Working Papers

FEDERAL RESERVE BANK OF BOSTON

No. 20-19

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

David Hao Zhang and Paul Willen

Abstract:
We use a new methodology to assess mortgage pricing discrimination faced by minority borrowers. We identify a "menu problem" that comes from the multidimensional nature of mortgage pricing: When getting a mortgage, borrowers can choose to avoid closing costs and pay a high interest rate or contribute to closing costs to get a lower rate. While data on both dimensions of mortgage pricing are by now often available, intuitively attractive metrics of lender pricing discrimination used in the literature can lead to both false and contradictory results. For example, it is sometimes observed that conditional on rate, minority borrowers pay the same closing costs as white borrowers, but conditional on closing costs, minority borrowers pay a higher rate. Though generally underappreciated, the menu problem is broadly relevant in economic assessments of differences in opportunity given data on outcomes. We develop a solution to the menu problem by defining (1) a test statistic for equality in menus and (2) a difference in menus (DIM) metric for assessing whether one group of borrowers would prefer to switch to another group's menus, both based on pairwise dominance relationships in the data. Our proposed solution is robust to arbitrary heterogeneity in borrower preferences across racial groups. We show how our metrics can be computed using methods from optimal transport and also devise a new procedure for hypothesis testing in this class of problems based on directional differentiation. Finally, we implement our methodology on a data set linking 2018–2019 Home Mortgage Disclosure Act (HMDA) data to Optimal Blue rate locks, and present novel results on mortgage pricing discrimination.

JEL Classifications: G21, G51
Keywords: mortgage pricing discrimination, minority borrowers, Home Mortgage Disclosure Act

David Hao Zhang is a PhD candidate in business economics at Harvard University; his email address is dzhang@hbs.edu. Paul Willen is a senior economist and policy advisor in the research department at the Federal Reserve Bank of Boston; his email address is paul.willen@bos.frb.org.

The authors thank Isaiah Andrews, John Campbell, Edward Glaeser, Robin Lee, Ariel Pakes, Adi Sunderam, and Elie Tamer for their continuous advice on this paper. They are further indebted to Alex Bell, Neil Bhutta, Gary Chamberlain, Chris Foote, Andreas Fuster, Camilo Garcia-Jimeno, Aurel Hizmo, Jun Ishii, Larry Katz, Xinwei Ma, Jonathan Roth, Andrei Shleifer, and attendees of the 2020 System Applied Micro Conference, the 2020 System Econometrics Conference, and the 2020 Econometric Society European Winter Meetings Conference for their valuable comments.

This version: December 2020 **https://doi.org/10.29412/res.wp.2020.19**

1 Introduction

Whether mortgage lenders discriminate against minority borrowers is an important question both in terms of academic research and in regard to its policy relevance.[1] However, the task of assessing whether lenders discriminate by offering minority borrowers worse prices is complicated by the fact there are two dimensions to mortgage pricing: the interest rate and the upfront fees charged by the lender. In particular, US mortgage borrowers can choose to pay higher upfront fees (in the industry referred to as paying discount points) in return for lower interest rates. Or, conversely, they can get the lender to pay some of their closing costs in exchange for a higher interest rate. We show that the availability of this choice between a higher upfront fee and a higher interest rate creates a fairly general "menu problem" that makes the detection of lender discrimination nontrivial, with the methods implemented in the literature susceptible to false and contradictory results. We then propose a novel identification argument and a new procedure for inference to deal with this menu problem, and apply it empirically to reassess racial discrimination in mortgage markets.

Many studies find that minority consumers pay higher interest rates compared with 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 pricing on their mortgages, another explanation raised in the literature is that minority consumers were simply more constrained in their choices of how many discount points to pay and how much lender credit, or how many negative discount points, to receive. The discount points explanation may still reflect structural disparities between racial groups, but it has policy implications that are very different from those of one in which the lenders themselves are systematically offering minority consumers worse menus of rates and discount point options. Given data on the borrowers' chosen mortgage rates and points (but not the menus borrowers faced, which are not typically observable), our objective is to

[1] Since the financial crisis, the Department of Justice has reached settlements of well over $500 million with lenders that overcharged Black and Hispanic borrowers in violation of the Fair Lending Act, as explained in Bhutta and Hizmo (Forthcoming). These settlements include $335 million with Bank of America (on behalf of Countrywide), $175 million with Wells Fargo, and $55 million with JP Morgan Chase. On June 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, for example, 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, Morse, Stanton, and Wallace (2019). Relatedly, Munnell, Tootell, Browne, and McEneaney (1996) and Tootell (1996) find that minority borrowers are more likely to be rejected for mortgages; Black, Boehm, and DeGennaro (2003) find that minority borrowers pay higher yield spreads when refinancing their mortgage; and Ambrose, Conklin, and Lopez (2020) find that minority borrowers pay more in broker fees particularly when faced with a white broker.

examine whether lenders discriminated against Black borrowers in the sense of offering them a distribution of menus that was worse than the one offered to observationally similar white consumers, a practice we call *discrimination in menus*.

Our first contribution is to point out that there exists a surprisingly nontrivial menu problem involved with assessing differences in the distribution of menus offered to minority and non-Hispanic white borrowers. As an example, researchers may be tempted to address a heterogeneity in preferences over rates and discount points by racial group by simply controlling for the discount points in their regressions. However, we show that the approach of controlling for rates and discount points can lead to false positives in the sense of detecting discrimination when none exists, and false negatives in the sense of failing to detect discrimination when it does exist. Furthermore, we show that even a seemingly foolproof comparisons of means—that is, checking if minority consumers on average pay both a higher interest rate and more in lender fees—can still lead to false positives and false negatives if interpreted as evidence of discrimination in menus.[3] These issues emerge when minority and non-Hispanic white borrowers are likely to make different decisions (that is, have different preferences) over menu items in an unknown manner, which is particularly relevant empirically because the possibility of heterogeneity in preferences across racial groups is usually the motivation researchers would seek to control for the choices of discount points in the first place.

The menu problem is important for the mortgage pricing discrimination literature. There are two main methods by which a large literature assesses discrimination in mortgage pricing given the rate and discount point trade-off, and differences in choice of existing methodology is partially responsible for the recent differences in findings by mortgage type. First, Courchane and Nickerson (1997) and Bhutta and Hizmo (Forthcoming) look at whether Black borrowers paid more in points conditional on rate in samples of FHA mortgages, for which Courchane and Nickerson (1997) find a differential in points paid by race, while Bhutta and Hizmo (Forthcoming) do not. In particular, Bhutta and Hizmo (Forthcoming) use a much larger sample and construct a more uniform sample of loans, which improves on the earlier literature.[4] Second, Woodward (2008), Woodward and Hall (2012), and Bartlett et al. (2019) compare the interest rates of minority and white borrowers after adjusting for points using a known range of rate-point trade-offs, and find that minorities consistently paid more for mortgages, even in the FinTech era, as shown in Bartlett et al. (2019). As we discussed, both

[3]It follows that adjusting by a known range of rate and lender fee trade-offs and then comparing means can be similarly misleading.

[4]Bhutta and Hizmo (Forthcoming)'s study also goes beyond the non-existence of pricing disparities at origination: It also finds that lenders received more revenue from loans that were made to minorities once points and secondary marketing revenue are added together.

of the existing methods used in the literature, (1) controlling for rates and (2) comparing mean levels of rates and discount points after adjusting by a known range of slopes, can lead to false positives and false negatives and can contradict one another. Indeed, we show that in our FHA sample the choice of either (1) or (2) does lead to apparently contradictory results on whether mortgage pricing discrimination exists. Our robust solution to the menu problem would therefore allow researchers to assess discrimination in mortgage markets in a more internally consistent and theoretically sound way.

The menu problem also extends well beyond the mortgage discrimination setting. Generally speaking, the problem is relevant whenever a researcher wishes to assess disparities in opportunity given data on choices while allowing for heterogeneous preferences across groups. For example, when workers make decisions that trade off wages and hours worked, researchers may wish to assess the extent to which the gender gap in pay may be explained by the choice of hours, as in the model of Goldin (2014). However, the menu problem implies that popular measures of gender inequality, such as the gender pay gap conditional on hours or even the gender pay gap after adjusting for all relevant average compensating differentials, are not necessarily informative about whether the data on wages can be explained by heterogeneous preferences over hours worked. Our methodology can be useful for this type of problem depending on the institutional details the researcher has access to. Furthermore, while the problems we point out are simple in hindsight, they highlights the role of seemingly innocuous assumptions, such as the restriction of productivity differences to a single component of unobserved heterogeneity, which is used in many structurally specified models of labor supply (e.g. Hwang, Reed, and Hubbard (1992) and Bell (2019)) and can in certain contexts lead to misleading inference about labor market inequality. Therefore, the problems we point out and the solution we propose may be of broad interest.

As a solution to the menu problem, we propose (1) a new metric for detecting whether there exists a difference in the distribution of menus offered to two groups and (2) 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. Both metrics are based on pairwise dominance relationships in the data (that is, a mortgage with a lower rate and paying fewer points dominates a mortgage with a higher rate and more points) that can be supplemented by industry knowledge. 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. Unlike the existing methodology used in the literature, our metrics are robust to any form of unobserved differences in preferences across borrower groups.

The sample counterparts to both of our metrics can be computed as solutions to optimal transport problems, which are computationally well understood and can be efficiently computed through linear programming. As a technical contribution, we also derive a new approach to uniformly valid inference for the value of optimal transport problems, which we implement for our metrics to distinguish between statistical noise and actual differences in menus. Conventional approaches to inference, such as bootstrapping, fail for optimal transport problems, because the objective function can be non-differentiable (Fang and Santos, 2018). We prove that optimal transport problems are directionally differentiable in the sense of Shapiro (1991) and Fang and Santos (2018). We then apply the asymptotic results of Fang and Santos (2018), which we combine with a Bonferroni correction following Romano, Shaikh, and Wolf (2014) and McCloskey (2017) to address sampling error in the directional derivatives. We show that this approach leads to asymptotically uniformly valid size control for hypothesis testing in the value of optimal transport problems, and test it in finite samples in a Monte Carlo simulation. Our new approach to inference in optimal transport may be useful for other researchers who wish to conduct inference on the value of optimal transport problems, many of which are described in Galichon (2016).

Empirically, 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 in the same county and within narrow covariate groups for conforming mortgages for both Black and Hispanic borrowers. Furthermore, we show that on average Black borrowers 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 more 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). Within conforming mortgages, the differences in menus we detect are not explained by loan originator compensation and are particularly concentrated among borrowers with lower loan-to-value (LTV) ratios and higher credit scores.[5] The fact that

[5]Relatedly, Ambrose, Conklin, and Lopez (2020) find in a pre-2008 sample period that higher-credit-score minorities pay more broker fees than observationally similar non-Hispanic white borrowers, even though the default risk of both whites and minorities are similar among this group. The Dodd-Frank Act of 2010 now forbids mortgage brokers from varying the fees they charge by borrower except as a function of the loan amount, shutting down this particular channel of disparity. So, the differences in menus we detect in our 2018–2019 sample are of a different nature than the broker fee differences in Ambrose, Conklin, and Lopez (2020). Nevertheless, they could both be driven by the same underlying mechanism of loan originators

the mortgage pricing discrimination we detect is concentrated among the more creditworthy conforming mortgage borrowers is consistent with the less risky non-Hispanic white borrowers being more likely to be offered discounts during the search and negotiation process.[6]

We focus on the important and basic question of whether we can detect differences in the rate and discount point menus faced by observably similar non-Hispanic white and minority mortgage borrowers in our empirical application. This implies that we capture not only pricing differences that result from taste-based discrimination, but also any disparate pricing impact from statistical discrimination or the search and negotiation process. Our approach is in line with the recent mortgage pricing discrimination literature (e.g. Bartlett et al. (2019) and Bhutta and Hizmo (Forthcoming)) and is justified by the unusual institutional details and the regulatory framework developed around this market.[7] First, as explained in Bartlett et al. (2019), the Fair Lending Act imposes a legal requirement of no pricing differentials by race conditional on observables in our setting, so our results are naturally interesting from a regulatory perspective. Second, for the types of mortgages we focus on, lenders are insured from the risk of default by either the GSEs or the FHA, which compensate investors for any losses of principal but do benefit from the more favorable prepayment risk of minorities. Therefore, in contrast to some stereotypes, mortgages from Black and Hispanic borrowers are likely significantly more valuable than those of observably similar white borrowers due to their lower prepayment risk, as shown in the simulations of Kau, Fang, and Munneke (2019) and in analyses of the actual mortgage-backed securities prices in Gerardi, Willen, and Zhang (2020). As a result, even if lenders did use unobservables that are correlated with race to price for expected loan performance in a possibly illegal manner, it would be unlikely to justify the unfavorable pricing to minorities we find.

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.

being more willing to offer discounts to more creditworthy non-Hispanic white borrowers relative to similarly creditworthy minority borrowers.

[6]Negotiation in mortgage markets is common. Studies that look at the search and negotiation process in mortgage markets include Allen, Clark, and Houde (2014), Allen, Clark, and Houde (2019), and Bhutta, Fuster, and Hizmo (2019). A reason why minority borrowers may behave as if they have higher search costs is in Agarwal, Grigsby, Hortaçsu, Matvos, Seru, and Yao (2020), where borrowers internalize a higher probability of rejection. Such price discrimination by effective search costs, if conducted within-lender, would violate the Fair Housing Act to the extent it results in disparate impact by race according to Bartlett et al. (2019).

[7]In fact, we expand on the earlier empirical literature by showing for the first time that the mortgage pricing differentials are not explained by loan originator compensation.

2 The menu problem

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 United States, an upfront fee/discount points and the interest rate, where each point is customarily worth 1 percent 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. We plot those choices from an example rate sheet in Figure 1. We also present a screenshot illustrating this trade-off from an online mortgage price comparison service in Appendix Figure A.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 (paying 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.

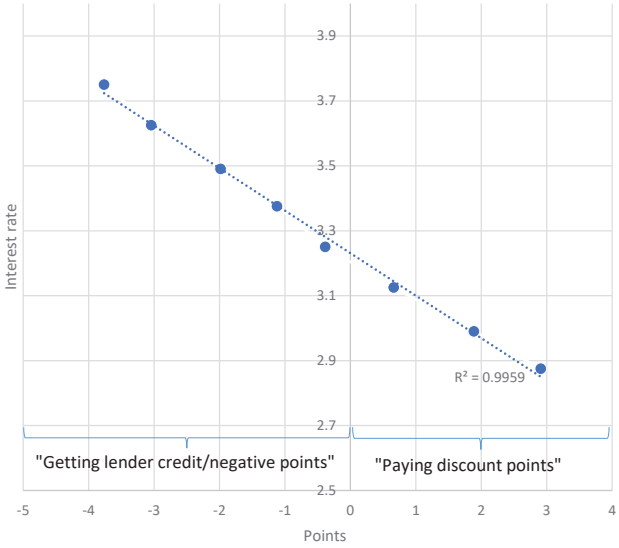


Figure 1: An example set of menu items from a lender rate sheet.

In practice, the researcher often observes the distribution of borrower choices x but not the underlying menu $\mathbf{m}$ where $x \in \mathbf{m}$ is chosen from. The menu problem then emerges as the problem of inference on the extent to which a matched group of borrowers who are by construction observationally similar in terms of covariates faced the same distributions of menus. More specifically, testing for equality of menus can be written as testing the null hypothesis $\mathcal{H}_0$ that the distributions of menus being offered to both groups are equal. That is, suppose $\mathbf{m}_1 \sim \mathbf{M}_1$ for borrowers in group 1 and $\mathbf{m}_2 \sim \mathbf{M}_2$ for borrowers in group 2, the

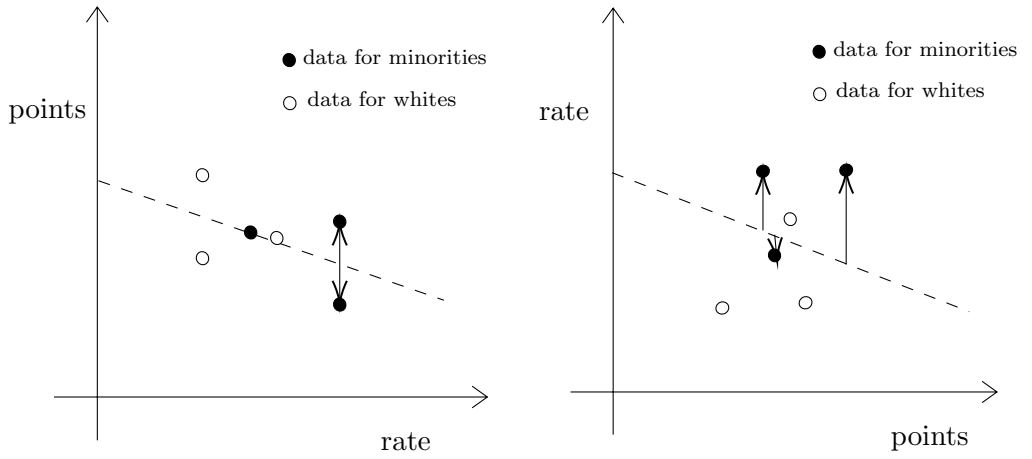
menu problem is the hypothesis testing problem where:

$$\mathcal{H}_0 : \mathbf{M}_1 = \mathbf{M}_2, \tag{1}$$

$$\mathcal{H}_1 : \neg\mathcal{H}_0. \tag{2}$$

The applied literature on mortgage pricing discrimination, and indeed the literature on differences in opportunity more broadly, can be viewed as testing for equality in menus as specified in Equation (1). Nevertheless, existing methodology used in this literature tends to fall short of being able to credibly assess inequality in menus, as we explain in the rest of this section.

One natural approach to assessing whether lenders offered minority and non-Hispanic white borrowers different menus is to control for one dimension of the menu, that is, conditional on the distribution of covariates, to estimate whether minority borrowers who received the same interest rate as white borrowers paid more in discount points. This approach was used in Courchane and Nickerson (1997) and Bhutta and Hizmo (Forthcoming). A problem with this approach, however, is that it can lead to contradictory estimates depending on whether the researcher chooses to control for rates or points. The situation in Figure 2 shows that it is possible for a regression of points on rate to show no discrimination against minorities while a regression of rate on points shows discrimination with the same example data. In this figure, we represent example data from minority and white borrowers using black and white dots, respectively; regression line by the dashed line; and the difference to the regression line by the solid arrows. Figure 2a shows that a regression of points on rate and borrower race would show a zero coefficient for minorities, with the two arrows balancing each other out. On the other hand, Figure 2b shows that, using the same data, a regression of rate on points would instead give a positive coefficient for minorities. This sort of contradiction is not particular to the linear regression case, and as we show in Appendix Figure A.2, it can appear with general conditional expectations.



(a) Points on rates shows no discrimination

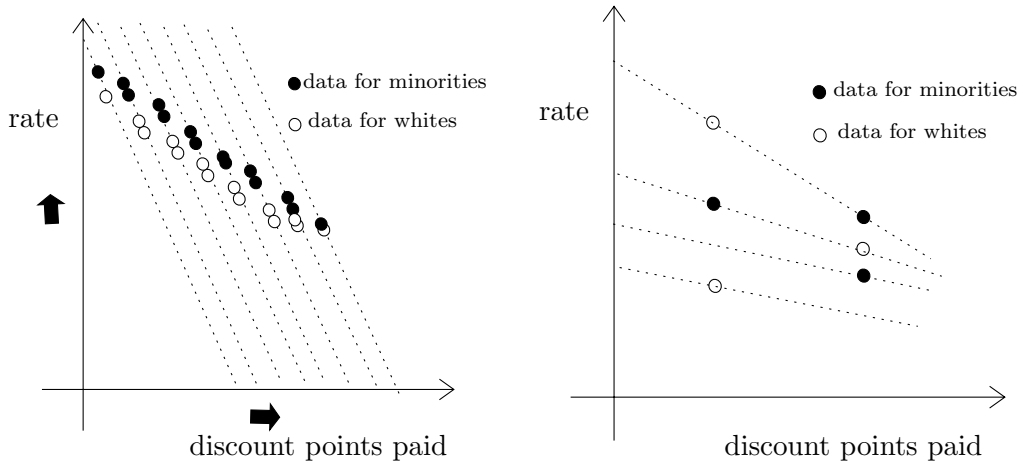
(b) With the same data, rates on points (flipped axes) shows discrimination

Figure 2: How the choice of which menu dimension to control for can lead to contradictory findings of discrimination

This possibility for contradiction as in Figure 2 can be viewed as a version of the reverse regression problem of Goldberger (1984). While the reverse regression problem illustrates how the heuristic of conditioning on one dimension of the menu can be unreliable, the problem with testing equality in menus goes far beyond the reverse regression problem, since there can be false positives and false negatives even when the forward and reverse regressions are consistent (and even when a simple comparison of means is consistent), which we will discuss in the rest of this section. Regardless, the example shows that the natural approach of controlling for a dimension of the menu is not a useful method of testing for the equality in menus problem specified in Equation (1).

Figure 3 shows how even when forward and reverse regressions consistently detect discrimination or no discrimination, the heuristic of controlling for one dimension of the menu can lead to false positives and false negatives if interpreted as a test for equality in menus as in Equation (1). In this figure, we represent example data from minority and white borrowers using black and white dots, respectively, and menus by dotted lines where each dotted line is one potential menu that a borrower may draw. The researcher wishes to evaluate whether the minority and white borrowers were offered the same distribution of menus. The left panel of Figure 3a 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 in borrower behavior by group is that minority consumers chose to pay fewer points on every menu, and yet it would appear as if they were discriminated against. That is, even though

$\mathbf{M}_1 = \mathbf{M}_2$, in reality, as represented by a common set of dotted lines facing both groups, it appears as if minorities are worse off, controlling for either dimension of the menu. In the right panel of Figure 3b, we illustrate a false negative situation in which minority consumers paid the same rate conditional on points but faced a worse distribution of menus, since the bottom menu (the most advantageous menu) was offered only to white borrowers while the second-to-bottom menu was offered only to minority borrowers. Furthermore, Figure 3b can be constructed in a way such that the variance of the rate and discount points paid are equal, such that forward and reverse regressions both give the same false negative but there does exist discrimination in menus. That is, $\mathbf{M}_1 \neq \mathbf{M}_2$ in reality, but controlling for either dimension of the menu finds no difference.



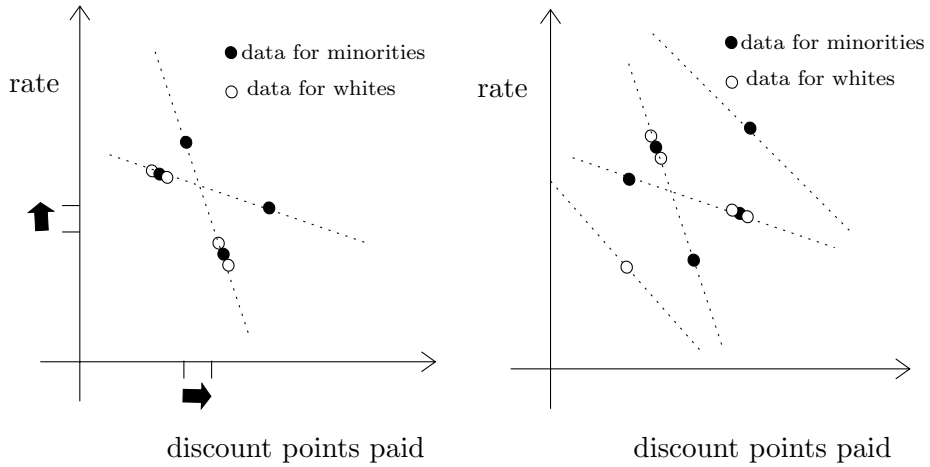
(a) **False positive**, minority borrowers paid more in rate (points), controlling for points (rate), but the menus were the same

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

Figure 3: False positives and false negatives from controlling for one direction of the menu

A second intuitively appealing approach for assessing discrimination in menus is to compare means, thus avoiding the problem of having to estimate menu slopes from the data. 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 with 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 is done in Woodward (2008), Woodward and Hall (2012), and Bartlett et al. (2019). While this avoids the problem that regressions may incorrectly estimate menu slopes, it can still lead to false positives and false negatives when slopes are not constant across menus, thus breaking the assumption (1) that all menus share the same shape. This is a realistic

problem, because we know from rate sheet data that an unobserved heterogeneity in slopes does exist in the mortgage setting, as the rate-point trade-offs do vary substantially across lenders and over time (Figure 8). Figure 4a illustrates how, when slopes differ across menus, 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 ($\mathbf{M}_1 = \mathbf{M}_2$). 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 ($\mathbf{M}_1 \neq \mathbf{M}_2$).



(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 borrowers paid more on average in both rates and points

More specifically, the mechanism for when a comparison of means would lead to false positives in the case of Figure 4a is that in the example, minority borrowers respond less to differences in the slopes of the rate-point menus compared with white borrowers. This possibility is empirically relevant, because there are two directions in which constraints can drive borrower choices of rates and closing costs. First, if borrowers are cash constrained when getting the loan, they may need to get a lower closing cost mortgage (pay fewer discount points) regardless of what rate-point trade-offs lenders offer. Second, if borrowers are debt-to-income (DTI) constrained, they may need to pay more points to buy down the rate to increase their borrowing limit.[8] Therefore, the finding that minority borrowers on average pay more than white borrowers may simply reflect the fact that minority borrowers are more

[8]See, for example: `https://www.thetruthaboutmortgage.com/dti-debt-to-income-ratio/`.

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.

While simple in hindsight, the situation of Figure 4a illustrates how a seemingly innocuous assumption that differences in menus faced by different agents can be summarized by a single monotonically additive term is actually a strong assumption when there are differences in both menu intercepts and slopes in reality. Perhaps due to tractability, such an assumption is popular in models of labor productivity (e.g. Hwang, Reed, and Hubbard (1992), Bell (2019)), where agents are assumed to face structurally defined menus of trade-offs indexed by a single productivity term, and estimation proceeds by devising a method to consistently estimate the distribution of that term. Our illustration implies that by ignoring the multidimensional way unobserved heterogeneity in trade-offs can enter into the model (for example, in terms of both levels and slopes), a researcher can be led to conclude that, for example, one group of workers is more productive than another group when one group is simply more flexible than another group in terms of their preferences over wage-amenities trade-offs. Therefore, whether a single index assumption is reasonable or not then depends on how much of a concern the situation illustrated in our Figure 4a would be in the specific empirical context in which they are applied.[9]

Fundamentally, the shortcomings of existing empirical methodologies, when applied to the menu problem, can be summarized as a combination of omitted variables bias and mis-specificiation bias. Intuitively, the assumptions underlying the heuristic approaches of assessing discrimination in menus are that (1) all menus share the same shape, with unobserved heterogeneity in menus being due to an additive error term, and (2) that this shape can be correctly estimated/inferred from other data. The omitted variables problem stems from the fact that the preferences of consumers, since they are unobserved, can lead to bias when the slopes of menus are estimated from data by controlling for them, thus breaking implicit assumption (2). This is the main problem with simply controlling for some dimensions of the menu, as illustrated in Figures 2 and 3. While it is less recognized in the mortgage pricing discrimination literature, much work in labor productivity and production function estimation is indeed aimed at estimating the slopes of trade-offs and addressing the omitted variables problem in (2). A more subtle problem is that correctly estimating the average slopes of trade-offs is not sufficient for assessing equality in menus $\mathcal{H}_0 : \mathbf{M}_1 = \mathbf{M}_2$. Models that assume away heterogeneity in slopes and interpret residuals to average trade-offs as

[9]Another setting where the single (log-)additive productivity assumption is popular is in the production function literature (e.g. Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015)). Therefore, while methods from the production estimation literature are useful for estimating average returns to capital and labor, they are also problematic if applied as a test of equality in menus as in our mortgage context.

differences in menus can generate a model mis-specification error when such heterogeneity does exist in reality, which can lead to misleading inference by breaking assumption (1). This is the problem illustrated in Figure 4, which is likely empirically relevant in the mortgage setting. Therefore, we point out a series of difficulties associated with using existing methodology to test $\mathcal{H}_0 : \mathbf{M}_1 = \mathbf{M}_2$, which we call the "menu problem."

3 Robust Metrics for Assessing Discrimination in Menus

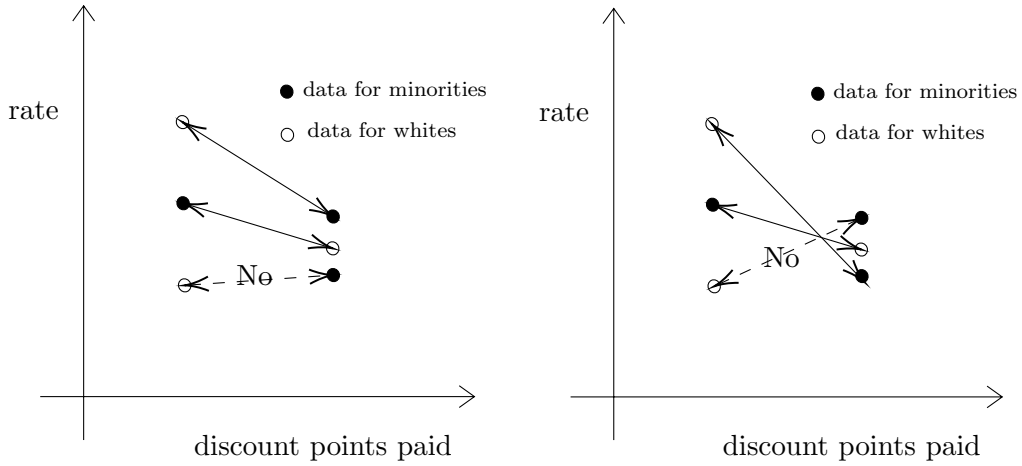
In this section, we define our robust metrics for assessing discrimination in menus. First, we present intuition for our approach in Section 3.1 that equality in menus should imply the existence in the data of a one-to-one "match" between minority and white consumers who could have faced the same menus. Next, we specify our model more formally in Section 3.2. We keep our model fairly simple; menus are treated simply as a collection of items. Then, we define a direct test metric for equality in menus in Section 3.3 and 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 3.4. We discuss the power of our identification results in Section 3.5. Finally, we leave inference on these metrics Section 4.

3.1 Intuition

Our new metrics for assessing discrimination in menus is 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 Figures 3a and 4a is that there does exists a possible common distribution of menus that rationalizes the choice distributions of both minority and white borrowers, in the sense of there being a possible one-to-one match between minority and white borrowers where within each match both of the borrowers' choices could have come from the same menu. This is the criteria we use for assessing equality in menus.

By construction, our metrics are robust to false positives, since we will detect discrimination in menus only when there is no way to rationalize the data under the assumption of a common distribution of menus, regardless of the nature of any preference heterogeneity between the two groups of borrowers. Our methodology can also detect discrimination in menus where the existing heuristic approaches to the menu problem fail to do so. In particular, we illustrate in Figure 5 how the situation of Figures 3b, in which regressions controlling for either rates or points, would show a false negative, but the data fail our one-to-one matching condition under the assumption that borrowers with pairwise strictly

dominated choices (that is, 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.



(a) Possibility #1 for matching fails (b) Possibility #2 for matching also fails

Figure 5: How data from Figure 3b fail a "perfect matching" condition

To summarize, we define new metrics based on whether a set of preferences can rationalize the data under equality in menus, which are robust to the false positives. Furthermore, in some situations, such as those in Figures 3b and 4b, our metrics can detect discrimination when existing, heuristic approaches fail to do so. Nevertheless, a drawback of our approach is that it still leaves 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 a weakness compared with experimental approaches 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.

3.2 Model

We now define our model more precisely. A menu item has values over k dimensions of attributes,[10] which we encode by $x \in \mathbb{X} \subset \mathbb{R}^k$. A menu $\mathbf{m} \subseteq \mathbb{X}$ is a set of such menu items that are presented to the borrowers. When borrower i is presented with a menu $\mathbf{m}$, we observe them making a choice that maximizes their utility over menu items $u_i(x)$. That

[10]In our context, the two dimensions of mortgage pricing are interest rates and discount points.

is, we observe choices x_i where:

$$x_i \in \arg\max_x \{u_i(x) : x \in \mathbf{m}\}. \tag{3}$$

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.*

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 (that is, complete and transitive) preferences over menu items.[11]

3.3 A robust test for inequality in menus

Suppose borrowers with a similar distribution of covariates in groups 1 and 2 face menus $\mathbf{m}_1 \sim \mathbf{M}_1$ 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 distributions of menus being offered to both groups are equal. That is:

$$\mathcal{H}_0 : \mathbf{M}_1 = \mathbf{M}_2, \tag{4}$$

$$\mathcal{H}_1 : \neg\mathcal{H}_0. \tag{5}$$

To go from data on choices to statements about menus, we place restrictions on the choices that 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 use 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 $r_1 \geq r_2, y_1 > y_2$, $u_i(x_1) < u_i(x_2), \forall i$.*

[11] Our Equation (3) that consumers maximize utility over menu items does rule out more behavioral representations of preferences over menus, such as in Gul and Pesendorfer (2001), where 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 (DIM) metric would be nuanced.

We illustrate in Figure 6 the restrictions on the observed choices that may come from 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 in 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, allows us to define a welfare-relevant metric for assessing differences in menus later on.

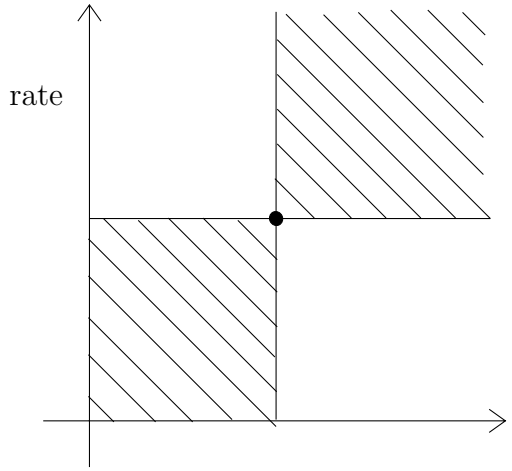


Figure 6: Restriction on what cannot lie on the same menus from dominance.

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 one-eighth to one-fourth 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 lie on the same menu* $\{x_1, x_2\} \subseteq \mathbf{m}$ *only if:*

$$a \leq \frac{r_2 - r_1}{y_2 - y_1} \leq b \text{ or } x_1 = x_2, \tag{6}$$

where $0 \leq a \leq b \leq \infty$*, and* $x_1 = [r_1, y_1], x_2 = [r_2, y_2]$*, with* 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, the range of possible choices that could have come from the same menu as that of 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 3.4.

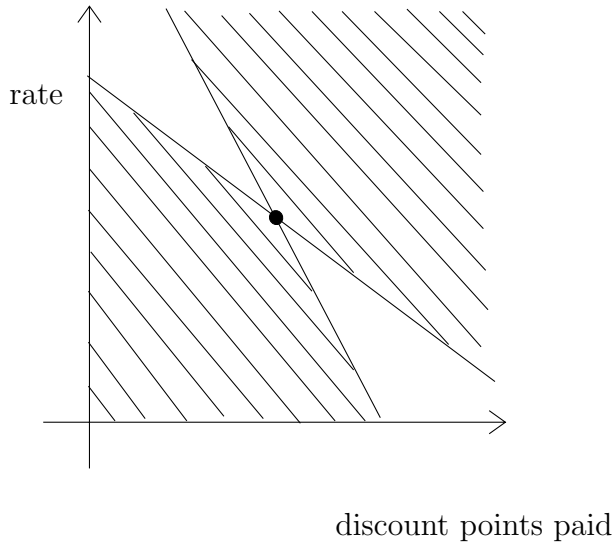
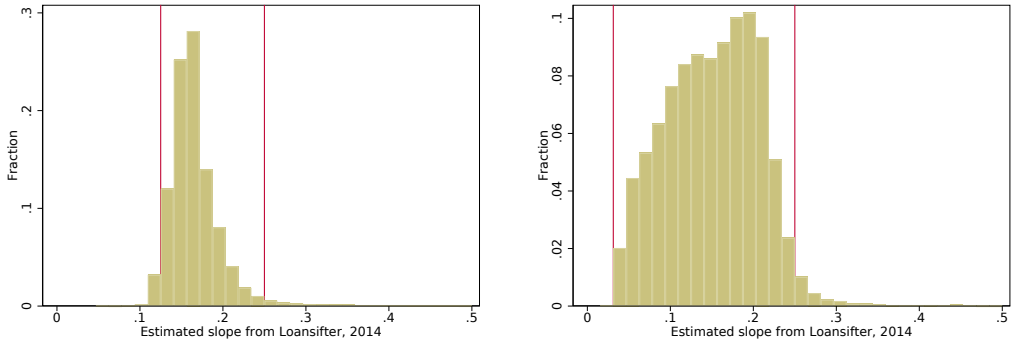


Figure 7: Restriction on what cannot lie on the same menus from the "rule of thumb" of Bartlett et al. (2019).

Empirically, we find that the industry rule of thumb from Bartlett et al. (2019), which motivated Assumption 3, covers a vast majority of menus based on data from a sample of lender rate sheets that enumerates the rate and discount point menus. Using the 2014 LoanSifter data from Fuster, Lo, and Willen (2019), we estimate slopes of menus within their sample of 30-year 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 rule of Bartlett et al. (2019) that each point paid is worth 1/8 to 1/4 of a point covers 94.4 percent of all rate sheet observations for conforming mortgages. For FHA mortgages, we use

an expanded rule that each point is worth 1/32 to 1/4 in rate, which covers 97.6 percent 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: Rate sheet 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 trade-off 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.

We see substantial heterogeneity between the menu slopes across lender-weeks in Figure 8, perhaps reflecting market power or lender- and time-specific costs. As we explained earlier, 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 Assumptions 2 and 3, let $x_1 = [r_1, y_1], x_2 = [r_2, y_2]$. We define an indicator function for whether choices x_1, x_2 could have come from the same menu:

$$\phi(x_1, x_2) = \begin{cases} 1 \text{ , if } a \leq \frac{r_2 - r_1}{y_2 - y_1}, \leq b \text{ or } x_1 = x_2 \\ 0 \text{ , otherwise.} \end{cases} \tag{7}$$

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* $\boldsymbol{\pi}(x_1, x_2) : \mathbb{X} \times \mathbb{X} \to [0, 1]$ *with implied marginal densities*

$\sum_{x_2} \boldsymbol{\pi} = \mathbf{p}_1, \sum_{x_1} \boldsymbol{\pi} = \mathbf{p}_2$ *such that:*

$$T \equiv 1 - E_{\boldsymbol{\pi}}\phi(x_1, x_2) = 0. \tag{8}$$

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 (8) 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 the extent to which this matching is 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:*

$$\hat{T} = \min_{\boldsymbol{\pi}(x_1,x_2)} 1 - E_{\boldsymbol{\pi}}\phi(x_1, x_2), s.t. \ \sum_{x_2} \boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1} \boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0, \tag{9}$$

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 (9) is a finite dimensional optimal transport objective that can be efficiently computed using a linear program. To deal with sampling error inherent in $\hat{T}$, we discuss in Section 4 how critical values for $\hat{T}$ from Equation (9), as an optimal transport objective, can be consistently simulated.

3.4 Metric for assessing differences in menus

The direct test of inequality in the distribution of menus in Section 3.3, 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 from the distribution of menus presented to another group may not be of welfare consequence, a problem that Abadie (Forthcoming) discusses in more detail. Rather, for the purposes of comparing menus across two distributions, we want a metric for assessing whether the distribution of 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 maps 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 (10):

$$\Delta W_{\mathcal{I}_1 \to \mathcal{I}_2, \boldsymbol{\pi}} = \sum_{i \in \mathcal{I}_1, j \in \mathcal{I}_2} \pi(i,j)(u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i)). \tag{10}$$

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 $d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j)$ that measures the extent to which consumer i would be willing to increase the interest rates on their loan in order to switch from menu $\mathbf{m}_i$ to menu $\mathbf{m}_j$:

$$u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i) = d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j) \equiv \sup\{\delta \in \mathbb{R} : u_i(\{x + \delta e^r, x \in \mathbf{m}_j\}) \geq u_i(\mathbf{m}_i)\}, \tag{11}$$

where e^r represents a basis vector that is equal to 1 at the location indexing interest rates. 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 $d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j)$ is directly proportional to the utility change for consumer i after switching to menu j. Even if consumers do not have constant marginal utility over interest rates, it is still meaningful as a "willingness to pay" metric, since $d_{i \to j}(\mathbf{m}_i, \mathbf{m}_j)$ measures the extent to which consumer i would be willing to increase the interest rates to switch from $\mathbf{m}_i$ to $\mathbf{m}_j$. 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.[12]

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 (for example, 0.134, 0.266, ...) rather than literally the full range, but the implications of such small gaps in menus are 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 (that is, 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.

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

[12] 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 that are likely binding for many consumers, constant marginal utility is very unlikely to hold for points, which makes for a weaker welfare interpretation.

The effect of 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 in the white borrower's menu dominate 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 the assumption of menu completeness in points sharpens the comparison of menus.

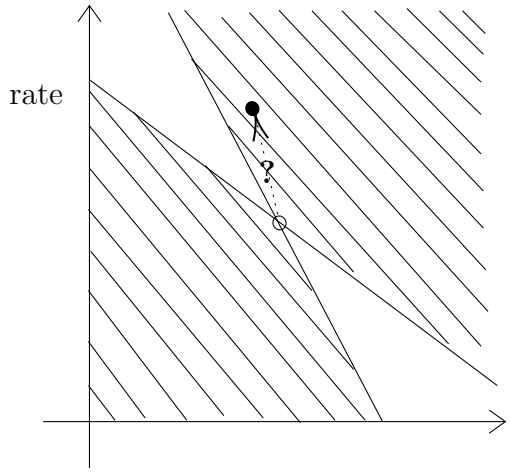


Figure 9: Impact of assuming that menus are complete in either rates or points

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, borrower 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 menu of borrower 2, who made a choice x_2 from a menu $\mathbf{m}_2$. This implies that, by revealed preference of borrower 1, the menu that borrower 2 faced is better than borrower 1's menu, or $\mathbf{m}_2 \succ_1 \mathbf{m}_1$. For borrower 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 borrower 1 is willing to pay at least $\underline{d}_{1\to 2}(x_1, x_2)$ more in interest rate in order to get borrower 2's menu. Similarly, the menu faced by borrower 2 would need to be shifted downward by at most the negative $\underline{d}_{3\to 2}$ before it dominates x_3's choice. Therefore, borrower 3, with choice x_3, would need to receive at most $-\underline{d}_{3\to 2}$, or pay at least $\underline{d}_{3\to 2}$, before being willing to switch

to borrower 2's menu.

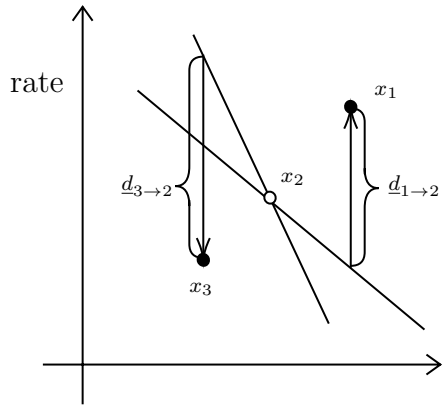


Figure 10: Lower bound for borrower 1's willingness to pay a higher interest rate in order to get borrower 2's menu.

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

$$\underline{d}_{1\to 2}(x_1, x_2) \equiv r_1 - r_2 + a\max(y_1 - y_2, 0) + b\min(y_1 - y_2, 0) \leq d_{1\to 2}(x_1, x_2), \tag{12}$$

where in the first line j indexes points, and in the second line we 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 borrower 1 would be willing to pay to switch to the menus of borrower 2, $\underline{d}_{1\to 2}(x_1, x_2)$, then allows us to define our differences in menus (DIM) metric:

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

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

where $\underline{DIM}_{1\to 2}$ *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. If borrowers have constant marginal utility in interest rate, then:*

$$\underline{DIM}_{1\to 2} \leq \Delta W_{\mathcal{I}_1\to\mathcal{I}_2,\boldsymbol{\pi}} \tag{14}$$

where utility is measured in terms of interest rate.

Proof. See Appendix A.1.2. □

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 $\Delta W_{\mathcal{I}_1,1\to2,\pi_{1\to2}}$. If instead consumers do not have constant marginal utility over interest rates, then the $\underline{DIM}_{1\to2}$ 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 $\underline{DIM}_{1\to2} \leq 0$, so a finding that $\underline{DIM}_{1\to2} > 0$ 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, $\underline{D\hat{I}M}_{1\to2}$, given choice probabilities $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2$, is:*

$$\underline{D\hat{I}M}_{1\to2} = \min_{\boldsymbol{\pi}(x_1,x_2)} E_{\boldsymbol{\pi}}\underline{d}_{1\to2}(x_1,x_2), s.t. \sum_{x_2}\boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1}\boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0 \tag{15}$$

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

3.5 When does our method have power?

Our metrics for differences in menus are robust in the sense that they are immune to the false positives problem from which the existing methods suffer, but they may still generate false negatives in that there exist scenarios where the data are rationalizable under equality in the distribution of menus, but in fact the distribution is different. Generally speaking, our method has power only to the extent that the borrowers' choices cannot be rationalized by an equal distribution of menus under the restrictions on preferences and menus that we have made. In this subsection, we give some examples of scenarios where this would or would not occur. In our later empirical analysis, we show that our methodology does have enough power to be useful in detecting discrimination in mortgage markets.

Figure 11 illustrates a scenario in which white and minority borrowers face the same default menu, but some white borrowers are sometimes offered a discretionary discount in terms of points. In the figure, the menu represented by the dashed line is shifted leftward for a white borrower, but all minority borrowers face the original menu. This shift makes the bottom-left choice by the white borrower not matchable to any of the minority borrowers' choices, so our one-to-one matching condition for equality in menus in Theorem 1 is broken. Our Theorem 1 would therefore have power to detect a difference in the distribution of menus offered to white and minority borrowers in this case. Furthermore, when the discretionary

discount being given to white borrowers is large enough relative to the range of permissible menus, our DIM metric defined in Theorem 2 would also show that minority borrowers would be willing to pay to switch to the white borrowers' menus.

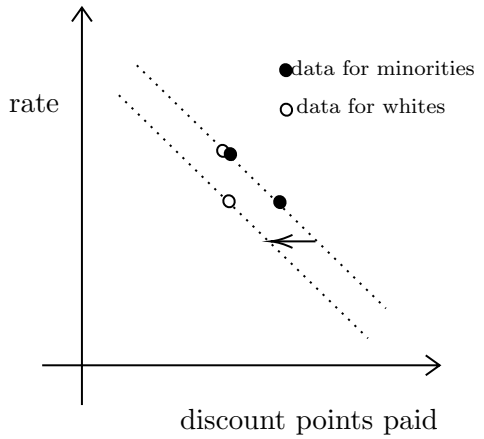


Figure 11: Power to detect discrimination when discretionary discounts offered to only some white borrowers make the data not rationalizable under equality in menus

Figure 12 gives an example situation in which minority and white borrowers were presented with different menus, with white borrowers on average paying less in both rates and points, but the data are rationalizable under equality in menus. More specifically, while the data were generated by menus represented by the dotted lines, which differ in distribution between white and minority borrowers, the data can be rationalized with one minority and one white borrower both choosing from the hypothetical menu represented by the dashed line and the other borrowers choosing from the correct dotted line menu. In this scenario, we cannot rule out that the true distribution of menus is represented by the dashed line plus a dotted line rather than the two dotted lines, and therefore our metrics from both Definition 1 and Definition 2 would fail to reject equality in menus.

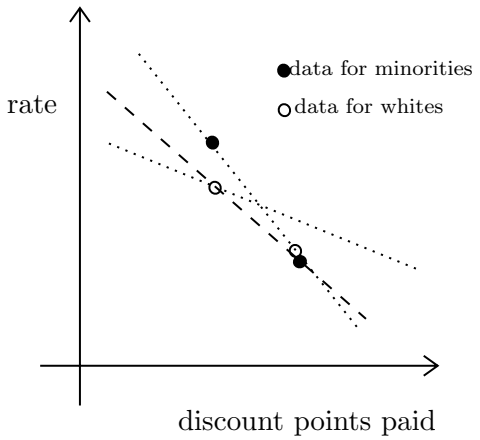


Figure 12: No power when the data can be rationalized by (incorrect but plausible) distribution of menus that are equal across the racial groups

While our approach cannot detect differences in menus in the situation of Figure 12, there is a good reason for this: It is possible that the borrowers' choices were truly generated by the dashed line and the dotted line such that the racial groups did in fact face the same distribution of menus. It would be difficult to rule out that possibility without more stringent assumptions. Therefore, in our quest to be robust to the false positives problem that existing methods suffer under heterogeneity in borrower preferences by group, we leave open the possibility of false negatives when the data are rationalizable by an equal distribution of menus across the racial groups. Empirically in Section 5, we show that we are able to reject equality in menus for conforming mortgages, which is evidence that our method does have enough power to be useful in our mortgage setting.

4 Inference for Optimal Transport

In this section, we devise a new procedure for conducting hypothesis testing on the values of optimal transport problems that 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). While some existing methods in the literature can be applied in more general contexts compared with optimal transport, they are either too conservative or converge too slowly based on our simulations, which we discuss at the end of this section. Our new procedure may be of independent interest for other researchers who wish to apply optimal transport methods to economics; some of those applications are listed in Galichon (2016).

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 $\mathbf{p}_1, \mathbf{p}_2$:

$$\hat{\varphi}(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) = \min_{\boldsymbol{\pi}(x_1,x_2)} E_{\boldsymbol{\pi}}\phi \text{ s.t. } \sum_{x_2} \boldsymbol{\pi} = \hat{\mathbf{p}}_1, \sum_{x_1} \boldsymbol{\pi} = \hat{\mathbf{p}}_2, \boldsymbol{\pi} \geq 0, \tag{16}$$

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 $\mathbf{p}_1, \mathbf{p}_2$ being less than or equal to some value φ_0:

$$\mathcal{H}_0 : \varphi(\mathbf{p}_1, \mathbf{p}_2) \leq \varphi_0, \tag{17}$$

$$\mathcal{H}_a : \varphi(\mathbf{p}_1, \mathbf{p}_2) > \varphi_0. \tag{18}$$

The form of the null hypothesis in Equation (17) 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 $DIM \leq DIM_0$) 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 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. To do 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 use the definition of Hadamard directional differentiation from Fang and Santos (2018), with some notational differences tailored 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. *(Fang and Santos, 2018) A map $\varphi : \mathbb{D}_\varphi \to \mathbb{R}$ is said 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:*

$$\lim_{n\to\infty} \left\| \frac{\varphi(\theta + t_n \mathbf{h}_n) - \varphi(\theta)}{t_n} - \varphi'_\theta(\mathbf{h}) \right\| = 0 \tag{19}$$

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 and the typical notion of differentiability is that t_n is restricted to be positive in Definition 3 but not in the standard definition of differentiability. That is, loosely speaking, the directional derivatives represent the change in the value of the function for a small change in its inputs "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 (the value of an optimal transport problem with the cost function restricted to distance metrics) on finite spaces is directionally differentiable.[13]

Theorem 3. *The value φ of an 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 derivative equal to:*

$$\varphi'_{\mathbf{p}_1,\mathbf{p}_2}(\mathbf{h}_1, \mathbf{h}_2) = \max_{\mathbf{u},\mathbf{v}\in\Psi^*(\mathbf{p}_1,\mathbf{p}_2)} \mathbf{h}_1^T\mathbf{u} + \mathbf{h}_2^T\mathbf{v} \tag{20}$$

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.1.3. □

Under i.i.d. sampling, we know that $\hat{\mathbf{p}}_1 - \mathbf{p}_1$ and $\hat{\mathbf{p}}_2 - \mathbf{p}_2$ approach a multivariate normal distribution:

$$\hat{\mathbf{p}}_1 - \mathbf{p}_1 \to^d \mathrm{N}\left(0, \begin{bmatrix} p_{1,1}(1-p_{1,1}) & -p_{1,1}p_{1,2} & \dots \\ -p_{1,2}p_{1,1} & p_{1,2}(1-p_{1,2}) & \dots \\ \dots & \dots & \dots \end{bmatrix}\right), \tag{21}$$

and likewise for $\hat{\mathbf{p}}_2 - \mathbf{p}_2$, such that by construction, Assumptions 2.1 and 2.2 of Fang and Santos (2018) are satisfied. Then, Theorem 2.1 of Fang and Santos (2018) immediately implies that:

$$r_n[\varphi(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) - \varphi(\mathbf{p}_1, \mathbf{p}_2)] = \varphi'_{\mathbf{p}_1,\mathbf{p}_2}(r_n[(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) - (\mathbf{p}_1, \mathbf{p}_2)]) + o_p(1), \tag{22}$$

[13] 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.

such that the asymptotic distribution of $\varphi(\{\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2\})$ can be obtained via the directional Delta method. The remaining challenge for inference is that the true $\mathbf{p}_1, \mathbf{p}_2$ used in $\varphi'_{\{\mathbf{p}_1,\mathbf{p}_2\}}$ in Equation 22 are not known, and thus must be estimated. While we could have used the plug-in analogue $\varphi'_{\{\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2\}}$, that would converge only pointwise and not uniformly, which is known in the moment inequalities literature to be a poor approximation to the finite sample properties of estimators for which there are discontinuities in the pointwise asymptotic distribution (Andrews and Soares, 2010). We deal with this problem using the logic of Romano, Shaikh, and Wolf (2014) and McCloskey (2017). More specifically, we construct a confidence band for $[\mathbf{p}_1, \mathbf{p}_2]$ at level β such that:

$$\limsup_{n\to\infty} \Pr([\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_{n,\beta}) \geq 1 - \beta. \tag{23}$$

Many uniform confidence bands satisfying Equation (23) are available, for example from Montiel Olea and Plagborg-Møller (2019). Then, we take our estimate of the directional derivative as the maximum directional derivative within this confidence band:

$$\hat{\varphi}'_\beta(\mathbf{h}_1, \mathbf{h}_2) = \max_{\mathbf{u},\mathbf{v}\in\Psi(\mathbf{p}_1,\mathbf{p}_2):[\mathbf{p}_1,\mathbf{p}_2]\in\hat{\mathcal{P}}_\beta} \mathbf{h}_1^T\mathbf{u} + \mathbf{h}_2^T\mathbf{v}, \tag{24}$$

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 define the critical value for φ by using our estimated maximum directional derivative from Equation (24) at level $1 - \alpha + \beta$:

$$\hat{c}_{1-\alpha+\beta} = \inf\{c \in \mathbb{R} : \Pr(\hat{\varphi}'_\beta(\mathbf{h}_1, \mathbf{h}_2) \leq c) \geq 1 - \alpha + \beta\}, \tag{25}$$

where the distribution of $\mathbf{h}_1, \mathbf{h}_2$ is the asymptotic distribution of $\hat{\mathbf{p}}_1 - \mathbf{p}_1, \hat{\mathbf{p}}_2 - \mathbf{p}_2$. In the following Corollary 1, we will prove uniform coverage for when the observed value $\hat{\varphi}$ is less than the critical value $\hat{c}_{1-\alpha+\beta}$, and suggest a computationally tractable version of it as a linear program with complementarity constraints (LPCC).[14]

Corollary 1. *Suppose we have uniform confidence bands for $[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta$ that provide uniform coverage as in Equation (23), then under $\mathcal{H}_0 : \varphi \leq \varphi_0$:*

$$\limsup_{n\to\infty} \Pr(\hat{\varphi} - \varphi_0 \geq \hat{c}_{1-\alpha+\beta}) \leq \alpha, \tag{26}$$

[14] 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.

where $\hat{c}_{n,1-\alpha+\beta}$ is computed as in Equation (25).

Proof. See Appendix A.1.4. □

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 (24). Computationally, directly maximizing over the directional derivative defined in Equation (24) is difficult, because it involves optimizing over a nonlinear 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 such as this can be found in standard texts on optimal transport/linear programming. In particular, Hsieh, Shi, and Shum (2017) use 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 $\mathcal{H}_a : \varphi > \varphi_0$ can be written as the following LPCC:

$$\hat{\varphi}'_\beta(\mathbf{h}_1, \mathbf{h}_2) = \max_{\mathbf{u},\mathbf{v},\mathbf{p}_1,\mathbf{p}_2,\mathbf{s},\boldsymbol{\pi}} \mathbf{h}_1^T\mathbf{u} + \mathbf{h}_2^T\mathbf{v}, \tag{27}$$

$$\sum_{x_2} \boldsymbol{\pi} = \mathbf{p}_1, \tag{28}$$

$$\sum_{x_1} \boldsymbol{\pi} = \mathbf{p}_2, \tag{29}$$

$$E_{\boldsymbol{\pi}}\phi \leq \varphi_0, \tag{30}$$

$$u(x_1) + v(x_2) + s(x_1, x_2) = \phi(\mathbf{p}_1, \mathbf{p}_2), \tag{31}$$

$$[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta, \tag{32}$$

$$\boldsymbol{\pi}, \boldsymbol{s} \geq 0, \tag{33}$$

$$\boldsymbol{\pi} \leq 0 \perp \boldsymbol{s} \geq 0. \tag{34}$$

To test our econometric approach, we conduct a Monte Carlo simulation with two possibilities for points $\{0, 1\}$ and five possibilities for rate $\{3, 3.25, 3.5, 3.75, 4\}$. Furthermore, Black and white borrowers choose each of the 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 $\beta = \frac{1}{10}\alpha$ following Romano, Shaikh, and Wolf (2014), and show the simulated probability that we reject equality in menus $\mathcal{H}_0 : \varphi_0 = 0$ at the 1 percent, 2.5 percent, 5 percent, and 10 percent levels in

Table 1. As Table 1 shows, our approach has the 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 with 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) have a novel projection method for parameter inference in mathematical programming problems, which is a broader set of problems than optimal transport, but their approach is conservative. In our empirical context, this conservativeness tends to make the confidence intervals uninformative. Another approach that is theoretically valid in this setting is the general m-out-of-n subsampling method of Politis and Romano (1994), but in addition to requiring the researcher to choose a subsample size m, it can require very large samples for convergence.

We look at the control of these competing approaches at the 5 percent significance level under our simulation setting in Appendix Table A.1. In that table, HSS (2017) refers 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 directional derivatives approach is replicated from Table 1 for comparison. Table A.1 shows that the HSS (2017) approach rejects the null with probability close to 0 percent, implying that it is conservative. On the other hand, the m-out-of-n subsampling approach tends to reject at rates greater than 5 percent for all values of m we tried, and appears to require more than 50,000 observations in order to converge to the correct rejection rate, which is significantly larger than our available sample size.[15]

[15]A method related to m-out-of-n subsampling is the numerical bootstrap of Hong and Li (2020), which can be more data efficient than the subsampling method, but we were not able to find a suitable choice of ϵ_n that converges to the correct coverage in our simulations using that method.

In Section 5, we also report one-sided confidence intervals from the inversion of our hypothesis test. Nothing in our econometric theory precludes us from also testing the other direction and reporting two-sided confidence intervals instead.[16] However, since our economic theory is focused on getting a *lower bound* for the existence and welfare effects of discrimination, one-sided confidence intervals are particularly suitable for our purposes.

In summary, in this section we devised a new procedure for inference in optimal transport that is uniformly valid and not overly conservative for our empirical context. Our empirical analysis in 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 data set constructed via matching the 2018–2019 public Home Mortgage Disclosure Act (HMDA) data to Optimal Blue rate locks. The public HDMA data contain information on borrower race and ethnicity, and we take the borrowers with an 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 of Hispanic borrowers, and borrowers with an HMDA-derived race of "White" along with a HMDA derived ethnicity of "Not Hispanic or Latino" as our sample of non-Hispanic white borrowers.

Starting from the uniquely matched HMDA-Optimal Blue matched data set, 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 include the complete set of mortgages originated in the United States with such characteristics, to our matched sample. We find that our matched sample has very similar average loan sizes, LTVs, rates, points, and the percentage composition of Black and Hispanic borrowers compared with the HMDA data, as can be seen from Table 2. One known caveat to this is that lenders using the Optimal Blue 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).

[16] The other direction is, however, computationally more challenging, since it would involve taking the minimum of a maximum.

Table 2: Comparison of means between the 2018–2019 HMDA data and our HMDA-Optimal Blue 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-Optimal Blue 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 rates 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 non-Hispanic 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 non-Hispanic white borrowers that are exactly matched on their covariates. If lenders offered Black and non-Hispanic white borrowers the same distribution of menus conditional on covariates, our covariate-matched sample of Black and non-Hispanic 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 eight categories of FICO scores and nine 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 are in Table 3. 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 halves, and we show in Table 3 that this step does not substantively change the mean differences in rates or points.[17]

[17]Both the HMDA data and the Optimal Blue data contain information about discount points paid that sometimes disagrees 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. More specifically, we find, as shown 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 with the Optimal Blue information on points. Furthermore, in a regression controlling for the HMDA points, the effect of Optimal Blue 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 Optimal Blue definition of points. Our

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 and rounded rates and points were demeaned by lender-county-month and covariate group and then rounded into the nearest eighths for rate and halves for points.

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 paid only 2.9 basis points more in interest rate and –3.2 basis points fewer 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 mortgages is different from the distribution of data for FHA mortgages. Nevertheless, we will show in Section 5.2 that this difference between the distributions of data is, however, not sufficient by itself to reconcile the results of Bartlett et al. (2019) and Bhutta and Hizmo (Forthcoming), the former of whom found evidence of discrimination in conforming mortgages while the latter did not in FHA mortgages. We find that the choice of heuristic used in analysis also contributes to the discrepancy in results.

5.2 Results from heuristic analyses

We show the results from the heuristic approaches we discussed in our sample in Table 4. In Table 4, Heuristic 1 in columns (1) and (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). Heuristic 2 in columns (3) through (8) refers to the comparing of 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) and (8) show results from an alternative form of Heuristic 1, where we control for points and compare rates instead.

results using the Optimal Blue data on points are qualitatively similar.

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	No	No	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 presents heuristic analyses of data for our Black 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 excluded, and rate spreads below –55 basis points and above 90 basis points were excluded.

Columns (1) and (2) of Table 4 show that the approach taken in Heuristic 1 would detect lender overcharge in conforming mortgages but not in FHA mortgages: with Black borrowers paying 4.1 basis points more discount points for conforming mortgages but a statistically insignificant –0.7 basis points *fewer* discount points for FHA mortgages. Thus, Column (2) replicates the findings of Bhutta and Hizmo (Forthcoming) that lenders appear to have not overcharged Black borrowers in terms of points, after controlling for rate in FHA mortgages. However, Columns (3) through (6) show that Heuristic 2 would consistently detect discrimination for both conforming and FHA mortgages, with magnitudes of 2.9 to 3.3 basis points for conforming mortgages and 2.4 to 2.9 basis points for FHA mortgages. Thus, our result using Heuristic 2 is consistent with Bartlett et al. (2019), who focus on conforming mortgages. Therefore, the result of no overcharge in FHA mortgages is sensitive to the choice of the heuristic that is used.

Furthermore, comparing Column (2) with Column (8) in Table 4 shows that the choice of which menu dimension to control for in implementing Heuristic 1 (that is, 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 are 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 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 between menus for Black borrowers and menus for white borrowers using our Definition 1. We find, as shown 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 versus non-Hispanic white matched sample, our test statistic in column (1) is $\hat{T} = 2.77$, which indicates that 2.77 percent of Black borrowers' choices in the data could not have been matched to those of non-Hispanic white borrowers' choices 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 from random chance is less than 0.1 percent. For Hispanic borrowers, our test statistic is $\hat{T} = 1.53$ in column (3), with a p-value of less than < 0.001. For FHA mortgages, on the other hand, we are unable to reject equality between the menus for Black and non-Hispanic white borrowers, with a test statistic of $\hat{T} = 0$ in column (2), but we are able to reject it 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, ∞)	[0.00, ∞)	[0.89, ∞)	[1.16, ∞)
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 computed 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 that 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 difference 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 $\hat{\boldsymbol{p}}_1$, $\hat{\boldsymbol{p}}_2$ using our procedure in Section 4, and confidence intervals are computed through inversion of the hypothesis test. The sample sizes for the conforming and FHA Black versus non-Hispanic white matched samples are equal by coincidence.

In summary, we find that lenders offered Black and Hispanic borrowers a distribution of menus that was different from the distribution they offered non-Hispanic white borrowers for conforming mortgages. In particular, 2.77 percent of Black borrowers' choices were unable to be matched to non-Hispanic white borrowers based on column (1) of Table 5, and for those unmatched Black borrowers their minimum willingness to pay to switch to unmatched non-Hispanic white borrowers' menus is at least 73 basis points (2.03/0.0277, combining with information in column (1) of Table 6). Similarly, column (3) of Table 5 shows that 1.53 percent of Hispanic borrowers' choices were unable to be matched to those of non-Hispanic white borrowers, with an average willingness to pay to switch to unmatched non-Hispanic white borrowers' menus being at least 99 basis points among those borrowers (1.52/0.0153, combining with information in column (3) of Table 6).

On the other hand, our results for FHA mortgages is more mixed: Column (2) of Table 5 shows that we cannot reject equality between the menus for Black and non-Hispanic white borrowers for FHA mortgages, and while column (4) of Table 5 shows that 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 6 in terms of the average increase in rate that Hispanic borrowers would be willing to pay in order to receive the menus of non-Hispanic white borrowers.

5.4 Further analyses of differences in menus

In Section 5.3 we showed that lenders offered Black and Hispanic borrowers a less advantageous distribution of menus compared with the distribution they offered observationally similar non-Hispanic white borrowers for conforming mortgages, but we were silent as to why this differential pricing appears. One possibility is that minority borrowers are more likely to get mortgages from loan originators that charge more for their services (that is, that they pay more in broker fees, in the sense of Ambrose, Conklin, and Lopez (2020)). The Optimal Blue data contain information on loan originator compensation for a subsample of lenders. We match on levels of loan originator compensation, rounded to the nearest 1 percent of the loan amount, and find, as shown in Table 7, that our results are qualitatively unchanged. In particular, the point estimate for the minority-to-white $\underline{D\hat{I}M}_{1\to 2}$ measure for conforming mortgages increased slightly, from 2.03 to 2.65, for Black borrowers and decreased slightly, from 1.52 to 1.19, for Hispanic borrowers. The confidence intervals are wider, reflecting a significantly smaller sample size.

Table 7: Analysis of differences in menus after further matching on loan originator compensation

	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 ($\hat{T}$)				
	4.43^{**}	2.53	1.64^{*}	3.58
95% CI	[0.51, ∞)	[0, ∞)	[-0.54, ∞)	[0.30, ∞)
Panel B: Difference in Menus ($D\hat{I}M$) Metric				
Minority to white ($\underline{D\hat{I}M}_{1\to 2}$)	2.65^{***}	-0.35	1.19^{**}	-1.23
95% CI	[1.09, ∞)	[-2.04, ∞)	[0.21, ∞)	[-2.59, ∞)
White to minority ($\underline{D\hat{I}M}_{2\to 1}$)	-6.49	-9.22	-6.11	-9.43
95% CI	[-8.03, ∞)	[-10.96, ∞)	[-7.05, ∞)	[-10.76, ∞)
N	1,818	1,818	5,290	2,438

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

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 ($\underline{D\hat{I}M}$) in units of basis points based on our Definition 2, with $\underline{D\hat{I}M}$ in units of basis points. We match on lender-county-month and covariate groups plus loan originator compensation 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 $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2$ using our procedure in Section 4, and confidence intervals are computed through inversion of the hypothesis test.

To further explore where differences in menus occurs, we divide the sample into different

LTV and FICO buckets, and detect more discrimination among the lower LTV and higher FICO (that is, more creditworthy) borrowers. In particular, columns (1) and (3) of Table 8 show that we detect large differences in menus for conforming mortgages in the LTV under 75 and LTV of 75 to 80 categories, such that Black and Hispanic borrowers would be willing to pay 6 and 5 basis points, respectively, to switch to non-Hispanic white menus. While we can detect some differences in menus in the 80 to 90 LTV range, it is smaller,and over a 90 LTV, we can no longer say that Black and Hispanic borrowers would be willing to switch to non-Hispanic white menus. Similarly, column (1) of Table 9 shows that we detect significant differences in menus for Black borrowers only in the FICO over 740, FICO 720 to 740, and FICO 700 to 720 categories, and smaller differences for the FICO categories below 700.

An internally consistent explanation for the fact that we primarily detect mortgage pricing discrimination among conforming, low-LTV and high-FICO borrowers is that lenders are more willing to offer discretionary discounts to the creditworthy non-Hispanic white borrowers, and less willing to do so for minority borrowers who are similarly creditworthy in terms of their underwriting variables. In particular, as shown in Ambrose, Conklin, and Lopez (2020), racial differences in default risk are very low among lower-LTV and higher-FICO borrowers, and in any case they are insured. Kau, Fang, and Munneke (2019) and Gerardi, Willen, and Zhang (2020) both show that the lower prepayment risk of minority borrowers makes the securities backed by their mortgages more valuable. Furthermore, it seems unlikely given the strict regulatory environment surrounding the mortgage market that lenders would condition their rate sheets and first offers based on race. Therefore, by process of elimination, we believe that the search and negotiation process, particularly for the more creditworthy borrowers, may play an important role in the within-lender disparate outcomes we find.

Table 8: Analysis of differences in menus comparing borrowers across categories of loan-to-value (LTV) ratio

	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 ($\hat{T}$)				
LTV ≤ 75 ($\hat{T}$)	8.04^{***}	-	7.88^{***}	-
95% CI	[3.09, ∞)	-	[4.68, ∞)	-
$75 <$ LTV ≤ 80 ($\hat{T}$)	7.23^{***}	-	7.06^{***}	-
95% CI	[5.00, ∞)	-	[5.39, ∞)	-
$80 <$ LTV ≤ 90 ($\hat{T}$)	2.56^{***}	-	0.94^{*}	-
95% CI	[-0.02, ∞)	-	[0, ∞)	-
$90 <$ LTV ≤ 95 ($\hat{T}$)	0.90^{**}	-	0.60	-
95% CI	[0.11, ∞)	-	[0, ∞)	-
LTV > 95 ($\hat{T}$)	1.27^{**}	0.17	0.63^{*}	1.26^{**}
95% CI	[0.19, ∞)	[0, ∞)	[0, ∞)	[0.22, ∞)
Panel B: Difference in Menus ($D\hat{I}M$) Metric				
LTV ≤ 75 ($\hat{T}$)	5.98^{***}	-	5.41^{***}	-
95% CI	[2.92, ∞)	-	[3.72, ∞)	-
$75 <$ LTV ≤ 80 ($\underline{D\hat{I}M}_{1\to 2}$)	6.19^{***}	-	5.18^{***}	-
95% CI	[4.60, ∞)	-	[4.23, ∞)	-
$80 <$ LTV ≤ 90 ($\underline{D\hat{I}M}_{1\to 2}$)	1.86^{**}	-	0.02	-
95% CI	[0.41, ∞)	-	[-1.00, ∞)	-
$90 <$ LTV ≤ 95 ($\underline{D\hat{I}M}_{1\to 2}$)	-0.35	-	0.12	-
95% CI	[-1.11, ∞)	-	[-0.45, ∞)	-
LTV > 95 ($\underline{D\hat{I}M}_{1\to 2}$)	-1.08	-2.39	-2.91	-1.19
95% CI	[-1.94, ∞)	[-3.21, ∞)	[-3.93, ∞)	[-1.85, ∞)
$N_{\text{LTV}\leq 75}$	840	-	3,142	-
$N_{75<\text{LTV}\leq 80}$	3,002	-	8,492	-
$N_{80<\text{LTV}\leq 90}$	1,714	-	3,988	-
$N_{90<\text{LTV}\leq 95}$	4,844	-	9,742	-
$N_{\text{LTV}>95}$	2,396	9,302	4,152	12,060

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

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 ($\underline{D\hat{I}M}$) in units of basis points based on our Definition 2, with $\underline{D\hat{I}M}$ in units of basis points. 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 $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2$ using our procedure in Section 4, and confidence intervals are computed through inversion of the hypothesis test.

Table 9: Analysis of differences in menus comparing borrowers across categories of FICO scores

	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 ($\hat{T}$)				
FICO $\geq$ 740 ($\hat{T}$)	2.92***	4.81	2.70***	0.61
95% CI	[2.13, ∞)	[0, ∞)	[1.67, ∞)	[0, ∞)
720 $\leq$ FICO $<$ 740 ($\hat{T}$)	2.94**	4.03	0.77	4.44*
95% CI	[0.04, ∞)	[0, ∞)	[0, ∞)	[0, ∞)
700 $\leq$ FICO $<$ 720 ($\hat{T}$)	6.21**	4.24	1.32	10.77***
95% CI	[0.81, ∞)	[0, ∞)	[0, ∞)	[5.37, ∞)
660 $\leq$ FICO $<$ 700 ($\hat{T}$)	1.49	1.57	1.96	6.15***
95% CI	[0, ∞)	[0, ∞)	[0, ∞)	[3.65, ∞)
FICO $<$ 660 ($\hat{T}$)	7.04	0.12	8.21*	0.03
95% CI	[0, ∞)	[0, ∞)	[0, ∞)	[0, ∞)
Panel B: Difference in Menus ($D\hat{I}M$) Metric				
FICO $\geq$ 740 ($\underline{D\hat{I}M}_{1\to2}$)	1.82***	-0.19	2.23***	-4.51
95% CI	[1.16, ∞)	[-3.49, ∞)	[1.88, ∞)	[-6.97, ∞)
720 $\leq$ FICO $<$ 740 ($\underline{D\hat{I}M}_{1\to2}$)	1.78*	-10.05	-0.45	-2.47
95% CI	[-0.06, ∞)	[-15.65, ∞)	[-1.74, ∞)	[-5.54, ∞)
700 $\leq$ FICO $<$ 720 ($\underline{D\hat{I}M}_{1\to2}$)	2.26*	-0.84	-1.96	3.81***
95% CI	[-0.12, ∞)	[-3.07, ∞)	[-3.59, ∞)	[1.24, ∞)
660 $\leq$ FICO $<$ 700 ($\underline{D\hat{I}M}_{1\to2}$)	-0.27	-1.69	0.49	0.51
95% CI	[-2.65, ∞)	[-3.01, ∞)	[-1.31, ∞)	[-0.57, ∞)
FICO $<$ 660 ($\underline{D\hat{I}M}_{1\to2}$)	-1.76	-2.86	-4.96	-2.59
95% CI	[-6.25, ∞)	[-3.89, ∞)	[-9.96, ∞)	[-3.55, ∞)
$N_{\text{FICO}\geq740}$	9,914	354	22,964	662
$N_{720\leq\text{FICO}<740}$	1,082	184	2,588	360
$N_{700\leq\text{FICO}<720}$	920	428	2,002	748
$N_{660\leq\text{FICO}<700}$	738	2,884	1,680	4,008
$N_{\text{FICO}<660}$	142	5,572	282	6,534

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

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 ($\underline{D\hat{I}M}$) in units of basis points based on our Definition 2, with $\underline{D\hat{I}M}$ in units of basis points. 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 $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2$ using our procedure in Section 4, and confidence intervals are computed through inversion of the hypothesis test.

6 Discussion

We identify a "menu problem" that confounds the estimation of mortgage pricing discrimination in the literature, which stems from the potential for unobserved preference heterogeneity across groups of borrowers. We also devise a new methodology for assessing differences in menus that is robust to such preference heterogeneity, and use it to produce new estimates of mortgage pricing discrimination. Empirically, we find that mortgage pricing differentials by race still exists, particularly among more creditworthy conforming borrowers.

Pinpointing the precise mechanisms driving the mortgage pricing differences, we find, would be a promising path for future research. We discuss how unobserved credit risk is an unlikely explanation for our findings, because default risk is insured and minority borrowers have an especially favorable prepayment profile (Kau, Fang, and Munneke (2019), Gerardi, Willen, and Zhang (2020)). We also rule out different levels of loan originator compensation as an explanation. A remaining possible explanation is the rejection margin: Perhaps minority borrowers who are offered lower prices are more likely to be ultimately rejected for loans. Because we condition on origination, selection on rejection could be a mechanism driving the differences in menus we detect among originated mortgages. Another plausible explanation is that the differences in menus emerges from the search and negotiation process where the more creditworthy non-Hispanic white borrowers may be more likely to be given discounts. It would be interesting to test whether such differences in discounts are true in a field-experiment setting.

At a higher level, our conceptual separation between menus and preferences is not without caveats. In many circumstances, external factors such as neighborhoods may influence both menus and preferences simultaneously (see, for example, Katz, Kling, and Liebman (2001), Chetty, Hendren, and Katz (2016), Chetty and Hendren (2018)), such that a distinction between menus and preferences may not always be sensible. Nevertheless, it is sometimes useful, particularly from a policy perspective, to be able to attribute differences in outcomes to either inequality in menus or heterogeneity in preferences. We do this for the mortgage market and find that, particularly for more creditworthy borrowers, lenders continue to offer minority borrowers worse menus in terms of rates and points compared with the menus they offer observationally similar white borrowers for conforming mortgages.

While the menu problem is broadly applicable, our specific methodology for addressing it is unlikely to be the final word. The main benefit of our approach is that it requires relatively few assumptions to be valid. However, if a researcher is willing to make more stringent assumptions, there may be other ways they can test for differences in menus with more statistical power. Another promising path for future research would involve the use of

other identifying assumptions for measuring differences in menus, or 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*.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83(6): 2411–2451. doi:10.3982/ECTA13408.
Available at `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13408`.

Agarwal, Sumit, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. 2020. "Searching for Approval." Working Paper 27341. National Bureau of Economic Research. doi:10.3386/w27341.
Available at `http://www.nber.org/papers/w27341`.

Allen, Jason, Robert Clark, and Jean-François Houde. 2014. "The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry." *American Economic Review* 104(10): 3365–96. doi:10.1257/aer.104.10.3365.
Available at `https://www.aeaweb.org/articles?id=10.1257/aer.104.10.3365`.

Allen, Jason, Robert Clark, and Jean-François Houde. 2019. "Search Frictions and Market Power in Negotiated-Price Markets." *Journal of Political Economy* 127(4): 1550–1598. doi:10.1086/701684.
Available at `https://doi.org/10.1086/701684`.

Ambrose, Brent W, James Conklin, and Luis A Lopez. 2020. "Does Borrower and Broker Race Affect the Cost of Mortgage Credit?" *Review of Financial Studies (Forthcoming)*.

Andrews, Donald W. K., and Gustavo Soares. 2010. "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection." *Econometrica* 78(1): 119–157. doi:10.3982/ECTA7502.
Available at `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7502`.

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.

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.

Hong, Han, and Jessie Li. 2020. "The numerical bootstrap." *Ann. Statist.* 48(1): 397–412. doi:10.1214/19-AOS1812.
Available at https://doi.org/10.1214/19-AOS1812.

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

Hwang, Hae-shin, W Robert Reed, and Carlton Hubbard. 1992. "Compensating wage differentials and unobserved productivity." *Journal of Political Economy* 100(4): 835–858.

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.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70(2): 317–341. ISSN 0034-6527. doi:10.1111/1467-937X.00246.
Available at https://doi.org/10.1111/1467-937X.00246.

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.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6): 1263–1297.

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 Suboptimal 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 distributions 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 $g(\mathbf{m}) = g_1(\mathbf{m}) = g_2(\mathbf{m})$. □

A.1.2 Proof of Theorem 2

Proof. Under constant marginal utility in interest rate, Equations (11) and (12) imply that $u_i(\mathbf{m}_j) - u_i(\mathbf{m}_i) = \beta d_{i\to j}(\mathbf{m}_i, \mathbf{m}_j) \geq \beta \underline{d}_{1\to 2}(x_1, x_2)$ for some constant β. Then, by property of minimum we must have $\beta \underline{DIM}_{1\to 2} \leq \Delta W_{\mathcal{I}_1\to\mathcal{I}_2,\boldsymbol{\pi}}, \forall \pi$, with $\beta \equiv 1$ if utility were measured in terms of interest rate. □

A.1.3 Proof of Theorem 3

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

$$\varphi'_{\mathbf{p}_1,\mathbf{p}_2}(\mathbf{h}_1, \mathbf{h}_2) = \lim_{t\to 0^+} \frac{\varphi(\mathbf{p}_1 + th_1, \mathbf{p}_2 + th_2) - \varphi(\mathbf{p}_1, \mathbf{p}_2)}{t} \tag{35}$$

exists for all and is equal to that given by Equation (20) 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:

$$\varphi^* = \min_{\boldsymbol{\pi}} E_{\boldsymbol{\pi}}\phi^* \text{ s.t. } E_{x_2}\boldsymbol{\pi} \geq \mathbf{p}_1, E_{x_1}\boldsymbol{\pi} \geq \mathbf{p}_2, \boldsymbol{\pi} \geq 0 \tag{36}$$

and then use Theorem 3.1 of Gal and Greenberg (2012), who set 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, if we let $\Pi^*(p_1, p_2)$ be the set of primal solutions to the linear programming problem in Equation (36), then the

set of primal solutions reachable from perturbations in the direction $\{h_1, h_2\}$ is non-empty, such that:

$$\Pi^{\infty}(\{\mathbf{p}_1, \mathbf{p}_2\}, \{\mathbf{h}_1, \mathbf{h}_2\}) = \{\boldsymbol{\pi} : \{\boldsymbol{\pi}^k\} \to \boldsymbol{\pi} \text{ for some } \{\epsilon_k \to 0^+\}, \tag{37}$$

$$\text{with } \boldsymbol{\pi}^k \in \boldsymbol{\Pi}^*(\mathbf{p}_1 + \epsilon_k \mathbf{h}_1, \mathbf{p}_2 + \epsilon_k \mathbf{h}_2)\} \neq \emptyset \tag{38}$$

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 (37) 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, such 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:

$$|\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2})| \leq M l_1(\{\mathbf{p}_{1,1}, \mathbf{p}_{1,2}\}, \{\mathbf{p}_{2,1}, \mathbf{p}_{2,2}\}) \tag{39}$$

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:

$$\varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^+, \mathbf{p}_2^+) \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2}) \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_1^-, \mathbf{p}_2^-) \tag{40}$$

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 (40) yields:

$$-M l_1 \leq \varphi(\mathbf{p}_{1,1}, \mathbf{p}_{1,2}) - \varphi(\mathbf{p}_{2,1}, \mathbf{p}_{2,2}) \leq M l_1 \tag{41}$$

which implies Equation (39) and that the mapping φ is Lipschitz.

□

A.1.4 Proof of Corollary 1

Proof. Suppose we knew the true $\mathbf{p}_1, \mathbf{p}_2$. Then, the "oracle" analogue of the critical value at level $1-\alpha+\beta$ conditional on the true $\mathbf{p}_1, \mathbf{p}_2$ is:

$$\hat{c}^o_{1-\alpha+\beta} = \inf\{c \in \mathbb{R} : \Pr(\hat{\psi}'_{\mathbf{p}_1,\mathbf{p}_2}(\mathbf{h}_1, \mathbf{h}_2) \leq c) \geq 1-\alpha+\beta\} \tag{42}$$

By construction, $\hat{c}_{1-\alpha+\beta} \geq \hat{c}^o_{1-\alpha+\beta}$ whenever $[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_\beta$. Therefore, if we let the event $[\mathbf{p}_1, \mathbf{p}_2] \in \hat{\mathcal{P}}_{n,\beta}$ be E and $[\mathbf{p}_1, \mathbf{p}_2] \notin \hat{\mathcal{P}}_{n,\beta}$ be $\neg E$, it follows that:

$$\begin{aligned}
\Pr(\hat{\psi} \geq \hat{c}_{1-\alpha+\beta}) &= \Pr(\hat{\psi} \geq \hat{c}_{1-\alpha+\beta}|E)\Pr(E) + \Pr(\hat{\psi} \geq \hat{c}_{1-\alpha+\beta}|\neg E)\Pr(\neg E) && (43)\\
&\leq \Pr(\hat{\psi} \geq \hat{c}^o_{1-\alpha+\beta})\Pr(E) + \Pr(\hat{\psi} \geq \hat{c}_{1-\alpha+\beta}|\neg E)\Pr(\neg E) && (44)\\
&\leq \Pr(\hat{\psi} \geq \hat{c}^o_{1-\alpha+\beta}) + \Pr(\neg E) && (45)
\end{aligned}$$

By property of limsup, we know that:

$$\limsup_{n\to\infty} \Pr_{\mathbf{p}_1,\mathbf{p}_2}(\hat{\varphi}_n \geq \hat{c}_{n,1-\alpha+\beta}) \leq \limsup_{n\to\infty} \Pr(\hat{\psi}_n \geq \hat{c}^o_{n,1-\alpha+\beta}) + \limsup_{n\to\infty} \Pr([\mathbf{p}_1, \mathbf{p}_2] \notin \hat{\mathcal{P}}_{n,\beta}) \tag{46}$$

By Equation (22) and the Portmanteau Theorem, we have:

$$\limsup_{n\to\infty} \Pr_{\mathbf{p}_1,\mathbf{p}_2}((\hat{\varphi}_n \geq \hat{c}^o_{n,1-\alpha+\beta})) \leq \alpha - \beta \tag{47}$$

By Equation (23) for the uniform confidence band, we know that:

$$\limsup_{n\to\infty} \Pr_{\mathbf{p}_1,\mathbf{p}_2}([\mathbf{p}_1, \mathbf{p}_2] \notin \hat{\mathcal{P}}_{n,\beta}) \leq \beta \tag{48}$$

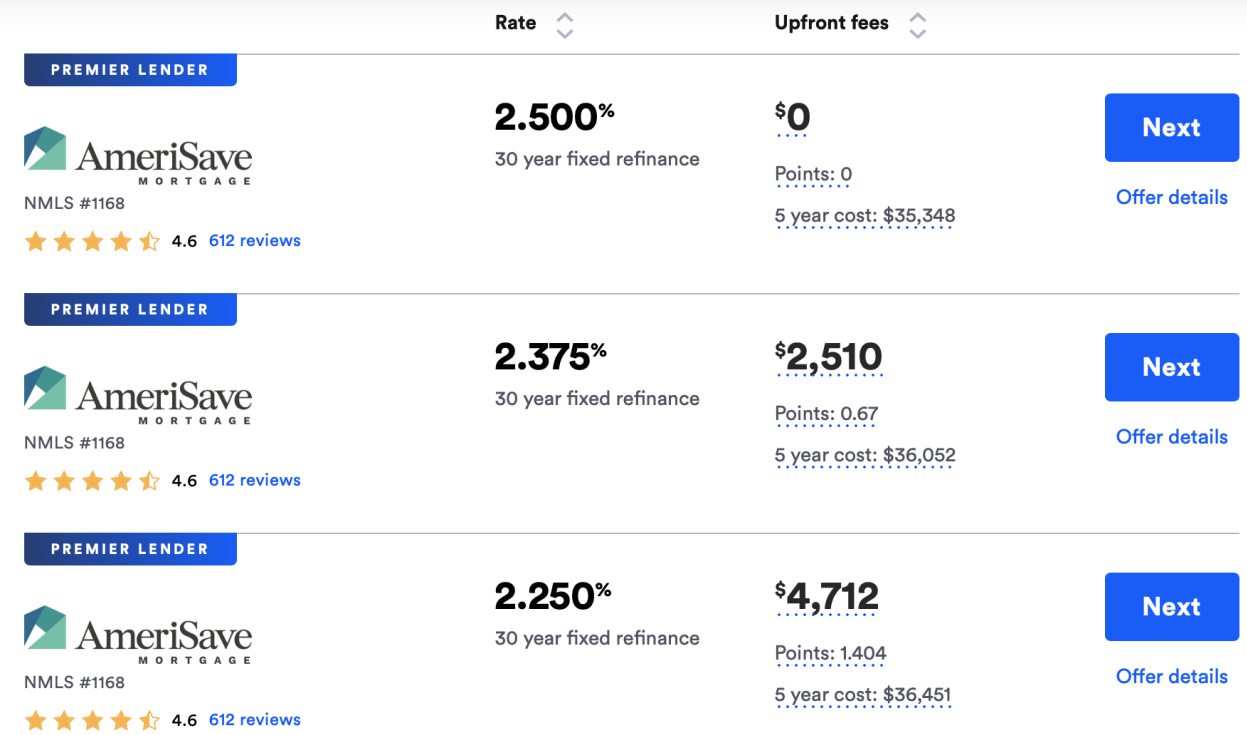
Combining the two parts, we have:

$$\limsup_{n\to\infty} \Pr_{\mathbf{p}_1,\mathbf{p}_2}(\hat{\varphi}_n \geq \hat{c}_{n,1-\alpha+\beta}) \leq \beta + (\alpha - \beta) = \alpha \tag{49}$$

as required. □

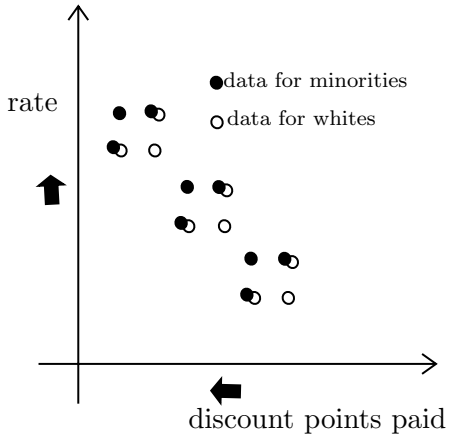
A.2 Additional Tables and Figures

Figure A.1: Screenshot of advertised rate and closing cost table



Note: This figure shows a screenshot obtained by the authors from Bankrate.com for a $300,000 refinancing mortgage with baseline characteristics on December 19, 2020. It shows how a borrower may choose to pay 0 points for a 2.500% interest rate mortgage, 0.67 points for a 2.375% interest rate mortgage, or 1.404 points for a 2.250% interest rate mortgage. Also, lender review websites such as NerdWallet.com and MortgageWaldo.com evaluate lenders based on both interest rates and discount points/closing costs.

Figure A.2: 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 it complements the linear regression case of Figure 2.

Table A.1: Other approaches to inference in optimal transport, compared with 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 the 5 [percent 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 was 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 percentage of the loan amount on HMDA's information on points (hmda_points) versus Optimal Blue'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-Optimal Blue 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 percentage of the loan amount below –3 percent or above 10 percent. 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 presents 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.