Empirical Analysis of Prediction Mistakes in New York City Pretrial Data *

Ashesh Rambachan[†] Jens Ludwig[‡]

February 8, 2024

1 Introduction

In the New York City pretrial release system, judges decide whether defendants should be detained or released before their trial based on predictions of whether defendants would fail to appear in court (FTA) if released. We know from a large body of behavioral science that this type of prediction exercise is enormously difficult for all people, especially given the difficulty with which human cognition deals with probabilistic thinking, drawing inferences, and making attributions (e.g., Kahneman, 2011). This is therefore an intrinsically challenging task even for judges with their extensive experience on the bench. In this report, we ask: Do judges in New York City make systematic prediction mistakes about FTA risk based on available defendant and case characteristics?

By a "systematic prediction mistake," we mean that a judge's observed pretrial release decisions are driven by mis-estimation of the probability that groups of defendants will FTA given their recorded characteristics, such as basic demographic information and current charge information, not their own preferences about what cases should have bail set. This raises two challenging methodological issues. First, we do not actually know the judge's own preferences. The judge, for example, may be more lenient towards younger defendants or defendants with unique family circumstances. Second, the judge may observe additional information about the defendant that helps them predict FTA risk, but this information is not recorded in the available data. The judge, for example, may learn new information about defendants in the courtroom, but we do not observe

*This data is provided by the New York State Division of Criminal Justice Services (DCJS) and the New York State Office of Court Administration (OCA). The opinions, findings, and conclusions expressed in this publication are those of the authors and not those of DCJS or OCA. Neither New York State nor DCJS or OCA assumes liability for its contents or use thereof. All errors are our own.

[†]MIT, Department of Economics.

[‡]University of Chicago, Harris School of Public Policy and NBER.

these courtroom interactions. Perhaps the judge's pretrial release decisions do not reflect systematic prediction mistakes about FTA risk, but may instead reflect their own preferences or some additional information that is unavailable in the data.

In this report, we address these methodological challenges by applying the econometric framework developed in Rambachan (2022) to investigate whether judges in New York City make systematic prediction mistakes about FTA risk based on defendant and case characteristics. Applied to the New York City pretrial release data analyzed originally in Kleinberg et al. (2018), this econometric framework models judges as making pretrial release decisions in order to maximize their own expected utility using accurate predictions of FTA risk given defendant and case characteristics that we observe in the data as well as some additional case information that is only observed by the judge. If a judge's observed pretrial release decisions are inconsistent with the predictions of this behavioral model, then we conclude that she is making systematic prediction mistakes. In this case, there exists no possible objective nor additional case information that is only available to the judge that would justify their observed choices.

More concretely, a judge is modeled as observing defendant and case characteristics (X_I, X_E, V). The defendant and case characteristics (X_I, X_E) are recorded in the available data, and include information such as a defendant's basic demographic information, current charge information, and prior criminal record. The defendant and case characteristics V summarize all other defendant and case information that is available to the judge at the time of her pretrial release decision, but not recorded in the available data. Based on the defendant and case characteristics (X_I, X_E, V), the judge predicts the FTA risk and then decides whether to release or detain the defendant. The judge's payoffs from releasing or detaining a defendant depend on how she assesses the cost of detaining a defendant that would not FTA relative to relative to the cost of releasing a defendant that would FTA. Our main assumption is that the judge's payoffs may only directly depend on the recorded defendant and case characteristics X_I, but may not directly depend on the other recorded characteristics X_E and the additional characteristics V. As an example, the recorded characteristics X_I may be defendant age, which captures that judges may prefer to be more lenient towards younger defendants, or the severity of the current charge, which captures that judges may be more strict towards defendants charged with more severe crimes. The other recorded characteristics X_E and the additional characteristics V are still relevant in the pretrial release decision as they affect the judge's predictions of FTA risk.

Given the judge's observed pretrial release decisions, we ask: does there exist any payoffs from releasing vs. detaining defendants that vary based on the recorded defendant and case characteristics X_I and any additional information V that could rationalize the judge's choices? If so, then the judge is acting as-if their predictions of FTA risk given the recorded characteristics (X_I, X_E) are accurate. Otherwise, there exists no configuration of payoffs that depend on the recorded characteristics X_I and additional information V such that the judge's observed pretrial release decisions could have been optimal at accurate predictions of FTA risk given the recorded characteristics (X_I, X_E). For this reason, we refer to such inconsistency with this behavioral model as evidence of a "systematic prediction mistake" based on the recorded characteristics (X_I, X_E) (Rambachan, 2022).

As mentioned, the main assumption required in this analysis is that only the recorded characteristics X_I (e.g., defendant age or charge severity) directly affect the judge's payoffs from releasing vs. detaining defendants. The other recorded characteristics X_E and the additional information V only affect the judge's predictions of FTA risk. This is a necessary "preference exclusion restriction" – without this assumption, any pretrial release decisions could be rationalized at accurate predictions of FTA risk given preferences that flexibly vary based on all defendant and case information (Rambachan, 2022). If this preference exclusion assumption is violated in practice, then rejections of this behavioral model could arise due to mis-specification of the judge's preferences.[1] Therefore, we must carefully interpret our results as testing the joint hypothesis that the judge makes systematic prediction mistakes and her payoffs satisfy the conjectured preference exclusion restriction.

To address this concern about the sensitivity of our analysis to this critical assumption, we emphasize two key features of our analysis. First, we test whether judges make systematic prediction mistakes under a range of assumptions about what recorded characteristics X_I directly affect their payoffs from releasing vs. detaining the defendants. This allows us to assess the robustness of our conclusions about systematic prediction mistakes to varying assumptions on how judge's payoffs may depend on recorded defendant and case characteristics. Second, the econometric framework asks whether there exists any additional information V that could rationalize the judge's choices. Since this additional information is by definition unobservable to us, we place as weak as possible assumptions on it. In other words, we apply an econometric framework that makes it as likely as possible that we are truly capturing what the judges themselves are optimizing rather than some artifact of our own modeling assumptions, and thereby generate insights that can be as useful as possible for informing judges and other interested parties.

By applying this econometric framework, we ask and answer several interesting questions under a range of assumptions, such as: what fraction of judges in New York City make systematic prediction mistakes about FTA risk given defendant and case characteristics? What types of sys-

[1] As an example, a defendant's particular family circumstances (e.g., whether the defendant is married) is not recorded in the data and judge's may learn this information in the courtroom. It is natural to imagine that a judge's payoffs from releasing vs. detaining defendants may depend on these circumstances. This concern would threaten our conclusions about systematic prediction mistakes if these unrecorded family circumstances simultaneously made the defendant more likely to FTA and raised the judge's relative costs of detention (or, conversely, made the defendant less likely to FTA and lowered the judge's relative costs of detention).

tematic prediction mistakes of FTA risk are being made by judges? How do these vary across judges? These questions are relevant as judges are increasingly given statistical risk assessments and recommendations in pretrial release systems across the United States. Answering these questions helps reveal exactly when and why these data tools would be particularly helpful in assisting judges to identify low risk and high risk defendants, and thereby improve pretrial release outcomes.

Among the top 25 judges that heard the most cases from November 1, 2008 to November 1, 2013, we find that at least 20% make systematic prediction mistakes about FTA risk based on defendant and case characteristics. Furthermore, we find that these systematic prediction mistakes occur because judges underreact to variation in FTA risk across defendants. That is, judges fail to differentiate between defendants that would are predicted to be low FTA risk based on observable characteristics and defendants that are predicted to be high FTA risk based on observable characteristics. These findings are robust to an array of assumptions about judges' preferences and extensions to our baseline empirical implementation. We again emphasize that these results must be interpreted as evidence of systematic prediction mistakes given our assumptions on preferences, and the limitations of the available data.

Based on these empirical findings, we document the effects introducing a statistical pretrial release rule in pretrial release decisions. In practice, judges will of course be the final deciders for pretrial hearings; these data tools are not intended to substitute for judges but rather to assist them. But, in order understand how helpful they could be, we compare what would happen under status quo judge decisions with what would happen if a judge followed the risk tool's recommendations with 100% fidelity as a thought exercise. We find that, under a range of assumptions on the social tradeoff between FTA risk and the cost of pretrial detention, replacing judges who make systematic prediction mistakes with a statistical decision rule weakly dominates the status quo. This suggests that these data tools may be particularly helpful on cases where we found that judges make systematic prediction mistakes.

Finally, we emphasize that our findings are based on cases heard in New York City from November 2008 to November 2013. In 2019, New York implemented significant reforms to its pretrial release system that significantly changed the conditions under which judges can set bail. Understanding the extent to which these results generalize to the post-reform practices in the New York City pretrial release system is an important direction for future research.

Roadmap: The rest of the report is organized as follows. Section 2 discusses the New York City pretrial data, sample restrictions and descriptive statistics. Section 3 describes our baseline empirical implementation of the general econometric framework developed in Rambachan (2022), the results under our baseline empirical implementation and the results from various extensions. Section 4 analyzes the effects of replacing judges with statistical decision rules. Section 5 briefly concludes.

2 New York City Pretrial Data

We analyze data on the New York City pretrial release system. We discuss the sample restrictions that are applied to construct the main estimation sample and provide descriptive statistics.

2.1 Sample Restrictions

We initially observe case information associated with all arrests made in New York City between November 1, 2008 and November 1, 2013. There are 1,460,462 cases in this initial universe.

Of these cases, only 758,027 cases were actually subject to a pretrial release decision. We construct the universe of cases that were subject to a pretrial release decision by following Kleinberg et al. (2018) and removing (i) cases involving desk appearance tickets, (ii) cases that were disposed of at arraignment, (iii) cases that were adjourned in contemplation of dismissal, and (iv) duplicate cases.

We apply further sample restrictions to the universe of cases that were actually subject to a pretrial release decision in order to construct the main estimation sample. We exclude (i) cases involving non-white and non-black defendants; (ii) cases that were assigned to judges that heard fewer than 100 cases over the sample period; (iii) cases that were heard in court-by-time cells (which are defined at courtroom by shift by day of week by month by year level) in which there were fewer than 100 cases or only one unique judge that heard cases. The main estimation sample consists of 569,265 cases heard by 265 unique judges. Appendix Tables A4-A5 compare the descriptive statistics for the main estimation sample against descriptive statistics for the initial universe of 758,027 cases that were subject to a pretrial release decision.

The empirical analysis focuses on the top 25 judges that heard the most cases in the main estimation sample from November 1, 2008 to November 1, 2013. These top 25 judges heard 243,118 cases in the main estimation sample, and each judge heard at least 5,000 cases (see Appendix Figure A1).

2.2 Descriptive Statistics

For each case in the main estimation sample, we observe: demographic information about the defendant such as their race, gender and age; detailed information about the current charge filed against them (e.g., the number of charge, the top charge and the severity of the top charge); detailed information about the defendant's prior criminal record (e.g., does the defendant have any prior misdemeanor arrests? felony arrests? misdemeanor convictions? felony convictions?); the defendant's prior record of pretrial misconduct (e.g., has the defendant ever failed to appear in court?). Importantly, we observe an identifier for the judge assigned to each case, and whether that judge decided to detain or release the defendant. We further observe whether the defendant failed to appear in court (FTA) or was re-arrested for a new crime if the judge decided to release them.

Appendix Table A1 provides a series of descriptive statistics about the defendant and case characteristics for both the main estimation sample and the cases heard by the top 25 judges. These descriptive statistics are broken out by the race of the defendant. Appendix Table A2 reports the same descriptive statistics for the main estimation sample and the cases heard by the top 25 judges broken out by whether the defendant was released or detained. Appendix Table A3 summarizes the observed pretrial misconduct rates among released defendants, broken out by the race of the defendant (i.e., what fraction of released defendants in the main estimation sample and cases heard by the top 25 judges FTA'd upon release?).

3 Estimates of Prediction Mistakes in New York City Pretrial Data

In this section, we empirically apply the general econometric framework developed in Rambachan (2022) to test whether judges make systematic prediction mistakes of FTA risk based on defendant characteristics in their pretrial release decisions. We report the fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs under a range of assumptions about what defendant characteristics directly affect their preferences. We further estimate whether judges overreact or underreact to variation in FTA risk based on defendant characteristics as well as the types of cases that systematic prediction mistakes are made on.

3.1 Estimating the Fraction of Judges that Make Prediction Mistakes

We will estimate the fraction of judges that make systematic prediction mistakes of FTA risk based on defendant characteristics in their pretrial release decisions. As mentioned, we estimate this quantity under various assumptions on judges' preferences, either assuming that (i) no defendant characteristics, (ii) only the defendant's race, (iii) only the defendant's race and age, or (iv) only the defendant's race and whether the defendant was charged with a felony offense directly affects the judges' utility function over releasing defendants. In our analysis, we discretize age into young and older defendants, where older defendants are those older than 25 years.

3.1.1 Baseline Empirical Implementation

Predicting FTA Risk Among Released Defendants: First, we construct a supervised machine learning prediction function to predict FTA among defendants released by the remaining held out judges. For each choice of the directly payoff relevant characteristics (either defendant race by age cells or defendant race by charge severity cells), we predict FTA among defendants released by the remaining held out judges by averaging the predictions of an estimated elastic net model and an estimated random forest. The elastic net model is tuned using three-fold cross-validation, and the random forest is estimated using the `R` package `ranger` (Wright and Ziegler, 2017). Appendix Figure A2 plots the ROC curves for the estimated prediction function when evaluated over defendants released by the top 25 judges.

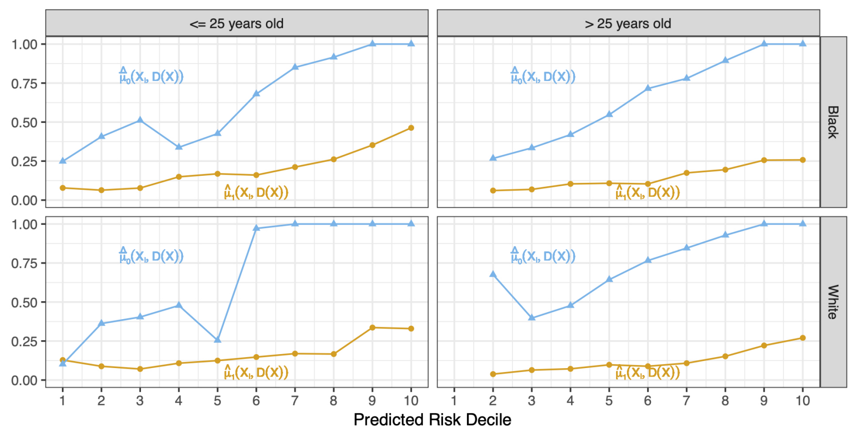
We use the constructed prediction functions to define partitions of the remaining defendant and case characteristics that are excluded from the judges' preferences within each possible value of the directly payoff-relevant characteristics. To do so, we construct the observed distribution of predicted FTA risk among defendants released by the remaining held out judges, and define deciles of predicted FTA risk with respect to this distribution. The partition maps predicted FTA risk among defendants released by the top 25 judges into deciles of predicted risk through this observed distribution. Using the general dimension reduction strategy for testing for systematic prediction mistakes in Rambachan (2022), we will test for violations of implied revealed preference inequalities for each judge in the top 25 across these constructed deciles of predicted risk. This step is useful empirically since otherwise the number of cases per characteristic cell is extremely small. If we were to discretize all demographic information, all current charge information and all prior criminal record information about defendants into binary values, there would be 134,062 unique characteristic cells with only 4.24 cases per characteristic cell on average.

Verifying Quasi-Random Assignment of Judges to Cases: Previous academic research such as Leslie and Pope (2017); Kleinberg et al. (2018); Arnold, Dobbie and Hull (2022) have shown that judges in New York City are as-if randomly assigned to cases within court-by-time cells defined at courtroom by shift by day of week by month by year level. We verify that this is also true in our main estimation sample through balance checks. We measure judge leniency on a case using the leave-one-out release rate among all other cases assigned to a particular judge (Dobbie, Goldin and Yang, 2018; Arnold, Dobbie and Yang, 2018; Arnold, Dobbie and Hull, 2022). Our balance checks regress judge leniency on defendant characteristics and court-by-time fixed effects. The results from these balance checks are reported in Appendix Tables A6-A8.

Bounding FTA Risk Among Detained Defendants: We use the quasi-random assignment of judges to cases in New York City to construct bounds on the FTA rate among defendants detained by each judge in the top 25.

First, we use the constructed leniency measure to group judges in the main estimation sample into "leniency quintiles." Second, we define the instrument to be the leniency quintile of the judge that was assigned to a particular case, and based on our previous balance checks we assume that this instrument is quasi-randomly assigned conditional on the court-by-time cells. Third, applying the general results from Rambachan (2022) to this setting, an upper bound on the FTA rate among detained defendants can be constructed by estimating the probability a defendant is released and FTA'd in a given cell of directly payoff-relevant characteristics and predicted FTA risk decile (adjusted for the court-by-time cells) and the probability a defendant is detained in a given cell of directly payoff-relevant characteristics and predicted FTA risk decile (adjusted for the court-by-time cells). We estimate these quantities by estimating specifications that regress (i) whether a defendant was detained and (ii) whether a defendant was released and FTA'd on saturated indicators for the directly payoff-relevant characteristic, predicted FTA risk decile and leniency quintile interaction and court-by-time fixed effects. We then use these estimated regressions to construct these quantities.

Figure 1: Observed failure to appear rate among released defendants and constructed bound on the failure to appear rate among detained defendants by race-and-age cells for one judge in New York City.



Notes: This figure plots the observed failure to appear rate among released defendants (orange) and the bounds on the failure to appear rate among detained defendants based on the judge leniency instrument (blue) at each decile of predicted risk and race-by-age cell for the judge that heard the most cases in the main estimation sample. See Section 3.1.1 for the estimation details.

Figure 1 plots the observed FTA rate among defendants released by the judge that heard the most cases in the sample period and the estimated bounds on the FTA rate among detained defendants using the most lenient quintile of judges over each race-by-age cell. Appendix Figure A3 plots the analogous quantities for each race-by-felony charge cell.

Implementation of Moment Inequality Test: We test the implied revealed preference inequalities derived in Rambachan (2022) over the deciles of predicted risk using the estimated bounds on the FTA rate among detained defendants. These implied revealed preference inequalities are a system of moment inequalities. We test these moment inequalities by first estimating the variance-covariance matrix of the observed FTA rate among released defendants and the estimated bounds on the FTA rate among detained defendants using the empirical bootstrap conditional on the payoff-relevant characteristics, the predicted FTA risk decile and the leniency quintile instrument. With the estimated moments and the estimated variance-covariance matrix, we will test the system of moment inequalities using the conditional least-favorable hybrid test developed in Andrews, Roth and Pakes (2023) and further analyzed in Rambachan and Roth (2022).

3.1.2 Results Under Baseline Empirical Implementation

Table 1: Fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk given defendant characteristics.

	Utility Functions			
	No Characteristics	Race	Race + Age	Race + Felony Charge
Unadjusted Rejection Rate	48%	48%	48%	56%
Adjusted Rejection Rate	24%	24%	20%	32%

Notes: This table summarizes the fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk given defendant characteristics and strict preference utility functions that either (i) do not depend on any characteristics, (ii) depend on the defendant's race, (iii) depend on both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. The unadjusted rejection rate reports the fraction of judges in the top 25 whose pretrial release decisions violate the moment inequalities derived in Rambachan (2022) at the 5% level. The adjusted rejection rate reports the fraction of rejections of the moment inequalities after applying the Holm-Bonferroni step down procedure for multiple hypothesis testing correction.

Under a range of alternative assumptions about what defendant characteristics directly affect their preferences, Table 1 reports the fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk given defendant characteristics. After applying a multiple hypothesis testing correction procedure, we estimate that the pretrial release decisions of at least 20% of judges violate the implied revealed preference inequalities.

3.1.3 Robustness of and Extensions to Baseline Empirical Implementation

We now investigate the robustness of our empirical findings under the baseline empirical implementation to several extensions.

Direct Imputation Bounds on FTA Risk Among Detained Defendants: Our baseline empirical implementation constructed bounds on the FTA rate among detained defendants using the quasi-random assignment of judges. To assess the robustness of our findings to this choice, we now construct bounds on the FTA rate among detained defendants using the FTA rate among released defendants. Rambachan (2022) calls this procedure, "direct imputation," and provides a formal econometric analysis of its properties.

To use direct imputation, we must specify a parameter κ, which specifies how much larger the FTA risk among detained defendants may be than the observed FTA risk among released defendants. We report results for $\kappa \in \{0, \ldots, 10\}$, where $\kappa = 0$ means that detained defendants have the same FTA risk as released defendants and $\kappa = 5$, for example, means that the FTA risk of detained defendants may be at most five times larger than the FTA risk of released defendants. Aside from changing the choice of bounds on the FTA risk among detained defendants, the rest of our empirical implementation is unchanged from the baseline empirical implementation. Appendix Figure A4 shows that a large fraction of judges' pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk given defendant characteristics under various assumptions on their preferences and a range of values $\kappa \in \{0, \ldots, 10\}$.

Release on Recognizance vs. Set Bail Conditions: The baseline empirical implementation ignored the fact that judges in New York City set different bail conditions and levels by defining the decision as a binary choice to either release or detain defendants. By instead defining the decision as either "release on recognizance" (i.e., defendant is released without any bail conditions) or "set bail conditions" (i.e., set either a secured or an unsecured bail amount), we may extend our analysis to test whether judges make systematic prediction mistakes about the ability of defendants to meet their bail conditions and failure to appear risk. The results of this extension are reported in Appendix Table A10. The findings are similar to our baseline empirical implementation.

Any Pretrial Misconduct Outcome Definition: The baseline empirical implementation defined the outcome as only whether defendants would FTA if they were released. To assess the robustness of our findings to this particular choice of outcome, we modify our baseline empirical implementation to now define the outcome as whether the defendant would either FTA or be re-arrested for a new crime (which we call an "any pretrial misconduct" outcome definition). We otherwise follow the same steps in our baseline empirical implementation by first predicting "any pretrial misconduct" among released defendants, and second bounding "any pretrial misconduct" risk among detained defendants using the quasi-random assignment of judges. The results from this robustness exercise are reported in Appendix Table A9. We find that at least 64% of judges in New York City make pretrial release decisions that are inconsistent with expected utility maximization behavior at accurate beliefs about any pretrial misconduct given defendant characteristics.

Multiple Misconduct Outcome Definition: The empirical implementations so far only considered a scalar outcome (either whether defendants would FTA or commit any pretrial misconduct if released). Of course, judges may separately consider multiple outcomes such as whether a defendant would FTA, be re-arrested for a non-violent crime or re-arrested for a violent crime (i.e., a violent felony offense, murder, rape, or robbery). We modify our baseline empirical implementation to now be this vector of outcomes, and again bounding the risk of these outcomes using the quasi-random assignment of judges. The results from this robustness exercise are reported in Appendix Table A11. We find that at least 20% of judges in New York City make pretrial release decisions that are inconsistent with expected utility maximization at accurate beliefs about failure to appear risk, non-violent crime risk, and violent crime risk given defendant characteristics.

3.2 Estimates of whether Judges Overreact or Underreact to FTA Risk

In this section, we additionally apply the general econometric framework developed in Rambachan (2022) to test whether judges' beliefs about FTA risk given defendant characteristics either overreact or underreact to true underlying variation in FTA risk based on defendant characteristics.

3.2.1 Baseline Empirical Implementation

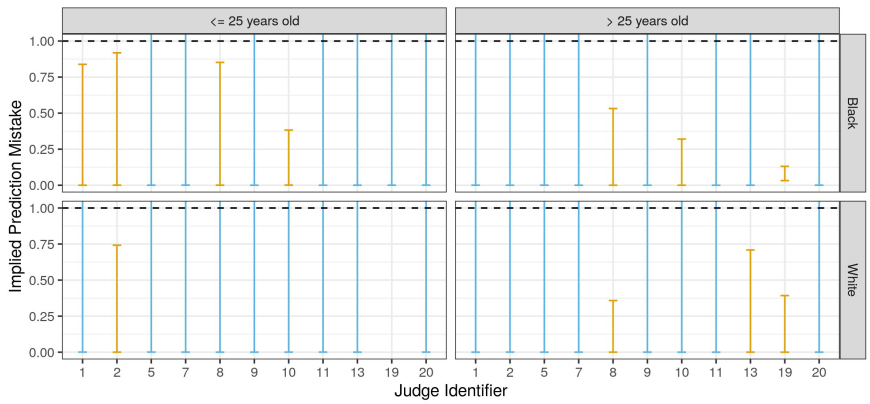
We now describe our baseline empirical implementation to estimate whether judges overreact or underreact to variation in FTA risk based on defendant characteristics.

Applied to the New York City pretrial release setting, the general results in Rambachan (2022) imply an identified set on the extent to which judges overreact or underreact to variation in FTA risk based on defendant characteristics. We first characterize an identified set for the judges' reweighted utility threshold, which depends on the FTA rate among released defendants and the bound on the FTA rate among detained defendants. The reweighted utility threshold summarizes both a judge's preferences as well as the judge's incorrect beliefs about FTA risk. By comparing the reweighed utility threshold across cells of defendant and case characteristics, Rambachan (2022) shows that we can infer the extent to which the judge's beliefs over-react or under-react to variation in true FTA risk across these observable characteristics.

To apply these results, we construct a 95% joint confidence set for the reweighted utility threshold between the bottom and top deciles of the predicted FTA risk distribution for each judge who was found to make systematic prediction mistakes about FTA risk whose choices were found to be inconsistent with expected utility maximization behavior at accurate beliefs. We do so by test inversion using the conditional-least favorable test developed in Andrews, Roth and Pakes (2023) as before. For each value of the reweighed utility thresholds that lie in the constructed confidence interval, we then compute the implied prediction mistake between the bottom and top deciles of the predicted FTA risk distribution.

3.2.2 Results under Baseline Empirical Implementation

Figure 2: 95% confidence intervals for the identified set of systematic prediction mistakes of FTA risk between lowest and highest predicted FTA risk deciles made by judges within each race-by-age cell.



Notes: This figure plots the 95% confidence interval for the identified set on each judge's prediction mistake between the lowest predicted risk decile and the highest predicted risk decile for the race-by-age directly payoff relevant characteristics. The estimated quantity measures the degree to which judges' beliefs about FTA risk underreact or overreact to actual variation in failure to appear risk. See Section 3.2 for discussion..

The estimates for race-and-age cells are summarized in Figure 2. When informative, we find that the estimated confidence intervals lie strictly below one, meaning that these judges' implied beliefs about FTA risk underreact to predictable variation in FTA risk between the lowest predicted risk decile and highest predicted risk decile. Analogous estimates for race-and-felony charge cells are summarized in Appendix Figure A5.

3.2.3 Robustness of and Extensions to Baseline Empirical Implementation

We now investigate the robustness of these empirical findings under the baseline empirical implementation to two extensions.

Direct Imputation Bounds on FTA Risk Among Detained Defendants: We report 95% confidence sets for each judge's implied prediction mistake on FTA risk between the bottom and top deciles of predicted FTA risk using bounds on the FTA rate of detained defendants using direct imputation with $\kappa = 2$. Otherwise, our empirical implementation remains the same. These results are plotted in Appendix Figure A6, and we again find that judges are underreacting to variation in FTA risk based on defendant characteristics.

Any Pretrial Misconduct Outcome Definition: We report 95% confidence sets for the judge's implied prediction mistake on "any pretrial misconduct" risk between the bottom and top deciles of predicted "any pretrial misconduct" risk using bounds constructed through the quasi-random assignment of judges. In other words, we modify our baseline empirical implementation by redefining the outcome to "any pretrial misconduct." These results are reported in Appendix Figure A7, and again our findings are unchanged.

3.3 Estimating the Location of the Largest Violations of Revealed Preference Inequalities

Table 2: Location of the maximum studentized violation of revealed preference inequalities among judges whose release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk given defendant characteristics.

	Utility Functions	
	Race and Age	Race and Felony Charge
Unadjusted Rejection Rate	48%	56%
White Defendants		
Middle Deciles	0%	0%
Tail Deciles	25%	7.14%
Black Defendants		
Middle Deciles	0%	0%
Tail Deciles	75%	92.85%

Notes: This table summarizes the location of the maximum studentized violation of revealed preference inequalities among judges whose release decisions are inconsistent with expected utility maximization behavior at accurate beliefs and preferences that depend on both the defendant's race and age as well as the defendant's race and whether the defendant was charged with a felony. Among judge's whose release decision violate the revealed preference inequalities at the 5% level, this table reports the fraction of judges for whom the maximal studentized violation occurs among white and black defendants on tail deciles (deciles 1-2, 9-10) and middle deciles (3-8).

Lastly, we estimate the cells of defendants on which cases the largest violations of the implied revealed preference inequalities occur. Among judges whose choices are inconsistent with expected utility maximization behavior at accurate beliefs, Table 2 reports the fraction of judges for whom the maximal studentized violation occurs over the deciles 1-2 and 9-10 of the predicted FTA risk distribution or deciles 3-8 of the predicted FTA risk distribution for black and white defendants. All of the largest studentized violations of the implied revealed preference inequalities occur over defendants whose observable characteristics place them in either deciles 1-2 or deciles 9-10 of the predicted FTA risk distribution. The majority of the largest studentized violations of the implied revealed preference inequalities also occur on cases involving black defendants. Intriguingly, we find that all judges' pretrial release decisions satisfy the implied revealed preference inequalities if we focus on only testing them over the defendants in the middle of the predicted risk distribution.

Robustness to Any Pretrial Misconduct Outcome Definition: In Appendix Table A12, we report analogous results when the outcome is instead defined to be "any pretrial misconduct" (i.e., FTA or be re-arrested for a new crime). We find similar results.

3.4 Estimates of the Cost and Share of Prediction Mistakes to FTA risk

In this section, we additionally apply the general econometric framework developed in Rambachan (2022) to bound how costly and how common are prediction mistakes about failure to appear risk. Applied to the New York City pretrial release setting, these general results imply that the magnitudes of the violations of the implied revealed preference inequalities can be used to calculate a bound on how costly are the prediction mistakes about failure to appear risk to judges and how common are they over judges' pretrial release decisions. Concretely, to estimate a bound on how costly are the prediction mistakes about failure to appear risk, we calculate the optimal value of a linear program using the violations of the implied revealed preference inequalities as a set of constraints. To estimate a bound on the share of prediction mistakes over a judges' decisions, we calculate the optimal value of a mixed integer linear program using the violations of the implied revealed preference inequalities as a set of constraints.

Table 3: Bound on the share of prediction mistakes about failure to appear risk across judges.

	Utility Functions $u(c, y^*; x_I)$	
	Race and Age	Race and Felony Charge
Unadjusted Rejection Rate	48%	56%
Prediction Mistake Share		
Minimum	6.30%	11.88%
Median	30.87%	24.45%
Maximum	42.26%	45.78%

Notes: This table calculates the estimated bound on the share of prediction mistakes among judges who were found to make prediction mistakes about failure to appear risk. Among judges' whose release decisions violate the revealed preference inequalities at the nominal 5% level, this is calculate by constructing the optimal value of a mixed-integer linear program over the violations of the implied revealed preference inequalities.

For each judge whose choices violated the implied revealed preference inequalities at the nominal 5% level, Table 3 summarizes the minimum, median, and maximum calculated bound on the share of prediction mistakes. We next calculate how costly are prediction mistakes for each judge whose choices violated the implied revealed preference inequalities at the nominal 5% level. By calculating the optimal value of the mentioned linear program, we found that the median judge's prediction mistakes lead to a 0.073 expected utility cost when both defendant race and age are allowed to affect utility, and a 0.113 expected utility cost when both defendant race and charge severity are allowed to directly affect utility. For the median judge, these expected utility costs correspond to a 9.92 percentage point reduction in the fraction of released defendants that would fail to appear in court when we allow both defendant race and age to affect utility, and an equivalent 12.1 percentage point reduction when both defendant race and charge severity are allowed to directly affect utility.

Robustness to Any Pretrial Misconduct Definition: In Appendix Table A13, we report analogous results on the share of prediction mistakes when the outcome is defined to be "any pretrial misconduct" (i.e., FTA or be re-arrested for a new crime). We find similar results. We find that the median judges' systematic prediction mistakes lead to a 0.252 expected utility cost when both defendant race and age directly affects utility and a 0.209 expected utility cost when both race and charge severity directly affect utility. This corresponds to an equivalent 25.06 percentage point reduction in the fraction of defendants that are released and would commit pretrial misconduct when both defendant race and age are allowed to directly affect utility. and 25.90 percentage point reduction when defendant race and charge severity are allowed to directly affect utility.

Robustness to Direct Imputation Bounds on FTA Risk Among Detained Defendants: In Appendix Figure A8, we report analogous results on the share of prediction mistakes about FTA risk using bounds on the FTA rate of detained defendants based on direct imputation with $\kappa \in \{0, 1, \ldots, 10\}$. In Appendix Figure A9, we also report analogous results on the expected utility cost of prediction mistakes using bounds on the FTA rate of detained defendants based on direct imputation with $\kappa \in \{0, 1, \ldots, 10\}$. We find similar results.

4 Estimates of the Effect of Statistical Release Rules

In this section, we apply the general econometric framework developed in Rambachan (2022) to estimate the effects of replacing judges with a statistical decision rule. This exercise is motivated by the earlier empirical findings on systematic prediction mistakes of FTA risk based on defendant characteristics.

4.1 Baseline Empirical Implementation

We now describe our baseline empirical implementation to estimate the effect of replacing judges with a statistical decision rule. To do so, we discuss how we may conduct inference on expected social welfare under a given statistical decision rule and then discuss how we construct such a statistical decision rule.

Conducting Inference on Expected Social Welfare Under a Known Decision Rule: We consider a policymaker who has preferences over whether a defendant is released and would FTA and whether a defendant is detained but would not FTA. This is called the "social welfare function." The policymaker evaluates a candidate statistical release rule that specifies the fraction of defendants at each observable characteristic that would be released based on the historical data. Because we do not observe the FTA rate among detained defendants, average social welfare under the statistical release rule is not point identified. However, Rambachan (2022) shows that average social welfare under the statistical release rule is partially identified and characterizes the identified set as a system of moment inequalities with nuisance parameters that enter linearly.

We will use these results to conduct inference on average social welfare under a constructed statistical decision rule (discussed below). We construct bounds on the FTA rate among detained defendants using the quasi-random assignment of judges to cases as before. We construct 95% confidence intervals for expected social welfare under the algorithmic decision rule and the judge's observed released decisions using the conditional least-favorable hybrid test for moment inequality models with linear nuisance parameters developed in Andrews, Roth and Pakes (2023).

Constructing the Statistical Decision Rule: Following the general analysis in Rambachan (2022), we again consider a policymaker with social welfare function given that summarizes the social cost of detaining a defendant that would not FTA and the social cost of releasing a defendant that would FTA. We construct a statistical decision rule for pretrial release decisions assuming that the policymaker make pretrial release decisions herself (i.e., fully removed the discretion of judges and automated pretrial release decisions). The statistical release rule that the policymaker selects is a threshold rule that decides whether to release or detain defendants based on the constructed bounds on the FTA rate given defendant characteristics. In our empirical analysis, we estimate expected social welfare of the judges' observed pretrial release decisions against an estimated version of this statistical decision rule. The statistical decision rule is based on constructed bounds on the FTA rate given defendant characteristics, where these bounds are constructed using the quasi-random assignment of judges to cases that we discussed earlier.

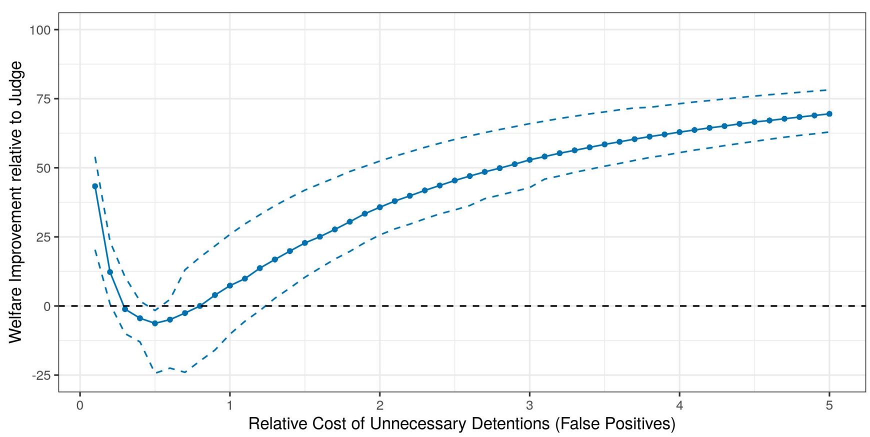
4.2 Results Under Baseline Empirical Implementation

As mentioned, we compare expected social welfare under the estimated statistical decision rule against the observed pretrial release decisions of judges. We do these comparisons under various assumptions on the social welfare costs of detaining a defendant would FTA (i.e., the social welfare cost of an "unnecessary detention"). We construct 95% confidence intervals for expected social welfare under the statistical decision rule and the judge's observed release decisions. We report the ratio of worst-case expected social welfare under the statistical decision rule against worst-case expected social welfare under the judge's observed release decisions. We report the median, minimum and maximum gain across judges at each considered social welfare cost of unnecessary detentions.

4.2.1 Effects over Judges Who Make Prediction Mistakes

We compare the statistical release rule against the release decisions of judges who were found to make systematic prediction mistakes about FTA risk in Figure 3 over the race-by-age cells. For most considered values of the social welfare cost of unnecessary detentions, the statistical decision rule improves upon these judges' decisions (i.e., the median, minimum and maximum of worst-case expected social welfare under the statistical decision rule across judges is larger than the worst-case expected social welfare under these judges' decisions).

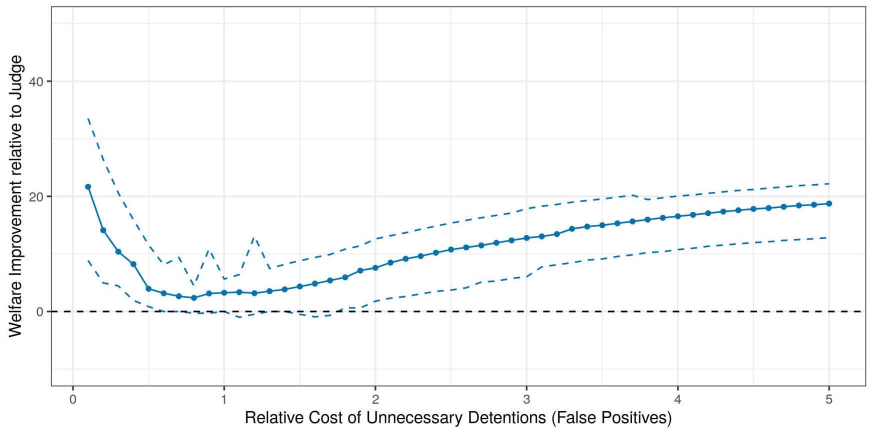
Figure 3: Ratio of expected social welfare under statistical decision rule relative to observed release decisions of judges that make systematic prediction mistakes about FTA risk.



Notes: This figure reports the change in worst-case expected social welfare under the statistical decision rule against the observed release decisions of judges who were found to make systematic prediction mistakes about FTA risk. Worst case expected social welfare under each decision rule is computed by constructing 95% confidence intervals for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the social welfare cost of detaining a defendant that would not FTA (i.e., an "unnecessary detention"). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for discussion.

For social welfare costs of unnecessary detentions ranging over $[0.3, 0.8]$, we found that the statistical decision rule either leads to no improvement or strictly lowers worst-case expected social welfare relative to these judges' decisions. Appendix Figure A10 compares worst-case expected social welfare under the statistical decision rule against the judges' decisions by the race of the defendant. Appendix Figure A11 compares the release rates of the statistical decision rule against the observed release rates of these judges.

Figure 4: Ratio of expected social welfare under statistical decision rule that only corrects systematic prediction mistakes relative to observed release decisions of judges that make systematic prediction mistakes about FTA risk.



Notes: This figure reports the change in worst-case expected social welfare under the statistical decision rule that only replaces the judge on cases in deciles 1-2, and deciles 9-10 of the predicted FTA risk distributed against the judge's observed release decisions. Worst case expected social welfare under each decision rule is computed by constructing 95% confidence intervals for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for discussion.

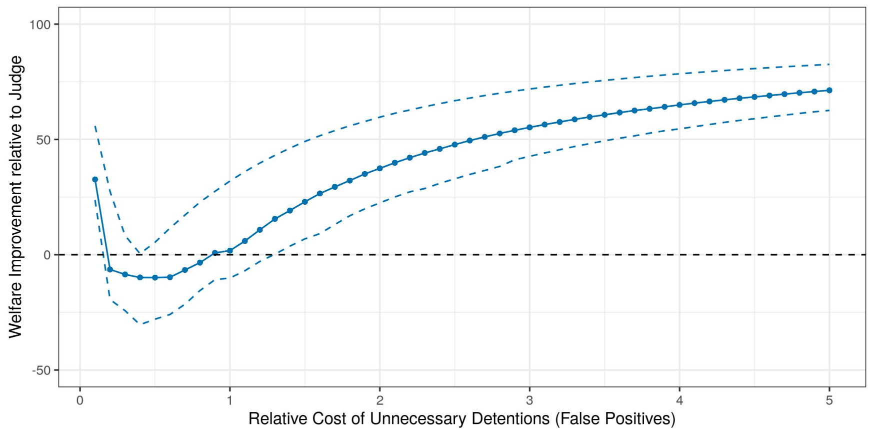
We now compare these judges' observed release decisions against a statistical decision rule that only replaces them over defendants in deciles 1-2 and 9-10 of the predicted FTA risk distribution. This is motivated by our earlier empirical results, which found that these judges primarily make systematic prediction mistakes over defendants in deciles 1-2 and 9-10 of the predicted FTA risk distribution. Figure 4 the results of this comparison.

Appendix Figures A12a-A12b plot the analogous quantities over the race-by-felony charge cells of directly payoff-relevant characteristics.

4.2.2 Effects over Judges Who Do Not Make Prediction Mistakes

We also compare the statistical decision rule against the release decisions of judges whose choices were found to be consistent with expected utility maximization behavior at accurate beliefs (i.e., not making systematic prediction mistakes about FTA risk) in Figure 5. Appendix Figure A13 plots the results broken out by defendant race and Appendix Figure A14 compares the release rates of the algorithmic decision rule against the observed release rates of these judges. Appendix Figure A15 also plots the same quantities for judges that do not make systematic prediction mistakes over the race-by-felony charge cells.

Figure 5: Ratio of expected social welfare under full automation decision rule relative to observed release decisions of judges that do not make systematic prediction mistakes about FTA risk.



Notes: This figure reports the change in worst-case expected social welfare under the algorithmic decision rule that fully automates decision-making against the judge's observed release decisions among judges whose choices were consistent with expected utility maximization behavior at accurate beliefs. Worst case expected social welfare under each decision rule is computed by constructing 95% confidence intervals for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for discussion.

5 Conclusion

In this report, we applied the general econometric framework developed in Rambachan (2022) to analyze whether judges in New York City make systematic prediction mistakes about FTA risk based on defendant and case characteristics. To do so, we tested whether judges observed pretrial release decisions were inconsistent with expected utility maximization behavior under a range of assumptions on which defendant characteristics directly affected their preferences. We found that a large fraction of judges in New York City make systematic prediction mistakes about FTA risk, and that these prediction mistakes primarily occur over defendants that would be predicted to be low FTA risk based on observable characteristics and defendants that would be high FTA risk based on observable characteristics. Finally, we also discussed the implications of these results for the design and use of statistical decision rules in pretrial release decisions.

References

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes. 2023. "Inference for Linear Conditional Moment Inequalities." 6.

Arnold, David, Will Dobbie, and Crystal Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Arnold, David, Will Dobbie, and Peter Hull. 2022. "Measuring Racial Discrimination in Bail Decisions." *American Economic Review*, 112(9): 2992–3038.

Dobbie, Will, Jacob Goldin, and Crystal Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–240.

Kahneman, Daniel. 2011. *Thinking Fast and Slow.* New York: Farrar, Strauss and Giroux.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*, 133(1): 237–293.

Leslie, Emily, and Nolan G. Pope. 2017. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *The Journal of Law and Economics*, 60(3): 529–557.

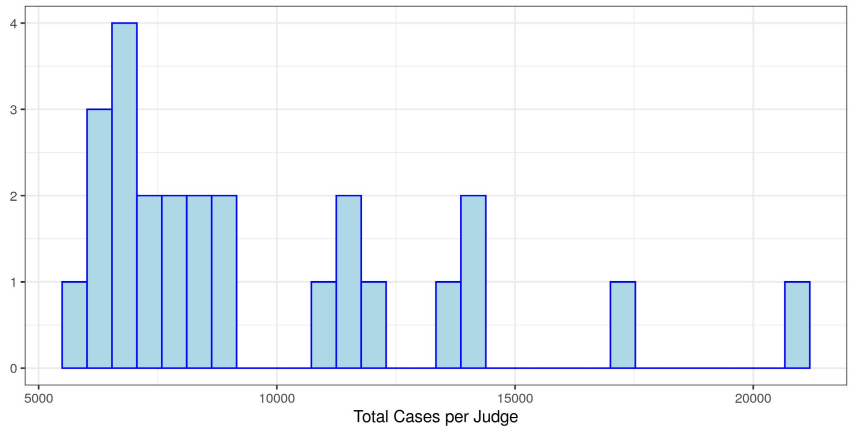
Rambachan, Ashesh. 2022. "Identifying Prediction Mistakes in Observational Data."

Rambachan, Ashesh, and Jonathan Roth. 2022. "A More Credible Approach to Parallal Trends." *The Review of Economic Studies*.

Wright, Marvin N., and Andreas Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software, Articles*, 77(1): 1–17.

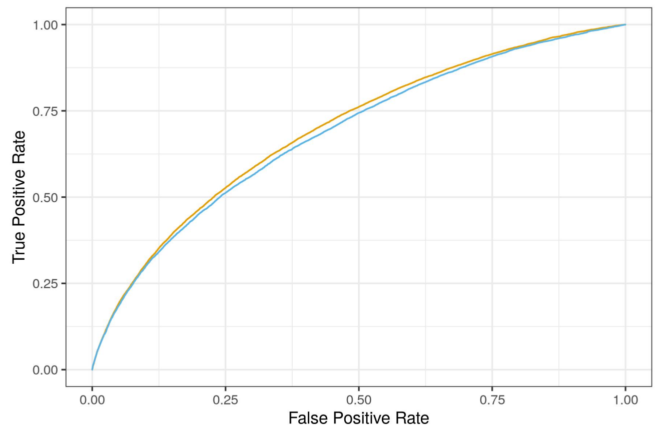
A Additional Figures

Figure A1: Histogram of number of cases heard by each judge in the top 25 judges.

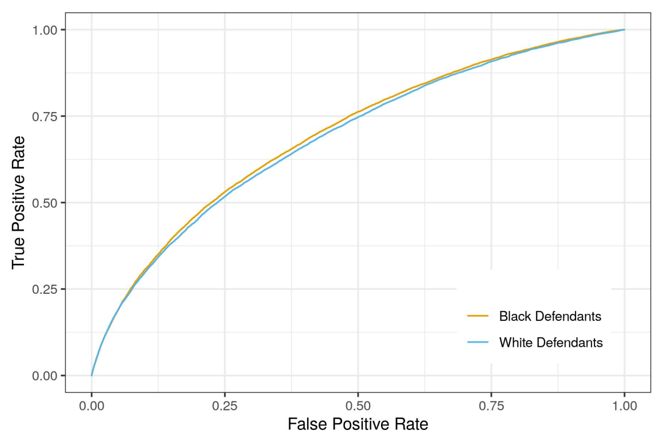


Notes: This figure plots a histogram of the total number of cases heard by each judge in the top 25 of judges that heard the most cases in the main estimation sample over the sample period from November 1, 2008 to November 1, 2013. See Section 2.1 for discussion.

Figure A2: Receiver-operating characteristic (ROC) curves for ensemble prediction functions



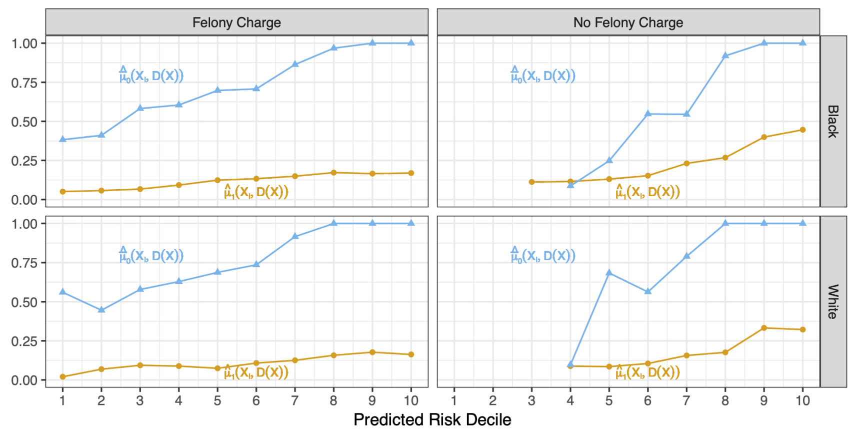
(a) Race-by-age cells



(b) Race-by-felony charge cells

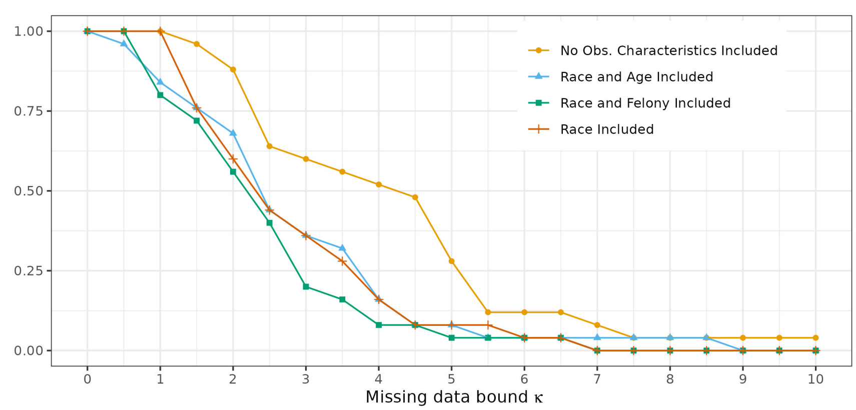
Notes: This figure plots the Receiver-Operating Characteristic (ROC) curves for the ensemble prediction function that predicts FTA among defendants that were released by the top 25 judges. The prediction function is constructed over cases heard by the remaining bail judges and evaluated out-of-sample on cases heard by the top 25 judges. It reports the ROC curve for the ensemble prediction function constructed within race-by-age cells and race-by-felony charge cells separately. Age is binarized into young and older defendants, where older defendants are defined as defendants older than 25 years. The ROC curve plots the false positive rate on the x-axis and the true positive rate on the y-axis. The out-of-sample area under the curve (AUC) on all defendants equals 0.693 for the ensemble prediction function constructed over race-by-age cells and 0.694 for the ensemble prediction function constructed over race-by-felony cells. See Section 3.1.1 for further discussion.

Figure A3: Observed failure to appear rate among released defendants and constructed bound on the failure to appear rate among detained defendants by race-and-felony charge cells for one judge in New York City.



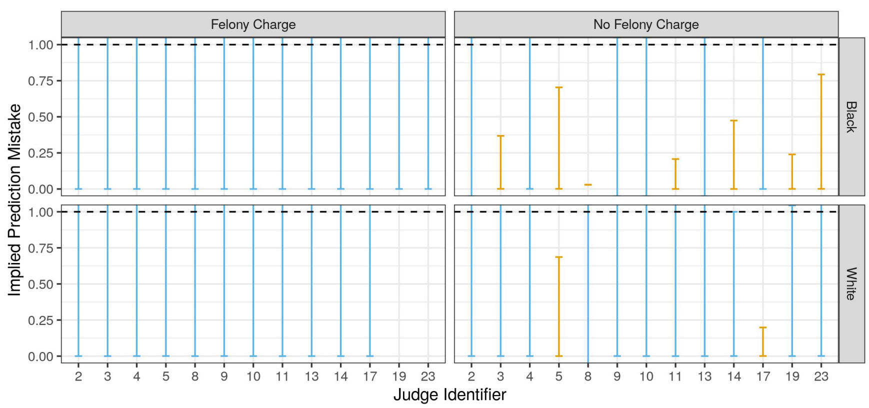
Notes: This figure plots the observed failure to appear rate among released defendants (orange) and the bounds on the failure to appear rate among detained defendants based on the judge leniency instrument (blue) at each decile of predicted risk and race-by-felony charge cell for the judge that heard the most cases in the main estimation sample. See Section 3.1.1 for the estimation details.

Figure A4: Fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk using direct imputation bounds on FTA risk among detained defendants.



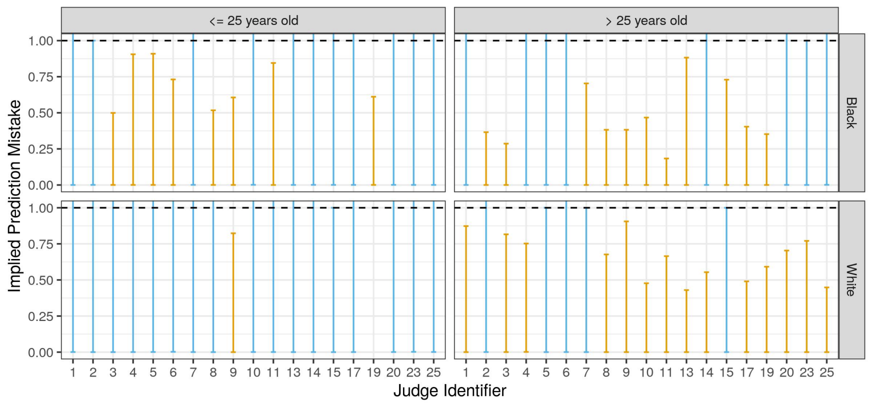
Notes: This figures reports the fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at strict preference utility functions that either (i) do not depend on any characteristics, (ii) depend on the defendant's race, (iii) depend on both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. Bounds on the FTA risk of detained defendants are constructed using the general "direct imputation" procedure discussed in **?**. The adjusted rejection rate reports the fraction of rejections correcting for multiple hypotheses using the Holm-Bonferroni step-down procedure. See Section 3.1.3 for further estimation details.

Figure A5: 95% confidence intervals for the identified set of systematic prediction mistakes of FTA risk between lowest and highest predicted risk deciles made by judges within each race-by-felony charge cell.

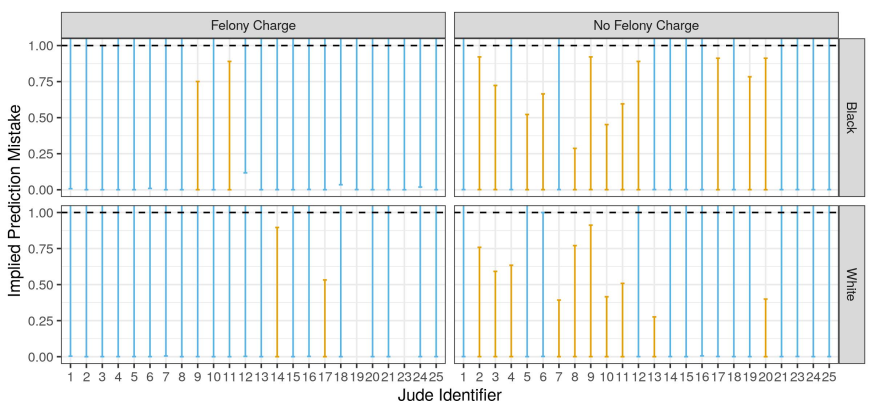


Notes: This figure plots the 95% confidence interval for the identified set on each judge's prediction mistake between the lowest predicted risk decile and the highest predicted risk decile for the race-by-felony charge cells of directly payoff relevant characteristics. See Section 3.2 for details.

Figure A6: 95% confidence intervals for the identified set of prediction mistakes of FTA risk between the highest and lowest predicted risk deciles using direct imputation bounds with $\kappa = 2$.



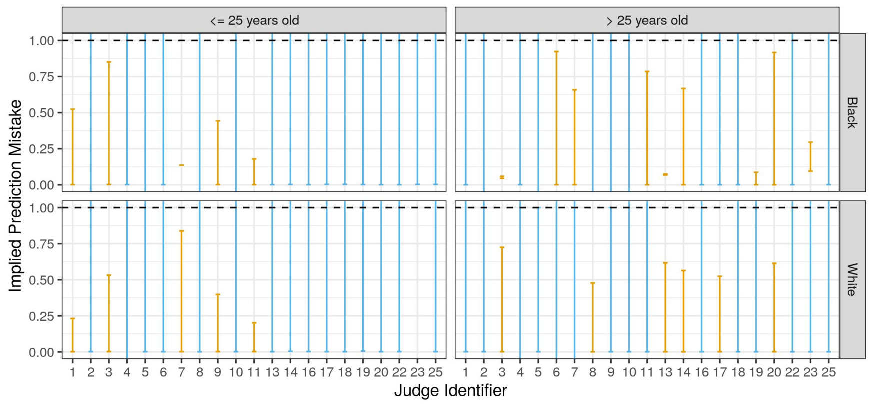
(a) Race-by-age W cells



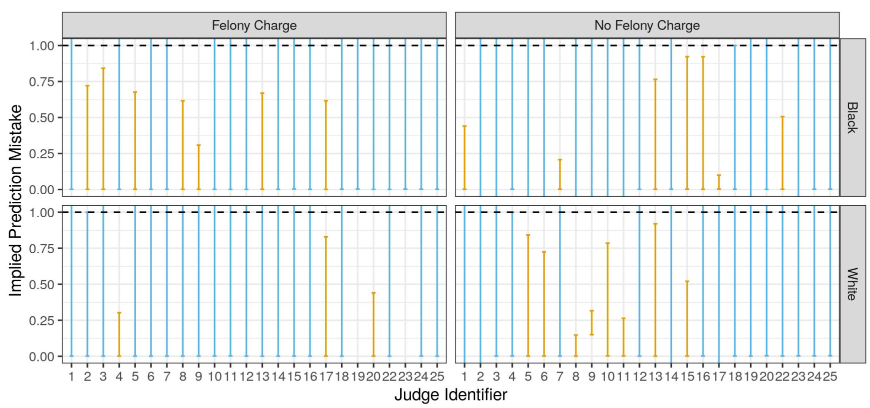
(b) Race-by-felony W cells

Notes: This figures plots 95% confidence intervals for the identified set on each judge's prediction mistake between the highest predicted risk decile and the lowest predicted risk decile within each race-by-age cell and race-by-felony charge cell. The identified set is constructed using direct imputation with $\kappa = 2$ and for each judge in the top 25 whose choices are inconsistent with expected utility maximization behavior at accurate beliefs about FTA risk based on defendant characteristics.. See Section 3.2 for further estimation details.

Figure A7: 95% confidence intervals for the identified set of prediction mistakes of "any pretrial misconduct" risk between the highest and lowest predicted risk deciles.



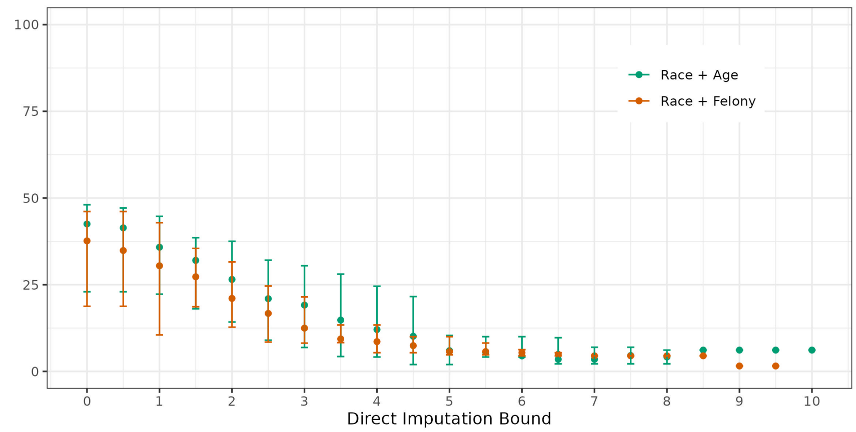
(a) Race-by-age W cells



(b) Race-by-felony W cells

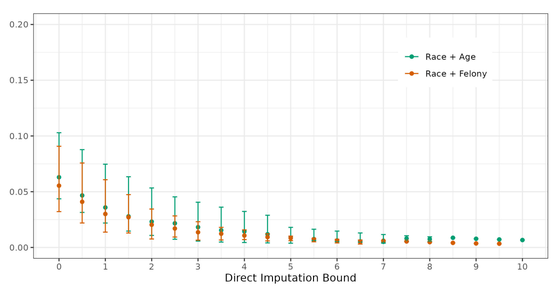
Notes: This figures plots the 95% confidence interval for the identified set on each judges' prediction mistake of "any pretrial misconduct risk" between the highest predicted risk decile and the lowest predicted risk decile within each race-by-age cell and race-by-felony charge cell. The outcome is whether the defendant would commit "any pretrial misconduct upon release" (i.e., either fail to appear in court or be re-arrested for a new crime). See Section 3.2 for further estimation details.

Figure A8: Bound on the share of prediction mistakes about FTA risk across judges using direct imputation bounds on FTA risk among detained defendants.

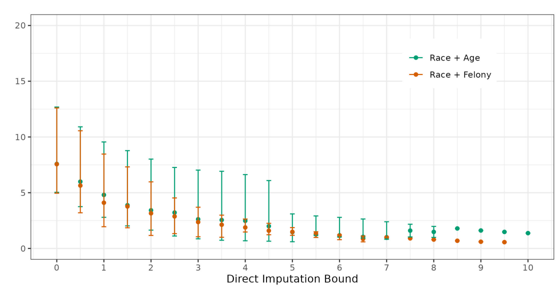


Notes: This figure reports how the share of systematic prediction mistakes about FTA risk across judges whose choices violated the revealed preference inequalities varies using direct imputation bounds on FTA risk among detained defendants. We report results allowing for either judges' utility functions to depend on (i) both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. The lower and upper error bars correspond to the minimum and maximum across judges whose choices violated the revealed preference inequalities at the nominal 5% level. The dot corresponds to the median across the same judges.

Figure A9: Bounds on the total expected utility costs of prediction mistakes about FTA risk using direct imputation bounds on FTA risk among detained defendants.



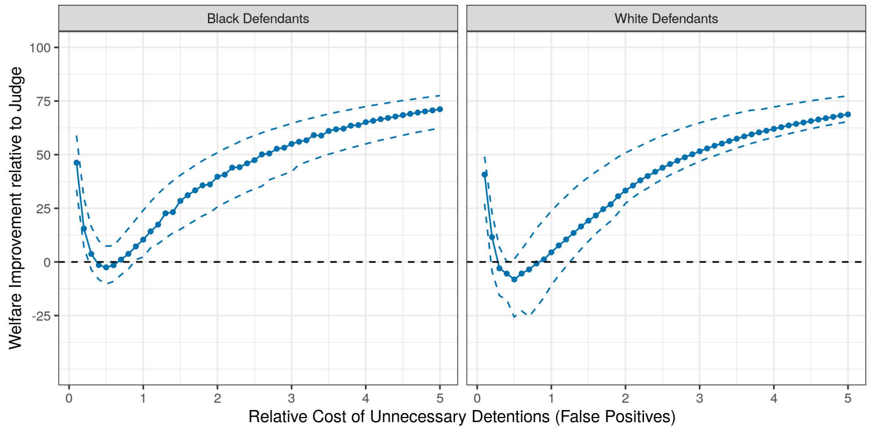
(a) Expected utility cost



(b) Equivalent reduction in fraction of defendants that are released and FTA

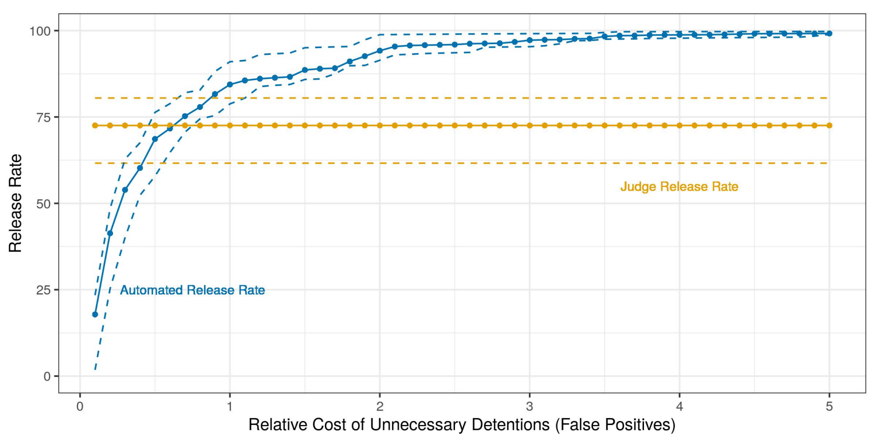
Notes: This figure reports how the total expected utility cost of prediction mistakes about FTA risk across judges whose choices violated the revealed preference inequalities varies using direct imputation bounds on FTA risk among detained defendants. We report results allowing for either judges' utility functions to depend on (i) both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. Panel (a) reports the total expected utility cost, and Panel (b) reports the equivalent reduction in the fraction of defendants that are released and FTA. The lower and upper error bars correspond to the minimum and maximum across judges whose choices violated the revealed preference inequalities at the nominal 5% level. The dot corresponds to the median across the same judges.

Figure A10: Ratio of expected social welfare under statistical decision rule relative to observed decisions of judges that make systematic prediction mistakes about FTA risk by defendant race.



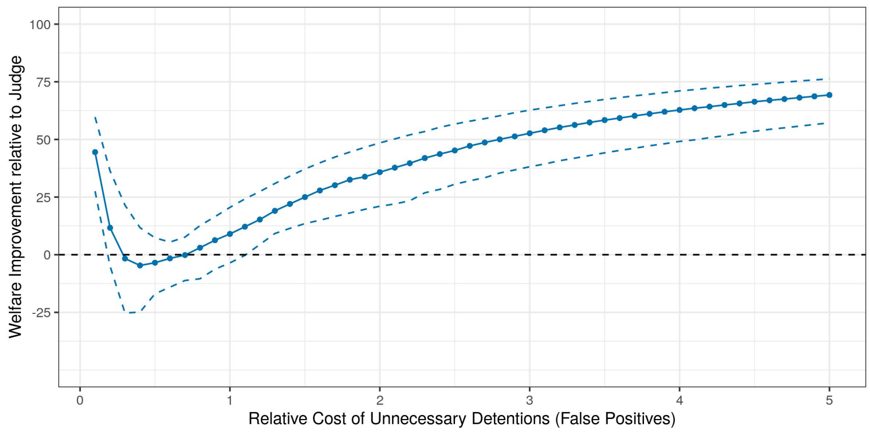
Notes: This figure reports the change in worst-case expected social welfare under the statistical decision rule against the judge's observed release decisions among judges who were found to make systematic prediction mistakes about FTA risk, broken out by defendant race. Worst-case expected social welfare under each decision rule is computed by first constructing a 95% confidence interval for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for further discussion.

Figure A11: Overall release rates under statistical decision rule relative to the observed release rates of judges that make systematic prediction mistakes about FTA risk.

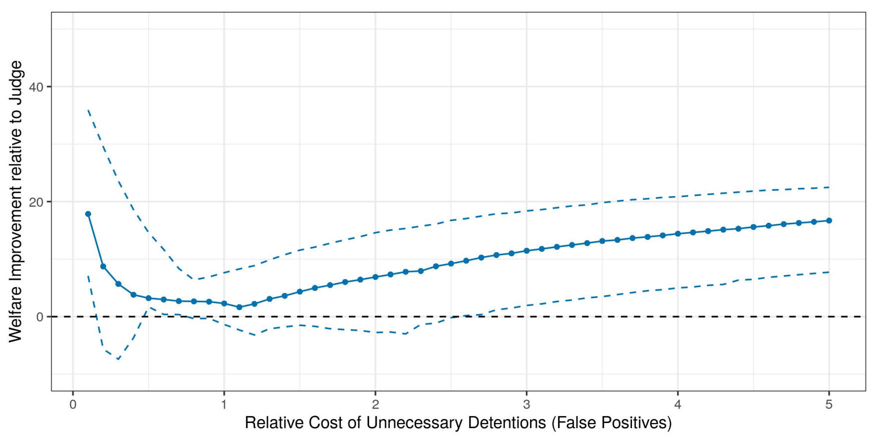


Notes: This figure reports the overall release rate of the statistical decision rule against the judge's observed release rates among judges who were found to make systematic prediction mistakes about FTA risk. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median release rate across judges that make mistakes, and the dashed lines report the minimum and maximum release rates across judges. See Section 4.2 for further discussion.

Figure A12: Ratio of expected social welfare under statistical decision rules relative to observed release decisions of judges that make systematic prediction mistakes about FTA risk over race-by-felony charge cells.



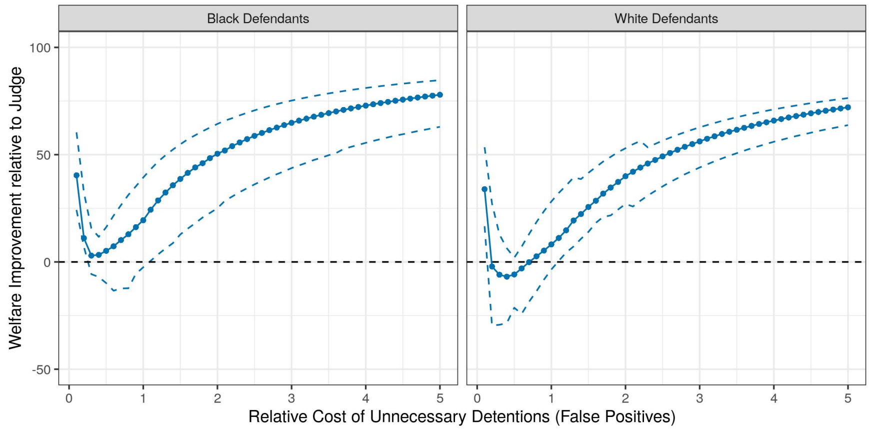
(a) Welfare improvement of full automation decision rule



(b) Welfare improvement of decision rule that corrects prediction mistakes

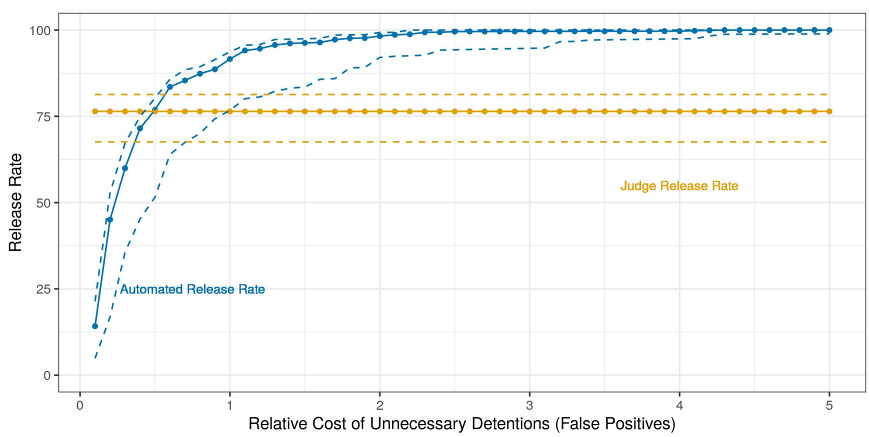
Notes: This figure reports the change in worst-case expected social welfare under two statistical decision rules against the judge's observed release decisions among judges who were found to make systematic prediction mistakes about FTA risk. Worst case expected social welfare under each decision rule is computed by first constructing a 95% confidence interval for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-felony cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for further discussion.

Figure A13: Ratio of expected social welfare under full automation decision rule relative to observed decisions of judges that do not make systematic prediction mistakes about FTA risk by defendant race.



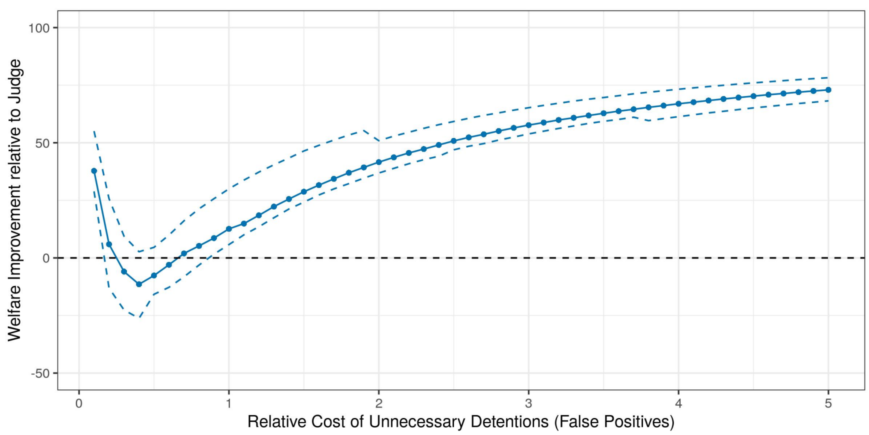
Notes: This figure reports the change in worst-case expected social welfare under the algorithmic decision rule that fully automates decision-making against the judge's observed release decisions among judges whose choices were consistent with expected utility maximization behavior at accurate beliefs, broken out by defendant race. Worst case expected social welfare under each decision rule is computed by first constructing a 95% confidence interval for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for further discussion.

Figure A14: Overall release rates under full automation decision rule relative to the observed release rates of judges that do not make systematic prediction mistakes about FTA risk.



Notes: This figure reports the overall release rate of the algorithmic decision rule that fully automates decisions against the judge's observed release rates among among judges whose choices were consistent with expected utility maximization behavior at accurate beliefs. These decisions rules are constructed and evaluated over race-by-age cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median release rate across judges that make mistakes, and the dashed lines report the minimum and maximum release rates across judges. See Section 4.2 for further discussion.

Figure A15: Ratio of expected social welfare under full automation decision rule relative to observed decisions of judges that do not make systematic prediction mistakes about FTA risk over race-by-felony charge cells.



Notes: This figure reports the change in worst-case expected social welfare under the algorithmic decision rule that fully automates decision-making against the judge's observed release decisions among judges whose choices were consistent with expected utility maximization behavior at accurate beliefs. Worst case expected social welfare under each decision rule is computed by first constructing a 95% confidence interval for expected social welfare under the decision rule, and reporting smallest value that lies in the confidence interval. These decisions rules are constructed and evaluated over race-by-felony cells and deciles of predicted risk. The x-axis plots the relative social welfare cost of detaining a defendant that would not FTA (i.e., an unnecessary detention). The solid line plots the median change across judges that make mistakes, and the dashed lines report the minimum and maximum change across judges. See Section 4.2 for further discussion.

B Additional Tables

Table A1: Descriptive statistics comparing the main estimation sample and cases heard by the top 25 judges, broken out by defendant race.

	All Defendants		White Defendants		Black Defendants	
	Estimation Sample	Top Judges	Estimation Sample	Top Judges	Estimation Sample	Top Judges
	(1)	(2)	(3)	(4)	(5)	(6)
Released before trial	0.720	0.736	0.757	0.777	0.687	0.699
Defendant Characteristics						
White	0.475	0.481	1.000	1.000	0.000	0.000
Female	0.173	0.173	0.154	0.152	0.190	0.192
Age at Arrest	31.95	31.75	32.03	31.88	31.87	31.63
Arrest Charge						
Number of Charges	1.152	1.167	1.187	1.217	1.119	1.121
Felony Charge	0.372	0.367	0.367	0.356	0.376	0.377
Any Drug Charge	0.253	0.224	0.253	0.217	0.253	0.230
Any DUI Charge	0.047	0.049	0.070	0.072	0.027	0.027
Any Violent Crime Charge	0.375	0.395	0.358	0.379	0.390	0.410
Property Charge	0.130	0.132	0.122	0.123	0.138	0.140
Defendant Priors						
Any FTA	0.516	0.497	0.443	0.419	0.582	0.570
Number of FTAs	2.177	2.034	1.633	1.492	2.670	2.537
Any Misdemeanor Arrest	0.683	0.667	0.615	0.596	0.744	0.734
Any Misdemeanor Conviction	0.383	0.368	0.334	0.315	0.427	0.418
Any Felony Arrest	0.581	0.566	0.503	0.482	0.652	0.644
Any Felony Conviction	0.285	0.271	0.234	0.215	0.331	0.323
Any Violent Felony Arrest	0.398	0.387	0.306	0.292	0.481	0.476
Any Violent Felony Conviction	0.119	0.114	0.084	0.078	0.150	0.147
Total Cases	569,256	243,118	270,704	117,073	298,552	126,045

Notes: This table provides descriptive statistics about defendant and case characteristics for the main estimation sample and the cases heard by the top 25 judges in the New York City pretrial release data, broken out separately for all defendants and by the race of the defendant. See Section 2.2 for discussion.

Table A2: Descriptive statistics for released and detained defendants in the main estimation sample and for cases heard by the top 25 judges

	All Defendants		Released Defendants		Detained Defendants	
	Estimation Sample	Top Judges	Estimation Sample	Top Judges	Estimation Sample	Top Judges
	(1)	(2)	(3)	(4)	(5)	(6)
Released before trial	0.720	0.736	1.000	1.000	0.000	0.000
Defendant Characteristics						
White	0.475	0.481	0.499	0.508	0.412	0.407
Female	0.173	0.173	0.199	0.197	0.107	0.106
Age at Arrest	31.95	31.75	31.22	31.20	33.82	33.29
Arrest Charge						
Number of Charges	1.152	1.167	1.148	1.162	1.161	1.182
Felony Charge	0.372	0.367	0.288	0.288	0.588	0.586
Any Drug Charge	0.253	0.224	0.229	0.204	0.314	0.279
Any DUI Charge	0.047	0.049	0.062	0.063	0.010	0.010
Any Violent Crime Charge	0.375	0.395	0.388	0.409	0.341	0.355
Property Charge	0.130	0.132	0.115	0.114	0.171	0.181
Defendant Priors						
Any FTA	0.516	0.497	0.409	0.395	0.793	0.784
Number of FTAs	2.177	2.034	1.362	1.295	4.284	4.103
Any Misdemeanor Arrest	0.683	0.667	0.610	0.598	0.871	0.863
Any Misdemeanor Conviction	0.383	0.368	0.284	0.278	0.637	0.621
Any Felony Arrest	0.581	0.566	0.487	0.477	0.824	0.814
Any Felony Conviction	0.285	0.271	0.200	0.194	0.505	0.487
Any Violent Felony Arrest	0.398	0.387	0.315	0.309	0.614	0.608
Any Violent Felony Conviction	0.119	0.114	0.081	0.080	0.216	0.210
Total Cases	569,256	243,118	410,394	179,143	158,862	63,975

Notes: This table provides descriptive statistics about defendant and case characteristics for the main estimation sample and the cases heard by the top 25 judges in the New York City pretrial release data, broken out separately for all defendants and by whether the defendant was released or detained. See Section 2.2 for discussion.

Table A3: Descriptive statistics of pretrial misconduct rates among released defendants in the main estimation sample and cases heard by the top 25 judges.

	All Defendants		White Defendants		Black Defendants	
	Estimation Sample	Top Judges	Estimation Sample	Top Judges	Estimation Sample	Top Judges
	(1)	(2)	(3)	(4)	(5)	(6)
Failure to Appear (FTA)	0.151	0.146	0.135	0.131	0.167	0.161
Rearrest (NCA)	0.261	0.257	0.230	0.225	0.292	0.289
Any Misconduct	0.331	0.324	0.297	0.290	0.366	0.359
Total Cases	410,394	179,143	205,174	91,026	205,220	88,117

Notes: This table summarizes the observed pretrial misconduct rates among released defendants in the main estimation sample and among released defendants by the top 25 judges in the New York City pretrial release data, broken out separately for all defendants and by the race of the defendant. See Section 2.2 for discussion.

Table A4: Descriptive statistics for the universe of all cases subject to a pretrial release decision and the main estimation sample in the New York City pretrial release data, broken out by defendant race.

	All Defendants		White Defendants		Black Defendants	
	Full Sample	Estimation Sample	Full Sample	Estimation Sample	Full Sample	Estimation Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Released before trial	0.736	0.720	0.765	0.757	0.691	0.687
Defendant Characteristics						
White	0.457	0.475	1.000	1.000	0.000	0.000
Female	0.169	0.173	0.153	0.154	0.184	0.190
Age at Arrest	32.06	31.95	32.06	32.03	31.88	31.87
Arrest Charge						
Number of Charges	1.165	1.152	1.176	1.187	1.114	1.119
Felony Charge	0.335	0.372	0.332	0.367	0.346	0.376
Any Drug Charge	0.244	0.253	0.251	0.253	0.252	0.253
Any DUI Charge	0.053	0.047	0.074	0.070	0.028	0.027
Any Violent Crime Charge	0.365	0.375	0.348	0.358	0.380	0.390
Property Charge	0.135	0.130	0.127	0.122	0.145	0.138
Defendant Priors						
Any FTA	0.499	0.516	0.442	0.443	0.586	0.582
Number of FTAs	2.099	2.177	1.635	1.633	2.707	2.670
Any Misdemeanor Arrest	0.668	0.683	0.616	0.615	0.747	0.744
Any Misdemeanor Conviction	0.371	0.383	0.335	0.334	0.430	0.427
Any Felony Arrest	0.565	0.581	0.502	0.503	0.654	0.652
Any Felony Conviction	0.273	0.285	0.232	0.234	0.334	0.331
Any Violent Felony Arrest	0.384	0.398	0.306	0.306	0.484	0.481
Any Violent Felony Conviction	0.114	0.119	0.084	0.084	0.152	0.150
Total Cases	758,027	569,256	347,006	270,704	370,793	298,552

Notes: This table provides descriptive statistics about defendant and case characteristics for the universe of all cases subject to a pretrial release decision and the main estimation sample in the New York City pretrial release data, broken out for all defendants and by the race of the defendant. See Section 2.2 for discussion.

Table A5: Descriptive statistics for released and detained defendants in the universe of all cases subject to a pretrial release decision and the main estimation sample in the New York City pretrial release data.

	All Defendants		Released Defendants		Detained Defendants	
	Full Sample	Estimation Sample	Full Sample	Estimation Sample	Full Sample	Estimation Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Released before trial	0.736	0.720	1.000	1.000	0.000	0.000
Defendant Characteristics						
White	0.457	0.475	0.476	0.499	0.406	0.412
Female	0.169	0.173	0.192	0.199	0.105	0.107
Age at Arrest	32.06	31.95	31.41	31.22	33.90	33.82
Arrest Charge						
Number of Charges	1.165	1.152	1.166	1.148	1.161	1.161
Felony Charge	0.335	0.372	0.258	0.288	0.549	0.588
Any Drug Charge	0.244	0.253	0.221	0.229	0.307	0.314
Any DUI Charge	0.053	0.047	0.068	0.062	0.011	0.010
Any Violent Crime Charge	0.365	0.375	0.377	0.388	0.333	0.341
Property Charge	0.135	0.130	0.119	0.115	0.179	0.171
Defendant Priors						
Any FTA	0.499	0.516	0.394	0.409	0.792	0.793
Number of FTAs	2.099	2.177	1.310	1.362	4.304	4.284
Any Misdemeanor Arrest	0.668	0.683	0.596	0.610	0.870	0.871
Any Misdemeanor Conviction	0.371	0.383	0.275	0.284	0.639	0.637
Any Felony Arrest	0.565	0.581	0.472	0.487	0.823	0.824
Any Felony Conviction	0.273	0.285	0.190	0.200	0.503	0.505
Any Violent Felony Arrest	0.384	0.398	0.302	0.315	0.612	0.614
Any Violent Felony Conviction	0.114	0.119	0.077	0.081	0.216	0.216
Total Cases	758,027	569,256	558,167	410,394	199,860	158,862

Notes: This table provides descriptive statistics about defendant and case characteristics for the sample of all cases subject to a pretrial release decision and the main estimation sample in the New York City pretrial release data, broken out for all defendants and by whether the defendant was released or detained. See Section 2.2 for discussion.

Table A6: Balance check regressions for the quasi-random assignment of judges: pooled across all defendants and separately by defendant race.

	All Defendants (1)	White Defendants (2)	Black Defendants (3)
Defendant Characteristics			
Black	−0.00011		
	(0.00008)		
Female	0.000003	0.00005	−0.00003
	(0.00013)	(0.00017)	(0.00017)
Age	−0.00001	−0.00002	−0.000002
	(0.000003)	(0.00001)	(0.000004)
Arrest Charge			
Number of Charges	−0.000003	−0.000003	0.000003
	(0.00001)	(0.00001)	(0.00003)
Felony Charge	0.00009	−0.00012	0.00027
	(0.00015)	(0.00017)	(0.00018)
Any Drug Charge	−0.00012	−0.00010	−0.00013
	(0.00013)	(0.00018)	(0.00016)
Any Violent Crime Charge	−0.00004	−0.00013	0.00004
	(0.00010)	(0.00015)	(0.00014)
Any Property Charge	−0.00033	−0.00029	−0.00035
	(0.00016)	(0.00019)	(0.00025)
Any DUI Charge	0.00044	0.00039	0.00028
	(0.00024)	(0.00027)	(0.00039)
Defendant Priors			
Prior FTA	−0.00004	−0.00011	0.00003
	(0.00010)	(0.00016)	(0.00013)
Prior Misdemeanor Arrest	0.00007	0.00003	0.00011
	(0.00010)	(0.00013)	(0.00015)
Prior Felony Arrest	0.00006	0.00003	0.00009
	(0.00014)	(0.00021)	(0.00019)
Prior Violent Felony Arrest	−0.00013	−0.00008	−0.00016
	(0.00011)	(0.00019)	(0.00016)
Prior Misdemeanor Conviction	0.00016	0.00021	0.00011
	(0.00013)	(0.00017)	(0.00016)
Prior Felony Conviction	−0.00019	0.00011	−0.00040
	(0.00012)	(0.00018)	(0.00015)
Prior Violent Felony Conviction	−0.00008	−0.00024	0.00002
	(0.00015)	(0.00021)	(0.00020)
Joint p-value	0.06953	0.15131	0.41840
Court × Time FE	✓	✓	✓
Cases	569,256	270,704	298,552

Notes: This table reports the OLS estimates for the balance check regressions of judge leniency on defendant characteristics in the main estimation sample. These balance check regressions are estimated over all defendants pooled together and separately by defendant race. Standard errors are clustered at the defendant and judge level, and reported in parentheses. The joint p-value reports the p-value associated with the F-statistic on whether all defendant characteristics are jointly significant. See Section 3.1.1 for further discussion.

Table A7: Balance check regressions for the quasi-random assignment of judges: separately by defendant race and age.

	White Defendants		Black Defendants	
	Young (1)	Older (2)	Young (3)	Older (4)
Defendant Characteristics				
Female	−0.00008	0.00017	−0.00007	−0.00005
	(0.00025)	(0.00019)	(0.00024)	(0.00024)
Age	−0.000004	−0.00001	−0.00006	−0.00001
	(0.00004)	(0.00001)	(0.00003)	(0.00001)
Arrest Charge				
Number of Charges	−0.00002	−0.000003	−0.00002	0.00001
	(0.00003)	(0.000005)	(0.00006)	(0.00003)
Felony Charge	0.00002	−0.00024	0.00019	0.00033
	(0.00023)	(0.00019)	(0.00023)	(0.00022)
Any Drug Charge	−0.00033	0.00004	−0.00046	0.00004
	(0.00033)	(0.00022)	(0.00025)	(0.00020)
Any Violent Crime Charge	−0.00025	−0.00010	−0.00016	0.00018
	(0.00026)	(0.00019)	(0.00024)	(0.00018)
Any Property Charge	−0.00005	−0.00046	−0.00017	−0.00045
	(0.00034)	(0.00023)	(0.00031)	(0.00029)
Any DUI Charge	0.00021	0.00042	−0.00160	0.00062
	(0.00045)	(0.00030)	(0.00072)	(0.00044)
Defendant Priors				
Prior FTA	−0.00013	−0.00015	0.00034	−0.00021
	(0.00026)	(0.00021)	(0.00022)	(0.00020)
Prior Misdemeanor Arrest	0.00026	−0.00018	−0.00008	0.00034
	(0.00021)	(0.00017)	(0.00022)	(0.00022)
Prior Felony Arrest	−0.00008	0.00018	0.00035	−0.00025
	(0.00026)	(0.00027)	(0.00030)	(0.00024)
Prior Violent Felony Arrest	−0.00024	−0.00001	−0.00020	−0.00019
	(0.00030)	(0.00023)	(0.00025)	(0.00021)
Prior Misdemeanor Conviction	0.00040	0.00023	0.00040	0.00004
	(0.00029)	(0.00025)	(0.00028)	(0.00018)
Prior Felony Conviction	0.00052	0.00005	−0.00094	−0.00016
	(0.00049)	(0.00019)	(0.00033)	(0.00017)
Prior Violent Felony Conviction	−0.00029	−0.00020	0.00113^{**}	−0.00012
	(0.00077)	(0.00022)	(0.00054)	(0.00021)
Joint p-value	0.85104	0.44370	0.038862	0.16062
Court × Time FE	✓	✓	✓	✓
Cases	99,536	171,168	119,156	179,396

Notes: This table reports the OLS estimates for the balance check regressions of judge leniency on defendant characteristics in the main estimation sample. These balance check regressions are estimated separately over subsamples defined by the race and age of the defendant, where "young" is defined as being less than or equal to 25 years and "older" is defined as older than 25 years. Standard errors are clustered at the defendant and judge level, and reported in parentheses. The joint p-value reports the p-value associated with the F-statistic on whether all defendant characteristics are jointly significant. See Section 3.1.1 for further discussion.

Table A8: Balance check regressions for the quasi-random assignment of judges: separately by defendant race and charge severity.

	White Defendants		Black Defendants	
	Felony Charge (1)	No Felony Charge (2)	Felony Charge (3)	No Felony Charge (4)
Defendant Characteristics				
Female	0.00003	0.00001	−0.00003	−0.00004
	(0.00023)	(0.00021)	(0.00026)	(0.00021)
Age	−0.00002	−0.00001	0.000004	−0.000004
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Arrest Charge				
Number of Charges	−0.000002	−0.00004	−0.000005	0.00003
	(0.00001)	(0.00003)	(0.00003)	(0.00007)
Any Drug Charge	−0.00022	−0.00008	−0.00012	−0.00008
	(0.00028)	(0.00024)	(0.00031)	(0.00023)
Any Violent Crime Charge	−0.00043	0.00001	0.00038	−0.00013
	(0.00030)	(0.00018)	(0.00026)	(0.00017)
Any Property Charge	−0.00038	−0.00038	0.00023	−0.00070
	(0.00027)	(0.00028)	(0.00029)	(0.00035)
Any DUI Charge	0.00047	0.00049	0.00100	0.00012
	(0.00057)	(0.00030)	(0.00093)	(0.00042)
Defendant Priors				
Prior FTA	−0.00014	−0.00005	0.00012	−0.00003
	(0.00023)	(0.00020)	(0.00024)	(0.00015)
Prior Misdemeanor Arrest	0.00024	−0.00012	0.00009	0.00010
	(0.00025)	(0.00017)	(0.00028)	(0.00018)
Prior Felony Arrest	−0.00007	−0.000005	−0.00043	0.00040
	(0.00036)	(0.00023)	(0.00032)	(0.00022)
Prior Violent Felony Arrest	−0.00042	0.00012	−0.00001	−0.00020
	(0.00029)	(0.00021)	(0.00025)	(0.00018)
Prior Misdemeanor Conviction	−0.00009	0.00050	0.00042	−0.00013
	(0.00030)	(0.00021)	(0.00027)	(0.00017)
Prior Felony Conviction	0.00010	0.00024	−0.00040	−0.00041
	(0.00034)	(0.00023)	(0.00025)	(0.00019)
Prior Violent Felony Conviction	0.00040	−0.00084	−0.00004	0.0000001
	(0.00036)	(0.00030)	(0.00028)	(0.00024)
Joint p-value	0.05623	0.27401	0.24607	0.24712
Court × Time FE	✓	✓	✓	✓
Cases	99,463	171,241	112,517	186,035

Notes: This table reports the OLS estimates for the balance check regressions of judge leniency on defendant characteristics in the main estimation sample. These balance check regressions are estimated separately over subsamples defined by defendant race and charge severity (i.e., whether the defendant was charged with a felony offense). Standard errors are clustered at the defendant and judge level, and reported in parentheses. The joint p-value reports the p-value associated with the F-statistic on whether all defendant characteristics are jointly significant. See Section 3.1.1 for further discussion.

Table A9: Fraction of judges whose pretrial release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about "any pretrial misconduct" risk given defendant characteristics.

	Utility Functions			
	No Characteristics	Race	Race + Age	Race + Felony Charge
Adjusted Rejection Rate	76%	72%	64%	92%

Notes: This table summarizes the results of our robustness exercise to assess whether the release decisions of judges are consistent with expected utility maximization behavior at strict preference utility functions that either (i) do not depend on any characteristics, (ii) depend on the defendant's race, (iii) depend on both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. The outcome is defined to be whether the defendant would commit "any pretrial misconduct" upon release (i.e., either fail to appear in court or be re-arrested for a new crime). The adjusted rejection rate reports the fraction of rejections after multiple hypothesis correction using the Holm-Bonferroni step down procedure. See Section 3.1.3 for discussion.

Table A10: Fraction of judges whose "release on recognizance" decisions are inconsistent with expected utility maximization behavior at accurate beliefs about behavior under bail conditions and FTA risk given defendant characteristics.

	Utility Functions			
	No Characteristics	Race	Race + Age	Race + Felony Charge
Adjusted Rejection Rate	32%	32%	32%	52%

Notes: This table summarizes the results of our robustness exercise to assess whether the "release on recognizance" decisions of judges are consistent with expected utility maximization behavior at strict preference utility functions that either (i) do not depend on any characteristics, (ii) depend on the defendant's race, (iii) depend on both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. The outcome is defined to be whether the defendant would be released under the chosen bail condition (i.e., either the judge decides to release the defendant on recognizance or the defendant satisfies the bail conditions set by the judge) and FTA if released. The adjusted rejection rate reports the fraction of rejections after multiple hypothesis correction using the Holm-Bonferroni step down procedure. See Section 3.1.3 for discussion.

Table A11: Estimated fraction of judges whose release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about failure to appear risk, non-violent criminal arrest risk, and violent criminal arrest risk given defendant characteristics.

	Utility Functions $u(c, y^*; x_I)$			
	No Characteristics	Race	Race + Age	Race + Felony Charge
Adjusted Rejection Rate	24%	24%	20%	32%

Notes: This table summarizes the results of our robustness exercise to assess whether the judges release decisions are consistent with expected utility maximization behavior at strict preference utility functions that either (i) do not depend on any characteristics, (ii) depend on the defendant's race, (iii) depend on both the defendant's race and age, and (iv) depend on both the defendant's race and whether the defendant was charged with a felony offense. The outcome is now defined to be vector-valued, and equal to whether a defendant would fail to appear in court, be re-arrested for a non-violent and be re-arrested for a violent crime if released. The adjusted rejection rate reports the fraction of rejections after multiple hypothesis correction using the Holm-Bonferroni step down procedure. See Section 3.1.3 for discussion.

Table A12: Location of the maximum studentized violation of implied revealed preference inequalities among judges whose release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about "any pretrial misconduct risk" given defendant characteristics.

	Utility Functions	
	Race and Age	Race and Felony Charge
Unadjusted Rejection Rate	84%	98%
White Defendants		
Middle Deciles	0.00%	0.00%
Tail Deciles	4.76%	4.16%
Black Defendants		
Middle Deciles	9.52%	16.66%
Tail Deciles	85.71%	79.16%

Notes: This table summarizes the location of the maximum studentized violation of the implied revealed preference inequalities among judges whose release decisions are inconsistent with expected utility maximization behavior at accurate beliefs about "any pretrial misconduct" (i.e., either fail to appear in court or be re-arrested for a new crime) risk and preferences that depend on both the defendant's race and age as well as the defendant's race and whether the defendant was charged with a felony. Among judges' whose release decision violate the revealed preference inequalities at the 5% level, this table reports the fraction of judges for whom the maximal studentized violation occurs among white and black defendants on tail deciles (deciles 1-2, 9-10) and middle deciles (3-8). See Section 3.3 for discussion.

Table A13: Bound on the share of prediction mistakes about pretrial misconduct risk across judges.

	Utility Functions $u(c, y; x_I)$	
	Race and Age	Race and Felony Charge
Unadjusted Rejection Rate	84%	98%
Prediction Mistake Share		
Minimum	54.87%	59.37%
Median	61.19%	71.30%
Maximum	72.55%	80.93%

Notes: This table calculates the estimated bound on the share of prediction mistakes among judges who were found to make prediction mistakes about pretrial misconduct risk. Among judges' whose release decisions violate the revealed preference inequalities at the nominal 5% level, this is calculate by constructing the optimal value of a mixed-integer linear program over the violations of the implied revealed preference inequalities.