

Chapter 4

Disparity Measures for Fair Lending Analyses

Jason Dietrich
March 2026

© 2026 by Jason Dietrich. All rights reserved.

The views and opinions expressed in this paper belong solely to me and do not represent the views or opinions of any employer, institution, or organization with which I have been affiliated.

This paper is for informational and educational purposes only and is not intended to serve as professional or legal advice. I specifically disclaim all responsibility for any liability, loss, or risk, personal or otherwise, which is incurred as a direct or indirect consequence of the use or application of the contents of this paper. Every effort has been made to ensure that the information in this paper is correct. However, I assume no responsibility for errors, inaccuracies, or omissions. The use of this paper implies the reader's acceptance of this disclaimer.

I. Introduction

The Interagency Fair Lending Examination Procedures (IFLET) define disparate treatment as, “... when a lender treats a credit applicant differently based on one of the prohibited bases,” and disparate impact as, “When a lender applies a racially or otherwise neutral policy or practice equally to all credit applicants, but the policy or practice disproportionately excludes or burdens certain persons on a prohibited basis”¹ The key to both of these definitions is that treatment and/or outcomes differ across groups. During fair lending reviews, disparity estimates from statistical analyses provide one important set of evidence of potentially different treatment or outcomes across groups.² One difficulty here is that there are many different disparity measures to choose from, and the IFLET and other regulatory guidance are silent on which specific disparity measure Economists should use to test for different treatment or outcomes. Therefore, the appropriate disparity measure for any given statistical fair lending analysis is an open question.

The three disparity measures Economists most commonly use for fair lending statistical analyses are the odds ratio, ratio, and difference (or marginal effect). Each measure has unique properties, characteristics, and interpretations, and most importantly, the qualitative conclusions of the statistical analysis can vary across measures. As one example, suppose the denial rates for minorities and non-minorities are 2 percent and 1 percent, respectively. Using the ratio as the disparity measure, the denial rate for minorities is twice as high as for non-minorities, which would likely raise concerns, all-else-equal. Alternatively, using the difference as the disparity measure, the denial rate for minorities is 1 percentage point (pp) higher than for non-minorities, which would likely not raise concerns, all-else-equal. This simple example highlights the

¹ See pages iii and iv of [2021-june-17-fairlend.pdf](#)

² File reviews and overt evidence are two other common pieces of evidence of different treatment or outcomes.

importance of understanding which specific disparity measure Economists used for the statistical analysis and then interpreting this disparity measure appropriately when drawing conclusions from the statistical results.

In this report we compare and contrast the three primary disparity measures (odds ratio, ratio, and difference) Economists use during fair lending statistical analyses to measure unconditional and conditional disparities. Overall, there are four key takeaways:

- There is not one disparity measure that is clearly appropriate for every statistical analysis. Instead, there are several reasonable measures available, each with strengths and weaknesses that need to be considered when choosing which specific measure to use for each given analysis.
- When presenting and discussing results from a statistical analysis, it is important to understand the details of the specific disparity measure used, and what specific conclusions can be drawn given the specific disparity measure.
- When presenting and discussing results from a statistical analysis, it is important to also include and consider the underlying average outcomes used to construct the disparities.
- For a given set of data, different disparity measures can lead to different conclusions about whether discrimination has occurred. When this occurs, relevant policy, analytical, and legal implications need to be considered when drawing overall conclusions and deciding on what actions to take.

The remainder of the report is structured as follows. Section II summarizes the characteristics of three disparity measures for constructing unconditional disparities. Section III explores the relationships between the three disparity measures for unconditional disparities. Section IV presents guidance on how to generate conditional disparities using each measure. Section V concludes the discussion.

II. Measures for Estimating Unconditional Disparities

There is not one correct disparity measure that Economists should use for every statistical fair lending analysis. Instead, there are a variety of acceptable disparity measures available, each with unique characteristics that need to be considered when deciding which specific measure to use, and when interpreting the statistical results. In this section, we provide details about three specific measures – odds ratio, ratio, and difference – for generating unconditional disparities.

Odds Ratio

As the name implies, the odds ratio is a ratio of two odds, where odds are the rate at which an event occurred relative to the rate at which the event did not occur. Specifically, the odds that an event occurred for individuals in group 1 is the percentage of individuals for which the event occurred divided by the percentage of individuals for which the event did not occur, or

$$Odds_1 = \frac{Percentage(event)_1}{100 - Percentage(event)_1} \quad (1)$$

The odds that the event occurred for individuals in group 2 is constructed similarly. The odds ratio is then just the ratio of the two odds, or

$$Odds Ratio = \frac{Odds_1}{Odds_2} \quad (2)$$

The odds ratio ranges from 0 to infinity. An odds ratio equal to 1 suggests the event was equally likely to occur for both groups, odds ratios less than 1 suggest the event was less likely to occur for group 1, and odds ratios greater than 1 suggest the event was more likely to occur for group 1. The conclusion that can be drawn with an odds ratio is that the odds that the event occurred for individuals in group 1 were X times higher than the odds that the event occurred for individuals in group 2.

Using an underwriting example, one possible definition of an event is that an application was denied. Suppose the denial rate for applicants in group 1 is 20%, so the odds of an applicant in group 1 being denied is $0.25 (= 20 / (100 - 20))$. Suppose further that the denial rate for applicants in group 2 is 10%, so the odds of an applicant in group 2 being denied is $0.11 (= 10 / (100 - 10))$. The odds ratio is then just the ratio of the two odds, or $2.25 (= .25 / .11)$. The conclusion that can be drawn in this example is that the odds of an applicant in group 1 being denied were 2.25 times higher than the odds of an applicant in group 2 being denied.

For formal statistical tests, the typical null hypothesis is that the true odds ratio equals 1, with an alternative hypothesis that the true odds ratio does not equal 1. There are several statistical approaches available to test this null hypothesis. We recommend a parametric test based on a logistic regression, since it is a standard approach, is easy to apply, and has a direct parallel to the test for conditional disparities discussed in Section IV below.³ Using an underwriting example to illustrate this approach, suppose we want to test the null hypothesis that the true unconditional denial rate disparity (constructed as an odds ratio) is equal to 1. The first step is to estimate a logistic regression with just a constant and a minority flag. Exponentiating the coefficient estimate on the minority flag from this regression provides the unconditional denial rate disparity estimate as an odds ratio, equivalent to using equation (2). The standard error of this odds ratio estimate is constructed as the odds ratio estimate times the standard error of the coefficient estimate.⁴ With the odds ratio estimate and its corresponding standard error estimate, we can then generate a standard z-statistic to test the null hypothesis that the true odds

³ One common non-parametric test is the chi-square goodness-of-fit test, which is standard output for cross-tabulations of two categorical variables for most statistical software packages.

⁴ The standard error estimate of the coefficient estimate is part of the standard output of a logistic regression for most statistical software packages. Two common approaches for estimating standard errors for the coefficient estimates in a logistic regression are the delta method (see Green (2012)), and expected and observed information matrices (See Van der Vaart (2000)).

ratio equals 1. An alternative version of this approach is to use the coefficient estimate on the minority flag directly from the logistic regression. Since $odds\ ratio = e^{(coefficient\ estimate)}$ and $e^{(0)} = 1$, the null hypothesis that the true odds ratio equals 1 is equivalent to the null hypothesis that the true parameter on the minority flag equals 0. Therefore, we can also test the null hypothesis that the true parameter is 0 using the z-statistic directly from the output from the logistic regression. The results of both versions will be exactly the same, and most statistical software packages readily provide both sets of estimates.⁵

The primary advantage of using odds ratios is ease of construction. Constructing an odds ratio for an unconditional disparity is a simple calculation as demonstrated above. Constructing an odds ratio for a conditional disparity is easy as well, since the logistic estimator is based on predicting log odds of events, so odds ratios are a standard output of logistic regressions. Using a logistic estimator to estimate a conditional disparity is discussed in more detail in Section IV below.

The primary disadvantage of using odds ratios is difficulty in interpretation. It is difficult to interpret odds ratios, since odds are often less intuitive than probabilities. In fact, there is a long history of Economists mis-interpreting odds as probabilities, and odds ratios as ratios. As one example, based on a review of every article published between 1998 and 1999 in the two leading journals in obstetrics and gynecology, Holcomb et al. (2001) found that of the 151 articles that used odds ratios, 39 (26%) inappropriately interpreted them as ratios. A secondary disadvantage of odds ratios is that they can only be used for 0/1 outcomes, such as approve/deny, and cannot be used for continuous outcomes, such as APR.⁶ This is important since fair lending

⁵ See [Using Stata for Logistic Regression](#) for an example.

⁶ Odds ratios can be constructed for discrete outcomes, which include 0/1 outcomes, as well as outcomes with more than two discrete values. Throughout this report, among discrete outcomes, we focus on only 0/1 outcomes, since it is the most common type of discrete outcome that arises during fair lending analyses.

analyses typically involve analyses of both 0/1 outcomes and continuous outcomes, and using odds ratios for 0/1 outcomes means two different types of disparity measures and two different types of interpretations would be needed, adding another layer of complexity.⁷

Ratio

The ratio measure is a ratio of two occurrence rates. Specifically, the ratio is the percentage of individuals in group 1 for which an event occurred divided by the percentage of individuals in group 2 for which the event occurred, or

$$Ratio = \frac{Percentage(event)_1}{Percentage(event)_2} \quad (3)$$

The ratio ranges from 0 to infinity. A ratio equal to 1 suggests the event was equally likely to occur for both groups, ratios less than 1 suggest the event was less likely to occur for group 1, and ratios greater than 1 suggest the event was more likely to occur for group 1. The conclusion that can be drawn with a ratio is that the event was X times more likely to occur for individuals in group 1 than for individuals in group 2.

Using an underwriting example again, suppose the denial rates for applicants in group 1 and group 2 are 20% and 10%, respectively. The ratio is then just 2 (= 20 / 10). The conclusion that can be drawn in this example is that applicants in group 1 were twice as likely to be denied as applicants in group 2.

For formal statistical tests, the typical null hypothesis is that the true ratio equals 1 and the alternative hypothesis is that the true ratio does not equal 1. There are several statistical

⁷ There are also other more technical disadvantages of the odds ratio that do not impact ratios or differences but that are beyond the scope of this paper. One example is that odds ratios are non-collapsible, which can result in biased odds ratio estimates depending on the other factors included in the logistic regression model (see Greenland and Pearl (2011) for more details).

approaches available to test this null hypothesis. We recommend a parametric test based on a log-binomial regression, since it is easy to apply and has a direct parallel to the test for conditional disparities discussed in Section IV below.⁸ In addition, it is very similar to the parametric test based on a logistic regression discussed above for odds ratios.⁹ Using an underwriting example again to illustrate this approach, suppose we want to test the null hypothesis that the true unconditional denial rate disparity (constructed now as a ratio) is equal to 1. The first step is to estimate a log-binomial regression with just a constant and a minority flag. Exponentiating the coefficient estimate on the minority flag from this regression provides the unconditional denial rate disparity estimate now as a ratio, equivalent to using equation (3). The standard error of this ratio estimate is constructed as the ratio estimate times the standard error of the coefficient estimate.¹⁰ With the ratio estimate and its corresponding standard error estimate, we can then generate a standard z-statistic to test the null hypothesis that the true ratio equals 1. An alternative version of this approach is to use the coefficient estimate on the minority flag directly from the log-binomial regression. Since for a log-binomial regression, the *Ratio* = $e^{(\text{coefficient estimate})}$ and $e^{(0)} = 1$, the null hypothesis that the true ratio equals 1 is equivalent to the null hypothesis that the true parameter on the minority flag equals 0. Therefore, we can also test the null hypothesis that the true parameter is 0 using the z-statistic directly from the

⁸ Two common non-parametric tests are a chi-square goodness-of-fit test and a z-test of equal proportions, both of which are standard output for cross-tabulations of two categorical variables for most statistical software packages. These two tests are equivalent in large samples, but differ in small samples.

⁹ A log-binomial estimator is very similar to a logistic estimator. The primary differences are that the logistic estimator uses a logit link function and defines the outcome as the log of the odds of an event while the log-binomial estimator uses a log link function and defines the outcome as the log of the probability of an event. As a result, exponentiating the coefficient estimates from a log-binomial regression yields ratios instead of odds ratios. See Greenland (2004), Chen et al. (2014), and Lindquist (available at [How can I estimate relative risk using glm for common outcomes in cohort studies?](#)) for details on the log-binomial estimator.

¹⁰ The standard error estimate of the coefficient estimate is part of the standard output of a logistic regression for most statistical software packages. Two common approaches for estimating standard errors for the coefficient estimates in a log-binomial regression are the delta method (see Green (2012)), and expected and observed information matrices (See Van der Vaart (2000)).

output from the log-binomial regression. The results of both versions will be exactly the same, and most statistical software packages readily provide both sets of estimates.¹¹

The primary advantage of using ratios is ease of interpretation. Ratios are functions of probabilities of events occurring, which are generally fairly intuitive to understand. In addition, ratios and the 4/5th rule were introduced in the early 1970s and were referenced by the Supreme Court in *Griggs vs Duke Power Co.*, which established legal precedent for disparate impact cases for employment discrimination, so the ratio is a long-standing measure that is well-known by fair lending attorneys and other stakeholders involved in fair lending reviews.¹²

There are three primary disadvantages to the ratio measure. First, ratios are only appropriate for 0/1 outcomes, such as approve/deny, and cannot be used for continuous outcomes, such as APR. Again, this is important since two different types of disparity measures and two different types of interpretations would typically be needed for fair lending analyses, adding another layer of complexity. Second, a disparity measured as a ratio cannot be interpreted without knowing the underlying probabilities used to generate the disparity. Using underwriting decisions as an example, denial rates of 2% and 1% for two groups, and denial rates of 60% and 30% for two groups, both yield a ratio disparity of 2. However, these two sets of denial rates are quite different and would likely lead to different conclusions and actions. Third, it is slightly more difficult to construct disparities as ratios. As discussed above, constructing disparities as ratios can involve estimation strategies, such as the log-binomial regression, that are not as standard or well known, and therefore more difficult to apply, check, and interpret.

¹¹ Log-binomial regressions may not always converge, so one possible alternative to generate relative risk ratios is the Poisson estimator with robust variances (see Greenland (2004), Chen et al. (2014), and Lindquist).

¹² The 4/5th rule, which the Equal Employment Opportunity Commission established, states that the selection rate of a treatment group should be at least 80 percent (4/5th) of the selection rate of the control group. Ratios below 80% suggest potential discrimination against the treatment group.

Difference

The difference measure is a difference of two average outcomes. Specifically, the difference is the average outcome for individuals in group 1 less the average outcome for individuals in group 2, or

$$\text{Difference} = \text{Average}(\text{Outcome})_1 - \text{Average}(\text{Outcome})_2 \quad (4)$$

Note that we explicitly use the broader term, average, here since the difference measure applies to both 0/1 outcomes and continuous outcomes. For 0/1 outcomes, the average will just be the percentage or rate, which were the terms we used above when discussing odds ratios and ratios.

The range of the difference measure depends on the type of outcome variable and its units. For example, if the outcome variable is a continuous variable such as fees, which are typically measured in dollars, the range of the difference measure will be a relatively large negative value in dollars to a relatively large positive value in dollars (-\$10,000 to \$10,000 for example). Alternatively, for 0/1 outcome variables, the range is -1 to 1. When using 0/1 outcome variables, it is common to multiply the difference measure by 100 to transform the units to percentage points, in which case the range is -100 to 100. Regardless of the type of outcome, assuming higher values of the outcome are worse, a difference equal to 0 suggests the average outcomes are equal for both groups, differences less than 0 suggest average outcomes are more favorable for group 1, and differences greater than 0 suggest average outcomes are less favorable for group 1. The conclusion that can be drawn with the difference is that the average outcome for individuals in group 1 is X units higher than the average outcome for individuals in group 2.

Using an underwriting example, one possible definition of an outcome is that an application is denied. In this example, the average outcome is the denial rate. Suppose the denial rates for applicants for groups 1 and 2 are 20% and 10%, respectively. The difference is then just

10 percentage points ($10 = 20 - 10$). The conclusion that can be drawn in this example is that the denial rate for applicants in group 1 is 10 percentage points higher than the denial rate for applicants in group 2.

For formal statistical tests, the typical null hypothesis is that the true difference equals 0 and the alternative hypothesis is that the true difference does not equal 0. There are several statistical approaches available to test this null hypothesis. Consistent with our recommendations above for odds ratios and ratios, we recommend a parametric test here based on either an Ordinary Least Squares (OLS) or logistic regression.¹³ For example, suppose we are conducting an analysis of pricing and are interested in the null hypothesis that the true difference in average APRs for two groups is 0. Based on a model with just a constant and a minority flag, an OLS estimate of the coefficient on the minority flag will be the difference measure of an unconditional disparity, equivalent to using equation (4). The t-statistic, which is part of standard OLS regression output provides a test of the null hypothesis that the true difference is 0. As a second example, suppose we are conducting an analysis of underwriting and are interested in the null hypothesis that the true difference in denial rates for two groups is 0. Although this is an example of a 0/1 outcome, we could test this null hypothesis exactly as we did for the pricing example using an OLS estimator. In addition, we could also test this hypothesis using a logistic estimator, although the process is a bit more complicated, requiring additional steps to generate estimates of unconditional disparities measured as a difference. Section IV below provides more details on this version of the test.

There are several advantages to using the difference measure. The first advantage is ease of construction. Constructing a difference for an unconditional disparity is a simple calculation

¹³ One common non-parametric test is a t-test of a difference of means.

as demonstrated above. Constructing a difference for a conditional disparity is a bit more involved, specifically for 0/1 outcomes as discussed in Section IV below, but still relatively easy since many statistical software packages now include built-in commands to generate these disparity estimates. A second advantage is that the difference has a solid, theoretical grounding in calculus. Specifically, it is based on the concept of derivatives and how changing one factor impacts another factor on the margin. These theoretical underpinnings make this measure more defensible. A third advantage is that it can be used for both 0/1 outcomes, such as approve/deny, and for continuous outcomes, such as APR. This is important since fair lending analyses typically involve analyses of both 0/1 outcomes and continuous outcomes, and using differences means that only one type of disparity measure and interpretation would be needed, which would eliminate one layer of complexity.

The primary disadvantage of using a difference measure is that for non-linear models typically used for 0/1 outcomes, there are multiple possible difference measures, which makes estimation and interpretation difficult. Section IV below presents some of the standard difference measures Economists commonly use for non-linear models during fair lending analyses.

III. Relationships Between Three Disparity Measures

This section analyzes the relationships between the odds ratio, ratio, and difference disparity measures.¹⁴ Since the odds ratio and ratio are not applicable to continuous outcomes, we focus on an outcome with only two values, such as approve/deny. In addition, all disparities in this section are unconditional disparities. Since unconditional disparities only provide signals of risk, all of the differences between disparity measures discussed in this section would

¹⁴ See Feng et al. (2016) for more details on the relationships between these three measures.

primarily impact decisions about where to focus the overall fair lending review rather than on conclusions of the review.

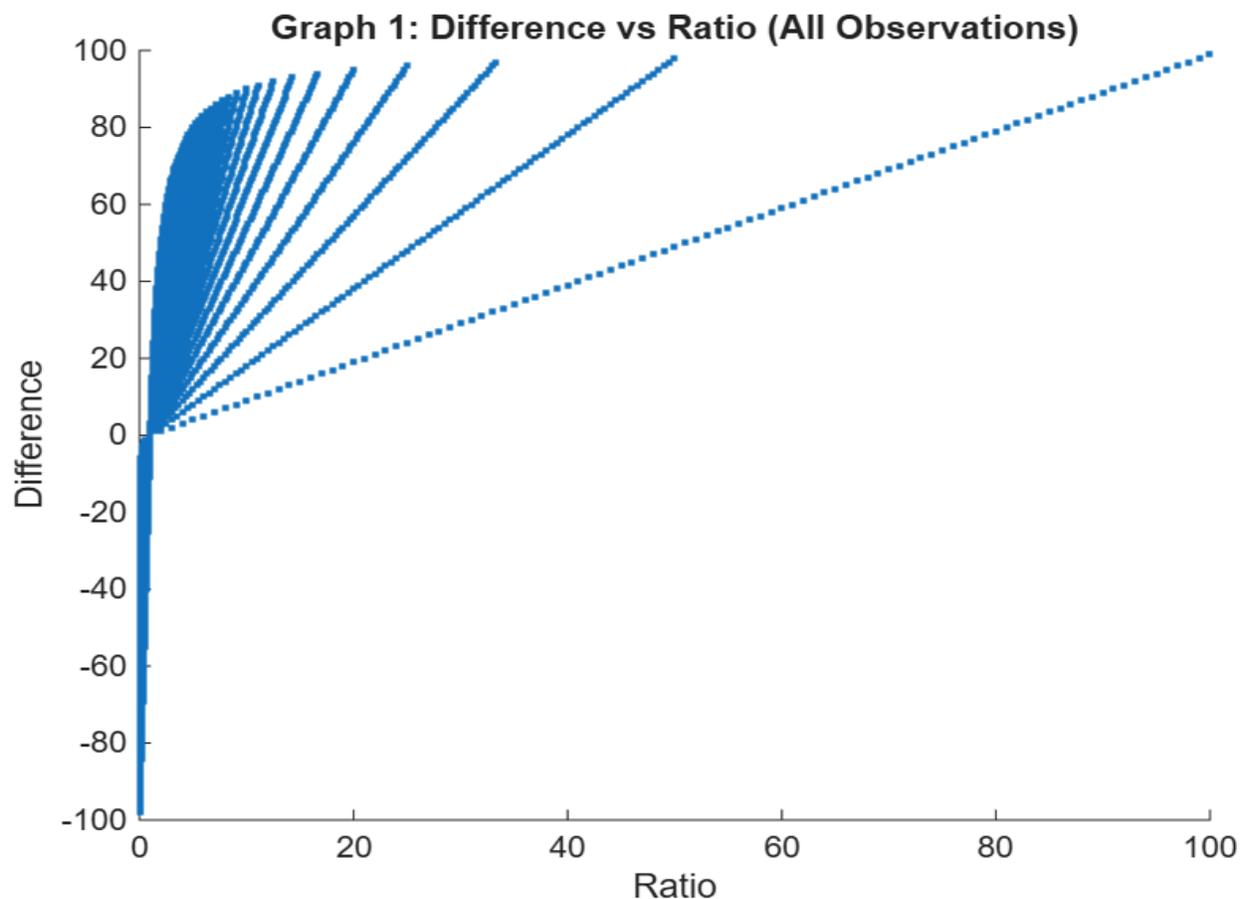
As a first step, we generate the data necessary for the analyses. All three disparity measures are a function of average outcomes for two groups, so we first generate all possible combinations of average outcomes for two groups. For an outcome with two values (specifically 0 and 1), average outcomes range from 0 to 1. We transform these averages to percentages by multiplying by 100, so the range for our analyses will be 0 to 100. Since outcome percentages can take on any value between 0 and 100, there is an infinite number of possible outcome percentages for each group and therefore an infinite number of combinations of outcome percentages for the two groups. To simplify the discussion and graphs below, we limit the possible outcome percentages to only integer values. In addition, since ratios go to infinity as the outcome percentage for group 2 nears 0 and odds ratios go to infinity as the outcome percentage for group 1 nears 100, we limit the possible outcome percentages for each group to all integers between 1 and 99, inclusive. With 99 possible outcome percentages for each group, there are 9,801 ($= 99 * 99$) possible combinations of outcome percentages for the two groups. These combinations will look like, (1%, 1%), (1%, 2%), (1%, 3%), (99%, 97%), (99%, 98%), and (99%, 99%). We then generate an odds ratio, ratio, and difference for each of these 9,801 combinations.

We begin with a bivariate comparison of the ratio and difference.¹⁵ Graph 1 presents a scatterplot of the ratio and difference for each of the 9,801 combinations of outcome percentages in our data. There are three items of note in the graph. First, because we do not allow for outcome percentages below 1% or above 99%, the range of the ratio is 0.01% to 99% instead of

¹⁵ Appendix A presents bivariate comparisons of the ratio and odds ratio, as well as the odds ratio and difference.

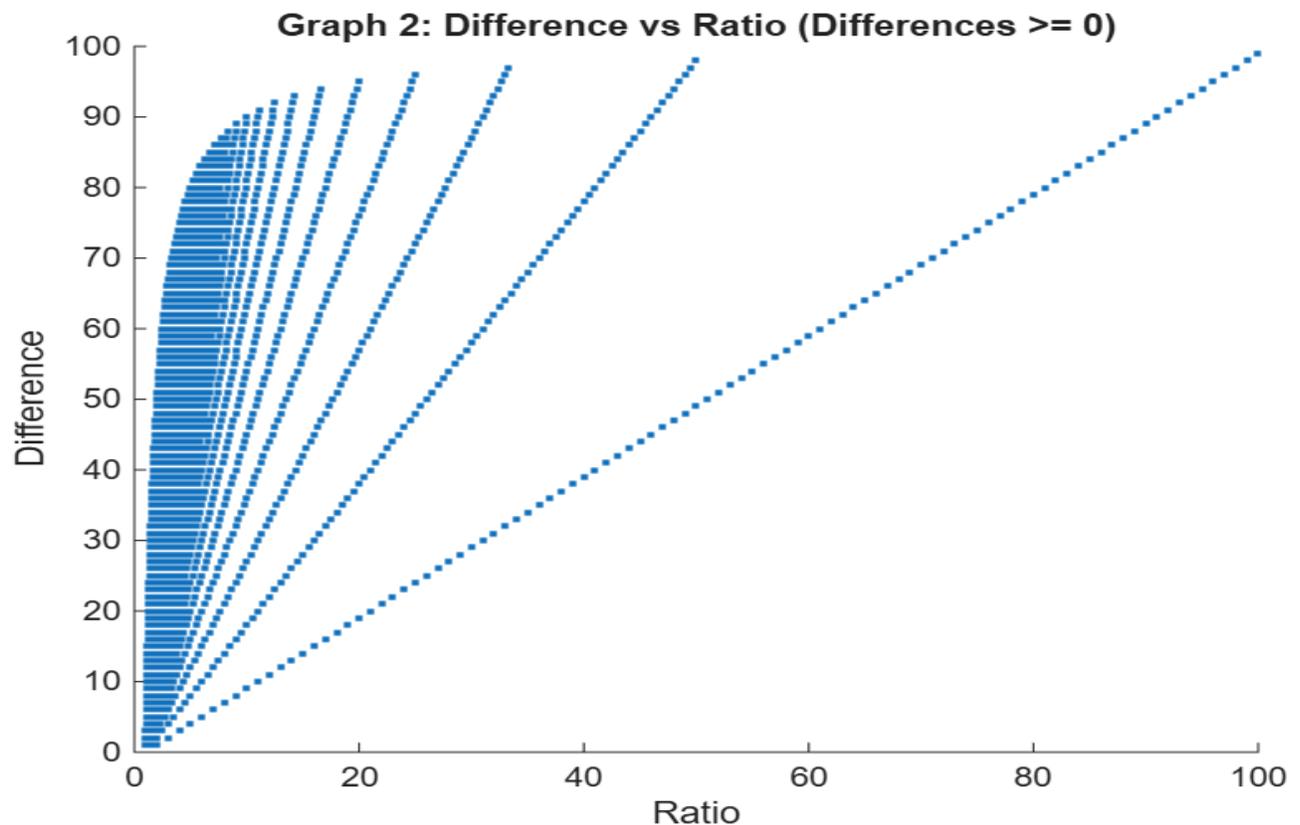
0% to infinity, and the range of the difference is -98% to 98% instead of -100% to 100%.

Second, for all differences less than 0, the ratio is very small, between 0.01% and 1%. Two examples of combinations of outcome percentages in this tail of the graph are (56%, 57%), i.e., a difference of -1 pp and a ratio of 0.98, and (2%, 91%), i.e., a difference of -89 pps and a ratio of 0.02. Third, for differences greater than 0 there are clear relationships between the ratios and differences as reflected in the upward-angled straight lines. We discuss these relationships in more detail below.



Given the limited variation in the ratio for differences less than 0, we generated a second graph (Graph 2), using only combinations of outcome percentages with a difference of 0 or

greater to better visualize the relationships between the ratio and difference. There are four interesting results in Graph 2. First, the distinct upward-angled straight lines indicate clear relationships between the ratio and difference. Each line corresponds to a different value of the outcome percentage for group 2. The line furthest to the right reflects the 99 combinations where the outcome percentage for group 2 is 1%, the second line from the right reflects the 98 combinations where the outcome percentage for group 2 is 2%, and so forth. The slope of each line is equal to the outcome percentage for group 2.¹⁶ These general relationships in Graphs 1 and 2, as well as in the graphs in Appendix A, provide background information and context when considering the statistical disparity results for a fair lending review.



¹⁶ Note that in a real-world application where outcome percentages can take any value between 0 and 100, there would be an infinite number of upward-angled straight lines filling in all of the space between the upward-angled lines for just integers in Graph 2.

Second, related to the upward-angled lines, almost every difference value and ratio value can be constructed from multiple outcome percentage combinations.¹⁷ For example, (60%, 50%) and (12%, 2%) both have a difference value of 10 pps, and (60%, 30%), (22%, 11%), and (2%, 1%) all have a ratio value of 2. As these examples suggest, the same disparity value may be interpreted differently for different outcome percentage combinations, especially for the ratio measure. This highlights the importance of including the underlying outcome percentages in addition to the disparity values when discussing and interpreting the statistical results.

Third, for each difference value except 98%, there are multiple unique ratio values. This can be seen in Graph 2 by choosing one difference value, 40 pps for example, and then drawing a horizontal line at that point in the graph. In our setup based on only integer values for outcome percentages, a difference of 40 pps is generated by the following 59 combinations, (99%, 59%), (98%, 58%),, and (41%, 1%). Each of these combinations has a different ratio value ranging from 1.68 for the (99%, 59%) combination to 41 for the (41%, 1%) combination.¹⁸ The overall point here is that for a given value of the difference measure, there will typically be many different values for the ratio measure that may lead to different interpretations and conclusions, so this needs to be considered when interpreting the statistical results and drawing conclusions.

Finally, very similar to the last point, for each ratio value, there can be multiple unique difference values. This can be seen in Graph 2 by choosing one ratio value, 20 for example, and

¹⁷ Based on our setup using only integer values of outcome percentages, each difference value other than 98% and each ratio value less than 50% can be constructed from multiple outcome percentage combinations. This can be seen for the difference by drawing a horizontal line at any given value except 98% and noting that this line crosses multiple upward-angled lines. Similarly, this can be seen for the ratio by drawing a vertical line at any given value less than 50% and noting that this line crosses multiple upward-angled lines. Note that in a real-world application where outcome percentages can take any value between 0 and 100, each difference value other than 100% and each ratio value other than infinity can be constructed from multiple outcome percentages.

¹⁸ In a real-world application where outcome percentages can take any value between 0 and 100, there will be an infinite number of combinations with a difference of 40 pps. Since these combinations would be bracketed by the combinations of (100%, 60%) and (41%, 1%), the corresponding ratio values would range from 1.66 to 41.

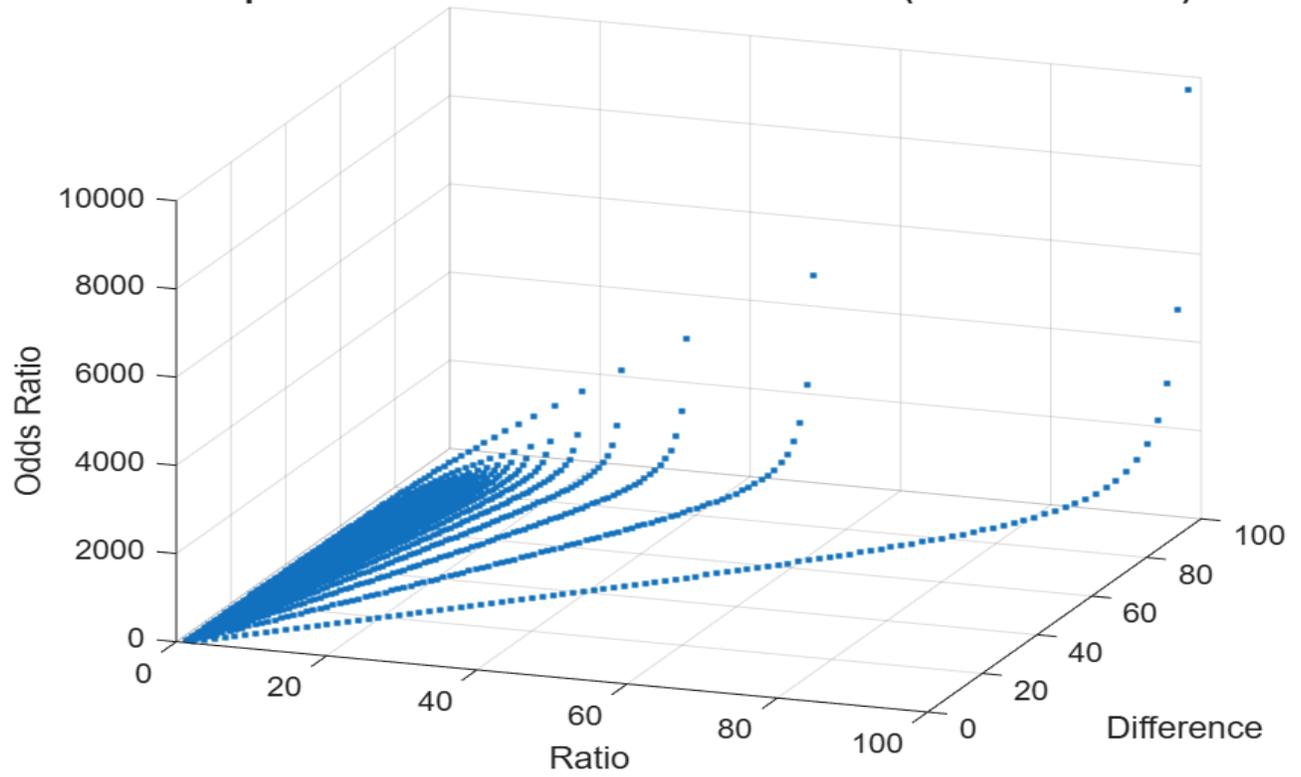
drawing a vertical line at that point in the graph. The number of unique difference values for a given ratio value will depend on the ratio value. In our setup based on only integer values for outcome percentages, for ratios above 49, there is just one difference value for each ratio value. For ratios between 33.33 and 49, there are two difference values for each ratio value. For ratios between 25 and 33.33 there are three difference values for each ratio value, and so forth. For example, a ratio of 5 can occur for 19 outcome percentage combinations: ((95%, 19%), (90%, 18%), (85%, 17%), (80%, 16%), (75%, 15%), (70%, 14%), ..., (10%, 2%), and (5%, 1%)). Although these 19 combinations have the same ratio value, they each have unique difference values ranging from 76 pps for the (95%, 19%) combination to 4 pps for the (5%, 1%) combination.¹⁹ Again, the overall point here is that for a given value of the ratio measure, there will typically be many different values for the difference measure that may lead to different interpretations and conclusions, so this needs to be considered when interpreting the statistical results and drawing conclusions.

We now move from bivariate analyses to a tri-variate analysis of the relationships between all three disparity measures. Graph 3 presents a scatterplot of the difference, ratio, and odds ratio. Following our approach for Graph 2, we only include difference values of 0 or greater. The base of Graph 3 is created by the difference and ratio and shows the same angled, straight-line pattern from Graph 2. Now, however, the graph also has height as shown by the odds ratios on the vertical axis. The primary notable result from Graph 3 is how the odds ratios vary from 0 to 10,000 with the most extreme odds ratios occurring for the largest difference values.

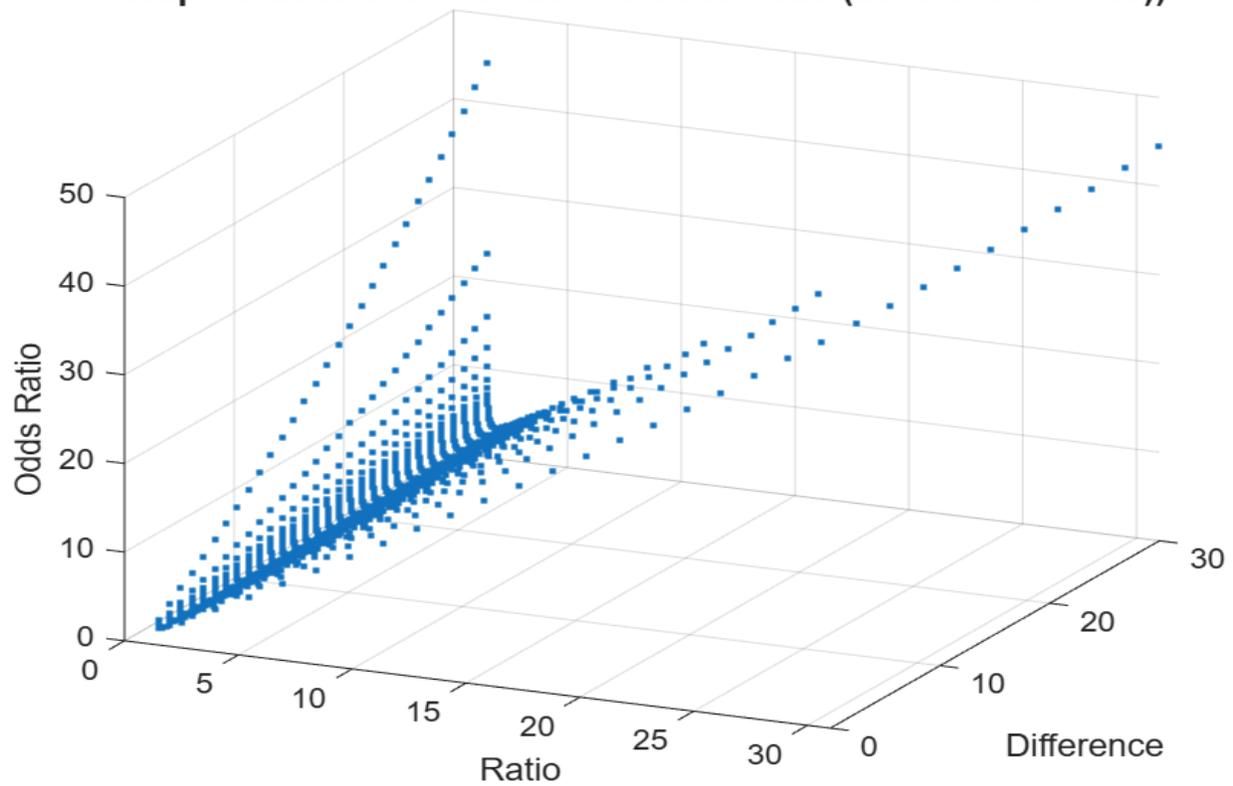
¹⁹ In a real-world application where outcome percentages can take any value between 0 and 100, there will be an infinite number of difference values for each ratio value.

Overall, the extreme odds ratio values in Graph 3 are not very interesting since they occur for difference values near 100, which are unlikely to occur during fair lending analyses. Given that the extreme odds ratio values dominate Graph 3 and mask patterns in the odds ratio values at the lower ends of the difference and ratio scales, we also generated Graph 4 using only difference values between 0 and 30. These difference values are more reflective of the types of difference disparities commonly found during fair lending statistical analyses. Once again, we see clear relationships between the disparity measures. In addition to many of the same types of patterns we identified and discussed in the bivariate analyses, Graph 4 shows two new patterns. First, as both the ratio and difference increase, the odds ratio increases monotonically as well. Although it is difficult to see in the graph, the odds ratio increases at a slightly increasing rate, which can be seen more clearly in Graph 3. Second, Graph 4 shows a clear v-shape pattern in the odds ratios. For any given difference value, the odds ratios are higher for higher ratios, decline steadily as ratio values decline, and then begin to increase steadily for the lowest ratio values. This v-shaped pattern provides another illustration of how different disparity measures can lead to different conclusions. As one example, a difference of 30 pps could be created by all combinations of outcome percentages from (99%, 69%) to (31%, 1%). The ratios for these combinations increase monotonically from 1.43 for the (99%, 69%) combination to 31 for the (31%, 1%) combination. At the same time, the odds ratios follow a v-shaped pattern starting at 44.48 for the (99%, 69%) combination, falling to 3.45 for the (65%, 35%) combination, and then increasing back to 44.48 for the (31%, 1%) combination. Graph 4 and this example show again that for a given value of a given disparity measure, other disparity measures may lead to different interpretations and conclusions, so this needs to be considered when interpreting statistical results and drawing conclusions.

Graph 3: Difference vs Ratio vs Odds Ratio (Differences ≥ 0)



Graph 4: Difference vs Ratio vs Odds Ratio (Differences ≤ 30)



IV. Conditional Disparities

All of the above discussions focused on an unconditional disparity, which provides only a signal of risk. To test for discrimination during statistical fair lending analyses, Economists typically focus on a conditional disparity, which is the disparity that remains after first accounting for the portion of the unconditional disparity that is driven by legitimate policy factors. Similar to unconditional disparities, conditional disparities can be measured as odds ratios, ratios, and differences. This section details how to use regression techniques to generate conditional disparities as an odds ratio, ratio, and difference. The details differ depending on the type of outcome (continuous or 0/1), so we discuss each separately.

Continuous Outcomes

Continuous outcomes, such as rate or fees, take on a large number of values over a continuous spectrum. The ratio and odds ratio are not applicable for continuous outcomes, so we focus only on the difference measure here. Generating the difference measure of a conditional disparity for continuous outcomes is straightforward using standard regression techniques. We illustrate this with a simple pricing example where a lender considers only FICO score when setting rate. Based on this decision-making process, the regression model to generate a conditional disparity and test for discrimination might look something like,

$$Rate_i = \beta_0 + \beta_1 FICO_i + \beta_2 Minority_i + \epsilon_i \quad (5)$$

where $Minority_i$ equals 1 if individual i is a minority applicant and 0 if individual i is a non-minority applicant, $FICO_i$ is the FICO score of applicant i , ϵ_i is a random error term, and the β s are the parameters that will be estimated during the regression analysis.²⁰

²⁰ See chapter 3 of Maddala (1992) for an explanation of the random error term in a regression model with a continuous outcome.

Using a standard estimator, such as OLS, to estimate the β s in equation (5), the difference measure of the conditional disparity will be the coefficient estimate on the minority flag variable. This coefficient estimate, along with the standard error estimate from the OLS regression, can then be used to test the null hypothesis that the true disparity equals 0. In our example, the conclusion that can be drawn with the difference measure is that the average rate for minority applicants is $\hat{\beta}_2$ pps higher than the average rate for non-minority applicants, after controlling for the legitimate policy factors the lender considered when setting rate.

0/1 Outcomes

A 0/1 outcome reflects two possible results, such as approve/deny. Generating conditional disparities when analyzing 0/1 outcomes is somewhat more challenging than for continuous outcomes, since many of the standard estimation strategies rely on non-linear relationships. For all of the discussions in this section, we focus on one common non-linear strategy, the logistic estimator.²¹

To make the discussion of the odds ratio, ratio, and difference measure of conditional disparities for 0/1 outcomes more concrete, we use a simple underwriting example. Suppose a lender considers only FICO score when deciding whether to approve an application. Based on this decision-making process, the regression model to generate a conditional disparity and test for discrimination might look something like,

$$P(\text{denied}_i = 1) = \beta_0 + \beta_1 FICO_i + \beta_2 \text{Minority}_i \quad (6)$$

²¹ As noted above, an OLS estimator can also be used to generate conditional disparity estimates and to test related null hypotheses for 0/1 outcomes.

where $Minority_i$ equals 1 if individual i is a minority applicant and 0 if individual i is a non-minority applicant, $FICO_i$ is the FICO score of applicant i , and the β s are the parameters that will be estimated during the regression analysis.²²

Odds Ratio

Generating the odds ratio measure of a conditional disparity for a 0/1 outcome is straightforward using regression techniques. The approach is exactly the same as for unconditional disparities discussed above except that equation (6) with control variables is estimated instead of a model with only a constant and a minority flag. After estimating equation (6) using a logistic estimator, the conditional disparity measured as an odds ratio is generated by exponentiating the coefficient estimate on the minority flag. The standard error of this odds ratio estimate is constructed as the odds ratio estimate times the standard error of the coefficient estimate. With the estimated odds ratio and standard error, we can then generate a standard z-statistic to test the null hypothesis that the true odds ratio equals 1. As noted above, the same result can also be achieved using the coefficient estimate and standard error for the minority flag directly from the logistic regression.

For our example, the conclusion that can be drawn with an odds ratio is that the odds of minority applicants being denied are X times higher than the odds of non-minority applicants being denied, after controlling for the policy factors that the lender considered when underwriting applications.

²² See pages 327 and 328 of Maddala (1992) for an explanation of the logistic estimator and why equation (6) does not include a random error term similar to equation (5).

Ratio

Generating the ratio measure of a conditional disparity for a 0/1 outcome is straightforward using regression techniques as well. The approach is exactly the same as for unconditional disparities discussed above except that equation (6) with control variables is estimated instead of a model with only a constant and a minority flag. After estimating equation (6) using a log-binomial estimator, the conditional disparity measured as a ratio is generated by exponentiating the coefficient estimate on the minority flag. The standard error of this ratio estimate is constructed as the ratio estimate times the standard error of the coefficient estimate. With the estimated ratio and standard error, we can then generate a standard z-statistic to test the null hypothesis that the true ratio equals 1. As noted above, the same result can be achieved using the coefficient estimate and standard error for the minority flag directly from the log-binomial regression.

For our example, the conclusion that can be drawn with a ratio is that the probability of minority applicants being denied are X times higher than the probability of non-minority applicants being denied, after controlling for the policy factors that the lender considered when underwriting applications.

Difference

Generating the difference measure of a conditional disparity for a 0/1 outcome is a bit more challenging. The difference measure focuses on how a small change in one variable impacts the outcome variable. For non-linear models, such as the logistic model, this estimated difference will not be a constant, but instead will vary across observations and depend on the values of the variable of interest as well as the values of all other variables in the model. Given

that each observation will have its own difference measure estimate, a strategy is needed to aggregate all of these difference estimates into one aggregate measure. There are four common aggregation approaches that Economists use. We summarize each of these approaches using the underwriting example from above.

Approach 1: Average Treatment Effect (ATE)

- Use a logistic estimator to estimate the β s in equation (6)
- Generate a predicted probability of denial for each application (\hat{y}_i)
- Save the original data
- Treat all applications as minority by setting $Minority_i = 1$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{min,i}$)
- Treat all applications as non-minority by setting $Minority_i = 0$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{nonmin,i}$)
- Generate averages separately of \hat{y}_i , $\hat{y}_{min,i}$, and $\hat{y}_{nonmin,i}$
- The difference of the averages of $\hat{y}_{min,i}$ and $\hat{y}_{nonmin,i}$ is the difference measure of the estimated conditional disparity (i.e. the ATE)
- As a check, the average of \hat{y}_i should equal the overall denial rate in the data
- Restore the data to its original state

Approach 2: Average Treatment Effect of the Treated (ATT)

- Use a logistic estimator to estimate the β s in equation (6)
- Generate a predicted probability of denial for each application (\hat{y}_i)
- Save the original data
- Treat all applications as minority by setting $Minority_i = 1$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{min,i}$)
- Treat all applications as non-minority by setting $Minority_i = 0$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{nonmin,i}$)
- Restore the data to its original state
- Generate averages separately of $\hat{y}_{min,i}$ and $\hat{y}_{nonmin,i}$ using only minority applications, i.e. $Minority_i = 1$
- The difference of the averages of $\hat{y}_{min,i}$ and $\hat{y}_{nonmin,i}$ is the difference measure of the estimated conditional disparity (i.e. the ATT)
- As a note, the average of \hat{y}_i will generally not equal the denial rate for minorities

Approach 3: Marginal Effects at the Means (MEM)

- Use a logistic estimator to estimate the β s in equation (6)
- Generate a predicted probability of denial for each application (\hat{y}_i)
- Save the original data
- Generate means for each variable in the model except for the outcome and minority flag
 - For our example, this would be only the FICO variable
- For each variable in the model except for the outcome and the minority flag, replace all applications with the mean
 - For our example, this would be only the FICO variable
- Treat all applications as minority by setting $Minority_i = 1$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{min,i}$)
- Treat all applications as non-minority by setting $Minority_i = 0$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{nonmin,i}$)
- Generate averages separately of \hat{y}_i , $\hat{y}_{min,i}$, and $\hat{y}_{nonmin,i}$
- The difference of the averages of $\hat{y}_{min,i}$ and $\hat{y}_{nonmin,i}$ is the difference measure of the estimated conditional disparity (i.e. the MEM)
- As a check, the average of \hat{y}_i should equal the overall denial rate in the data
- Restore the data to its original state

Approach 4: Marginal Effects at Representative Values (MER)

- Use a logistic estimator to estimate the β s in equation (6)
- Generate a predicted probability of denial for each application (\hat{y}_i)
- Save the original data
- For each variable in the model except for the outcome and the minority flag, replace all applications with a representative value that you choose
 - For our example, this would be only the FICO variable
- Treat all applications as minority by setting $Minority_i = 1$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{min,i}$)
- Treat all applications as non-minority by setting $Minority_i = 0$ for all applications
- Generate a new predicted probability of denial for each application ($\hat{y}_{nonmin,i}$)
- Generate averages separately of \hat{y}_i , $\hat{y}_{min,i}$, and $\hat{y}_{nonmin,i}$
- The difference of the averages of $\hat{y}_{min,i}$ and $\hat{y}_{nonmin,i}$ is the difference measure of the estimated conditional disparity (i.e. the MER)
- As a check, the average of \hat{y}_i should equal the overall denial rate in the data
- Restore the data to its original state

For each of these specific difference measures of estimated conditional disparities,

estimated standard errors can be generated using the delta method. With the disparity measure

and standard error estimates, a standard z-test can then be used to test the null hypothesis that the true difference equals 0.²³

For our example, the conclusion that can be drawn with a difference is that the probability of minority applicants being denied are X pps higher than the probability of non-minority applicants being denied, after controlling for the policy factors that the lender considered when underwriting applications.

V. Conclusion

Discussions of fair lending statistical analyses often focus on the “disparity estimate” and whether it is statistically significant. However, there are several different types of disparity measures, each with strengths and weaknesses and each with unique interpretations. In this report we explored three specific measures Economists commonly use for fair lending analyses, the odds ratio, ratio, and difference. We described the details behind each measure, explored the relationships across measures in the context of unconditional disparities, and then explained how to generate conditional disparity estimates for each measure.

There were four main takeaways. First, there is not one disparity measure that is clearly appropriate for every statistical analysis. Instead, there are several reasonable measures available, each with strengths and weaknesses that need to be considered when choosing which specific measure to use for each given analysis. Second, when presenting and discussing results from a statistical analysis, it is important to understand the details of the specific disparity measure used, and what specific conclusions can be drawn given the specific disparity measure. Third, when

²³ All of these calculations can be done manually. However, many statistical software packages include built in commands that generate these measures, as well as the corresponding standard errors and z-statistics. As one example, the margins command in STATA with no options generates ATEs. The “atmeans” option generates MEMs, the “at” option generates MERs, and the “over” and ”subpop” options generate ATTs.

presenting and discussing results from a statistical analysis, it is important to also include and consider the underlying average outcomes used to construct the disparities. Finally, for a given set of data, different disparity measures can lead to different conclusions about whether discrimination has occurred. When this occurs, relevant policy, analytical, and legal implications need to be considered when drawing overall conclusions and deciding on what actions to take.

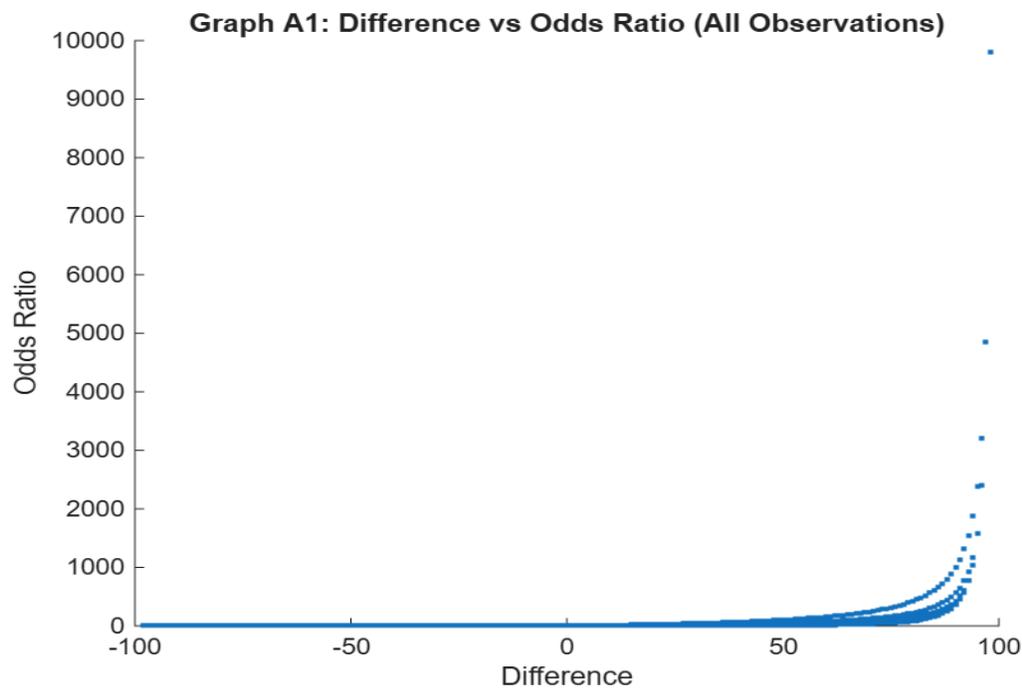
Overall, when discussing statistical fair lending results it is important to be clear about what specific disparity measure was used and then to interpret the specific disparity measure appropriately.

References

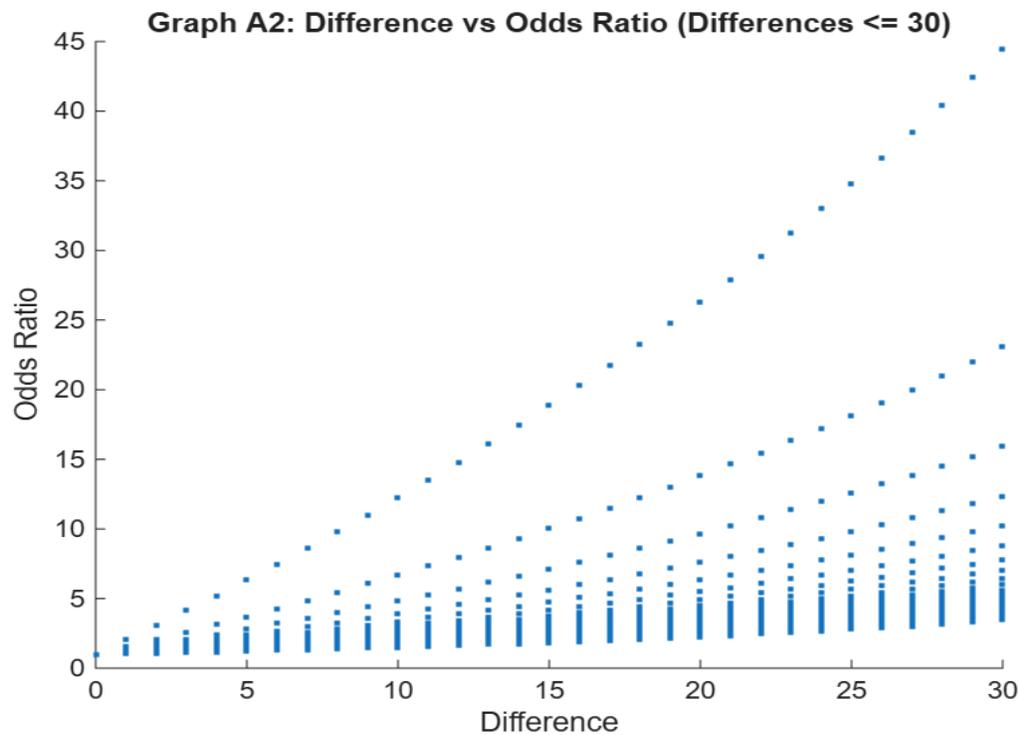
- Chen, Wansu, Jiaxiao Shi, Lei Qian, and Stanley P Azen. 2014. "Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study," *BMC Medical Research Methodology*, 14:82.
- Feng, Chanyong, Hongyue Wang, Bokai Wang, Xiang Lu, Hao Sun, Xin M Tu. 2016. "Relationships among three popular measures of differential risks: relative risk, risk difference, and odds ratio," *Shanghai Arch Psychiatry*, Feb 25; 28(1):56–60.
- Greene, W. H. 2012. *Econometric Analysis*. 7th Edition, Pearson Education, Inc., New York.
- Greenland, Sander. 2004. "Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies," *American Journal of Epidemiology*, 160(4):301–305.
- Greenland, Sander, and Judea Pearl. 2011. "Adjustments and their Consequences—Collapsibility Analysis using Graphical Models," *International Statistical Review*, 79(3):401–26.
- Holcomb, W. L., T. Chaiworapongsa, D. A. Luke, and K. D. Burgdorf. 2001. "An odd measure of risk: use and misuse of the odds ratio," *Obstet. Gynecol.* 98(4):685–88.
- Lindquist, Karla. "How Can I Estimate Relative Risk Using GLM for Common Outcomes in Cohort Studies," *UCLA Advanced Research Computing: Statistical Methods and Data Analytics*.
- Maddala, G. S. 1992. *Introduction to Econometrics*, Second Ed., Macmillan, University of California.
- Van der Vaart, A. W. 2000. *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Appendix A: Bivariate Analyses of Differences vs Odds Ratios and Ratios vs Odds Ratios

Graph A1 presents a scatterplot of the differences and odds ratios for all 9,801 combinations of outcome percentages in our data. The two defining features of Graph A1 are the non-linear relationship between the difference and odds ratio and the extreme odds ratio values as the difference values near 100.



Given how the extreme odds ratios mask variation in the odds ratios for smaller difference values in Graph A1, we generated a second graph (Graph A2), using only combinations of outcome percentages with a difference value between 0 and 30 to better visualize the relationships between the odds ratios and differences. These difference values reflect the types of difference disparities commonly found during fair lending statistical analyses. Similar to Graph 2 above, which compares differences to ratios, Graph A2 also shows upward-



angled lines, although the lines here are slightly non-linear. Also similar to Graph 2, each line here corresponds to a different value of the outcome percentage for group 2. The line furthest to the left reflects the 31 combinations where the outcome percentage for group 2 is 1%, the second line from the left reflects the 31 combinations where the outcome percentage for group 2 is 2%, and so forth.²⁴ Most importantly, all of the key results from the bivariate analysis of the ratio and difference discussed in Section III above also hold here as well. Specifically, each difference value and each odds ratio value can be constructed from multiple outcome percentage combinations; for each difference value, there are multiple unique odds ratio values; and for each odds ratio value, there can be multiple unique difference values. Again, these graphs highlight the importance of including the underlying outcome percentages along with the disparity values

²⁴ Note that in a real-world application where outcome percentages can take any value between 0 and 100, there would be an infinite number of upward-angled lines filling in all of the space between the upward-angled lines for just integers in Graph A2.

when evaluating the statistical results, and understanding that for a given value of a given disparity measure, other disparity measures may lead to different interpretations and conclusions.

Graph A3 presents a scatterplot of the ratios and odds ratios for all 9,801 combinations of outcome percentages in our data. The patterns and takeaways from this graph are very similar to those for the comparisons of differences and ratios, and differences and odds ratios, already discussed.

