## Revising Statistical Evidence Required for Establishing Disparate Impact Claims

Equal opportunity in employment and housing is a core right in the United States. This right is enforced through anti-discrimination cases brought under Title VI/VII of the Civil Rights Act and the Fair Housing Act. Recently, there has been a surge in Fair Housing and Equal Opportunity complaints while concurrently, less than 1% of employment cases brought to courts have established discrimination.[[1]](#footnote-1)Increasingly, these cases stem from ostensibly neutral actions that disproportionately affect protected classes. For example, landlords use former-incarceration status to screen out tenants, which may cause racial disparities in housing access.[[2]](#footnote-2) Additionally, some justice programs use algorithmic decision-making, potentially leading to discriminatory harms by perpetuating existing biases found in historical data.[[3]](#footnote-3)

These cases may be challenged through disparate impact claims, which are resolved by courts using a three-part test.[[4]](#footnote-4) The first step requires demonstrating a disparity using statistical evidence. This requires showing a protected class is adversely affected at a four-fifths rate relative to a comparator population (the four-fifths rule) and showing the effect is statistically significant (the significance rule). [[5]](#footnote-5) The intent of these rules is to standardize the identification of disparities by providing clear guidelines.

While initially these rules appear to apply neutrally to cases, meeting the bar set by the four-fifths rule and significance rule can vary substantially based on the statistical choices made by an analyst and the properties of data such as sample size and available covariates. For instance, finding a 20% effect size depends on the analyst’s chosen model and controls, while statistical significance depends on statistical power, sample size, and the size of an effect.[[6]](#footnote-6) These factors can make determining disparities highly variable under seemingly static rules and place an undue burden on plaintiffs. To better align with the intent of disparate impact cases, courts may need to adjust rules to account for additional measurement factors.

This is particularly important to investigate since more policies are now based on algorithmic data mining techniques, which present unique challenges for disparate impact cases.[[7]](#footnote-7) For example, in 2018, HUD filed a complaint against Facebook, arguing the company’s advertisement system permitted advertisers to restrict the availability of housing-related ads presented to specific racial groups.[[8]](#footnote-8) Facebook responded to this complaint by stating their algorithm did not include an variable for “race”, and as such, had no ability to target users based on their race. Nonetheless, this algorithm led to disparities in the ads served to users due to correlations between the factors included in the algorithm and race.

Detecting a disparate impact in this instance will depend critically on the analyst’s choice of a model and the model’s covariates. The typical model used by courts to demonstrate a disparity uses an adjusted regression, interpreting the effect size and significance of a beta coefficient on a protected class covariate, including all other available covariates as controls. This has been described as a “kitchen sink” approach and the inclusion of additional covariates can lead to a substantial underestimate of the true disparity due to included-variable bias.[[9]](#footnote-9) This bias occurs since many of the variables used as controls are causally related to race and their inclusion absorbs the effect that likely generated the disparate impact. Subsequently, in many cases like those of *HUD v. Facebook,* conventional models used within disparate impact cases will fail to meet the four-fifths rule.

Researchers have proposed several alternative models to identify disparate impacts which address the included-variable bias in the kitchen sink approach. For instance, Jung et al., 2023 proposes a risk-adjusted regression, where an analyst uses the available data to first determine an individual’s risk for a given outcome the decision maker is trying to prevent (for example, what is the risk of an individual defaulting on their loan?).[[10]](#footnote-10) Then, the analyst regresses the decision action (was a loan extended to an individual or not?) on the estimated risk scores and a protected class-covariate. Then, the analyst interprets the estimated coefficient on the protected class covariate to determine if there was a disparate impact. Several alternative procedures have also been proposed, such as using error-rates and hypothetical risk thresholds.[[11]](#footnote-11) However, these procedures all require additional information and larger samples to estimate than standard approach used in court rooms.

In this study, we propose to compare how well these different approaches can capture disparities given the properties of available data. To investigate this, we will construct synthetic datasets which match those used in case law, allowing us to examine model performance given known truths underlying of both the data generating process and true disparity ratios. We will then vary properties of the synthetic datasets – such as sample size and underlying risks by groups – and construct rules used by decision makers to determine selection rates. We will then compare the relative bias and precision of different procedures for estimating disparate impact based on selection rates using monte carlo simulations.

We provide a small example to illustrate how this method will work in practice.[[12]](#footnote-12) We start by specifying a data generating process which aligns with observed real-world data. We simulate a baseline population of N individuals which contains simulated variables for the individual’s race, monthly earnings, and former incarceration status. We simulate race so that it follows the population proportion of Black individuals in the United States, former-incarceration status so that the variable is correlated with race, with underlying probabilities derived from DOJ-OJP data and previous academic research. We then simulate monthly earnings so that earnings are correlated with both former-incarceration status and race, following the empirical distribution of earnings from the US Census.

We then use this synthetic dataset to compare the performance of the kitchen-sink regression in recovering a disparity ratio across three decision rules. The broader scenario we are considering is a landlord deciding to rent a unit to an applicant. In all three scenarios, we assume a landlord is more likely to rent an apartment to a tenant when they have a higher monthly income and have not been incarcerated. In the first scenario, we modify this probability so that the landlord is 25% more likely to reject a black applicant (direct discrimination scenario). In the second scenario, the landlord is 75% more likely to reject every third black applicant (higher variance scenario). In the last scenario, we increase the probability the landlord will reject an applicant, if they were formerly incarcerated (disparate impact scenario). Critically, we set probabilities across all scenarios such that black applicants are selected only 60% as often as white applicants.

We then simulate synthetic datasets and decision rules across 500 iterations, varying the population size between 200 and 4000 people. For each simulate, we estimate a logit regression regressing the decision outcome on race, monthly earnings, and former-incarceration status. We then extract the estimated beta coefficient on the race variable, along with the coefficient’s p-value. We transform the beta coefficient into a disparity ratio: what percentage of black to white applicants re estimated to have been chosen based on their race?

We plot the estimated disparity ratio averaged across iterations for the three scenarios in figure 1. This figure shows that across decision rules the kitchen sink regression model underestimates the true disparity ratio. More concerningly, in the scenario where the decisionmaker uses the “disparate impact” rule, comparable to the Facebook case discussed above, the kitchen sink regression does not estimate a disparity at all, due to included variable bias. For this scenario, an alternative procedure would be needed, or the analyst would have to not include both income and former-incarceration status as covariates in the model.

Figure 1: Estimated disparity ratio by scenario and sample size.

Chart, line chart

Description automatically generated

We show the average simulated p-value for the disparity ratio by scenario and sample size in figure 2. This figure shows that the higher variance scenario requires a 20% larger sample to detect an effect than the direct discrimination scenario. Alternative disparate impact models will also require a larger sample to estimate a disparity ratio to a given level of precision since these approaches require more information and incorporate additional modeling steps which further propagate uncertainty. Determining to what extent these alternative procedures require additional sample to find known effects at a given level of precision will be a central aim of this project.

Figure 2: P-value of estimated disparity ratio, by scenario and sample size.

Chart, line chart

Description automatically generated

To proceed, this project will develop simulations like this example, emulating datasets and decision rules found in disparate impact case law. We will additionally develop code to implement the alternative disparate impact models discussed above. We will then use these simulations to assess the bias and precision of these models under a range of data conditions. Ultimately, these results will demonstrate to what extent the current approach in statistically assessing disparate impact is biased and how the four-fifths rule, or statistical significance rule may need to modified to permit less unbiased methods without increasing burden on the plaintiff.

### Impact Potential and Products

We anticipate the results from this project will be used to inform court decision making on which statistical tests should be used within disparate impact cases. We will use the results from this project to develop an academic report which will be submitted to a peer-reviewed journal, along with creating an interactive visualization to explore the results from this project. Our hope is the interactive visualization may provide an intuitive way to understand the results from this project, demonstrating the sensitivity of statistical models estimating disparate impacts to changes in data properties and decision rules.

### Timelines, Budget, and Personnel

We anticipate this project will take 12 months to complete, starting from April 2024 and cost $150,000 (budget estimate). The project will be co-PI’ed by Max Griswold (60 days) and Osonde Osoba (15 days). We will also seek a legal researcher (5 days) to provide expertise on the legal questions involved in this project and to assist in drafting the final manuscript detailing the results from this project. The budget also contains a line-item for distributed computing resources ($5000) and coverage for QA review of the project’s code and the interactive visualization ($10,000).

1. [*State of Fair Housing – Annual Report to Congress,* 2022](https://www.hud.gov/sites/dfiles/FHEO/documents/FHEO%20Annual%20Report%20FY%202022.pdf); [Jameel, Shapiro, and Yerardi, 2019](https://www.washingtonpost.com/graphics/2019/business/discrimination-complaint-outcomes/) [↑](#footnote-ref-1)
2. [Griswold et al., 2023](https://www.rand.org/pubs/research_reports/RRA2689-1.html) [↑](#footnote-ref-2)
3. [Osoba & Welser, 2017](https://www.rand.org/pubs/research_reports/RR1744.html). [↑](#footnote-ref-3)
4. [Title VI Legal Manual, Section VII, Sec. C](https://www.justice.gov/crt/fcs/T6Manual7#C) [↑](#footnote-ref-4)
5. [Title VI Legal Manual, Section VII, 1 (C) ii - iv](https://www.justice.gov/crt/fcs/T6Manual7#G) [↑](#footnote-ref-5)
6. [Cohen, 2013](https://www.jstor.org/stable/pdf/24758720.pdf?refreqid=fastly-default%3A24937b1ae04c24243be4456ee301795d&ab_segments=&origin=&initiator=&acceptTC=1); [Kim & Choi, 2019](https://onlinelibrary.wiley.com/doi/full/10.1111/abac.12172); [Spanos, 2017](https://onlinelibrary.wiley.com/doi/full/10.1111/joes.12200); [Grossman, Nyarko, & Goel, 2023](https://5harad.com/papers/disparate-impact.pdf) [↑](#footnote-ref-6)
7. [Barocas & Selbst, 2016](https://www.jstor.org/stable/pdf/24758720.pdf?refreqid=fastly-default%3A24937b1ae04c24243be4456ee301795d&ab_segments=&origin=&initiator=&acceptTC=1) [↑](#footnote-ref-7)
8. [*HUD v. Facebook,* 2018](https://www.hud.gov/sites/dfiles/PIH/documents/HUD_01-18-0323_Complaint.pdf) [↑](#footnote-ref-8)
9. [Jung et al., 2018](https://5harad.com/papers/included-variable-bias.pdf); [Ayres, 2010](https://ianayres.yale.edu/sites/default/files/files/Testing%20for%20Discrimination.pdf). [↑](#footnote-ref-9)
10. [Jung et al., 2018](https://5harad.com/papers/included-variable-bias.pdf) [↑](#footnote-ref-10)
11. [Grossman, Nyarko, and Goel, 2024](https://5harad.com/papers/disparate-impact.pdf); [Arnold, Dobbie, and Hull, 2022](https://www.aeaweb.org/articles?id=10.1257/aer.20201653); [Elzayn et al., 2023.](https://drive.google.com/file/d/1kA7CG3cLq6eWmwBVgTDOIMhxuGZwRJ5O/view) [↑](#footnote-ref-11)
12. More specific details on this approach can be found in a [coding notebook](https://htmlpreview.github.io/?https://github.com/maxgriswold/disparate_impact/blob/main/code/proposal/disparate_impact_synthetic_scenario_example.html) which accompanies this proposal. [↑](#footnote-ref-12)