by offering fewer discretionary discounts on top of ratesheets for borrowers in certain neighborhoods. To test this explanation, we compare borrowers in the top vs bottom quartiles in terms of their share of minority (Black + Hispanic) mortgage originations in Table 7. The top quartile has at least 27% of borrowers in the tract being Black or Hispanic, while the bottom quartile has less than 3%. We find very large differences in menus by neighborhood for Conforming mortgages: Column (1) of Table 7 shows that as a test statistic for equality in menus we are unable to match 12% of borrowers in the top quartile to menus in the bottom quartile, and that borrowers in the top quartile are willing to pay at least 9.8 basis points in order to switch menus with borrowers in the bottom quartile. Therefore, we find strong evidence that lenders present borrowers in different neighborhoods different menus.

---

[16]Due to the history of redlining, pricing/discounting by neighborhood as a response to competition, while seemingly innocuous, is likely illegal if it has a disparate impact on minority racial groups (Bartlett et al., 2019). This illegality may explain why the differences we do find are small in magnitude (though still consequential in aggregate).

Table 7: Our metrics for detecting differences in menus comparing borrowers in the highest vs lowest quartile of minority percent by Census tract

	Highest vs Lowest Quartile	
	(1)	(2)
	Conforming	FHA
Panel A: Test of Equality in Menus ($\hat{T}$)		
Test statistic ($\hat{T}_{n_1,n_2}$)	11.76***	2.27
95% CI	$[9.82, \infty)$	$[0, \infty)$
p-value	<0.001	0.223
Panel B: Difference in Menus ($\hat{D\!I\!M}$) Metric		
Highest to Lowest Quartile ($\hat{D\!I\!M}_{1\to2}$)	9.70***	-1.40
95% CI	$[8.74, \infty)$	$[-3.44, \infty)$
Lowest to Highest Quartile ($\hat{D\!I\!M}_{2\to1}$)	-13.87	-8.50
95% CI	$[-14.84, \infty)$	$[-10.76, \infty)$
N	6,398	1,048

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Panel A shows results for our test of equality in menu ($\hat{T}$) in units of percentage points based on our Definition 1 and Panel B shows our metric for differences in menus ($\hat{D\!I\!M}$) in units of basis points based on our Definition 2, with $\hat{D\!I\!M}$ in units of basis points. We compare borrowers in the highest vs lowest quartile of minority (Black + Hispanic) percent by Census tract in the 2018-2019 HMDA data. We match on lender-county-month and covariate groups for 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded. P-values were computed using 2,000 draws from the asymptotic Normal distribution implied by $\hat{p}_1, \hat{p}_2$ using our procedure in Section 4, and confidence intervals are through inversion of the hypothesis test.

Furthermore, we check if lenders offer borrowers of different races within a neighborhood type different menus by controlling adding the Black and Hispanic share of mortgage originations in the Census tract (rounded to the nearest 5%) as a variable to match on for our analysis of differences in menus by race. Table 8 shows that once we control for the racial composition of neighborhoods, we can no longer detect a racial difference in the distribution of menus between minority and non-Hispanic White borrowers. We note that our inability to reject equality in menus within neighborhoods does not mean that there is necessarily no discrimination in menus within neighborhoods. It only means that we do statistically reject the possibility that there exists a set of menus, equally distributed between across racial groups within neighborhoods, that borrowers chose from: the existence of such a possibility given our data does not not mean that the possibility is the reality. Nevertheless, we can say from Table 8 that we get much lower point estimates for our test statistic of equality in

menus $\hat{T}$ compared to Tables 5,[17] and that we no longer find a statistically significant deviation from equality in menus after matching on neighborhood racial composition. Therefore, the inequality in menus by race we *detect* can be explained by the racial composition of neighborhoods.

Table 8: Our metrics for detecting differences in menus after matching on tract-level minority percentage

	Black vs Non-Hispanic White		Hispanic vs Non-Hispanic White	
	(1)	(2)	(3)	(4)
	Conforming	FHA	Conforming	FHA
Panel A: Test of Equality in Menus ($\hat{T}$)				
Test statistic ($\hat{T}$)	0.75	0.77	0.63	1.40
95% CI	$[0, \infty)$	$[0, \infty)$	$[0, \infty)$	$[0, \infty)$
p-value	0.112	0.372	0.154	0.231
Panel B: Difference in Menus ($\hat{DIM}$) Metric				
Minority to white ($\underline{\hat{DIM}}_{1\to2}$)	-0.44	-1.52	0.27	-0.94
95% CI	$[-1.41, \infty)$	$[-3.02, \infty)$	$[-0.39, \infty)$	$[-2.26, \infty)$
White to minority ($\underline{\hat{DIM}}_{2\to1}$)	-4.47	-7.74	-4.77	-8.23
95% CI	$[-5.44, \infty)$	$[-9.20, \infty)$	$[-5.40, \infty)$	$[-9.63, \infty)$
N	3,374	1,588	8,148	2,280

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Panel A shows results for our test of equality in menu ($\hat{T}$) in units of percentage points based on our Definition 1 and Panel B shows our metric for differences in menus ($\underline{\hat{DIM}}$) in units of basis points based on our Definition 2, with $\underline{\hat{DIM}}$ in units of basis points. We match on percent of minority mortgage originations in the Census tract (to the nearest 5%), lender-county-month and covariate groups for 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded. P-values were computed using 2,000 draws from the asymptotic Normal distribution implied by $\hat{p}_1, \hat{p}_2$ using our procedure in Section 4, and confidence intervals are through inversion of the hypothesis test.

In summary, we found in Section 5.3 that lenders offered minority borrowers worse menus in terms of rates and points for conforming but not necessarily FHA mortgages. In this section we further found that this can be explained by large differences in menus across neighborhoods of differing racial compositions, and that we do not detect differences in menus after controlling for the differences in minority share of borrowers across neighborhoods.

---

[17]With one exception for Black FHA mortgages in Column (2) which was zero to begin with.

6 Discussion

At a high level, our conceptual separation between opportunity and preferences is not without caveats. In many circumstances, the line between opportunity and preferences may be blurred when preferences itself may be influenced by factors such as neighborhoods (see, e.g. Katz, Kling, and Liebman (2001), Chetty, Hendren, and Katz (2016), Chetty and Hendren (2018)). Therefore, a distinction between preferences and opportunity may not always be sensible. Nevertheless, it is sometimes useful, particularly from a policy perspective, to be able to designate differences in outcomes to either inequality in opportunity or heterogeneity in preferences. We do this for the mortgage market, and find that lenders continue to offer minorities borrowers worse menus in terms of rates and points compared to observationally similar white borrowers, largely by varying their menus across neighborhoods.

Our methodology is unlikely to be the final word on the menu problem. While our approach has the advantage of requiring few assumptions to be valid, there may be other assumptions that can be used for identifying differences in menus. A promising path for future research would be to use other identifying assumptions for measuring differences in menus, or by running experiments to address the menu problem and assess its relevance in specific contexts.

References

Abadie, Alberto. Forthcoming. "Statistical Non-significance in Empirical Economics." *American Economic Review: Insights*.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2019. "Consumer-Lending Discrimination in the FinTech Era." *Working Paper*.

Bell, Alex. 2019. "Job Amenities & Earnings Inequality." *Working Paper*.

Bhutta, Neil, Andreas Fuster, and Aurel Hizmo. 2019. "Paying Too Much? Price Dispersion in the US Mortgage Market." *Working Paper*.

Bhutta, Neil, and Aurel Hizmo. Forthcoming. "Do Minorities Pay More for Mortgages?" *Review of Financial Studies*.

Black, Harold A., Thomas P. Boehm, and Ramon P. DeGennaro. 2003. "Is there discrimination in mortgage pricing? The case of overages." *Journal of Banking & Finance* 27(6): 1139 – 1165. ISSN 0378-4266. doi:https://doi.org/10.1016/S0378-4266(02)00250-9. Available at http://www.sciencedirect.com/science/article/pii/S0378426602002509.

Black, Harold A, and Robert L Schweitzer. 1985. "A Canonical Analysis of Mortgage Lending Terms: Testing for Lending Discrimination at a Commercial Bank." *Urban Studies* 22(1): 13–19. ISSN 0042-0980.

Bocian, Debbie Gruenstein, Keith S. Ernst, and Wei Li. 2008. "Race, ethnicity and subprime home loan pricing." *Journal of Economics and Business* 60(1): 110 – 124. ISSN 0148-6195. doi:https://doi.org/10.1016/j.jeconbus.2007.10.001. Financing Community Reinvestment and Development. Available at http://www.sciencedirect.com/science/article/pii/S0148619507000847.

Boehm, Thomas P., Paul D. Thistle, and Alan Schlottmann. 2006. "Rates and race: An analysis of racial disparities in mortgage rates." *Housing Policy Debate* 17(1): 109–149. doi:10.1080/10511482.2006.9521563. Available at https://doi.org/10.1080/10511482.2006.9521563.

Cheng, Ping, Zhenguo Lin, and Yingchun Liu. 2015. "Racial discrepancy in mortgage interest rates." *The Journal of Real Estate Finance and Economics* 51(1): 101–120.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*." *The Quarterly Journal of Economics* 133(3): 1163–1228. ISSN 0033-5533. doi:10.1093/qje/qjy006.
Available at `https://doi.org/10.1093/qje/qjy006`.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106(4): 855–902. doi:10.1257/aer.20150572.
Available at `https://www.aeaweb.org/articles?id=10.1257/aer.20150572`.

Courchane, Marsha, and David Nickerson. 1997. "Discrimination Resulting from Overage Practices." *Journal of Financial Services Research* 11(1): 133–151. ISSN 0920-8550.

Crawford, Gregory S, Rachel Griffith, and Alessandro Iaria. 2019. "Preference estimation with unobserved choice set heterogeneity using sufficient sets." *Discussion Paper 11675, Center for Economic Policy Research.*

Fang, Zheng, and Andres Santos. 2018. "Inference on Directionally Differentiable Functions." *The Review of Economic Studies* 86(1): 377–412. ISSN 0034-6527. doi:10.1093/restud/rdy049.
Available at `https://doi.org/10.1093/restud/rdy049`.

Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen. 2019. "The Time-Varying Price of Financial Intermediation in the Mortgage Market." *Working Paper.*

Gal, Tomas, and Harvey J Greenberg. 2012. *Advances in sensitivity analysis and parametric programming*, vol. 6. Springer Science & Business Media.

Galichon, Alfred. 2016. *Optimal transport methods in economics.* Princeton University Press.

Gerardi, Kristopher, Paul Willen, and David Zhang. 2020. "Mortgage Prepayment, Race, and Monetary Policy." *Working paper.*

Ghent, Andra C., Rubén Hernández-Murillo, and Michael T. Owyang. 2014. "Differences in subprime loan pricing across races and neighborhoods." *Regional Science and Urban Economics* 48: 199 – 215. ISSN 0166-0462. doi: https://doi.org/10.1016/j.regsciurbeco.2014.07.006.
Available at `http://www.sciencedirect.com/science/article/pii/S0166046214000751`.

Goldberger, Arthur S. 1984. "Reverse Regression and Salary Discrimination." *The Journal of Human Resources* 19(3): 293–318. ISSN 0022166X.
Available at `http://www.jstor.org/stable/145875`.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4): 1091–1119. doi:10.1257/aer.104.4.1091.
Available at `http://www.aeaweb.org/articles?id=10.1257/aer.104.4.1091`.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica* 69(6): 1403–1435. doi:10.1111/1468-0262.00252.
Available at `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00252`.

Haughwout, Andrew, Christopher Mayer, Joseph Tracy, Dwight M. Jaffee, and Tomasz Piskorski. 2009. "Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing." *Brookings-Wharton Papers on Urban Affairs* 33–63. ISSN 15287084.
Available at `http://www.jstor.org/stable/25609559`.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12(2): 101–116. doi:10.1257/jep.12.2.101.
Available at `http://www.aeaweb.org/articles?id=10.1257/jep.12.2.101`.

Hsieh, Yu-Wei, Xiaoxia Shi, and Matthew Shum. 2017. "Inference on Estimators defined by Mathematical Programming."

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment*." *The Quarterly Journal of Economics* 116(2): 607–654. ISSN 0033-5533. doi:10.1162/00335530151144113.
Available at `https://doi.org/10.1162/00335530151144113`.

Kau, James B, Lu Fang, and Henry J Munneke. 2019. "An unintended consequence of mortgage financing regulation–a racial disparity." *The Journal of Real Estate Finance and Economics* 59(4): 549–588.

McCloskey, Adam. 2017. "Bonferroni-based size-correction for nonstandard testing problems." *Journal of Econometrics* 200(1): 17 – 35. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2017.05.001.
Available at `http://www.sciencedirect.com/science/article/pii/S0304407617300556`.

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2019. "Simultaneous confidence bands: Theory, implementation, and an application to SVARs." *Journal of Applied Econometrics* 34(1): 1–17. doi:10.1002/jae.2656.
Available at `https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2656`.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynn E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *The American Economic Review* 86(1): 25–53. ISSN 00028282.
Available at `http://www.jstor.org/stable/2118254`.

Nothaft, Frank E, and Vanessa G Perry. 2002. "Do mortgage rates vary by neighborhood? Implications for loan pricing and redlining." *Journal of Housing Economics* 11(3): 244 – 265. ISSN 1051-1377. doi:https://doi.org/10.1016/S1051-1377(02)00103-1.
Available at `http://www.sciencedirect.com/science/article/pii/S1051137702001031`.

Politis, Dimitris N., and Joseph P. Romano. 1994. "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *Ann. Statist.* 22(4): 2031–2050. doi:10.1214/aos/1176325770.
Available at `https://doi.org/10.1214/aos/1176325770`.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2014. "A Practical Two-Step Method for Testing Moment Inequalities." *Econometrica* 82(5): 1979–2002. doi:10.3982/ECTA11011.
Available at `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11011`.

Santambrogio, Filippo. 2015. "Optimal transport for applied mathematicians." *Birkäuser, NY* 55(58-63): 94.

Shapiro, Alexander. 1990. "On concepts of directional differentiability." *Journal of Optimization Theory and Applications* 66: 477–487. doi:10.1007/BF00940933.

Shapiro, Alexander. 1991. "Asymptotic analysis of stochastic programs." *Annals of Operations Research* 30: 169–186.

Sommerfeld, Max, and Axel Munk. 2018. "Inference for empirical Wasserstein distances on finite spaces." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80(1): 219–238. doi:10.1111/rssb.12236.
Available at `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12236`.

Tameling, Carla, Max Sommerfeld, and Axel Munk. 2019. "Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications." *The Annals of Applied Probability* 29(5): 2744–2781. doi:10.1214/19-AAP1463.

Tootell, Geoffrey M. B. 1996. "Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?*." *The Quarterly Journal of Economics* 111(4): 1049–1079. ISSN 0033-5533. doi:10.2307/2946707.
Available at `https://doi.org/10.2307/2946707`.

Villani, Cédric. 2008. *Optimal transport: old and new*, vol. 338. Springer Science & Business Media.

Woodward, Susan E. 2008. "A Study of Closing Costs for FHA Mortgages." *U.S. Department of Housing and Urban Development Report*.

Woodward, Susan E., and Robert E. Hall. 2012. "Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence." *American Economic Review* 102(7): 3249–76. doi:10.1257/aer.102.7.3249.
Available at `http://www.aeaweb.org/articles?id=10.1257/aer.102.7.3249`.

A.1 Proofs of results

A.1.1 Proof of Theorem 1

Proof. In the forward direction, if such a $\pi(x_1, x_2)$ exists, then it is possible for there to be a series of menus $\mathbf{m} = \{x_1, x_2\}$ where $\phi(x_1, x_2) = 1$, each appearing with probability $g_1(\mathbf{m}) = g_2(\mathbf{m}) = \pi(x_1, x_2)$, in which group 1 consumers chose x_1 and group 2 consumers chose x_2. Under this construction, the distribution of menus across the two groups are equal such that $\mathbf{M}_1 = \mathbf{M}_2$, and the choice probabilities are rationalized. The reverse direction follows from the fact that, denoting $c(x_1, x_2|\mathbf{m})$ by the probability that group 1 consumers choose $x_1 \in \mathbf{m}$ and group 2 consumers choose $x_2 \in \mathbf{m}$ given a menu $\mathbf{m}$, we can compute such a $\pi(x_1, x_2) = \sum_{\mathbf{m}} c(x_1, x_2|\mathbf{m}) g(\mathbf{m})$ for any $c(x_1, x_2|\mathbf{m}), g(\mathbf{m})$ where $g(\mathbf{m}) = g_1(\mathbf{m}) = g_2(\mathbf{m})$. $\square$

A.1.2 Proof of Theorem 3

Proof. First, we show that φ is Gâteaux directionally differentiable in the sense that the limit:

$$\varphi'_{\mathbf{p}_1, \mathbf{p}_2}(\mathbf{h}_1, \mathbf{h}_2) = \lim_{t \to 0^+} \frac{\varphi(\mathbf{p}_1 + t h_1, \mathbf{p}_2 + t h_2) - \varphi(\mathbf{p}_1, \mathbf{p}_2)}{t} \tag{32}$$

exists for all and is equal to that given by Equations (18) for $\{\mathbf{p}_1, \mathbf{p}_2\} \in \mathbb{D}_\varphi$ and $\{\mathbf{h}_1, \mathbf{h}_2\} \in \mathbb{D}_0$. To do this, without loss of generality letting $\phi^* = \phi + M, M = \sup |\phi|$ such that $\phi^* \geq 0$ and $\varphi = \varphi^* - M$, we transform the problem to standard linear programming form:

$$\varphi^* = \min_{\boldsymbol{\pi}} E_{\boldsymbol{\pi}} \phi^* \text{ s.t. } E_{x_2} \boldsymbol{\pi} \geq \mathbf{p}_1, E_{x_1} \boldsymbol{\pi} \geq \mathbf{p}_2, \boldsymbol{\pi} \geq 0 \tag{33}$$

and then use Theorem 3.1 of Gal and Greenberg (2012) which sets out conditions for the Gâteaux directional differentiability of standard linear programs with inequality constraints. In particular, we need to check primal and dual stability in the sense that, let $\Pi^*(p_1, p_2)$ be the set of primal solutions to the linear programming problem in Equation (33), then the set of primal solutions reachable from perturbations in the direction $\{h_1, h_2\}$ is non-empty, such that:

$$\Pi^\infty(\{\mathbf{p}_1, \mathbf{p}_2\}, \{\mathbf{h}_1, \mathbf{h}_2\}) = \{\boldsymbol{\pi} : \{\boldsymbol{\pi}^k\} \to \boldsymbol{\pi} \text{ for some } \{\epsilon_k \to 0^+\}, \tag{34}$$

$$\text{with } \boldsymbol{\pi}^k \in \mathbf{\Pi}^*(\mathbf{p}_1 + \epsilon_k \mathbf{h}_1, \mathbf{p}_2 + \epsilon_k \mathbf{h}_2)\} \neq \emptyset \tag{35}$$

and analogously for dual solutions. This can be done by referencing existing results:

1. Since $\{\mathbf{p}_1 + \epsilon_k \mathbf{h}_1, \mathbf{p}_2 + \epsilon_k \mathbf{h}_2\}$ are a series of probability measures for $\epsilon_k \leq 1$, primal stability in the sense of Equation (34) is guaranteed by Theorem 5.19 in Villani (2008).

2. Similarly, dual stability is guaranteed by Theorem 1.52 in Santambrogio (2015), in particular by taking the sequence of c-concave Kantorovich potentials corresponding to $(\mathbf{p}_1 + \epsilon_k \mathbf{h}_1, \mathbf{p}_2 + \epsilon_k \mathbf{h}_2)\}$.

Second, we show that φ is Lipschitz, so that the Gâteaux directionally differentiability of φ is equivalent to Hadamard directionally differentiability following Shapiro (1990). For the l_1 norm $l_1(\{\mathbf{p}_{1,1}, \mathbf{p}_{1,2}\}, \{\mathbf{p}_{2,1}, \mathbf{p}_{2,2}\}) = \sum_x |\mathbf{p}_{1,1}(x) - \mathbf{p}_{1,2}(x)| + |\mathbf{p}_{2,1}(x) - \mathbf{p}_{2,2}(x)|$, we will show that:

$$|\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2})| \leq M l_1(\{\mathbf{p}_{1,1}, \mathbf{p}_{1,2}\}, \{\mathbf{p}_{2,1}, \mathbf{p}_{2,2}\}) \tag{36}$$

More specifically, let $\mathbf{p}_1^- = \min\{\mathbf{p}_{1,1}, \mathbf{p}_{1,2}\}, \mathbf{p}_1^+ = \max\{\mathbf{p}_{1,1}, \mathbf{p}_{1,2}\}$, and analogously for $\mathbf{p}_2^-, \mathbf{p}_2^+$. By construction we know that $\varphi(\mathbf{p}_1^-, \mathbf{p}_2^-) \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}), \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2})$ and $\varphi(\mathbf{p}_1^+, \mathbf{p}_2^+) \geq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}), \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2})$, and therefore:

$$\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^+, \mathbf{p}_2^+) \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2}) \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^-, \mathbf{p}_2^-) \tag{37}$$

Furthermore, we know that $\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^-, \mathbf{p}_2^-) \leq M l_1$ since taking the optimal plan from $\pi^- = \varphi(\{\mathbf{p}_1^-, \mathbf{p}_2^-\})$ and then constructing a plan $\pi^{-,*} = \pi^- + (\mathbf{p}_{1,1} - \mathbf{p}_1^-)(\mathbf{p}_{1,2} - \mathbf{p}_2^-)$ yields an upper bound for the value value $\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) \leq E_{\pi^{-,*}}\phi = \varphi(\{\mathbf{p}_1^-, \mathbf{p}_2^-\}) + \sum(\mathbf{p}_{1,1} - \mathbf{p}_1^-)(\mathbf{p}_{1,2} - \mathbf{p}_2^-)\phi \leq \varphi(\{\mathbf{p}_1^-, \mathbf{p}_2^-\}) + M\sum(\mathbf{p}_{1,1} - \mathbf{p}_1^-)\sum(\mathbf{p}_{1,2} - \mathbf{p}_2^-) \leq \varphi(\{\mathbf{p}_1^-, \mathbf{p}_2^-\}) + M l_1$. Similarly, we have that $\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^+, \mathbf{p}_2^+) \geq -M l_1$. Substituting into Equation (37) yields:

$$-M l_1 \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2}) \leq M l_1 \tag{38}$$

Which implies Equation (36) and that the mapping φ is Lipschitz.

□

A.1.3 Proof of Corollary 1

Proof. Suppose otherwise, then there exists a subsequence n_k with a corresponding sequence of $[\mathbf{p}_{1,k}, \mathbf{p}_{2,k}] \in \mathcal{D}_\varphi$ such that:

$$\lim_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} (\hat{\varphi}_{n_k} \geq \hat{c}^a_{n_k, 1-\alpha+\beta}) > \alpha \tag{39}$$

44

However, we know that:

$$\lim_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} \left(\hat{\varphi}_{n_k} \geq \hat{c}^a_{n_k, 1-\alpha+\beta} \right) \tag{40}$$

$$\leq \limsup_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} \left([\mathbf{p}_{1,k}, \mathbf{p}_{2,k}] \in \hat{\mathcal{F}}_{\beta,k} \right) + \limsup_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} \left((\hat{\varphi}_{n_k} \geq \hat{c}^a_{n_k, 1-\alpha+\beta}) \cap [\mathbf{p}_{1,k}, \mathbf{p}_{2,k}] \in \hat{\mathcal{P}}_{\beta,k} \right)$$

$$\tag{41}$$

$$\leq \beta + \alpha - \beta = \alpha \tag{42}$$

Where $\limsup_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} ([\mathbf{p}_{1,k}, \mathbf{p}_{2,k}] \leq \beta$ by assumption and:

$$\limsup_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} \left((\hat{\varphi}_{n_k} \geq \hat{c}^a_{n_k, 1-\alpha+\beta}) \cap [\mathbf{p}_{1,k}, \mathbf{p}_{2,k}] \in \hat{\mathcal{F}}_{\beta,k} \right) \leq 1 - \alpha + \beta \tag{43}$$

by Fang and Santos (2018) since $\hat{c}^a_{n_k, 1-\alpha+\beta}$ will dominate the true $c^a_{n_k, 1-\alpha+\beta}$ at each $[\mathbf{p}_{1,k}, \mathbf{p}_{2,k}]$, and Theorem 3.3 in Fang and Santos (2018) shows (under a subadditivity assumption on φ' which is satisfied here) $\limsup_{n_k \to \infty} \Pr_{\mathbf{p}_{1,k}, \mathbf{p}_{2,k}} ((\hat{T}_{n_k} \geq \hat{c}^a_{n_k, 1-\alpha+\beta}) \leq \alpha$. Which is a contradiction. $\square$

A.2   Additional Tables and Figures

Figure A.1: How the choice of which menu dimension to condition on can yield contradictory results
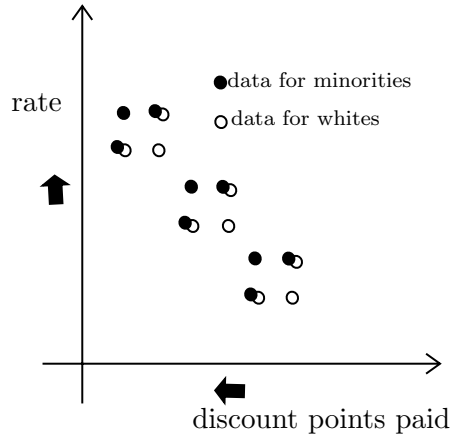
Note: this Figure shows that a "contradictory" assessment of discrimination can appear nonlinearly, and complements the linear regression case of Figure 3.

Table A.1: Other approaches to inference, compared to our size-corrected directional derivatives approach

| | HSS (2017) | m out of n subsampling | | | Size-corrected |
		$m = n^{2/3}$	$m = n^{1/2}$	$m = n^{1/3}$	directional derivatives
$n_1 = n_2 = 500$	0.0	31.8	21.9	16.2	4.4
$n_1 = n_2 = 1000$	0.0	28.3	18.4	12.5	4.7
$n_1 = n_2 = 5000$	0.0	20.0	11.3	8.6	4.8
$n_1 = n_2 = 10000$	0.1	18.3	10.3	8.4	5.8
$n_1 = n_2 = 50000$	0.0	13.7	8.6	6.5	5.0

Note: Entries represent the probability of rejecting the null at 5% level. The control of the subsampling approach is taken with subsample size m as indicated, with $n = \frac{n_1 n_2}{n_1 + n_2} = \frac{1}{2} n_1$. The control of our size-corrected directional derivatives approach were computed via 2000 sample draws and 500 draws of h_1, h_2 from the estimated asymptotic multivariate Normal distribution for p_1, p_2 within each sample draw.

Table A.2: Regressions of origination costs and total loan costs as a percent of the loan amount on HMDA's information on points (hmda_points) versus OptimalBlue's information on points (ob_points)

| | Origination costs | | | Total loan costs | | |
	(1)	(2)	(3)	(4)	(5)	(6)
hmda_points	0.915***		0.902***	0.940***		0.941***
	(0.00277)		(0.00308)	(0.00372)		(0.00463)
ob_points		0.469***	0.0200***		0.474***	0.00531
		(0.00377)	(0.00220)		(0.00434)	(0.00361)
_cons	0.580***	0.560***	0.575***	2.194***	2.174***	2.190***
	(0.000432)	(0.00129)	(0.000591)	(0.000577)	(0.00149)	(0.000927)
N	1224911	1221338	1221208	1224417	1220921	1220715
R^2	0.553	0.231	0.553	0.233	0.093	0.232

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: The sample consists of the conforming and FHA purchase mortgages originated within the Retail channel within our 2018-2019 HMDA-OptimalBlue matched sample. In each regression we excluded observations with extreme outliers for points (below -4 or above 4) and for origination costs and total loan costs as a percent of the loan amount below -3% or above 10%. All regressions include lender by county by year by product type fixed effects. Standard errors were also clustered at the lender by county by year by product type level.

Table A.3: Assessments of lender discrimination using two heuristic approaches in the Hispanic and Non-Hispanic White matched sample

	Heuristic 1		Heuristic 2				Alternate Heuristic 1	
	Conforming	FHA	Conforming		FHA		Conforming	FHA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	points	points	rate_1/8	rate_1/4	rate_1/32	rate_1/4	rate	rate
hispanic	5.153***	2.021*	2.945***	3.437***	3.449***	3.331***	2.623***	3.334***
	(0.633)	(1.113)	(0.225)	(0.432)	(0.254)	(0.369)	(0.221)	(0.370)
Rate Decile FE	Yes	Yes	No	No	No	No		
Points Decile FE	No	No	No	No	Yes	Yes		
N	28273	12055	28273	12055	28273	12055	28273	12055

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table conducts heuristic analyses of data for our Hispanic and Non-Hispanic lender-county-month and covariate matched sample of 30 year, new purchase, fixed rate, first lien mortgages on owner-occupied site-built properties without prepayment penalties, balloon, interest only, negative amortization, or non-amortization features. Outliers for points above 4 and below -4 were excluded, and rate spreads below -55 basis points and above 90 basis points were excluded.