“Statistical Approaches for Assessing Disparate Impact in Fair Housing Cases” Methods and Normative Considerations

Mackie Jackson

25 October, 2024

Hurricane Helene decimated thousands of homes in Western North Carolina. As resources begin to be distributed to individuals in order to start rebuilding, we might consider a housing discrimination case that happened in the wake of Hurricane Katrina. The 2010 civil suit *Greater New Orleans Fair Housing Action Center v. United States Department of Housing & Urban Development* exemplifies our in-class discussion of unfairness as inequality. The plaintiffs alleged that HUD gave homeowners in predominantly Black neighborhoods less money to rebuild than individuals in predominantly white neighborhoods (Greater New Orleans Fair Housing Action Center v. United States Department of Housing & Urban Development, 2010). They successfully settled the suit in 2011 and were compensated by the state for withheld funding, but ultimately failed to convince judges of disparate impact. This case should lead us to contemplate why accurate, persuasive methods of assessing disparate impact are of moral importance as the climate crisis wreaks increasing havoc on disadvantaged communities.

In their article “Statistical Approaches for Assessing Disparate Impact in Fair Housing Cases”, authors Aigner, del Ángel, and Wiles introduce a plethora of legal jargon and definitions necessary to understand statistical inference techniques related to disparate impact, and I feel it is necessary to summarize a bit of their Fair Housing Act exposition. The Fair Housing Act (FHA) of 1968 prohibits discrimination against individuals (including the following actions but not limited to) renting, buying, applying for a mortgage, or seeking housing assistance. Protected classes defined by the legislation include race, color, national origin, religion, sex, familial status, and disability. In order to pursue legal action under the FHA, plaintiffs must prove that a defendant’s ongoing policy causes a sufficiently sizable disparate impact on the plaintiff’s protected class. Following the establishment of disparate impact, plaintiffs must also prove discriminatory causation, which is to isolate the given policy from other reasons that might have resulted in perceived discrimination. While a defendant may be able to provide legitimate causation for an offending policy, a plaintiff can still prove successful in litigation by proving that a policy that results in decreased disparate impact would have sufficed for the defendant housing entity.

Following an FHA Supreme Court case in 2015, a set of four guidelines were established to determine whether a policy meets the standard for disparate impact. The plaintiff's statistical evidence must focus on the protected class group affected by the alleged discriminatory policy, prove via comparison that other groups were "less harmed" by the policy, show relative (not just absolute) impact, and show impact of a significant size. The disparity ratio, which is the typical method of statistically proving disparate impact, compares the effects of a given policy on a protected class ($P1$) to a non-protected class ($P2$) and is given by the simple equation $R = P1/P2$. When addressing housing discrimination, a measure called the adverse impact ratio is often utilized. This is the proportion of the percentage of harmful outcomes, or rejections, for the protected class ($Q1$) compared to the non-protected class ($Q2$) and is given by $R^* = Q1/Q2$. Disparate impact is established if the adverse impact ratio is greater than 1.25.

Aigner, del Ángel, and Wiles seek to improve upon methods for statistically proving disparate impact by testing and comparing three inference techniques. Our disparity and adverse impact ratio definitions are given above, but statistical inference techniques must be used to prove that a dataset meets their criteria. Aigner et al. first examine law literature and introduce the non-overlapping confidence intervals test, in which the lower bound of the confidence interval for P1 is compared to the upper bound of the confidence interval for P2. It tests whether the difference between numerator and denominator in $R = P1/P2$ is greater than or equal to zero, thus yielding a disparity ratio of greater than or equal to one. The authors then describe the more popular direct difference test, which they posit as similar but algebraically varying. The direct difference test is preferred among law practitioners because the lower bound of a confidence interval for $(P1 - P2)$ will always exceed the lower bound of the differences between $P1$ and $P2$ respectively, and is consistent with normal binomial distribution.

Aigner and his colleagues next define the direct ratio estimator method, in which the value of R given by $(P1/P2)$ and its corresponding confidence interval is used to determine disparate impact. The third method defined is the log-ratio method, which is essentially the same but the natural log is taken of R. They explain that while the direct ratio and log-ratio return similar results, the lower confidence interval bound of the log-ratio will always be greater than the direct ratio and that there is no work concretely determining the efficacy of one over the other. We can note that any lower bound value of $R > 1$ resulting from a ratio test suggests disparate impact. Ultimately the direct difference, direct ratio, and log-ratio methods for determining disparate impact are assessed and compared using Section 8 housing voucher eligibility data based on income and classified by race.

Federal law prohibits discrimination against Section 8 voucher holders in government housing, but this does not extend to private landlords. Aigner and his colleagues replicate past experiments and collect data on the number of Black and white households eligible for Section 8 vouchers based on income. They display results from the District of Columbia with a sample size of $n = 1263$ from Black households and $n = 1345$ from white households, returning consistent proof of disparate impact using lower confidence interval bounds from all three aforementioned methods. The disparity ratio for all five income brackets below 50,000 dollars annually (the cutoff for Section 8 vouchers in D.C. as of 2014) considerably exceeds 1.25. The direct difference test lower bounds are consistently positive, suggesting that we can reject the null hypothesis that $(P1 - P2) \leq 0$, and all ratio test results are greater than 1.

Aiger and his colleagues conclude that the preferred method for proving disparate impact remains the direct difference test due to its amount of legal precedent and its normal distribution adherence. However, there are some cases in which a ratio of disparity will be of greater or equal interest than an absolute difference. Such an example would be if the disparity ratio is large but the results of the difference test are small, or vice versa. From the results of this test, the authors recommend use of the log-ratio method when a population sample n is less than or equal to 50 due to inconsistent results yielded from the direct ratio method. From a legal perspective that emphasizes conservative methods of analysis to shield from defense critiques, the authors recommend the direct difference or ratio method for large sample sizes. For the purposes of replicating this work with a novel dataset I will not discuss the authors' following experiment using more complex incremental disparity methods.

Why is statistical evidence generally necessary? There is legal precedent for the use of a 4/5ths (or 0.8 disparity ratio) difference in effects on a protected class versus a non-protected class without the use of further statistical evidence like a confidence interval, but this allows a defendant to argue that the observed proportions are a mere product of chance. We have introduced the increasing need for concrete statistical methods of proving disparate impact in the face of climate change-induced housing crises, but why else might they be of importance given present statistics on housing discrimination? As recently as 2016, the probability of a Black individual receiving a response from an initial housing inquiry is eight percent lower than that of a similarly qualified white individual. Black borrowers also pay approximately 5-11 percent more on monthly home loans than white borrowers (Lincoln, Lee, & Brandon, 2020).

In political philosopher John Rawls' 1985 essay "Justice as Fairness: Political not Metaphysical", he describes a conception of justice based on both liberty and equality. He writes of liberty as all individuals having certain equal, inviolable rights, and one such right is arguably access to housing. Rawls also argues for equality of opportunities; homeownership, tenancy, and the ability to access resources after a natural disaster might be considered among these. A Rawlsian view of justice is that which asks individuals to blindly choose the structure of the society they live in, not knowing whether they will be on top or bottom of the social hierarchy (Rawls, 1985). This belies the moral importance of determining an accurate measure of disparity a housing policy might inflict upon a disadvantaged group of people. When we can quantify them, we will be better able to remedy violations of individual liberties and basic rights in a court of law.

References

Aigner, Dennis J., Marco del Ángel, and Joel Wiles. 2024. "Statistical Approaches for Assessing Disparate Impact in | Fair Housing Cases." Statistics and Public Policy 11 (1): 2263038. https://doi.org/10.1080/2330443X.2023.2263038.

Greater New Orleans Fair Housing Action Center v. United States Department of Housing & Urban Development. 2010. United States District Court, D. Columbia.

Lincoln, Quillian, John J. Lee, and Honoré Brandon. 2020. "Racial Discrimination in the U.S. Housing and Mortgage Lending Markets: A Quantitative Review of Trends, 1976–2016." Race and Social Problems 12 (1): 13–28. https://doi.org/10.1007/s12552-019-09276-x.

Rawls, John. 1985. "Justice as Fairness: Political Not Metaphysical." Philosophy & Public Affairs 14 (3): 223–51.