# Supplemental Appendix:
# Building Non-Discriminatory Algorithms in Selected Data

David Arnold, Will Dobbie, and Peter Hull

## Table of Contents

# A Appendix Figures and Tables

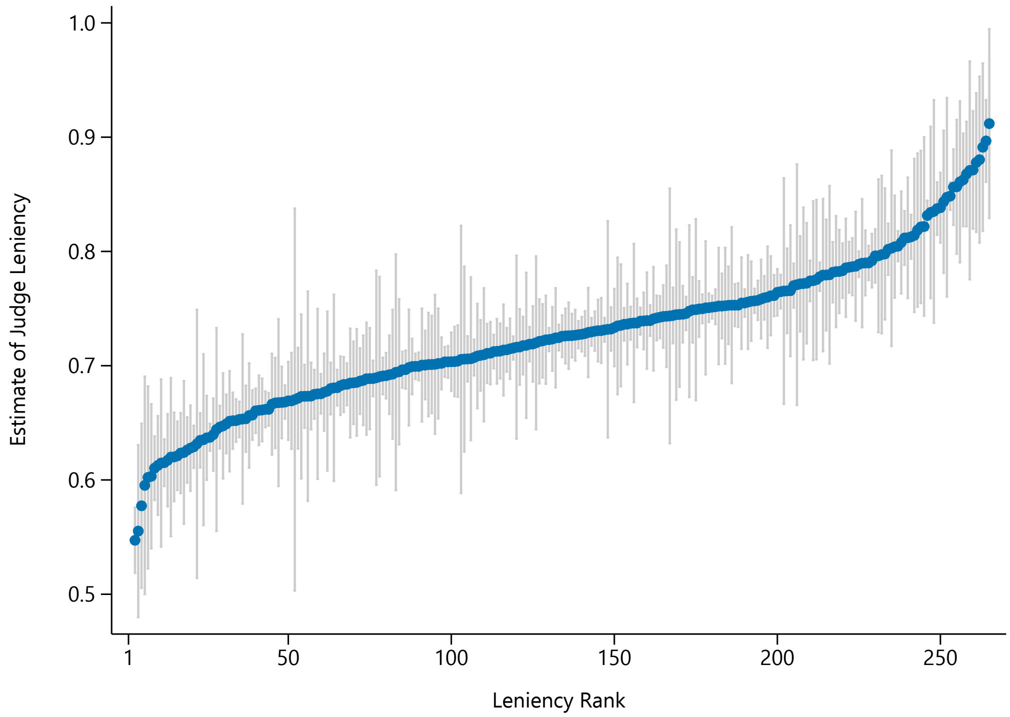
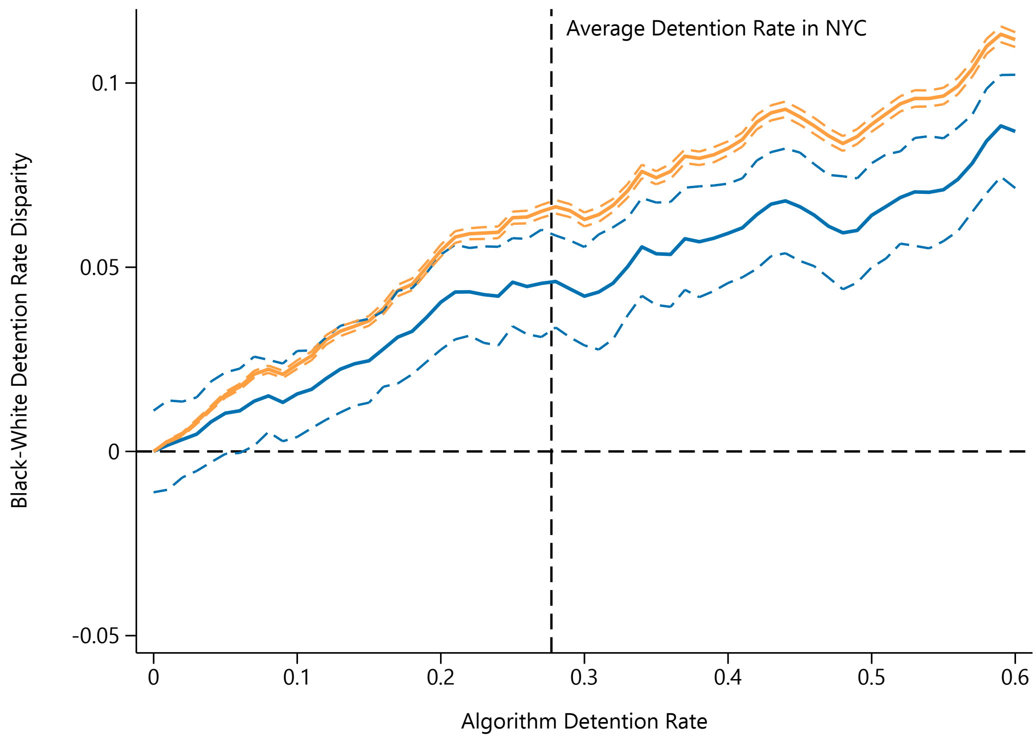
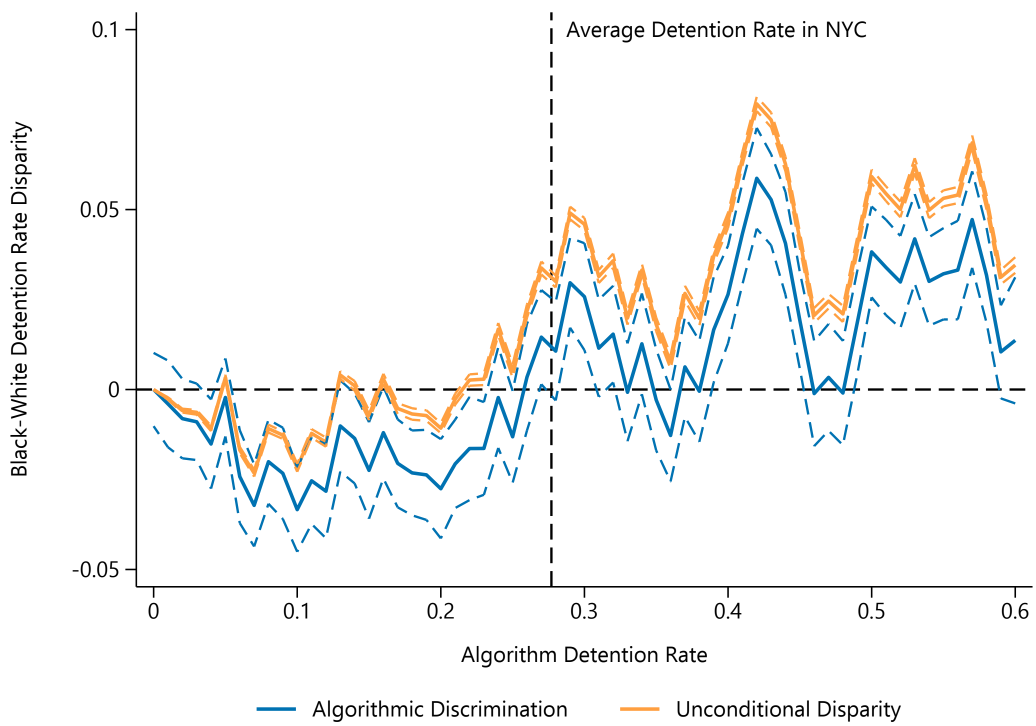



Figure A1: Estimates of Judge Leniency

*Notes.* This figure plots estimates of judge leniency in increasing rank for the 265 judges in our sample. Gray bars indicate pointwise 95% confidence intervals.




A. Conventional Model




B. Pre-Processed Model

Figure A2: Unconditional Disparities and Algorithmic Discrimination in Detention Rates

*Notes.* This figure plots unconditional disparities and algorithmic discrimination in detention rates for various thresholds of detention. In Panel A, we first estimate a conventional model and then rank defendants by predicted risk of misconduct. The unconditional disparity is the difference in detention rates for Black defendants relative to white defendants for a given detention rate in the population. Algorithmic discrimination is defined as the difference in detention rates for Black defendants relative to white defendants, conditional on the same objective misconduct potential. Panel B uses the pre-processed model to rank defendants by predicted risk of misconduct. Dashed lines indicate pointwise 95 percent confidence intervals obtained from a bootstrapping procedure.

Table A1: Tests of Quasi-Random Judge Assignment

| | All Defendants | White Defendants | Black Defendants |
|---|---|---|---|
| | (1) | (2) | (3) |
| Male | 0.00008 | 0.00010 | 0.00007 |
| | (0.00012) | (0.00016) | (0.00016) |
| Age 22 to 39 | -0.00013 | -0.00020 | -0.00006 |
| | (0.00012) | (0.00017) | (0.00017) |
| Age Above 40 | -0.00022 | -0.00028 | -0.00015 |
| | (0.00012) | (0.00019) | (0.00016) |
| Prior Rearrest | -0.00003 | 0.00015 | -0.00018 |
| | (0.00009) | (0.00015) | (0.00012) |
| Prior FTA | -0.00027 | -0.00018 | -0.00032 |
| | (0.00019) | (0.00026) | (0.00025) |
| Any Drug Charge | -0.00007 | -0.00008 | -0.00008 |
| | (0.00013) | (0.00018) | (0.00016) |
| Any DUI Charge | 0.00041 | 0.00046 | 0.00014 |
| | (0.00023) | (0.00026) | (0.00037) |
| Any Violent Charge | 0.00008 | -0.00013 | 0.00022 |
| | (0.00018) | (0.00026) | (0.00019) |
| Any Prior Felony Conviction | -0.00022 | 0.00001 | -0.00038 |
| | (0.00013) | (0.00020) | (0.00015) |
| Any Prior Violent Conviction | -0.00015 | -0.00027 | -0.00007 |
| | (0.00015) | (0.00020) | (0.00019) |
| Any Prior Misdemeanor Conviction | 0.00018 | 0.00013 | 0.00020 |
| | (0.00011) | (0.00014) | (0.00014) |
| Any Property Charge | -0.00029 | -0.00029 | -0.00030 |
| | (0.00015) | (0.00017) | (0.00023) |
| Black | -0.00011 | | |
| | (0.00008) | | |
| Joint p-value | [0.12304] | [0.44736] | [0.09443] |
| Court x Time FE | Yes | Yes | Yes |
| Cases | 567,687 | 270,188 | 297,499 |

*Notes.* This table reports OLS estimates of regressions of judge leniency on various defendant and case characteristics. The regressions are estimated on the sample described in Table 1. Judge leniency is estimated using data from other cases assigned to a given bail judge, following the procedure in Arnold, Dobbie and Hull (2021). All regressions control for court-by-time fixed effects. The p-values reported at the bottom of each column are from $F$-tests for joint significance of the variables listed in the rows. Robust standard errors, two-way clustered at the individual and the judge level, are reported in parentheses.

Table A2: Extensions and Robustness Checks

| | | Alternative Outcomes | | | Alternative Extrapolations | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Baseline | Failure to Appear | Any Rearrest | Social Cost | Linear Extrapolation | Quadratic Extrapolation | By Courtroom | With Judge x Time FE | Released Sample |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Conventional | 0.025 | 0.011 | 0.024 | 243.145 | 0.024 | 0.027 | 0.026 | 0.025 | 0.025 |
| | (0.003) | (0.001) | (0.002) | (20.927) | (0.002) | (0.003) | (0.003) | (0.005) | (0.003) |
| Unselected | 0.027 | 0.012 | 0.027 | 121.354 | 0.021 | 0.042 | 0.025 | 0.029 | 0.025 |
| | (0.009) | (0.006) | (0.008) | (126.719) | (0.004) | (0.014) | (0.010) | (0.016) | (0.003) |
| Pre-Processed | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 |
| | – | – | – | – | – | – | – | – | (0.004) |
| In-Processed | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.004 |
| | – | – | – | – | – | – | – | – | (0.002) |
| Post-Processed | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.003 |
| | – | – | – | – | – | – | – | – | (0.003) |
| Cases | 567,687 | 567,687 | 567,687 | 567,687 | 567,687 | 567,687 | 567,342 | 565,534 | 567,687 |

*Notes.* This table shows the main results from our baseline specification along with various extensions and robustness checks. Column 1 reports algorithmic discrimination in pretrial misconduct, our main outcome of interest, where algorithmic discrimination is the coefficient from regressing predicted pretrial misconduct on a Black indicator, controlling for true misconduct potential. Pretrial misconduct is predicted for all defendants, not only those that are released. Columns 2, 3, and 4 report algorithmic discrimination when using alternative outcomes: FTA, any rearrest, and the social cost of misconduct. We construct our social cost measure as described in Angelova, Dobbie, and Yang (2023). Column 5 reports algorithmic discrimination when using a simple linear extrapolation method, relative to our baseline local linear approach. Column 6 reports algorithmic discrimination when using a quadratic extrapolation method. Column 7 reports algorithmic discrimination when applying the baseline local linear extrapolation separately by borough and averaging together the resulting unselected moment estimates by borough case share. For this analysis we limit the sample to judges who see at least 25 cases in each borough. Column 8 reports algorithmic discrimination as in Column 7 while additionally controlling for judge x year-month fixed effects in the initial borough-specific estimation of Equations (11) and (12). Here, we additionally limit the sample to judges for which judge-specific time effects are identified. Column 9 reports algorithmic discrimination when selection correction is done using only the sample of released defendants. We report bootstrapped standard errors in parentheses. Dashes represent standard errors that are zero by construction. See Table 1 for details on the sample.

# B Data and Setting Appendix

## B.1 Judge Assignment and Decisions in NYC

In the main text, we describe judges as making a binary release or detain decision. However, in practice, judges generally have a few options. They can choose to release-on-recognizance (ROR), in which case the defendant is released without conditions. The judge can also charge a monetary bail that a defendant must pay in order to be released. The bail amount is up to the judge's discretion. The bail amount will be returned if the defendant appears at all future mandated court dates. A defendant may also go through a bail bondsman, who will post bail for a fee. If the defendant is unable or unwilling to pay this fee, then the defendant will remain detained until trial. Finally, a judge can deny the possibility of bail altogether and remand the defendant into custody. Misconduct outcomes are unobserved for both defendants that are remanded and defendants that are unable to pay their monetary bail.

During a case, a judge is presented with a variety of information about the defendant, including details of the arrest and charge. Since 2003, judges have also been given a risk assessment tool that predicts whether a defendant would fail to appear in court. This risk assessment was updated in November 2019 (Luminosity & The University of Chicago's Crime Lab New York, 2020). While the two risk assessments vary in terms of the algorithmic inputs and weights associated with them, both are linear in a small number of characteristics.

Cases are assigned to judges in NYC using a rotation calendar system in each of the five county courthouses, generating quasi-random variation in bail judge assignment for defendants arrested at the same time and in the same place. Each county courthouse employs a supervising judge to determine the schedule that assigns bail judges to the day (9 a.m. to 5 p.m.) and night arraignment shifts (5 p.m. to 1 a.m.) in one or more courtrooms within each courthouse. Individual judges can request to work certain days or shifts but in practice, there is considerable variation in judge assignments within a given arraignment shift, day-of-week, month, and year cell. Our assumption is that within these court-by-time cells (i.e., assigned courtroom, shift, day-of-week, month, and year cells), the judge assigned to a given defendant is randomly selected.

To test this assumption, Appendix Table A1 reports coefficients from an ordinary least squares (OLS) regression of judge leniency on various defendant and case characteristics, controlling for court-by-time fixed effects. We measure leniency using the leave-one-out average release rate among all other defendants assigned to a defendant's judge.[1] Most coefficients in this balance table are small and not statistically significantly different from zero, both overall and by defendant race. A joint $F$-test fails to reject the null of quasi-random assignment at conventional levels of statistical significance.

Appendix Figure A1 verifies that judge assignment meaningfully affects the probability that a defendant is released before trial, with a strong relationship between the predicted leniency of a defendant's judge and the probability of release. First stage regressions show that a one percentage point increase in the predicted leniency of a defendant's judge is associated with a 1.17 percentage point increase in the probability of release, after accounting for court-by-time fixed effects.

[1] Following the standard approach in the literature (e.g., Arnold, Dobbie and Yang, 2018; Dobbie, Goldin and Yang, 2018; Arnold, Dobbie and Hull, 2022), we construct the leave-one-out measure by first regressing pretrial release on court-by-time fixed effects and then using the residuals from this regression to construct the leave-one-out residualized release rate. By first residualizing on court-by-time effects, the leave-one-out measure captures the leniency of a particular judge relative to that of judges assigned to the same court-by-time cells.

### B.2 Sample-Selection Criteria

We make six key restrictions to arrive at our estimation sample, broadly following Arnold, Dobbie and Hull (2022). First, we drop cases where the defendant is not charged with a felony or misdemeanor ($N$=26,057). Second, we drop cases that were disposed at arraignment ($N$=364,051) or adjourned in contemplation of dismissal ($N$=230,517). Third, we drop cases in which the defendant is assigned a $1 cash bail ($N$=1,284). Cash bail is assigned if the defendant is already serving time in jail on an unrelated charge. The $1 cash bail is set so that the defendant receives credit for served time and does not reflect a new judge decision. Fourth, we drop defendants who are non-white and non-Black ($N$=45,529). Fifth, we drop cases for which a defendant received a desk appearance ticket since a desk appearance ticket does not require an arraignment hearing ($N$=76,232). Finally, we drop defendants assigned to judges with fewer than 100 cases ($N$=3,637) and court-by-time cells with fewer than 100 cases or only one unique judge ($N$=143,062), where a court-by-time cell is defined by the assigned courtroom, shift, day-of-week, month and year (e.g., the Wednesday night shift in Courtroom A of the Kings County courthouse in January 2012). The final sample consists of 567,687 cases, 353,422 defendants, and 265 judges. Relative to the full sample of cases, our estimation sample has a somewhat lower release rate, although the ratio of release rates by race is similar. Our estimation sample is also broadly representative in terms of defendant and charge characteristics, with a slightly lower share of defendants with prior FTAs, a slightly higher share of defendants with prior rearrests, and a lower share of defendants charged with drug and property crimes.

## C Econometric Appendix

### C.1 Mean Squared Error Calculation

The mean squared error (MSE) of the predictions of a linear algorithm $\hat{Y}_i = \boldsymbol{X_i'\beta}$ is given by:

$$E[(Y_i^* - \hat{Y}_i)^2] = E[Y_i^{*2}] - 2E[\boldsymbol{X_i'Y_i^*}]\boldsymbol{\beta} + \boldsymbol{\beta'}E[\boldsymbol{X_iX_i'}]\boldsymbol{\beta}, \tag{C1}$$

where the first and second terms are elements of $\boldsymbol{\Theta}$. These are affected by the selective observability of $Y_i^*$ and are estimated as part of our main analysis. The third term is not affected by selective observability of $Y_i^*$ and is directly estimable. We use this formula to compute the MSE of the conventional, unselected , and in-processed models.

The MSE of the pre-processed model predictions $\hat{Y}_i^{Pre} = \boldsymbol{\tilde{X}_i'\tilde{\beta}}$, where $\boldsymbol{\tilde{X}_i} = \boldsymbol{X_i} - \boldsymbol{\Gamma}G_i$, is given by:

$$\begin{aligned} E[(Y_i^* - \hat{Y}_i^{Pre})^2] =& E[Y_i^{*2}] - 2E[\boldsymbol{\tilde{X}_i'Y_i^*}]\boldsymbol{\tilde{\beta}} + \boldsymbol{\tilde{\beta}'}E[\boldsymbol{\tilde{X}_i\tilde{X}_i'}]\boldsymbol{\tilde{\beta}} \\ =& E[Y_i^{*2}] - 2E[\boldsymbol{X_i'Y_i^*}]\boldsymbol{\tilde{\beta}} - 2E[G_iY_i^*]\boldsymbol{\Gamma'\tilde{\beta}} \\ &+ \boldsymbol{\tilde{\beta}'}E[\boldsymbol{X_iX_i'}]\boldsymbol{\tilde{\beta}} - \boldsymbol{\tilde{\beta}'\Gamma}E[\boldsymbol{G_iX_i'}]\boldsymbol{\tilde{\beta}} - \boldsymbol{\tilde{\beta}'}E[\boldsymbol{X_iG_i}]\boldsymbol{\Gamma'\tilde{\beta}} + \boldsymbol{\tilde{\beta}'\Gamma}E[G_i]\boldsymbol{\Gamma'\tilde{\beta}}, \end{aligned} \tag{C2}$$

which is again a function of the elements of $\boldsymbol{\Theta}$ (both directly estimable and selection-affected) and the directly estimable $E[\boldsymbol{X_iX_i'}]$. Finally, the MSE of the post-processed model predictions $\hat{Y}_i^{Post} =$

$\boldsymbol{X_i'}\boldsymbol{\beta^U} - \Delta^U G_i$ is given by:

$$E[(Y_i^* - \hat{Y}_i^{Post})^2] = E[(Y_i^* - \boldsymbol{X_i'}\boldsymbol{\beta^U})^2] + 2\Delta E[G_i(Y_i^* - \boldsymbol{X_i'}\boldsymbol{\beta^U})] + \Delta^2 E[G_i]$$
$$= E[Y_i^{*2}] - E[\boldsymbol{X_i'}\boldsymbol{Y_i^*}]\boldsymbol{\beta^U} + 2\Delta\left(E[G_i Y_i^*] - E[\boldsymbol{X_i'}\boldsymbol{G_i}]\boldsymbol{\beta^U}\right) + \Delta^2 E[G_i], \tag{C3}$$

where we use the fact that $\boldsymbol{\beta^U} = E[\boldsymbol{X_i}\boldsymbol{X_i'}]^{-1}E[\boldsymbol{X_i}\boldsymbol{Y_i^*}]$ to simplify in the second line. This is also a function of the elements of $\boldsymbol{\Theta}$ (both directly estimable and selection-affected).

## C.2 Algorithmic Discrimination in Released-Sample Models

Column 9 of Appendix Table A2 shows estimates of algorithmic discrimination for versions of the pre-processed, in-processed, and post-processed models that use released-sample observations of $Y_i^*$ to adjust the conventional model instead of our baseline quasi-experimental selection-correction approach. We detail these calculations below.

Released-sample pre-processed model predictions are given by $\boldsymbol{\tilde{X}_i^{R\prime}}\boldsymbol{\beta^R}$, where $\tilde{X}_{ik}^R = X_{ik} - \Gamma_k^R G_i$. $\Gamma_k^R$ is the coefficient on $G_i$ from running regression (3) in the $D_i = 1$ subpopulation. The $\boldsymbol{\beta^R}$ coefficient vector is obtained from regressing $Y_i^*$ on the set of $\tilde{X}_{ik}^R$ in the $D_i = 1$ subpopulation. The level of algorithmic discrimination in this model is given by $(\boldsymbol{\Gamma} - \boldsymbol{\Gamma^R})'\boldsymbol{\beta^R}$, where $\boldsymbol{\Gamma}$ collects the coefficients on $G_i$ from the full-population regression (3), estimated using quasi-experimental variation.

Released-sample in-processed model predictions are given by $\boldsymbol{X_i'}\boldsymbol{\beta^{*R}}$, where $\boldsymbol{\beta^{*R}}$ is given by a released-sample version of Equation (6):

$$\boldsymbol{\beta^{*R}} = \boldsymbol{\beta} - E[\boldsymbol{X_i}\boldsymbol{X_i'}]^{-1}\boldsymbol{\Gamma^R}\left(\boldsymbol{\Gamma^{R\prime}}E[\boldsymbol{X_i}\boldsymbol{X_i'}]^{-1}\boldsymbol{\Gamma^R}\right)^{-1}\boldsymbol{\Gamma^{R\prime}}\boldsymbol{\beta}, \tag{C4}$$

which replaces $(\boldsymbol{\beta^U}, \boldsymbol{\Gamma})$ with $(\boldsymbol{\beta}, \boldsymbol{\Gamma^R})$. $\boldsymbol{\beta}$ is the coefficient from regressing $Y_i^*$ on the set of $X_{ik}$ in the $D_i = 1$ subpopulation. The level of algorithmic discrimination in this model is $\boldsymbol{\Gamma'}\boldsymbol{\beta^{*R}}$.

Released-sample post-processed model predictions are given by $\boldsymbol{X_i'}\boldsymbol{\beta} - \Delta^R G_i$, where $\Delta^R = \boldsymbol{\Gamma^R}\boldsymbol{\beta}$ is a released-sample measure of algorithmic discrimination for the conventional model, i.e., the coefficient from regressing $\boldsymbol{X_i'}\boldsymbol{\beta}$ on $G_i$ controlling for $Y_i^*$ in the $D_i = 1$ subpopulation. The level of algorithmic discrimination in this model is given by $\boldsymbol{\Gamma'}\boldsymbol{\beta} - \Delta^R$.