American Economic Journal: Economic Policy 3 (August 2011): 206–231
http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.3.206

Implementing Anti-Discrimination Policies in Statistical Profiling Models[†]

By Devin G. Pope and Justin R. Sydnor*

How should statistical models used for assigning prices or eligibility be implemented when there is concern about discrimination? In many settings, factors such as race, gender, and age are prohibited. However, the use of variables that correlate with these omitted characteristics (e.g., zip codes, credit scores) is often contentious. We provide a framework to address these issues and propose a method that can eliminate proxy effects while maintaining predictive accuracy relative to an approach that restricts the use of contentious variables outright. We illustrate the value of our proposed method using data from the Worker Profiling and Reemployment Services system. (*JEL* C53, J15, J65, J71)

There is a growing debate about how predictive statistical models should be implemented in settings where discrimination is a concern. Anti-discrimination policies typically ensure that predictive models exclude characteristics such as race, and (depending on the situation) age, gender, sexual orientation, etc. However, there is a large and contentious gray area concerning other variables that may be directly predictive but may also serve as proxies for these omitted characteristics.

One prominent example of the debate on this issue came in the passage of California's Proposition 103 in 1988, which limited the variables that insurance companies could use in pricing automobile insurance. Most of the debate surrounded the use of location controls (e.g., zip code) and credit scores. The insurers argued that these types of variables were directly predictive of losses. On the other side, consumer advocates argued that these variables were clearly correlated with race and income (which are banned from use in insurance pricing) and were serving as proxies for the omitted characteristics. Of course, both arguments have merits—these variables are likely neither solely predictive nor purely proxies for omitted characteristics. In the case of Proposition 103, the arguments that the variables were proxies won the day, and the state mandated that insurers price their policies on a very small range of variables that excluded zip codes or credit scores.

* Pope: University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637 (e-mail: devin.pope@chicagobooth.edu); Sydnor: University of Wisconsin School of Business, 5287 Grainger Hall, 975 University Avenue, Madison, WI 53706 (e-mail: jsydnor@bus.wisc.edu). We thank Dan Black, David Card, Stefano DellaVigna, Jesse Leary, Jaren Pope, Matthew Rabin, Stephen Ross, Jesse Rothstein, Paige Skiba, Jeff Smith, Justin Wolfers, and seminar participants at UC Berkeley for helpful comments and suggestions. We are also grateful to David Card for generously sharing data used for this project. All errors are our own.
[†] To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.3.206.

In the intervening years, though, these restrictions have been largely relaxed and the debate continues. While this issue arises frequently in major insurance markets, it is also relevant in a host of other settings, such as the administration of public-assistance programs, college admissions, and mortgage pricing.

In this paper, we argue that in many settings these types of debates can be reframed. Policy debates focus on the choice of either allowing or banning the use of contentious variables, yet there is a middle ground available. A simple procedure can allow the direct predictive power of contentious variables (e.g., zip codes, credit scores) to be captured, while at the same time ensuring that their predictive strength does not capture a proxying effect for omitted characteristics.

Econometrically the problem here is simply classic omitted variable bias. If a variable (e.g., zip code) in the model is correlated with a predictive characteristic that is left out of the model (e.g., race), the included variable will partially proxy for the omitted characteristic and the estimated impact of the included variable will be biased. What makes this situation more interesting is that in these cases the omitted-variable bias arises from the conscious decision to exclude characteristics from the model that may be predictive. It is the fact that in most settings these omitted characteristics are deliberately excluded from the model, and are not inherently unobservable (or simply not observed), that provides a solution to the debates.

What we propose in these settings is the use of a simple procedure that involves first estimating a full model that includes *all* of the characteristics that may be predictive. This means both the usual set of variables *as well as* the protected individual characteristics, such as race, age, gender, etc. Using this full model as a starting point ensures that the predictive power of the other variables does not in any way come from their correlation with the protected characteristics. The key step in the process, however, is that only the coefficients from the non-sensitive predictors are used when producing individuals' predicted values. This can be achieved by using predicted values that for each person are integrated over all of the potential values of the protected characteristics. In linear settings, such as the common OLS model, this is quite simple to do by replacing the individual values of the protected characteristics (e.g., race) with their population means. For example, if a standard OLS model is estimated with a simple dummy variable for minority status, then in the predictive stage the value of the dummy variable for each individual can be set to the population proportion of minorities (rather than the traditional 0 or 1). This approach ensures that the weight given to each variable that is used to form predicted values reflects its estimated direct effect on the outcome of interest. Importantly, though, while the characteristics protected by anti-discrimination laws are used in the estimation, they are not used in forming predicted values, and thus the procedure maintains race-blindness, gender-blindness, etc. in the predictive values.

In Section II, we present an extended discussion of this procedure relative to the alternative approaches that are typically seen in real-world settings. After discussing the implementation of the procedure, we show formally for the OLS case that this procedure will improve predictive accuracy, as measured by the sum of squared errors, relative to a Propostion 103 case where the contentious variables (e.g., zip codes) are banned from use. This section also makes clear that a situation in which the contentious variables are allowed to proxy for the omitted characteristics (what we call the "common" procedure) will have better predictive accuracy than our proposed procedure, but that this improvement in predictive accuracy comes only from the ability of the model to capture some of the influence of the protected characteristics. We also discuss the method applied to discrete-choice settings and provide simulation results showing the same patterns for predictive accuracy in the probit case.

While these different procedures have predictable impacts on predictive accuracy, their effects on economic efficiency are more nuanced. We discuss economic efficiency at the end of Section II and point out that there is a general tension in these different procedures between a) achieving accurate predictions of the outcome of interest and b) giving appropriate relative weight to the different predictors in the model. Relative to a restricted model that eliminates contentious predictors, the proposed method does better on both criteria and can be expected to improve economic efficiency. Relative to the common approach that allows contentious predictors to proxy for omitted characteristics, though, our proposed approach does worse on a) and better on b). We discuss examples of situations where our proposed method may improve economic efficiency relative to this common approach. These are situations in which the distortions in economic decisions that arise from the predictions being wrong "at the margins" (criteria b) are more important than the distortions caused by predictions being wrong "on average" (criteria a).

In Section III, we provide an empirical example to demonstrate the effects of this procedure using an application to data on the use of unemployment benefits from the federally mandated Worker Profiling and Reemployment Services (WPRS) system. Since 1993, states have been required by the federal government to develop and use statistical models that predict which unemployment-insurance (UI) claimants are most likely to exhaust their UI benefits. For those profiled as likely to exhaust, UI benefits are withheld unless the claimant attends a reemployment service workshop.

Using a dataset that contains a 100 percent sample of UI claimants from the state of New Jersey between 1995 and 1997, we analyze the effect of different modeling approaches. We find that the composition of the group of claimants required to attend the reemployment-service workshops is indeed affected by the ability of variables in the standard models to partially proxy for omitted characteristics. The largest effect is for the composition by race. For example, under a system in which the 10 percent of claimants with the highest predicted exhaustion rates are required to attend workshops, the procedure proposed here decreases the proportion of blacks sent relative to the typical approach by over 25 percent.[1] We show that this difference mainly arises due to the strong correlation between race and geographic controls (i.e., zip codes). The bias on the zip-code coefficients in the standard approach results in claimants of any race being more likely to be assigned to a workshop if they come from areas with high black concentration relative to their counterparts from highly white areas. We also show that our proposed procedure has much better predictive accuracy than a model with no zip code controls, which demonstrates the value of our procedure when situations like that of CA Proposition 103 arise and political/societal pressure builds to ban the use of predictors like zip codes.

[1] The percent of workshop participants that are black would fall from 21.9 percent to 16.0 percent.

To our knowledge, the type of procedure outlined here was first discussed by Stephen Ross and John Yinger (2002) in their book, *The Color of Credit*.[2] They argue that omitting minority status when predicting loan default leads to omitted variable bias that may harm minorities.[3] Although they do not have a dataset that includes the type of predictors (e.g., credit score or zip code) that might generate this bias in a substantively significant way, they present a simulation that shows how omitted variable bias might dramatically alter the loan rates charged to minority borrowers. Our paper builds on their work in several directions. First, we provide a formal framework that not only describes the impact of omitted variable bias, but also theoretically outlines the difference in predictive validity across model specifications. Our theoretical analysis also allows for more general statistical procedures (e.g., probit) and includes a framework for evaluating the relative efficiency of the various models relative to a model that excludes the worrisome socially acceptable predictors from the model. Finally, a key contribution of our paper is providing an empirical example using non-simulated data to gauge the potential impact of these various models in an important setting.

The procedure is also closely related to recent efforts by the Federal Trade Commission to quantify the extent to which credit scores used in auto-insurance pricing are proxying for race (Federal Trade Commission 2007). Congress was concerned enough about the impact that the use of variables like credit score have on minorities that in 2003 they passed The Fair and Accurate Transaction Act directing the FTC to investigate whether the use of credit scores has a disparate impact on insurance for minorities. In a study using loss data from auto-insurance companies and controls for race based in part on the racial makeup of census tracts, the FTC study found that credit scores were highly predictive even after controlling for race. On the other hand, they also concluded that the ability of credit scores to proxy for race in the standard models causes an increase in rates for African Americans of 1.1 percent. The problem, of course, is that both those who support and those who oppose the use of credit scores in pricing could cite these results in favor of their cause. It is this type of situation where the compromise provided by the proposed predictive method has a lot of appeal.

This paper also relates to some of the literature on affirmative action in labor markets and college admissions. For instance, both Jimmy Chan and Erik Eyster (2003) and Roland Fryer Jr., Glenn C. Loury, and Tolga Yuret (2003) discuss how universities may react to bans on affirmative action by altering the weights they give to various admissions criteria in a way that increases demographic diversity. Both papers suggest that quota-based affirmative-action policies will be more economically efficient than such a distorted admissions process. Jesse Rothstein (2005) discusses the opposite problem of a university that is not trying to sway the minority

[2] Ross and Yinger's (2002) study was partially motivated by the theoretical work of Shelly J. Lundberg (1991), who examined the effectiveness of equal opportunity laws in labor markets that exhibit statistical discrimination. Lundberg considers how policy makers can enforce these laws when firms use measures such as test scores or height as partial proxies for race or gender.

[3] In the law literature there is a debate about whether this type of harm should be classified as "disparate impact" discrimination. See Ross (2005) and Ross and Yinger (2002), especially pages 313–324, for a discussion of relevant court cases and a review of the legality of disparate-impact discrimination. Also see the Supreme Court ruling in Griggs v. Duke Power Co. (401 US 424 [1971]) for an early ruling on disparate-impact discrimination.

representation and is simply using a race-blind statistical model to determine admittance criteria. He finds that including demographic and socioeconomic characteristics in a regression of college success on SAT and high school GPA reduces the importance of the SAT relative to high school GPA. Thus if universities base admissions on statistical models without accounting for demographics, the importance given to SAT scores relative to GPA in the admission process may be greater than is warranted by its ceteris paribus direct effect on college achievement. If monitored by a regulatory body, the procedure outlined here could address both of these cases and ensure that the weight given to characteristics in admissions policies reflect their direct predictive power.

The remainder of the paper proceeds as follows: Section I gives our theoretical framework, focusing primarily on the simple OLS case. Section II describes the worker profiling system and data. Section III presents the results of our analysis on the WPRS data. Finally, Section IV concludes with a brief discussion about the appropriateness of the procedure in settings where one may be concerned that protected characteristics, such as race, might themselves be proxies for other unobserved characteristics as well as the practical issues that might arise when trying to implement this procedure.

I. Theoretical Framework

For this paper we consider three types of variables. The first are variables, such as race, that have been explicitly determined to be socially unacceptable for use in predictive models (SUPs). The second type of variable is non-contentious variables that are widely considered socially acceptable for use in predictive models (SAPs). The third type of variable is the contentious predictors (CPs), such as zip codes or credit scores, which are in principle socially acceptable, yet are correlated with the socially unacceptable variables. It is these contentious variables that may be banned by policymakers concerned with proxies for socially unacceptable variables.

We start by assuming a data-generating process (a true model) satisfying the standard OLS assumptions:

$$y_i = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}_i^{SAP} + \boldsymbol{\beta}_2 \mathbf{X}_i^{CP} + \boldsymbol{\beta}_3 \mathbf{X}_i^{SUP} + \varepsilon_i, \tag{1}$$

where y_i is the observed outcome for person i who has vectors of socially acceptable predictors $\mathbf{X}_i^{SAP}$, contentious predictors $\mathbf{X}_i^{CP}$, and socially unacceptable predictors $\mathbf{X}_i^{SUP}$. The outcome y_i could be any outcome of interest such as the fraction of UI benefits exhausted, the dollar amount of insurance claims, or a college grade-point average. By their definition, we assume that the contentious variables are correlated with the socially unacceptable predictors, while the socially acceptable predictors are uncorrelated with the socially unacceptable predictors and (for simplicity) are also uncorrelated with the contentious predictors. More formally, we assume an artificial regression model for the SUPs of the form:

$$\mathbf{X}_i^{SUP} = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_{CP} \mathbf{X}_i^{CP} + \mathbf{v}_i, \tag{2}$$

where the elements in the error vector $\mathbf{v}$ are by definition mean zero and uncorrelated with both $\mathbf{X}^{CP}$ and $\mathbf{X}^{SAP}$.

A. *OLS Estimation Procedures*

Estimating the baseline model, which we label the full model, generates unbiased coefficient estimates and provides the following OLS predicted values:

$$\hat{y}_i^{full} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{X}_i^{SAP} + \hat{\boldsymbol{\beta}}_2 \mathbf{X}_i^{CP} + \hat{\boldsymbol{\beta}}_3 \mathbf{X}_i^{SUP}. \tag{3}$$

By their definition, the use of SUPs in generating $\hat{y}_i^{full}$ is not allowed for the final predicted values. The common approach to dealing with this issue is to achieve "SUP-blindness" by simply omitting $\mathbf{X}_i^{SUP}$ from the estimation. This results in a new OLS regression model:

$$\hat{y}_i^{common} = \hat{\gamma}_0 + \hat{\boldsymbol{\gamma}}_1 \mathbf{X}_i^{SAP} + \hat{\boldsymbol{\gamma}}_2 \mathbf{X}_i^{CP}. \tag{4}$$

If the SUPs are predictive in the true model, with $\boldsymbol{\beta}_3 \neq 0$, and there is a correlation between these variables and the contentious predictors, as we have assumed, then naturally the coefficient estimate vector $\hat{\boldsymbol{\gamma}}_2$ will partly reflect the ability of the CPs to proxy for the omitted SUPs. The effect on this coefficient can be obtained simply from the standard omitted-variable-bias formula. By plugging the artificial regression equation for $\mathbf{X}_i^{SUP}$ from equation (2) into the full model in equation (1) and rearranging terms we see that:

$$y_i = (\beta_0 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_0) + \boldsymbol{\beta}_1 \mathbf{X}_i^{SAP} + (\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp}) \mathbf{X}_i^{CP} + (\varepsilon_i + \boldsymbol{\beta}_3 \mathbf{v}_i). \tag{5}$$

This formula shows that rather than consistently estimating $\boldsymbol{\beta}_2$, the coefficient estimate $\hat{\boldsymbol{\gamma}}_2$ will converge to $(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp})$, as the contentious predictors serve as partial proxies for the omitted socially unacceptable predictors in the common estimation procedure.

If society is concerned enough about this residual influence of the socially unacceptable predictors on the predictive process, there may be pressure to also eliminate the contentious predictors from the estimation model as well. This would result in a very restricted model with the following predicted values:

$$\hat{y}_i^{restricted} = \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}_1 \mathbf{X}_i^{SAP}. \tag{6}$$

Because we have assumed that the socially acceptable predictors are uncorrelated with both the socially unacceptable and contentious predictors, each of these different estimation procedures should provide consistent estimates for the coefficient on the socially acceptable predictors. That is we can expect $\hat{\boldsymbol{\alpha}}_1$, $\hat{\boldsymbol{\gamma}}_1$, and $\hat{\boldsymbol{\beta}}_1$ to all converge to the true coefficient estimate $\boldsymbol{\beta}_1$. The problem with this restricted model is that while it completely eliminates the influence of the socially unacceptable predictors, it also loses the predictive power of the contentious variables from the true model that is orthogonal to the socially unacceptable predictors (i.e., $\boldsymbol{\beta}_2 \mathbf{X}_i^{CP}$). The proposed model that we consider avoids the omitted-variable-bias problem inherent in the common method, but does not eliminate the direct predictive influence of the contentious predictors. The proposed method begins by estimating the full model to obtain the coefficient estimates from equation (3). However, unlike the predicted values $\hat{y}_i^{full}$ from equation (3), which include $\hat{\boldsymbol{\beta}}_3 \mathbf{X}_i^{SUP}$ in the generation of predicted values, the proposed method averages across the potential SUP values in the population to form predicted values:

$$\hat{y}_i^{proposed} = \frac{1}{N} \sum_j (\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{X}_i^{SAP} + \hat{\boldsymbol{\beta}}_2 \mathbf{X}_i^{CP} + \hat{\boldsymbol{\beta}}_3 \mathbf{X}_j^{SUP}). \tag{7}$$

Because of the linearity of the OLS case, this clearly simplifies to:

$$\hat{y}_i^{proposed} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{X}_i^{SAP} + \hat{\boldsymbol{\beta}}_2 \mathbf{X}_i^{CP} + \hat{\boldsymbol{\beta}}_3 \overline{\mathbf{X}}^{SUP}, \tag{8}$$

where $\overline{\mathbf{X}}^{SUP}$ is the average vector of socially unacceptable predictors for the population. Because the predicted values are generated on this average vector of SUPs, rather than the individual's vector of SUPs, the procedure is "SUP-blind," while still ensuring that the *variation* in the predicted values comes from the unbiased coefficients on the other variables in the model.

It is important to note that this process is "SUP-blind" but does not ensure equal treatment on average across SUP groups. It is "SUP-blind" because two individuals who differ only in their SUP characteristics will receive the same predicted value from the model. However, the average predicted values will vary across SUP groups because of differences in other characteristics (i.e., CPs) across groups. So, for example, it may be that minorities pay higher rates for auto-insurance because they have lower credit scores on average compared to whites, and credit scores predict loss rates (even after controlling for other characteristics such as race). That effect would still remain under the proposed procedure, but the weight on credit scores would not be influenced by an ability to proxy for race. In this sense, then, it may be useful to think of our procedure as capturing the variation in predictions based on available characteristics from estimations run *within* SUP groups.[4]

B. *Predictive Accuracy*

In most applications of predictive modeling where the proposed procedure might be relevant, the key issue is obtaining predictive accuracy, subject to the regulations on predictors imposed by society. As our primary measure of predictive accuracy we will consider the sum of squared errors, which is a common measure of predictive fit and is the underlying deviance measure that is minimized by OLS estimation. We show in this section that the different predictive processes can be ranked in terms of predictive accuracy from highest to lowest as: full method, common method, proposed method, restricted method. The key result here is that the proposed method of

[4] Our procedure will differ from a within procedure, however, when there are interactions between SUPs and other characteristics in the true model.

generating predicted values will increase predictive accuracy relative to a world in which society bans the use of the contentious predictors.

Consider first the predictions from the estimation of the full model in equation (3):

$$SSE^{full} = \sum (y_i - \hat{y}_i^{full})^2 = \sum \varepsilon_i^2. \tag{9}$$

By definition, the full estimation generates the smallest possible SSE. Of course, while predictive modelers would ideally like to obtain this predictive accuracy, they are limited by the regulations society puts on the variables that can be used in the predictive process. If the common method is used and the SUPs are simply eliminated from the estimation, the resulting predicted values from equation (4) will generate a higher sum of squared errors, because the method does not fully exploit the predictive power of the SUPs.

An equation for the SSE for the common method can be obtained by noting that the common estimation provides predicted values that are based on estimates of the first three terms from equation (5), implying that:

$$SSE^{common} = \sum (y_i - \hat{y}_i^{common})^2 = \sum (\varepsilon_i + \boldsymbol{\beta}_3 \mathbf{v}_i)^2. \tag{10}$$

So here the errors include not only the underlying model errors (ε_i), but also the impact of the variation in the SUPs that cannot be proxied for by the contentious predictors, which is given by $\mathbf{v}_i$ in the artificial equation for the SUPs in equation (2). Since we have assumed that the true model satisfies the usual OLS assumption that the error term (ε_i) is mean zero and orthogonal to each of the independent variables, it follows that the expected value of $\varepsilon_i \mathbf{v}_i$ in the population will also be zero for all elements of the vector. As such, we can rewrite equation (10) as:

$$\begin{aligned} SSE^{common} &= \sum (y_i - \hat{y}_i^{common})^2 = \sum \varepsilon_i^2 + \sum (\boldsymbol{\beta}_3 \mathbf{v}_i)^2 \\ &= SSE^{full} + \sum (\boldsymbol{\beta}_3 \mathbf{v}_i)^2. \end{aligned} \tag{11}$$

Here we can see directly that the decrease in predictive accuracy from the common method relative to the full method is determined by the importance SUPs have on the true outcome values (i.e., $\boldsymbol{\beta}_3$) and the partial inability of the common method to account for variation in SUPs.

If society determines that the proxy effect inherent in this common method is unacceptable and requires that the contentious predictors also be eliminated from the predictive process, the predictive accuracy of the model will be further reduced. To obtain the formula for the SSE for this restricted method, it is useful to rewrite equation (5) as:

$$\begin{aligned} y_i = &\left[\beta_0 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_0 + (\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp})\overline{\mathbf{X}^{CP}}\right] + \boldsymbol{\beta}_1 \mathbf{X}_1^{SAP} \\ &+ \left[(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp})\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right) + (\varepsilon_i + \boldsymbol{\beta}_3 \mathbf{v}_i)\right], \end{aligned} \tag{12}$$

where $\overline{\mathbf{X}^{CP}}$ is the vector of average contentious predictor values in the population. This equation has three large terms that correspond to the terms that will be estimated if the restricted model is estimated. The constant term α_0, will estimate the first large term, while $\boldsymbol{\alpha}_1$ will provide a consistent estimate of $\boldsymbol{\beta}_1$. The final term in equation (12) will be the mean-zero error to the restricted estimation equation. Hence the formula for the SSE for the restricted model is:

$$(13)\qquad SSE^{restricted} = \sum \left(y_i - \hat{y}_i^{restricted}\right)^2$$

$$= \sum \left[(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp})\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right) + (\varepsilon_i + \boldsymbol{\beta}_3 \mathbf{v}_i)\right]^2.$$

Recalling that the ε, $\mathbf{v}$, and $\mathbf{X}^{CP}$ are all orthogonal to one another, the cross products in this equation will all sum to zero, leaving a simplified formula:

$$(14)\qquad SSE^{restricted} = SSE^{common} + \sum \left[(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp})\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right)\right]^2.$$

The predictive accuracy of the restricted method is lower (i.e., higher SSE) than the common method for two reasons. First, the restricted method does not capture the direct effect that the contentious predictors have on the outcome variable. This is reflected by the $\boldsymbol{\beta}_2\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right)$ term in equation (14). Second, the restricted method does not capture the effect of the contentious variables as a proxy for the SUPs and loses that predictive power as well, which is reflected by the $\boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp}\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right)$ term in the equation.

Now consider the proposed method. Recall that we can rewrite the true model by plugging in the artificial regression model for $\mathbf{X}_i^{SUP}$ from equation (2), which gives us:

$$(15)\qquad y_i = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}_i^{SAP} + \boldsymbol{\beta}_2 \mathbf{X}_i^{CP} + \boldsymbol{\beta}_3 (\boldsymbol{\delta}_0 + \boldsymbol{\delta}_{cp} \mathbf{X}_i^{CP} + \mathbf{v}_i) + \varepsilon_i.$$

The proposed method uses the unbiased coefficient estimates for the $\boldsymbol{\beta}$ coefficients from the full model and then plugs in the average value of $\mathbf{X}_i^{SUP}$, which can be obtained from equation (2) as $\boldsymbol{\delta}_0 + \boldsymbol{\delta}_{cp} \overline{\mathbf{X}^{CP}}$. Hence

$$(16)\qquad \hat{y}_i^{proposed} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{X}_i^{SAP} + \hat{\boldsymbol{\beta}}_2 \mathbf{X}_i^{CP} + \hat{\boldsymbol{\beta}}_3\left(\hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_{cp} \overline{\mathbf{X}_i^{CP}}\right).$$

Subtracting equation (16), with consistent $\boldsymbol{\beta}$ estimates, from equation (15) gives the formula for the SSE for the proposed method:

$$(17)\qquad SSE^{proposed} = \sum (y_i - \hat{y}_i^{proposed})^2$$

$$= \sum \left[\boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp}\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right) + (\varepsilon_i + \boldsymbol{\beta}_3 \mathbf{v}_i)\right]^2.$$

Again recalling the orthogonality of ε, $\mathbf{v}$, and $\mathbf{X}^{CP}$, this can be rewritten as:

$$SSE^{proposed} = SSE^{common} + \sum \left[\boldsymbol{\beta}_3 \boldsymbol{\delta}_{cp}\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right)\right]^2. \tag{18}$$

The primary result here is that the predictive accuracy of the proposed method is less than the predictive accuracy of the common method, but is greater than the predictive accuracy of the restricted method. The result is intuitive. The common method allows for the model to capture any direct effect of the contentious variables on the outcome along with the part of the influence of the SUPs that can be proxied for by the contentious predictors. Eliminating the contentious predictors from the model, the restricted approach, increases the SSE by eliminating both the direct predictive influence and the proxy power of the contentious variables. The proposed method allows the model to retain the predictive influence of the contentious variables that is orthogonal to the influence of the SUPs. A comparison of equation (18) and equation (14) makes this difference clear.

$$\mathrm{SSE}^{restricted} = \mathrm{SSE}^{proposed} + \sum \left[\boldsymbol{\beta}_2\left(\mathbf{X}_i^{CP} - \overline{\mathbf{X}^{CP}}\right)\right]^2. \tag{19}$$

The increase in the SSE in the restricted model relative to the proposed comes entirely from the fact that the restricted model loses the direct effect (i.e., $\boldsymbol{\beta}_2$) of the contentious predictors.

C. *Non-OLS Estimation*

The procedure here can be used mechanically with a range of estimation models. All that is required is that the SUPs are included in the estimation procedure and that predicted values from the model are based on sample averages across the SUPs, rather than on the individual SUP values.

However, working out analytical solutions for the exact effect that the various procedures will have on predictive accuracy is generally more complicated outside of the simple OLS estimation framework. Considering the type of applications where our study is likely to be relevant, the most important alternative class of models to consider is probably discrete-choice models. For example, consider statistical estimation of the probability of default on a loan, the likelihood of an accident, or the probability of dropping out of school. In this subsection we discuss the application of our procedure to a probit model and provide simulations of the effect the different procedures would have on predictive accuracy for the probit under a range of parameter values.

We follow the basic latent-variable approach to the probit model as described in Adonis Yatchew and Zvi Griliches (1985). Let y_i^* be an unobserved latent variable that depends linearly on the three types of variables we discuss above and generates the observable index function y_i. We then have:

$$y_i^* = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}_i^{SAP} + \boldsymbol{\beta}_2 \mathbf{X}_i^{CP} + \boldsymbol{\beta}_3 \mathbf{X}_i^{SUP} + \varepsilon_i$$

$$\begin{aligned} y_i &= 1 \qquad \text{if } y_i^* \geq 0 \\ &= 0 \qquad \text{if } y_i^* < 0, \end{aligned}$$

where ε_i is the standard normal distribution. As above we assume the SUPs are uncorrelated with the SAPs but are correlated with the contentious predictors via the artificial regression equation:

$$\mathbf{X}_i^{SUP} = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_{cp}\mathbf{X}_i^{CP} + \mathbf{v}_i, \tag{20}$$

where the elements of $\mathbf{v}_i$ also come from the standard normal distribution. As in the OLS case, if the SUPs are simply dropped from the probit estimation (the common approach), the estimate of $\boldsymbol{\beta}_2$ will be biased as the CPs partially proxy for the omitted SUPs.[5] The predicted values for the full, common, and restricted procedures are simply the Probit predicted values for the probit analogs to the regressions discussed above. The proposed procedure estimates the full probit model, controlling for SAPs, CPs, and SUPs. The proposed predicted values are then given by:

$$\hat{y}_i^{proposed} = \frac{1}{N}\sum_j \Phi(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1\mathbf{X}_i^{SAP} + \hat{\boldsymbol{\beta}}_2\mathbf{X}_i^{CP} + \hat{\boldsymbol{\beta}}_3\mathbf{X}_j^{SUP}). \tag{21}$$

Notice that because of the nonlinearity of the standard normal cdf (Φ), there is no simple analog to equation (8) using the average SUP vector for the probit case. Instead it is necessary for each predicted value to plug in the individual's SAP and CP values and then to average over the range of potential SUP values in the population.

In Table A1 of the Appendix, we present results from a simulation of the effects of these different procedures in a probit model under a range of parameter estimates. The outcome variable of interest here is the SSE from each estimation, and to keep the simulations invariant to the number of observations, we use the average SSE. For the simulations we assume one SAP and one CP that are both independently drawn from the standard normal distribution. We then generate a SUP using equation (20) with a standard normal error, setting $\boldsymbol{\delta}_0$ to zero, and allowing $\boldsymbol{\delta}_{cp}$ to take on values in different simulations of 0, 0.58, and 2.07, which correspond to correlations between the SUP and CP of 0, 0.5, and 0.9 respectively. We then simulate a data generating process as above, allowing the beta coefficients to take a range of different values in order to generate models with varying importance of the different predictors. The results correspond to our analytical discussion of predictive accuracy in the OLS case. In particular, in each case the $SSE^{proposed}$ is less than the $SSE^{restricted}$. We also confirm that the proposed procedure always has at least as high of a sum of squared errors as the common procedure. Consistent with the results from the OLS case, the proposed procedure obtains substantially better predictive accuracy only in specifications where the SUP is strongly predictive of the final outcome and there is a high correlation between the SUP and the CP (refer to models 3, 6, and 9 in the table).

[5] The estimate of $\boldsymbol{\beta}_2$ will converge to $(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3\boldsymbol{\delta}_{\mathrm{CP}})/\sqrt{/\boldsymbol{\beta}_3^2\sigma_\nu^2 + \sigma_\varepsilon^2}$ (Yatchew and Griliches 1985).

D. *Economic Efficiency versus Predictive Accuracy*

To this point we have discussed the effects of these various procedures on predictive accuracy, which is generally the goal of statistical modelers working within applications such as insurance pricing. However, for economic policy the question of interest is often what effect these different procedures will have on economic efficiency. The effects on economic efficiency will generally depend on the particular context in question, but our framework here helps to highlight the issues at play. There is a general tension in these situations between a) achieving accurate predictions of the outcome of interest and b) giving appropriate relative weight to the different predictors in the model. We consider our proposed method superior to a restricted model because (as we showed above) the proposed method gets the relative weights of predictors "right" (i.e., matching the true model) and achieves better predictive accuracy than the restricted model. Comparing our proposed method to the common method, though, highlights the tension—the common method does a better job at predicting outcomes, but achieves those predictions by distorting the weight on the contentious predictors in the model. So the question becomes, for economic efficiency is it more important to predict the outcomes correctly ("get it right on average") or to weight the different characteristics properly ("get it right at the margins")? From the standpoint of economic efficiency, our proposed procedure will be less attractive in situations where the former dominates and more attractive where the latter dominates.

Consider the example of applying these approaches to insurance pricing in settings where insurance purchases are mandated, such as auto insurance. If all companies have to use the same statistical procedures (i.e., no cherry picking), whether or not there is a great deal of predictive accuracy in the prices is largely a question of transfers from one set of customers to another. On the other hand, the effect of various characteristics on prices may cause people to alter their behaviors in ways that are economically inefficient. For example, if zip codes proxy for socio-economic characteristics in the common model and get distorted effects relative to their "true" impact on insurance losses, inefficiencies could arise as people move away from these zip codes to improve their insurance rates or choose (not) to own cars because of the (high) low insurance rates in their zip codes. In these situations, our proposed procedure is likely to improve economic efficiency relative to the common approach.

When insurance is not mandatory, it is likely more important to get the prices more accurate across groups. In particular, if individuals opt out because they can self-insure, then those with "good" SUPs (in terms of effect on the likelihood of a loss) may perceive prices under our proposed procedure to be too high and may opt out, leading to the classic adverse selection problem. The common procedure will alleviate some of the transfers from "good" SUP groups to "bad" SUP groups and will thereby help to mitigate some of the adverse selection problem. If the inefficiencies caused by distorting the relative prices of the CPs are small relative to the benefits of alleviating adverse selection, then the common procedure will be preferable to our proposed procedure.

In settings other than insurance, where for example the purpose is to determine admissions or eligibility for an institution or program, the relative effects of the approaches on economic efficiency will generally be related to the distinction between getting a "good match" and providing the right incentives. Take the example of using statistical modeling for admissions to college, where for instance the outcome of the estimation might be college GPA. The common approach will do a better job than our proposed procedure at ensuring that those most likely to succeed at college get in. However, the common method may also distort the incentives for investing in achieving "good" levels of contentious predictors, such as high SAT scores, and if these distortions are particularly wasteful, then the proposed procedure could improve economic efficiency.

As this discussion highlights, economic efficiency is a nuanced subject when discussing statistical profiling models once society has fairness concerns and limits the use of certain variables in the modeling process. In all of these settings there are likely to be tensions between predictive accuracy, economic efficiency, and perceived social fairness. Our hope is that our framework here may help to clarify some of those discussions and that this subsection will have highlighted the types of situations where our proposed procedure is likely to improve economic efficiency. Furthermore, in situations where society might opt for a restricted model (one that throws away contentious predictors), our procedure will provide a clear improvement in economic efficiency relative to the restricted approach without sacrificing concerns regarding fairness.

II. Data

We apply this method to the setting of unemployment-insurance benefits. In 1993, the Unemployment Compensation Amendments (P. L. 103–152) to the Social Security Act created the Worker Profiling and Reemployment Services (WPRS) system, which required all states to develop and implement statistical profiling models of unemployment insurance (UI) claimants. The stated goal was to identify "who will be likely to exhaust regular compensation and will need job search assistance services." Those individuals who are predicted to be most likely to exhaust their benefits must attend reemployment service classes/workshops in order to continue receiving UI benefits.[6] The percentage of claimants who are sent to the classes/workshops and the type of classes/workshops that are taught vary by state and depend in part on capacity constraints. Similar worker profiling and reemployment service systems are in effect in Australia, Canada, and the United Kingdom (Organisation for Economic Co-Operation and Development 1998).[7]

The Department of Labor gave suggestions to states on potentially useful modeling variables and also specified certain variables that according to Section 188 of

[6] It is ambiguous in this instance whether the requirement to attend reemployment service classes/workshops is considered a "benefit" or a "cost." For example, the results in Dan Black et al. (2003a) shed some light on how claimants are affected by the system. They show that the WPRS system decreases the mean weeks of UI benefit receipt among claimants who are required to attend a reemployment workshop, suggesting that these workshops are effective and beneficial. However, they also conclude that a large part of this reduction is a direct result of claimants being notified of the requirement to attend a workshop, as opposed to actually attending. This seems to suggest that claimants find the prospect of attending the workshop costly and undesirable. We do not take a stand on this issue in the paper.

[7] Further detail regarding the Worker Profiling and Reemployment Services program in the US can be found in David Balducchi (1996) and Stephen A. Wandner (1997).

the Workforce Investment Act of 1998 are not to be used in the predictive process. Specifically, it was determined that "a worker profiling system is not permitted to produce results which discriminate against groups of people. … For this reason, the following variables may not be used in the worker profiling: age, race, ethnic group, sex, color, national origin, disability, religion, political affiliation, and citizenship" (Wandner and Jon C. Messenger 1999). We thus classify these variables as SUPs. While states currently do not include any of these variables in their predictive models, most of them do have data for at least age, race, sex, and citizenship.

In a comprehensive report, Black et al. (2003b) provided recommendations to states on how to develop and implement a Worker Profiling model. While states vary in the complexity of the models that they use, most states have implemented a profiling model similar to what Black et al. (2003b) proposed. We use Black et al.'s recommendations as the foundation for our baseline model. First, they suggest that the model should be estimated using OLS rather than alternatives such as logit or tobit. Second, they argue that the dependent variable should be the fraction of potential benefits that an individual exhausts rather than a discrete measure of benefit exhaustion. Furthermore, they argue that richer models do better, although there is an obvious tradeoff between tractability and predictive accuracy. In the end, the model that Black et al. describe as their preferred model includes the following predictors: Education dummies, job tenure, job tenure squared, 1-digit occupation codes, pre-unemployment wages, local-unemployment-office fixed effects, welfare receipt measures (food stamps, public transit use, etc.), and measures of previous UI benefit exhaustion if applicable.

We analyze a dataset that contains a 100 percent sample of UI claimants from the state of New Jersey between 1995 and 1997.[8] We include as SAPs in our baseline model all variables that the New Jersey data has available that were suggested by Black et al. (2003b). In the end, we are able to include education dummies, job tenure, job tenure squared, 1-digit industry codes, pre-unemployment wages, and local-unemployment-office fixed effects in the model.[9] The New Jersey dataset does not include measures for welfare or previous UI benefit exhaustion. While clearly imperfect, these variables are similar to those used in practice by many states. The New Jersey data also contains information on the race, gender, citizenship, and age of the UI claimants. These are the SUPs that we analyze.

We begin with 732,980 UI-claimant observations between 1995 and 1997. We drop 39,923 observations (5.4 percent) that have data problems or extreme values.[10] Following the directions of Black et al. (2003b) and the Department of

[8] See Card and Phillip B. Levine (2000) for a more detailed description of the data.

[9] As studied by Card and Levine (2000), New Jersey had a special program during a portion of our sample, which allowed claimants to receive extra weeks of benefits. We include fixed effects denoting whether a claimant was eligible for this program in our analysis.

[10] Observations are dropped if they are missing an industry code, gender, age < 18, age > 80, average weekly wage in top 1 percent or bottom 1 percent, or if the race was undefined (not defined as black (non-Hispanic), white (non-Hispanic), or Hispanic).

TABLE 1—SUMMARY STATISTICS

Variable	Full sample	Race: White	Race: Black	Race: Hispanic	Gender: Female	Gender: Male
Benefit exhaustion measures						
Fraction of benefits exhausted	0.74	0.72	0.78	0.76	0.75	0.73
Fraction who exhausted completely	0.48	0.44	0.56	0.50	0.50	0.46
SUPs						
Race						
White	0.62	1.00	0.00	0.00	0.63	0.62
Black	0.19	0.00	1.00	0.00	0.20	0.18
Hispanic	0.19	0.00	0.00	1.00	0.18	0.20
Age						
Age < 30	0.27	0.24	0.31	0.32	0.24	0.29
Age 30–45	0.42	0.41	0.46	0.43	0.41	0.43
Age > 45	0.31	0.35	0.23	0.25	0.35	0.28
Gender						
Female	0.42	0.43	0.44	0.40	1.00	0.00
Male	0.58	0.57	0.56	0.60	0.00	1.00
Citizenship						
US citizen	0.88	0.95	0.92	0.59	0.88	0.87
Non US citizen	0.12	0.05	0.08	0.41	0.12	0.13
SAPs						
Employment information						
Pre-UI weekly wage ($100s)	4.9	5.5	4.1	3.7	4.2	5.4
Tenure (years)	0.95	1.06	0.84	0.71	1.00	0.90
Education						
Less than high school	0.19	0.13	0.17	0.43	0.17	0.21
High school diploma	0.46	0.47	0.48	0.38	0.46	0.45
Some college	0.21	0.22	0.25	0.13	0.23	0.19
College degree	0.14	0.18	0.09	0.06	0.14	0.14
Graduate degree	0.002	0.003	0.001	0.001	0.002	0.002
Industry						
Agriculture	0.05	0.04	0.02	0.09	0.01	0.07
Mining	0.001	0.001	0.0003	0.0004	0.0002	0.002
Construction	0.09	0.11	0.04	0.06	0.02	0.14
Manufacturing	0.18	0.14	0.14	0.32	0.18	0.17
Transportation	0.07	0.07	0.09	0.05	0.06	0.08
Wholesale trade	0.08	0.09	0.07	0.09	0.08	0.09
Retail trade	0.16	0.17	0.15	0.11	0.17	0.15
Finance, insurance, real estate	0.06	0.06	0.06	0.03	0.08	0.04
Public administration	0.03	0.03	0.05	0.01	0.04	0.02
Services	0.29	0.28	0.39	0.22	0.36	0.23
Observations	590,924	367,794	110,239	112,891	250,156	340,768

Notes: The data includes the maximum number of weeks of benefits for which the claimant was eligible. The fraction of benefits exhausted for each claimant is computed by dividing the number of weeks of benefits used by the total number available. Graduate degrees include masters, professional, and doctoral degrees. The industry groups are based on the 1-digit, US Standard Industrial Classification (SIC) industry codes from 1987.

Labor, we also drop 102,133 individuals (13.9 percent) who indicated they were union members.[11]

Summary statistics for the remaining 590,924 observations are displayed in Table 1. The table lists measures of benefit exhaustion, SUP characteristics, and SAP characteristics. The SUP classifications are race (black (non-Hispanic), white

[11] Black et al. (2003b) actually drop claimants who find work through a union hiring hall, or who have a known recall date. Since these variables are not available in our dataset we drop all union workers.

Table 2—Different Methods of Predicting Exhaustion

Dependent variable: Fraction of benefits exhausted (columns 1–3) and an indicator for exhaust (columns 4–6)

	OLS			Logit		
Variable [Percent white, percent black, percent Hispanic]	Full method	Common method	Restricted method (without location controls)	Full method	Common method	Restricted method (without location controls)
Black	0.0538***			0.0942***		
	(0.0012)			(0.0019)		
Hispanic	0.0220***			0.0419***		
	(0.0014)			(0.0021)		
30 < Age < 45	0.0392***			0.0628***		
	(0.0010)			(0.0016)		
45 < Age	0.0559***			0.0952***		
	(0.0012)			(0.0018)		
Female	0.0077***			0.0231***		
	(0.0009)			(0.0014)		
US citizen	0.0340***			0.0620***		
	(0.0014)			(0.0022)		
Prior weekly wage ($100s)	−0.0001	0.0002	−0.0004**	0.0009***	0.0009***	0.0004
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0002)
Tenure (years)	−0.0055***	0.0003	−0.0037***	−0.0053***	0.0055***	−0.0025***
	(0.0005)	(0.0005)	(0.0005)	(0.0009)	(0.0008)	(0.0009)
Tenure squared	0.0007***	0.0004***	0.0007***	0.0011***	0.0006***	0.0011***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

(*Continued*)

(non-Hispanic), Hispanic), age (<30, 30–45, >45), and gender.[12] Looking at the dependent variable measure (fraction of benefits exhausted), UI claimants on average exhausted 74 percent of their available benefits. This exhaustion rate appears to vary across race and gender groups and also age and citizenship (not shown in the table). For instance, on average whites, blacks, and Hispanics exhausted 72 percent, 78 percent, and 76 percent of available benefits respectively. Table 1 also provides the first glimpse of correlations that exist between SUP and SAP characteristics. While not reported due to space constraints, some of the largest correlations occur between SUP characteristics (especially race) and local-office controls.

III. Empirical Results

A. *Changes in Regression Coefficients*

Following the specification suggested by Black et al. (2003b), columns 1–3 of Table 2 present OLS regressions and standard errors—robust to heteroskedasticity—using the fraction of potential benefits exhausted as the dependent variable. The first

[12] The race designations are those available in the data. The data includes exact age, however, to ease the interpretation of the results and after examining the nonparametric relationship between age and benefit exhaustion, we chose to use three discrete age groups in the analysis.

Table 2—Different Methods of Predicting Exhaustion (*Continued*)

Dependent variable: Fraction of benefits exhausted (columns 1–3) and an indicator for exhaust (columns 4–6)

	OLS			Logit		
Variable [Percent white, percent black, percent Hispanic]	Full method	Common method	Restricted method (without location controls)	Full method	Common method	Restricted method (without location controls)
Less than high school	−0.0011	−0.0013	0.0056***	−0.0129***	−0.0138***	−0.0047***
	(0.0012)	(0.0012)	(0.0011)	(0.0018)	(0.0018)	(0.0018)
Some college	−0.0088***	−0.0112***	−0.0184***	−0.0091***	−0.0129***	−0.0252***
	(0.0011)	(0.0011)	(0.0011)	(0.0017)	(0.0017)	(0.0017)
College degree	−0.0349***	−0.0395***	−0.0505***	−0.0440***	−0.0520***	−0.0695***
	(0.0014)	(0.0014)	(0.0013)	(0.0021)	(0.0021)	(0.0020)
Master's degree	−0.0541***	−0.0569***	−0.0633***	−0.0753***	−0.0801***	−0.0901***
	(0.0095)	(0.0095)	(0.0096)	(0.0142)	(0.0142)	(0.0142)
Construction [79, 9, 12]	−0.0190***	−0.0359***	−0.0465***	−0.0878***	−0.1200***	−0.1361***
	(0.0017)	(0.0016)	(0.0016)	(0.0025)	(0.0024)	(0.0024)
Public admin. [58, 35, 7]	−0.0221***	−0.0159***	−0.0132***	−0.0513***	−0.0410***	−0.0368***
	(0.0026)	(0.0026)	(0.0026)	(0.0039)	(0.0039)	(0.0039)
Local office #10 [9, 85, 6]	0.0555***	0.0843***		0.0741***	0.1224***	
	(0.0032)	(0.0031)		(0.0049)	(0.0046)	
Local office #12 [27, 41, 32]	0.0572***	0.0674***		0.0687***	0.0860***	
	(0.0029)	(0.0029)		(0.0045)	(0.0044)	
Local office #21 [97, 1, 2]	−0.0541***	−0.0609***		−0.0873***	−0.0997***	
	(0.0040)	(0.0040)		(0.0061)	(0.0061)	
Local office #26 [97, 1, 1]	−0.0969***	−0.1016***		−0.1376***	−0.1459***	
	(0.0073)	(0.0073)		(0.0111)	(0.0110)	
Other industry codes	X	X	X	X	X	X
Other local offices	X	X	X	X	X	X
Observations	590,924	590,924	590,924	590,924	590,924	590,924
Adjusted and psuedo R^2	0.034	0.026	0.0093	0.032	0.025	0.012

Notes: Coefficient values and robust standard errors are presented from the OLS and Logit regressions using fraction of benefits exhausted (an indicator for exhaustion in the Logit case) as the dependent variable. Columns 1 and 4 use all of the available covariates in the model (including SUPs such as race)—the full method. Columns 2 and 5 report the results of the common method where the SUPs are excluded from the model. Columns 3 and 6 report the results of the restricted method where the SUPs and the local office controls are excluded from the model. White is the racial base group. Age < 30 is the age base group. Having a high school diploma is the education base group. All local offices (#17 = base group) and all industry groups (services = base group) are included in the regression, but only six coefficients are presented. The brackets next to these variables presents the percent of white, black, and Hispanic claimants, respectively, for which each local office or industry group is composed.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

column provides the estimated coefficients from the OLS regression using both the SAPs and the SUPs as predictors (*full method*, equation (1)). Of particular interest, we find that the SUPs do have an effect on benefit exhaustion outcomes even after controlling for the SAPs in the model. Females, older individuals, US citizens, blacks, and Hispanics are all more likely to use up their UI benefits than their observationally equivalent counterparts. While we include all industry and local-unemployment-office dummies in the regression, due to space constraints we only present the coefficient values on two industry dummies and four local-unemployment-offices dummies. To aid the discussion that follows, we highlight those controls that are most correlated with race. Specifically, we report the coefficient on public administration, which is the industry that has the highest percentage of black claimants, and construction, which is the industry that has the highest percentage of white claimants. Similarly, we present the coefficients for local-unemployment office 10 and 12, which have the highest percentage of black UI claimants, and office 21 and 26, which have the highest percentage of white UI claimants. In brackets next to each of these offices and industries we indicate the racial breakdown.[13]

The second column of Table 2 presents the coefficient values that are generated by regressing fraction of potential benefits exhausted on only the SAPs (*common method*). The changes in the SAP coefficients between the *full method* and the *common method* reflect the correlations between the SAPs and omitted SUPs. For example, in the *full method* the coefficient on tenure is negative and statistically significant. However, once the SUPs are dropped from the model, the coefficient on tenure switches sign and is indistinguishable from zero. This is likely the result of the positive correlation between a claimant's age and his/her tenure with previous employer. Indeed, the *full method* reveals that all else equal claimants over the age of 45 exhaust on average 5.6 percent more of their available benefits than do those under the age of 30.

Some of the biggest effects of such omitted variable bias are seen in the changes to the coefficients for the local-unemployment-offices controls. For instance, the predicted fraction of benefits exhausted by claimants from local office #10 (which has the highest concentration of black claimants) rises by more than 2.8 percentage points when SUP controls are dropped from the model. In addition, the coefficient for local office #21 (which has the highest concentration of white claimants) falls by approximately 0.7 percentage points. The combined effect of these two changes (3.5 percent) indicates that in the *common method*, these highly segregated local unemployment office controls are able to absorb a substantively significant fraction of the black-white difference that was found when using the *full method* (5.4 percent). In contrast, the *proposed method* uses the same SAP coefficients as the *full method* and simply eliminates the indicated 5.4 percent black-white difference during the generation of the final predicted values.

As discussed in the introduction, when variables such as local office controls are highly correlated with race, pressure may rise to exclude these variables from the predictive model due to the potential they have of proxying for SUPs. Column 3 of Table 2 provides the results from a regression where both the SUPs and local office controls are omitted from the model, the *restricted method* from Section II. While this model eliminates the ability of local office controls to proxy for race, it now increases the pressure on other variables in the model to start proxying for SUPs. For example, the coefficients on the education variables change in both a statistically and economically significant way, suggesting that they are now being estimated with bias in order to partially account for the variables left out of the model. The adjusted R^2 in column 3—which is significantly lower than the adjusted R^2s in columns 1 and 2—indicates that a substantial reduction in

[13] The local-unemployment office base-group (office #17) and the industry base group (services) were chosen to be groups that had racial compositions most similar to the makeup of the overall population.

predictive accuracy is a further result of omitting these variables—which is something we will return to at the end of the paper.

Columns 4–6 report analogous results to columns 1–3 when using a logit specification with marginal effects (evaluated at the mean). In this case, exhaust all benefits is the binary dependent variable. These results are qualitatively similar to the OLS results presented in columns 1–3 and illustrate that this procedure can be easily generalized to more nonlinear modeling approaches.[14]

B. *Changes in Workshop Composition*

We are interested in how the different coefficients on the SAPs might affect which claimants are required to attend reemployment service classes. This is probably the effect of interest for those concerned with the social "fairness" of the use of different modeling approaches. We ask what the composition of the class (by race, gender, age, citizenship) would be under the alternative profiling methods. In Table 3, for each method we consider two hypothetical cutoff rules: 90 percent and 80 percent. We use these cutoff rules as examples, but they are in line with practice, as in 2005 New Jersey required 16.9 percent of all UI claimants to attend workshops. Under these rules, the top 10 percent and 20 percent of predicted exhausters, respectively, would be required to attend a workshop. Claimants are ranked by their predicted benefit exhaustion for each of the three profiling methods discussed in this paper. The table then compares the SUP makeup of the top 10 percent or 20 percent of predicted exhausters using these rankings.[15]

Looking at the 90 percent-cutoff rule, if the *full method* is used (which includes explicit controls for racial categories), 30.5 percent of the class is white, 41.3 percent is black, and 28.2 percent is Hispanic. In the *common method* (which drops the racial controls from the estimation), the class composition is 45.2 percent white, 21.9 percent black, and 32.9 percent Hispanic. Finally, in the *proposed method* (which includes race in the estimation, but not in the predicted values), the makeup is 47.6 percent white, 16.0 percent black, and 36.4 percent Hispanic. It is worth noting that in the absence of omitted variable bias, the predictions from the *common* and *proposed method*s would be identical. Table 3 reveals, however, that the fractions of the required workshop participants that are white and Hispanic are lowered by 2.4 percent and 3.5 percent, respectively, under the *common method* relative to the proposed alternative. Blacks, on the other hand, make up 5.9 percent more of the workshop under the *common method*.[16] Thus, using the *proposed method*, instead

[14] In an earlier version of the paper, we also included estimates from a Tobit specification that suggested similar results.

[15] One potential concern for this type of analysis is that the actual treatment of various individuals by the program could contaminate our analysis (especially if the state of New Jersey was using a model similar to the *common method* when assigning treatment). For this reason, it is helpful that we are looking at various percentile cutoffs (see the discussion of Figure 1 below that shows the results for all percentile cutoffs). As long as we look at a percentile cutoff that is far away from that used by the state of New Jersey during our sample period, the amount of bias if any will be small since none of the marginal claimants for our various methods will have likely been actually treated.

[16] Given that the coefficient on the Hispanic dummy in Table 2 suggests that being Hispanic increases the probability of exhaustion, it may be surprising to find that Hispanics are more likely to get sent to attend workshops under the *proposed method*. If the regressions only included Hispanics and whites, then Hispanics would be sent to classes less often under the *proposed method*. However, this effect is being overpowered by the much larger change

Table 3—Composition of Claimants Required to Attend a Reemployment Workshop

	90 percent cutoff rule				80 percent cutoff rule			
	Full	Common	Proposed	Difference (common–proposed) mean (SE)	Full	Common	Proposed	Difference (common–proposed) mean (SE)
Race								
Percent white (not Hispanic)	30.5	45.2	47.6	−2.5 (0.6***)	31.6	43.4	46.4	−3.0 (0.4***)
Percent black (not Hispanic)	41.3	21.9	16.0	5.9 (0.5***)	42.3	24.5	18.6	5.9 (0.6***)
Percent Hispanic	28.2	32.9	36.4	−3.5 (0.8***)	26.1	32.1	35.0	−2.9 (0.4***)
Gender								
Percent male	50.3	57.0	58.1	−1.1 (0.3***)	51.1	54.5	55.7	−1.2 (0.2***)
Percent female	49.7	43.0	41.9	1.1 (0.3***)	48.9	45.5	44.3	1.2 (0.2***)
Age								
Percent < 30	9.1	26.9	28.8	−1.9 (0.2***)	12.4	26.8	27.9	−1.1 (0.2***)
Percent 30−45	43.8	40.5	40.5	0.0 (0.1)	44.7	40.5	40.9	−0.4 (0.1***)
Percent 45+	47.1	32.6	30.7	1.9 (0.2***)	42.9	32.7	31.2	1.5 (0.2***)
Citizenship								
Percent US citizen	94.5	86.8	85.5	1.3 (0.5**)	91.9	82.8	81.7	1.1 (0.2***)
Percent not US citizen	5.5	13.2	14.5	−1.3 (0.5**)	8.1	17.2	18.3	−1.1 (0.2***)

Notes: For each cutoff rule and profiling method, the values in the table give the proportion of the reemployment workshop that have a given SUP characteristic. Standard errors are generated using a bootstrap routine with 200 repetitions and 100 percent samples with replacement.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

of the *common method*, generates more than a 25 percent reduction in the fraction of the class that is black.[17] A bootstrap routine reveals that these differences are significant at the 1 percent level.

For gender, age, and citizenship the differences between the workshop composition under the *common* and *proposed method*s are in the directions that one would predict given the omitted variable bias inherent in the *common method*. Females, older claimants, and US citizens all represent a higher fraction of the workshop when using the *common* rather than the *proposed method*. The bootstrapped standard errors reveal that in all but one case these differences are statistically significant at the 5 percent level, with most significant at the 1 percent level. In general, however, the difference between the two methods is larger for breakdowns by race than age, gender, or citizenship.

in the coefficients on highly black zip codes, since blacks are relatively more likely to exhaust benefits than both whites *and* Hispanics.

[17] $(21.9 \text{ percent} - 16.0 \text{ percent})/21.95 = 0.269$.

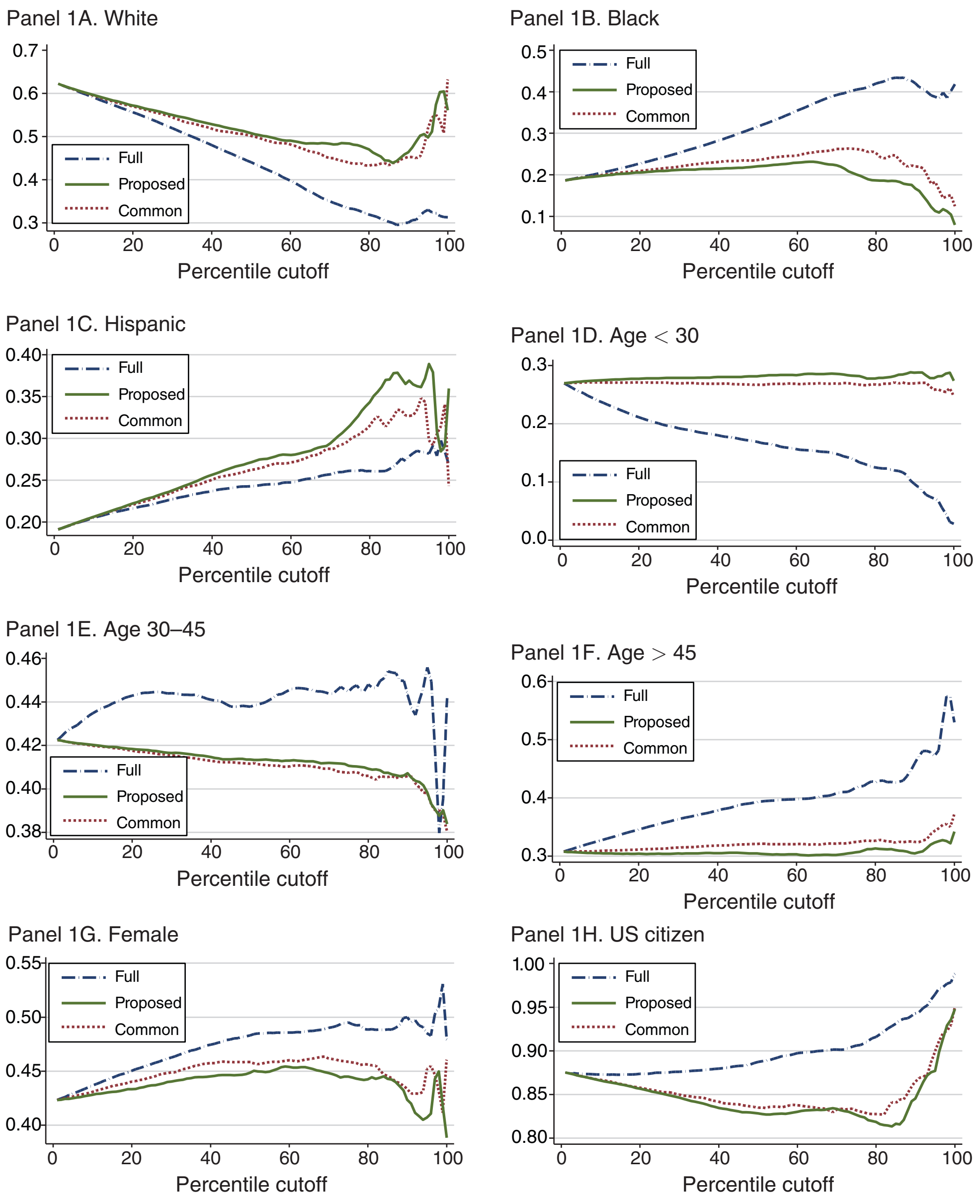

Figure 1. Composition of Reemployment Services Class by Percentile Cutoff of Predicted Exhaustion

Notes: For any predicted exhaustion cutoff, the figures illustrate the composition of the people required to attend a reemployment services class. For instance, if the predicted exhaustion cutoff is at the 80th percentile, the people with the highest 20 percent of predicted exhaustion are required to attend a workshop. The subfigures present a curve for each of the three methods discussed in the text (*full*, *common*, and *proposed*). For example, Panel 1F shows that when the predicted exhaustion cutoff is set at the 60th percentile, using the *full method* 40 percent of the class is older than 45.

All of these results are similar for the 80 percent-cutoff rule. One might worry, however, that the results may vary depending on the chosen cutoff rule. Figure 1 illustrates the breakdowns from Table 3 for any possible assignment rule. For each

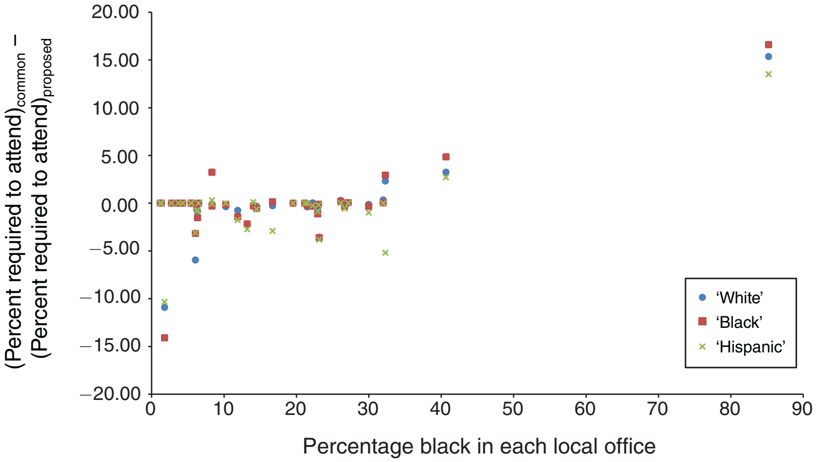

FIGURE 2. DIFFERENCE IN THE PERCENT OF EACH RACIAL GROUP REQUIRED TO ATTEND WORKSHOP BETWEEN THE COMMON AND PROPOSED METHODS BY LOCAL OFFICE (90% CUTOFF RULE)

Notes: For each racial group, the fraction of claimants mandated to attend a workshop in each local office under a 90 percent-cutoff rule was calculated for each profiling method. The *y*-axis gives the difference in this measure between the *common* and the *proposed* methods while the *x*-axis is the percentage of applicants from the local office that are black.

SUP characteristic, the graphs show the fraction of the workshop with that characteristic at each cutoff level. The results highlighted in Table 3 hold across a wide range of potential cutoff rules and again the largest differences between the *proposed* and *common methods* are found across racial categories.

C. *Effects within a SUP Group*

Thus far we have highlighted how implementing the *proposed method* rather than the *common method* might change the composition of a reemployment services workshop. Focusing solely on these results, however, understates the total number of people that would be affected by such a change. The differences between the two methods are manifested not only between, but within SUP groups as well. Because in the *common method* the SAP coefficients are allowed to proxy for the omitted SUPs, claimants who have SAP characteristics that easily identify them as being a member of a particular SUP group are treated differently than otherwise identical claimants. For example, relative to the *proposed method*, under the *common method* black claimants who live in areas with a high black concentration are assigned higher predicted exhaustion levels relative to otherwise similar blacks who live in high white concentration areas. Thus, while the *common method* sends more blacks in general than the *proposed method*, it sends especially more blacks whose SAPs "identify" them as likely being black. The same effects also exist for whites and Hispanics who happen to live in highly concentrated black areas. Figure 2 shows this clear pattern by comparing the difference in the percent of each racial group that is required to attend a workshop under a 90 percent-cutoff rule between the *common* and *proposed methods* across local offices with varying levels of black-claimant

Table 4—Predictive Accuracy of the Different Models

	90 percent cutoff rule				80 percent cutoff rule			
	Full method	Common method	Proposed method	Restricted method (no local office controls)	Full method	Common method	Proposed method	Restricted method (no local office controls)
OLS								
Percent of benefits exhausted by selected claimants	83.9	83.2	83.2	79.3	81.9	81.1	81.0	78.2
Logit								
Percent of benefits exhausted by selected claimants	83.2	82.8	82.6	78.9	81.6	80.8	80.7	77.9
Tobit								
Percent of benefits exhausted by selected claimants	83.8	83.3	83.1	79.2	81.8	81.1	80.9	77.9

Notes: For each cutoff rule and profiling method, the values in the table give the percentage of claimants that actually exhaust all of their benefits. Thus, these percentages indicate the predictive accuracy of each of the models using the two different cutoffs.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

concentration. It is these types of distortions, which could alter economic incentives and choices, that could lead to improved economic efficiency using our proposed method (see the discussion in Section II).

One easy way to see the differences in predictive accuracy between the models is by comparing the adjusted R^2 values in Table 2.[18] Comparing the R^2 values in columns 1–3 indicates that, as expected, the *full method* is more predictive (Adjusted $R^2 = 0.034$) than the more constrained *common method* (Adjusted $R^2 = 0.026$). The *proposed method* (not listed in the table) yields an Adjusted R^2 value of 0.025. Since the *proposed method* is more constrained than the *common method* the R^2 is lower. However, the predictive power of the *proposed method* is significantly better than the model presented in column 3 which eliminates the local office controls (Adjusted $R^2 = 0.009$).

Table 4 illustrates an additional way to view the predictive accuracy of the various methods. Using a 90 and 80 percent cutoff rule, we report the percentage of people who fully exhaust their benefits for the claimants that would be selected using each of the methods discussed in the paper. The first row reports the percentage of selected claimants who exhaust their benefits when using the various methods and OLS. As must be true, the full method has the greatest predictive accuracy. 83.9 percent of the claimants that the *full method* suggested should be sent to reemployment classes (using a 90 percent cutoff) exhausted their UI benefits. The *common* and proposed methods are both significantly less predictive than the *full method*, but are very similar in predictive accuracy to each other—although, as expected, the *common method* is slightly more accurate than the *proposed method*. Consistent with the results in

[18] Comparing adjusted R^2 values provides the same results as comparing sum of squared errors.

Section II, perhaps the most interesting statistic is that if local office controls are eliminated altogether from the model (due to worries about them serving as proxies), predictive accuracy drops substantially (to 79.3 percent). Analyzing predictive accuracy when using logit and tobit models yields similar qualitative results across the different methods.

IV. Conclusion

The prediction method discussed in this paper is quite straightforward, and as we have demonstrated can have a meaningful impact on the way in which different groups are treated by statistical models. We have also shown that the method provides an alternative in situations where discrimination concerns might otherwise lead to bans on the use of a range of variables in statistical models due to their correlation with characteristics like race, gender, age, etc.

While the method has some clear appeal in this regard, there are some concerns with practical implementation. For example, in our application to the WPRS system, information about race, age, and gender were all available. However, in many other cases such as mortgage or insurance markets, data on race and other characteristics may not be collected. Furthermore, a mandate to begin collecting this information has the obvious problem that companies may begin to use these data for predictive purposes rather than for the purpose of eliminating discrimination. Thus, it is worth emphasizing the importance of regulatory and legal institutions to help to ensure that these data are used appropriately. A potential benefit of collecting this information, however, is the increased transparency of this system. With data on race and other characteristics, it would be much easier to understand how and when these SUPs are affecting predictive models. The procedure outlined here would provide a basic statistical framework, and firms could be required to demonstrate that they used these procedures if a discrimination case arose.

An additional complication to this framework exists if there are SAPs which have a direct impact on y_i, but are unobserved to those implementing statistical models. As a somewhat contrived example, imagine that height (a SAP) helps to determine how quickly a worker is hired, and therefore the fraction of UI benefits exhausted. Furthermore assume that there is a correlation between height and (say) race, but that height is unobservable to the modeler. Thus, part of the effect attributed to race in the *full-method* regression would actually be a reflection of race proxying for height. In such a situation, the *common method* allows zip-code factors to (partially) proxy for height, through the correlation between race and zip code. As long as the only variable (SAP or SUP) that was directly correlated to height was race, the *proposed method* would eliminate the predictive power of height rather than the (inappropriate) predictive influence of race.

While this is certainly a valid concern, we would argue that the *proposed method* is likely the more appropriate and cautious approach in the face of concerns about discrimination. For example, in the context of unemployment duration, given the literature on discrimination in labor markets (e.g., Glen G. Cain 1987) it is entirely feasible that race, age, and citizenship status have direct effects on benefit exhaustion. Even if this is not the case, and these variables are proxying for some

Table A1—Simulation of the Effects of Procedures on Predictive Accuracy for the Probit Case

Influence of predictors on data generating process	Correlation between SUP & CP	β_1	β_2	β_3	SSE^{full}	SSE^{common}	$SSE^{restricted}$	$SSE^{proposed}$
1 Equal contributions of SAP, CP, SUP	0	1	1	1	0.11	0.16	0.21	0.16
2 Equal contributions of SAP, CP, SUP	0.5	1	1	1	0.10	0.14	0.22	0.15
3 Equal contributions of SAP, CP, SUP	0.9	1	1	1	0.07	0.09	0.24	0.16
4 SUPs with strong effect	0	1	1	3	0.06	0.22	0.24	0.22
5 SUPs with strong effect	0.5	1	1	3	0.05	0.17	0.24	0.20
6 SUPs with strong effect	0.9	1	1	3	0.03	0.09	0.25	0.21
7 SUPs with very strong effect	0	1	1	10	0.02	0.25	0.25	0.25
8 SUPs with very strong effect	0.5	1	1	10	0.02	0.20	0.25	0.24
9 SUPs with very strong effect	0.9	1	1	10	0.01	0.09	0.25	0.24
10 CPs with strong effect	0	1	3	1	0.06	0.09	0.24	0.09
11 CPs with strong effect	0.5	1	3	1	0.06	0.08	0.24	0.08
12 CPs with strong effect	0.9	1	3	1	0.04	0.06	0.25	0.08
13 CPs with very strong effect	0	1	10	1	0.02	0.03	0.25	0.03
14 CPs with very strong effect	0.5	1	10	1	0.02	0.03	0.25	0.03
15 CPs with very strong effect	0.9	1	10	1	0.02	0.03	0.25	0.03
16 SAPs with strong effect	0	3	1	1	0.06	0.09	0.11	0.09
17 SAPs with strong effect	0.5	3	1	1	0.06	0.09	0.13	0.09
18 SAPs with strong effect	0.9	3	1	1	0.05	0.07	0.18	0.12
19 SAPs with very strong effect	0	10	1	1	0.02	0.03	0.04	0.03
20 SAPs with very strong effect	0.5	10	1	1	0.02	0.03	0.05	0.03
21 SAPs with very strong effect	0.9	10	1	1	0.02	0.03	0.07	0.06

Notes: This table shows the average sum of squared errors for different statistical approaches based on 10,000 simulated observations. See the text in Section II for a description of these simulations and a discussion of the results.

other unobserved characteristic, it is not clear whether that unobserved characteristic would be considered socially acceptable for use if it were observable. Ultimately, it is impossible to prove that the effect of the SUPs on benefit exhaustion is direct, but it is possible to show that the effect is not a direct effect. Thus, in this situation, we would argue that the logical default is to use the *proposed method* and place the burden on the profiler to find additional SAPs if it is believed that the SUPs do not have direct effects. To return to our (contrived) example, if one thinks race proxies for height, the solution is to find data on height.

The debates regarding discrimination in predictive models are often heated and can be confusing to those unfamiliar with statistical methods. One side will argue that the models are estimated without knowledge of race, gender, etc., and are thus nondiscriminatory. The other side will argue that because of correlations, there is still discrimination in the model. Both sides are, in a very real sense, correct. We hope this paper will provide a framework for clarifying some of these debates, and that the method outlined here can provide a reasonable compromise that helps to satisfy demands for fair treatment without drastically reducing the predictive power of statistical models.

REFERENCES

Balducchi, David, ed. 1996. *Worker Profiling and Reemployment Services (WPRS) Systems: National WPRS Colloquium: Selected Papers and Materials.* Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Black, Dan A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel. 2003a. "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System." *American Economic Review*, 93(4): 1313–27.

Black, Dan A., Jeffrey A. Smith, Miana Plesca, and Suzanne Shannon. 2003b. "Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States." U.S. Department of Labor, Employment and Training Administration Report. Washington, DC, March.

Cain, Glen G. 1987. "The Economic Analysis of Labor Market Discrimination: A Survey." In *Handbook of Labor Economics*, Vol 1., ed. Orley C. Ashenfelter and Richard Layard, 693–785. Amsterdam: Elsevier.

Card, David, and Phillip B. Levine. 2000. "Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Extended Benefit Program." *Journal of Public Economics*, 78(1–2): 107–38.

Chan, Jimmy, and Erik Eyster. 2003. "Does Banning Affirmative Action Lower College Student Quality?" *American Economic Review*, 93(3): 858–72.

Federal Trade Commission. 2007. *Credit-Based Insurance Scores: Impacts on Consumers of Automobile Insurance*. Federal Trade Commission Report to Congress. Washington, DC, July. http://www.ftc.gov/os/2007/07/P044804FACTA_Report_Credit-Based_Insurance_Scores.pdf.

Fryer, Roland, Jr., Glenn C. Loury, and Tolga Yuret. 2003. "Color-Blind Affirmative Action." National Bureau of Economic Research Working Paper 10103.

Lundberg, Shelly J. 1991. "The Enforcement of Equal Opportunity Laws under Imperfect Information: Affirmative Action and Alternatives." *Quarterly Journal of Economics*, 106(1): 309–26.

Organisation for Economic Co-Operation and Development (OECD). 1998. *Early Identification of Jobseekers at Risk of Long-term Unemployment: The Role of Profiling*. Paris: OECD.

Pope, Devin G., and Justin R. Sydnor. 2011. "Implementing Anti-Discrimination Policies in Statistical Models: Dataset." *American Economic Journal: Economic Policy.* http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.3.206.

Ross, Stephen L. 2005. "The Continuing Practice and Impact of Discrimination." University of Connecticut, Department of Economics Working Paper 2005-19.

Ross, Stephen, and John Yinger. 2002. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press.

Rothstein, Jesse. Forthcoming. "SAT Scores, High Schools, and Collegiate Performance Predictions." In *Rethinking Admissions for a New Millenium: Moving Past the SAT for Social Diversity and Academic Excellence*, ed. Joseph A. Soares. Teacher's College Press.

Wandner, Stephen A. 1997. "Early Reemployment for Dislocated Workers in the United States." *International Social Security Review*, 50(4): 95–112.

Wandner, Stephen A., and Jon C. Messenger, ed. 1999. *Worker Profiling and Reemployment Services Policy Workgroup: Final Report and Recommendations*. Final Report of the Worker Profiling and Reemployment Services Policy Workgroup. Washington, DC, February.

Yatchew, Adonis, and Zvi Griliches. 1985. "Specification Error in Probit Models." *Review of Economics and Statistics*, 67(1): 134–39.

This article has been cited by:

1. Devin G. Pope, , Justin R. Sydnor. 2011. Implementing Anti-Discrimination Policies in Statistical Profiling ModelsImplementing Anti-Discrimination Policies in Statistical Profiling Models. *American Economic Journal: Economic Policy* **3**:3, 206-231. [Abstract] [View PDF article] [PDF with links]