Adverse Impact Discrimination and the Fallacy of Composition

November, 2013

Anthony M. Yezer
Department of Economics
George Washington University

Abstract

A fallacy of composition arises when an inference about a part is extended to the whole. Observing that the price of a particular stock always moves up and down with the general market does not mean that every, or even most, stocks are highly correlated with the broad market indexes.

In adverse impact discrimination cases a fallacy of composition problem arises when a test for adverse impact considers only effects on a single protected group of a particular incumbent screening mechanism used by a firm. Firms screen based on statistical model of the relation between applicant characteristics and the likelihood of costly outcomes (consider beneficial outcomes to be negative costs). In a world where alternative statistical models are equally plausible, it is certainly possible changing a screening model can cause it to predict lower costs for a particular protected group compared to an unprotected group. When this is taken as evidence of adverse impact discrimination, there is a fallacy of composition because there are many protected groups and an alternative screening model that favors one protected group does not necessarily favor ALL protected groups. Indeed, an incumbent model that adversely impacts one group compared to an alternative may beneficially impact other groups compared to the alternative. Thus the interests of various protected groups in various possible statistical models and the screening mechanisms that they produce are not necessarily consistent. This raises the possibility that firms which switch to the alternative model in response to complaints from one protected group may invite litigation by those protected groups who are made worse off by the change.

The problem raised here was implicitly recognized when Avery, Brevoort, and Canner (2009) analyzed current credit scoring models for several protected groups and concluded that the models were efficient and effective in predicting credit risk. In this paper, the issue is demonstrated formally for those who value such demonstrations but the logical argument for the source of the fallacy should be evident to those who lack a background in statistics. The conclusion contains a few thoughts on what to do about concerns over the potential for adverse impact discrimination.

Adverse Impact Discrimination and the Fallacy of Composition

Introduction

A fallacy of composition arises when an inference about a part is extended to the whole. Observing that the price of a particular stock always moves up and down with the general market does not mean that every, or even most, stocks are highly correlated with the broad market indexes.

In adverse impact discrimination cases a fallacy of composition problem arises when a test for adverse impact considers only effects on a single minority group of a particular incumbent screening mechanism used by a firm. Firms screen based on statistical models of the relation between applicant characteristics and the likelihood of costly outcomes (consider beneficial outcomes to be negative costs). In a world where alternative statistical models are equally plausible, it is certainly possible for a different model to predict lower costs for a particular protected group relative to an unprotected group. When this is taken as evidence of adverse impact discrimination, there is a fallacy of composition because there are many protected groups and an alternative screening model that favors one protected group does not necessarily favor ALL protected groups. Indeed, an incumbent model that adversely impacts one group compared to an alternative may beneficially impact other groups compared to the alternative. Thus the interests of various protected groups in various possible statistical models and the screening mechanisms that they produce are not necessarily consistent. This raises the possibility that firms changing to an alternative model in response to complaints from one protected group may prompt litigation by other protected groups who did better under the incumbent model. The possibility for endless litigation suggests that the fallacy of composition inherent in the concept of adverse impact discrimination is more than an academic curiosity.

This note begins with a stylized model of the firm's problem. The firm engages in behavior that would generally be characterized as adverse impact discrimination using current approaches for testing. The fallacy of composition is then demonstrated as alternative approaches to the resolution of the problem of adverse impact discrimination result in changes that result in adverse impact discrimination against some other protected group. The conclusion offers some thoughts on problems associated with current models used to provide support for statistical screening systems.

Stylized Model of Adverse Impact Discrimination

A firm needs to make a decision that involves estimating the future behavior of individual applicants. The firm is aware that behavior varies significantly in the population and that this behavior can impose significant costs on the firm (consider benefits to be negative costs). In making its estimate of future behavior, the firm has access to data on past behavior of individuals who had a similar contractual relation with the firm. Furthermore, there is reason to believe that past behavior will be replicated in the future.

In the data from past experience with individuals, let C_j be a measure of the cost imposed on the firm by individual j with j running from 1 to N, the total number of observations in the data. Let a series of X_i's represent variable characteristics of the individuals. Specifically X_{ij} is an observation of the the i^{th} characteristic of the j^{th} individual. Assume that the data available to the firm includes a large number of characteristics so that i runs from 1 to M and M is a substantial number but it is far smaller than N.

The firm then estimates a statistical model of the relation between C_i and the X_{ij}'s. Let's begin with the simplest specification of such a model:

$$C_j = \alpha_1 X_{1j} + \alpha_2 X_{2j} + \alpha_3 X_{3j} + \alpha_4 X_{4j} + \ldots\ldots\ \alpha_M X_{Mj} + \varepsilon_i \qquad (1)$$

In equation (1), the α_i's are parameters reflecting the relation between the individual X_i's and C_i and ε_i is an error term reflecting the effects of measurement error and, most importantly, other factors, call these Y_k's that are not available to the firm. To understand ε_i better, consider an alternative form of equation (1).

$$\varepsilon_i = C_j - \alpha_1 X_{1j} + \alpha_2 X_{2j} + \alpha_3 X_{3j} + \alpha_4 X_{4j} + \ldots\ldots\ \alpha_M X_{Mj} \qquad (2)$$

Equation (2) makes clear that the ε_i is the residual that explains the difference between the actual value of C_j and the value of C would be predicted if the firm had data on the X_i's for a particular applicant and used the estimated model, i.e. the α_i's to predict C_j. The difference between the actual and predicted values of C_j would be ε_i. Obviously, the firm wants ε_i to be small.

One approach would be to include every X_i in the statistical analysis. This would create three types of problems. First, some of the X_i's might be unrelated to C and their inclusion in the model would increase errors in forecasts. Second, forcing in all the X_i's would result in estimated values of the α_i's that were not statistically significant and this would lead to charges that the firm was estimating costs using factors that were inappropriate. Third, forcing in many X_i's would result in what is termed "included variable bias" in which variables that truly belong in the equation might appear to be statistically non-significant. Included variable bias is best explained with an example. Let's say that $X_7 \approx X_{12} + X_{14} - X_4$, i.e. that variable X_7 is approximately the sum of X_{12} and X_{14} less X_4. Under these circumstances, it is very likely that conventional statistical tests of a model where all 4 variables were included would find that X_7 is non-significant because its effect is already reflected in the other three variables. Actually all four variables could be non-significant. Yet we know that X_7 is an important cause of cost differences.

For all three reasons, the firm will need to decide which X_j's to omit from the statistical analysis used to estimate the α_i's in equation (1). In doing this, it is important for the firm to identify the X_j's that are the true cause of costs and to estimate the true α_j's. Why is this so important? Let's say that the firm has the correct X_j's but that one of the α_j's, call it $\alpha_{9,}$ is too small. This means that the firm will systematically underestimate the costs imposed by individuals with high values of X_9. If this happens, and other firms have the correct value of α_9, they will correctly identify the high cost individuals with large X_9 values. In such a world, individuals with high X_9 will identify the firm with the biased equation as the best place to do business and impose high costs on that firm. The firm will either change its forecasting equation, experience losses and go out of business, or pass its higher costs on to all the individuals with whom it deals. The market consequences of lending with a biased forecasting model are very important when considering adverse impact discrimination.

Testing for Adverse Impact Discrimination

Assume that the firm has selected a set of X_i's to use in its forecasting equation. Let's renumber the X_i's so that the firm uses i = 1….L where L < M. This means that X_{L+1}…..X_M are included in the error term. However, to the extent that these omitted variables are correlated with the included X_i's their effects will be included in the estimates of cost.

In testing for adverse impact discrimination, the population is divided into a group called non-minority or unprotected, U, and several minority or protected groups, P_k. It is then possible to test the following hypothesis. Let the current credit scoring scheme be given by:

$$C_j = \alpha_1 X_{1j} + \alpha_2 X_{2j} + \alpha_3 X_{3j} + \alpha_4 X_{4j} + \ldots\ldots\ \alpha_L X_{Lj} + \upsilon_i \qquad (3)$$

This is simply equation (1) with X_{L+1}…..X_M omitted. The error term has changed because υ_i includes the effects of the omitted X_i's. The test for adverse impact discrimination involves trying to find an alternative forecasting equation of the form:

$$C_j = \beta_8 X_{8j} + \beta_9 X_{9j} + \beta_{10} X_{10j} + \beta_{11} X_{11j} + \ldots + \beta_L X_{Lj} + \beta_{L+1} X_{L+1j} + \beta_{L+2} X_{L+2j} + \kappa_i \qquad (4)$$

In equation (4) the first seven X_i's have been dropped and two additional X_i's (specifically X_{L+1} and X_{L+2}) have been added. The estimated coefficients of variables like X_8 which appear in both equations (3) and (4) are noted by β's because they will be different in the two specifications of the cost equation and the error term is now noted κ because the forecast errors from (4) will be:

$$\kappa_i = C_j - \beta_8 X_{8j} + \beta_9 X_{9j} + \beta_{10} X_{10j} + \beta_{11} X_{11j} + \ldots + \beta_L X_{Lj} + \beta_{L+1} X_{L+1j} + \beta_{L+2} X_{L+2j\ i} \qquad (5)$$

which is clearly different than the forecast errors from equation (2).

The demonstration of adverse impact would then involve comparing the predicted values of cost from equation (4) with those from (3) and showing that the relative cost of the protected group is lower using estimates from (4) that from (3). This may be stated formally using a bit of convenient notation. Taking a sample of applicants from the unprotected group, compute expected cost using the estimated values of the coefficients from equations (3) and (4) and call these estimates: $C_U(\alpha_i, X_1,..X_L)$ and $C_U(\beta_i, X_8\ldots X_{L+2})$ then do the same for a random sample of applicants from the k^{th} protected group. Now perform the following comparison:

$$C_U(\alpha_i, X_1,..X_L)/C_U(\beta_i, X_8\ldots X_{L+2}) \text{ versus } C_k(\alpha_i, X_1,..X_L)/C_k(\beta_i, X_8\ldots X_{L+2}) \qquad (6)$$

Three possibilities arise but the one associated with a finding that the incumbent screening equation has an adverse impact on minority group k is:

$$C_U(\alpha_i, X_1,..X_L)/C_U(\beta_i, X_8\ldots X_{L+2}) < C_k(\alpha_i, X_1,..X_L)/C_k(\beta_i, X_8\ldots X_{L+2}) \qquad (7)$$

Equation (7) states that the incumbent screening equation produces relatively higher cost estimates for the protected group k than the alternative estimates based on equation (4). As the inequality becomes larger and more statistically significant, the case for adverse impact of the incumbent system is more compelling.[1]

In adverse impact cases, the firm is allowed a "business purpose" defense. If the firm can show that the absolute value of the errors from (5) are significantly larger than the errors from (2), i.e. show that $|\upsilon_i| > |\varepsilon_i|$, it can claim that it needs to use equation (3) in order to avoid losses or in order to avoid being at a serious competitive disadvantage. As noted above, because use of a screening equation that does not predict well will tend to result in higher losses and higher prices, the protected group may be better off with a more precise screening equation than one that lowers their relative cost estimate. Consider, an obvious non-discriminatory screening device such as flipping a fair coin and having the applicant call heads or tails. Firms could use this device but they would attract a disproportional share of high cost applicants and either have to charge very high prices or go out of business. All this is obvious and should be well understood.

The fallacy of composition enters because it is sometimes assumed that a finding of adverse impact against one protected group based on the inequality in equation (7) can be taken as a finding of adverse impact discrimination. If there are many protected groups, then the test applied in equation (7) must hold for all or, at least, the inequality cannot be reversed for any protected group. If the inequality is reversed, then the protected group is more advantaged under

---

[1] The results of this particular test may well depend on the X characteristics of the unprotected and protected groups that are used to compute the costs. This is a further problem with the test for adverse impact, because the results may depend on how the particular applicants chosen to populate the U and k groups are chosen. There is an additional problem because the characteristics of applicants should vary with the nature of the screening model used. Put another way, changing the screening model will change the X characteristics of the applicants at the firm.

the incumbent screening scheme and the alternative makes them relatively worse off. The remedy to adverse impact discrimination, screening based on equation (4) rather than equation (3), would then be a change which had an adverse impact on some protected groups. Obviously it is a logical contradiction for the remedy for adverse impact discrimination to be an act of adverse impact discrimination.

Given the large number of protected groups, consider the intersection of race, gender, age, and marital status categories, i.e. black, female, elderly, single households as a category, it is very difficult to imagine that one could find an incumbent screening equation that would fail to advantage at least one protected group in comparison to an alternative that predicted as well. The one exception to this would be if the incumbent equation was not well specified or estimated, i.e. the case of gross firm incompetence. Such incompetence should be evident in firm performance due to the process of adverse selection noted above.

Conclusions

Screening of applicants is important in markets. Modern business practice uses screening mechanisms supported by statistical models relating individual characteristics to subsequent performance, called “cost” in this essay. There is discretion in the formulation of the equations used to estimate and support screening efforts. Judgments about what is included in the models and how they are estimated can influence the relative cost estimated for protected or unprotected groups as well as overall model performance.

There is a well understood business purpose defense based on the need for screening to be based on models that predict with precision.

The point of this essay is that a finding of adverse impact against one protected group cannot be used as a finding of adverse impact discrimination unless that finding is not reversed for any other protected group. It is perfectly possible for an alternative model to favor one protected group at the expense of another protected group. The assumption that what is good for one group is good for all is a classic case of the fallacy of composition. The effect of an alternative screening model must be analyzed for each individual protected group before there is a finding of adverse impact discrimination if circular cases in which the remedy for one group is a disease to another are to be avoided.

One final observation is that the focus on the possibility that screening systems could possibly result in adverse impact discrimination tends to distract from the search for more precise statistical models and screening mechanisms. In retrospect, the screening models used in mortgage markets during the 2004–2008 performed badly. This has been documented by research in which the model estimated for the period prior to 2004 have been used to generate predictions of default during the subsequent period.[2] In retrospect, all groups, protected and unprotected, in the population have suffered from this failure of default models. There are sophisticated statistical reasons for the under prediction of default loss models during this period. The primary source of the problem is that individuals who know that they are going to be screened learn about the screening process and prepare in advance for the classification system that they will face. Unfortunately, current statistical models used to estimate cost equations, such as equations (3) and (4) above, assume that the applicant does not know the screening process. This means that better informed applicants find ways to raise their scores artificially, which tends to invalidate a screening process based on these equations. Improving credit scoring models, and particularly dealing with this problem of strategic applicant behavior, should be given greater priority so that all applicants, and the firms that serve them, can benefit.

[2] As a practical model, estimated cost models that are used for applicants in one year must be based on performance data from previous years. Thus, loans underwritten in 2004 were screened and priced based on cost models estimated using mortgages endorsed in 1998, 1999, and perhaps 2000 because it takes at least three years to observe default performance of a cohort of mortgages.