Racial Disparities in Criminal Sentencing Vary Considerably across Federal Judges

by

Nicholas Goldrosen, Christian Michael Smith, Maria-Veronica Ciocanel, Rebecca Santorella, Shilad Sen, Shawn Bushway, and Chad M. Topaz*

Substantial race-based disparities exist in federal criminal sentencing. We analyze 380,000 recent (2006–2019) sentences in the JUSTFAIR database and show that these disparities are large and vary considerably across judges. Judges assign White defendants sentences 13% shorter than Black defendants' and 19% shorter than Hispanic defendants' sentences, on average, conditional on case characteristics and district. Judges one standard deviation above average in their estimated Black–White disparity give Black defendants sentences 39% conditionally longer than White defendants' sentences, vis-à-vis average disparity of 13%. Judges one standard deviation above average in their estimated Hispanic–White disparity give Hispanic defendants sentences 49% conditionally longer than White defendants' sentences, compared to the average disparity of 19%.

Keywords: sentencing, sentencing guidelines, racial disparities, judicial discretion

JEL classification code: J15, K14, K42

* Nicholas Goldrosen (corresponding author): University of Cambridge, United Kingdom. Christian Michael Smith: University of California, Merced (CA), USA. Maria-Veronica Ciocanel: Duke University, Durham (NC), USA; Institute for the Quantitative Study of Inclusion, Diversity, and Equity, Williamstown (MA), USA. Rebecca Santorella: Brown University, Providence (RI), USA. Shilad Sen: Macalester College, St. Paul (MN), USA; Microsoft Corporation, Redmond (WA), USA. Shawn Bushway: RAND Corporation and State University of New York at Albany (NY), USA. Chad M. Topaz: Institute for the Quantitative Study of Inclusion, Diversity, and Equity, Williamstown (MA), USA; University of Colorado, Boulder (CO), USA. The authors are grateful for the helpful comments of Christoph Engel, Daniel Nagin, and all the participants at the Seminar on Judicial Decision-Making sponsored by the Max Planck Institute for Research on Collective Goods. Author NG was a member of the advisory public user group for the Administrative Office of the Courts' electronic public access programs, which include PACER. This work does not make use of any nonpublic information, access, or data obtained in the course of that service.

Journal of Institutional and Theoretical Economics 179, 92–113 – ISSN 0932-4569
DOI: 10.1628/jite-2023-0005 –

1 Introduction

Racial inequalities in U.S. federal sentencing outcomes have decreased substantially over the last decade, from an average difference between Black and White[1] defendants of three years in 2009 to less than six months in 2018 (Light, 2022). This declining racial gap has been attributed to three factors: (1) Black and White defendants now have more similar case characteristics, (2) the federal system has reduced the punishments associated with drug crimes for all defendants, and (3) prosecutors now pursue mandatory minimum sentences in fewer cases (Light, 2022). None of these facts, however, address racially disparate treatment by individual actors within the courtroom workgroup, treatment that can still cause significant differences in sentencing outcomes. Indeed, the U.S. Sentencing Commission (USSC) found that "Black male offenders continued to receive longer sentences than similarly situated White male offenders" (USSC, 2017, p. 2), with sentences that are 19.1% longer, on average, after controlling for observable characteristics. The difference between similar Hispanic and White males is 5%.[2]

As Black and White defendants become more equal on case characteristics due to statutory, charging, and other factors, racially disparate treatment by individual actors like judges becomes even more important to identify and eradicate (Fisch-

[1] We recognize that the capitalization of the term "White" is debated. Capitalization can reflect that Whiteness is a discrete sociopolitical construct; simultaneously, not capitalizing white can reflect the lack of shared cultural context and historical discrimination against white people as compared to other racial and ethnic groups. We have opted to capitalize White in keeping with the style specifications of this journal.

[2] Racial disparities are a common, and well-researched, issue in the sentencing practices of U.S. federal courts. Most research on aggregate disparities in federal sentencing has focused on the disparity between Black and White defendants. Black defendants consistently receive harsher sentences than White defendants (Feldmeyer and Ulmer, 2011; Mustard, 2001; Rachlinski and Wistrich, 2017). Young Black men, in particular, are sentenced most harshly (Doerner and Demuth, 2010). Yang (2015) finds that Black–White disparities in sentencing increased after *United States v. Booker*, in which the Supreme Court held the U.S. Sentencing Guidelines to be advisory rather than mandatory. However, others find that no such increase occurred (Starr, 2013). While scholars debate whether *Booker* increased Black–White disparities in sentencing, there is little debate on whether such disparities have existed both before and after the decision. The literature on Hispanic–White sentencing disparities is more variable. Young Hispanic men do receive particularly harsh sentences (Doerner and Demuth, 2010). Still, some argue that the Hispanic–White sentencing disparity can be explained as a function of noncitizens being sentenced more harshly and Hispanic defendants being disproportionately noncitizens (Light, 2014; Light, Massoglia, and King, 2014). What research does exist on other groups finds that Asian American defendants are sentenced similarly to White defendants (Johnson and Betsinger, 2009). Due to unique federal jurisdiction in Indian Country, Native American defendants face much harsher sentences for crimes prosecuted federally than defendants of other races who commit crimes elsewhere and who would only face state prosecution (Droske, 2008). On the whole, though, Native American defendants do not receive harsher sentences than other defendants in the federal courts (Ulmer and Bradley, 2018). Young Native men, though, receive the harshest sentences of almost any group (Franklin, 2013).

man and Schanzenbach, 2012). This concern over disparity is long-standing: Bipartisan unease about variation in sentencing among judges was the main driver of the Sentencing Reform Act of 1984 that created then-mandatory sentencing guidelines to constrain judicial discretion (USSC, 2019; Frankel, 1973). Subsequent research has showed that some of this variation among judges was correlated with the racial composition of the district's bench (e.g., Schanzenbach, 2005).

The USSC has published three reports about variation in sentencing across judges – one in 1999 (Hofer, Blackwell, and Ruback, 1999), one in 2012 (USSC, 2012), and another in 2019 (USSC, 2019). All three revolve around the impact of changes in the guidelines and changes in the guidelines' legal status (e.g., the *Booker* decision making them advisory) on aggregate measures of variation among judges. Hofer, Blackwell, and Ruback (1999) found that the introduction of the USSC guidelines reduced the amount of variation in sentencing outcomes explained by judges by half. Somewhat ironically, the absolute size of the variation in prison time remained the same. Because the guidelines increased the average sentence length, though, the variation was a smaller proportion of the average sentence. The 2012 USSC report, which focused on the impact of the 2005 *Booker* Supreme Court decision, found that the variation in the rate at which judges granted downward departures increased once the guidelines were no longer mandatory. This decrease occurred in ways not attributable to case characteristics. This report was framed in part by an earlier report (USSC, 2010), which found that racial disparities had increased since the *Booker* decision in 2005.

Although in some sense this increase in variation was expected, there was considerable concern amongst academic researchers that increased disparity would be used to call for a return to the stricter guidelines and dramatically increase sentences while reducing disparity (Fischman and Schanzenbach, 2012; Ulmer, Light, and Kramer, 2011). The 2019 study on judicial discretion (USSC, 2019) was a renewed attempt to improve upon the 2012 study and to respond directly to some of the pointed concerns about the limitations of that report. This study looked at deviations from the guideline minimum rather than simply at the existence of a departure and dug more deeply into the interjudge differences within 30 courthouses with sufficient data to support the more comprehensive analysis. The main conclusion was striking: "In most cities, the length of a defendant's sentence increasingly depends on which judge in the courthouse is assigned to his or her case" (USSC, 2019, p. 7).

Not all the cities considered had high intra-courthouse variation – in some cities the range of the variation across judges was relatively minor. For example, Oklahoma City was found to have the smallest range across judges, with the harshest judge and the least harsh judge only varying by 6.9 percentage points in percent deviation from the guideline minimum, and the average judge providing sentences that were 19.3% below the guideline minimum. In contrast, the range for judges in Philadelphia was huge, with a spread of 63.8 percentage points. To put this in perspective, one judge had an average sentence that was about 6% higher than the guideline minimum, and one judge had an average sentence that was 58% below the guideline minimum. On average, judges in Philadelphia were almost 14 percentage points from the district average. To make this result even more concrete, consider that the average sentence minimum was 60 months (the national average). That means that the harshest judge in the Eastern District of Pennsylvania gave an average sentence of 66 months, while the most lenient judge gave an average sentence of 25 months. The average judge's deviation from the mean would be about seven months. These differences are tremendous and call to mind Judge Marvin Frankel's condemnation of existing judge disparities in the federal system as "horrible" and "lawless" (Frankel, 1973).

In interpreting these results, it is important to note that judges are not always randomly assigned to cases in the federal system. The USSC's (2019) analysis did not attempt to control for characteristics of cases or defendants that might be consequently unbalanced between judges. The study also did not attempt to identify interjudge differences in race-based sentencing disparities, only interjudge differences in overall severity. While this wide variation in judges' overall severity may be cause for concern on its own, it is particularly problematic if it is correlated with demographic characteristics of defendants. We attempt to address these limitations by studying the correlation between judicial variation in sentences that largely deviate downward from the guideline minimums and the race/ethnicity of the defendant, controlling for case characteristics. The main challenge is that neither the USSC nor the Federal Judicial Center (FJC) identifies the sentencing judge in their publicly available case-level data. According to Schanzenbach and Tiller (2008, p. 728), the "unavailability of judge data is one of the most frustrating aspects of the study of federal sentencing and has significantly impeded scholarly evaluation of the Guidelines' efficacy."

We will start in the next section with a brief review of the previous attempts to address this problem by finding alternative sources of data, before providing details on our attempt using data that are all publicly available. We then present our answers to four primary questions:

(1) What is the average disparity between the sentences of observationally equivalent Black and White defendants?
(2) What is the average disparity between the sentences of observationally equivalent Hispanic and White defendants?
(3) How much do conditional Black–White sentencing disparities vary across judges?
(4) How much do conditional Hispanic–White sentencing disparities vary across judges?

In the conclusion, we will discuss whether the limitations of our data undercut the data's ability to add value to the discussion around judge discretion, particularly in the context that the USSC does not release the information on judges that the USSC itself uses to highlight a fairly significant problem in cities like Philadelphia.

1.1 Literature on Interjudge Differences in Disparities

We focus on differences across federal judges in their sentencing disparities based on race. The literature on *interjudge* differences in racial disparities is much sparser than the literature on *aggregate* racial disparities. Some relevant evidence exists among judges in nonfederal courts – for example, Abrams, Bertrand, and Mullainathan (2012) found that judges in the Circuit Court of Cook County in Illinois differ substantially with regards to incarceration rates but not with regards to sentence lengths.

Both economics and criminology have begun to make widespread use of random assignment to judges to study variation in sentence outcomes (for reviews see Harding et al., 2018, and Humphries et al., 2022). These studies have relied on the fact that different judges have different underlying propensities in sentencing similar defendants, reflected in their average sentences, and then use that quasi-random variation to study the impact of sentencing on outcomes like employment and recidivism. In general, these studies have not focused on racial variation in the sentence imposed across judges but rather have identified causal effects of sentencing on other outcomes. Although these judge assignments as instrumental variable papers have been used in several cities (Philadelphia, DC), states (MI, PA, VA), and countries (Norway), we are not aware of any study that has made use of judicial variation in the U.S. federal data.

Within federal sentencing research, the original work on judicial variation was done by Waldfogel (1991) in three districts, relying on hand coding of cases amongst the 30 judges who worked in these three districts. Waldfogel assumed that judges are assigned to cases via random assignment. If this assumption is true, identical judges have similar average sentences for their dockets. He then interpreted the large variation he found across judges both before and after the onset of the Sentencing Reform Act and mandatory guidelines (on the order of 30–40 months) as evidence of substantial judicial discretion. He also concluded that the guidelines reduced that discretion, given some observed reduction in the differences between sentences imposed by the judges after the guidelines were promulgated. This work was extended by Payne (1997) in the same three districts. Payne focused on how the guidelines had dramatically increased sentence length, and she questioned whether a decrease in disparity could justify the increase in punishment.

Schanzenbach (2005) tried to make progress in the federal system in the absence of information on individual cases by looking at the variation among districts and differences in judge characteristics, particularly race and gender. While insightful, this approach conflates between- and within-district variation, which can be problematic in the absence of sufficient controls. Numerous studies have found evidence of substantial variation across districts in the federal system (USSC, 2020).

Schanzenbach and Tiller (2008) made a subsequent attempt at studying judge variation when they used data from PACER, the public access system for federal court records. PACER has information on the sentencing judge, although it does not reliably include structured sentencing data such as guideline calculations or enhancements. Accessing case files from PACER manually, Schanzenbach and Tiller looked at a random selection of 2000 cases to study variation across judges, controlling for the political party of the president that nominated the judge. They found that the observed variation in sentences corresponded to political ideology, a finding that led them to call for more transparency on judge identification to allow for further explorations of judge bias.

The judiciary did not heed this call for transparency, so Yang (2014, 2015) and Cohen and Yang (2019) created another federal sentencing dataset by merging the Sentencing Commission data with proprietary case-level data from the Transactional Records Access Clearinghouse (TRAC). The TRAC data are assembled through FOIA requests. Yang merged cases on the basis of date, sentence length, and other case characteristics to identify judges, with a match rate of 50–60% from the USSC data to an identifiable judge and a resulting dataset of well over 500,000 cases.

With this dataset, Yang found that both interjudge variation and racial disparity correlated with judges increased post-*Booker* (Yang, 2015). Like Schanzenbach and Tiller (2008), Cohen and Yang (2019) found the variation in average sentence length to be correlated with political affiliation of the judge, with Republican-appointed judges sentencing Black and male defendants more harshly than similar White and female defendants as compared to Democrat-appointed judges.

In what follows, we leverage a public federal dataset with judge identifiers called JUSTFAIR (Judicial System Transparency through Federal Archive Inferred Records), created using all public sources (Ciocanel et al., 2020). That dataset follows an approach similar to Yang's approach – matching USSC records to judge identifiers by unique case information – and has a similar match rate and size. Because we use different data sources for judge identity and characteristics, our dataset will vary in meaningful, if largely unknowable, ways from these previous analyses. We believe our analysis using the JUSTFAIR dataset nonetheless contributes to the understanding of federal sentencing disparities and can be triangulated with findings from other datasets and researchers.

2 *Methods*

2.1 Data Source

We analyze the JUSTFAIR database of criminal sentencing decisions in federal courts (Ciocanel et al., 2020). JUSTFAIR is compiled from five public sources, includes almost 600,000 records from fiscal years 2001–2018, and links information about defendants, their federal crimes, their sentences, and the sentencing judges. Here we briefly summarize the JUSTFAIR data pipeline presented by Ciocanel et al. (2020). Ciocanel and colleagues obtained information about criminal cases, defendants, and sentences from the USSC and merged this dataset with the FJC Integrated Database, another dataset maintained by the federal judiciary, to obtain court docket numbers for each case. The docket numbers allowed Ciocanel et al. to access the PACER system and retrieve the initials of the sentencing judges. Finally, they connected these initials with name, demographic, and education information of the judges from Wikipedia and from the FJC Biographical Directory of Article III Federal Judges. As a result, the records in JUSTFAIR contain variables pertaining to the demographic characteristics of the sentenced individual; sentence characteristics, date, and federal district location; and sentencing judge name, appointment, and education information. Crucially to the current work, JUSTFAIR also includes the sections of the law relevant to the conviction and factors influencing the guideline sentence.

We have augmented the JUSTFAIR dataset to include the available 2018–2019 fiscal year federal sentencing data, which is updated yearly in the USSC datafile for individual offenders and quarterly in the FJC Integrated Database. Our merging of offender, sentence, and judge information proceeded largely as described above and by Ciocanel et al. (2020). Several USSC variables denoting the total prison sentence, the coding of the offense type, and the post-*Booker* reporting categories changed in the USSC database during this period; therefore, we adjusted the data-processing approach of Ciocanel et al. (2020) to remain consistent with the variables in JUSTFAIR. This extended the dataset we analyze in this study by over 30,000 cases.

In our analysis, we include cases only from 2006 and later. We are principally concerned with racial disparities observed after the *Booker* decision, and so cases before 2006 lie outside of our target of study. We further exclude immigration cases because of the unique use of fast-track sentencing in these cases and the extreme concentration of these cases in several southwestern district courts (Hartley and Tillyer, 2012). We also do not consider cases for which we cannot infer the sentence length; specifically, in the JUSTFAIR database, there are about 13,000 cases that resulted in a sentence length of zero according to the continuous sentence total variable but resulted in prison time according to the categorical variable of imprisonment type. These two features contradict each other; we therefore remove these cases from the analysis. After these pre-processing steps, the analytic sample represents about 380,000 cases corresponding to 1116 sentencing judges. The median number of cases per judge is 263 (mean = 358.7, SD = 389.6).

Ciocanel et al. (2020) detail several measures they took to ensure the quality of the JUSTFAIR database. They validated merged records by checking for additional common variables (such as offense codes) in the original datasets, used two independent sources to identify judge names from judge initials, and excluded records where sentencing dates fall outside the judges' activity periods. In addition, they carried out a manual data validation procedure for randomly sampled cases from the assembled dataset (Ciocanel et al., 2020). The validation set confirms the accuracy of the merging procedure: Very few cases violate the assumption that the judge involved in a proceeding is the same as the sentencing judge. Nevertheless, while they corrected these cases in the validation set, there could be rare instances where a judge other than the sentencing judge became involved in proceedings post-sentencing, leading to an incorrect inference in the database.

While JUSTFAIR is, to our knowledge, the largest publicly available database of federal sentences currently available, it nevertheless remains an incomplete database of the cases sentenced in 2001–2018 (extended in this work to 2019) due to issues pertaining to data quality and merging challenges. For example, when merging USSC and FJC sentence and defendant information, JUSTFAIR follows Yang (2015) and only retains cases where there is a unique matching. Similarly, when retrieving sentencing judge initials from PACER, criminal cases may include more than one defendant, and therefore JUSTFAIR only keeps records where all defendants with the same court docket number are associated with the same judge. JUSTFAIR also cannot include information on sealed federal cases.

Other factors that prevent JUSTFAIR from being complete include the inability to distinguish between judges who have both the same initials and district, and the complete or partial lack of judge initials in certain districts or records. In particular, due to such data quality issues, JUSTFAIR contains no sentencing data from the Eastern District of North Carolina, the Southern District of West Virginia, the Southern District of Texas, the Middle District of Tennessee, the Northern District of Illinois, the District of Guam, and the District of the Northern Mariana Islands. The database also includes limited sentencing data (less than 33% of the starting USSC cases) from the Northern District of Texas, the Southern District of California, the District of Oregon, the District of New Mexico, the Western District of Oklahoma, and the Northern District of Florida.

Nonrandom assignment of cases to judges presents another limitation, not just in the case of the JUSTFAIR database but whenever federal judge effects are of interest. The Administrative Office of the U.S. Courts claims that "The majority of courts use some variation of a random drawing," meaning "to assure equitable distribution of caseloads and avoid judge shopping" (Administrative Office of the United States Courts, 2020), the district courts have a rotation plan for cases assigned to judges; however, the Administrative Office of the Courts mentions that special expertise and geography are also considered in case assignment. Previous studies have exploited the random assignment of cases to judges in their analyses of sentencing equity (Anderson, Kling, dan Stith, 1999; Abrams, Bertrand, and Mullainathan, 2012; Cohen and Yang, 2019). Establishing random assignment has been considered key for these studies since it ensures that each judge receives a similar combination of cases and defendants (in terms of observable characteristics), so that unobservable characteristics can also be assumed to be similar across judges.

For instance, Abrams, Bertrand, and Mullainathan (2012) analyze racial sentencing disparities in felony cases from Cook County, Illinois. To verify random assignment, they use a Monte Carlo simulation methodology for felony data from 1995 to 2001 to construct a randomly assigned dataset across characteristics such as defendant race, age, gender, and crime category (Abrams, Bertrand, and Mullainathan, 2012). Similarly, the study of the proprietary federal sentence dataset in Cohen and Yang (2019) relies on the assumption that cases are randomly assigned to judges in the same district court, focusing on observed case and defendant characteristics across Republican-appointed and Democrat-appointed judges. Motivated by whether the political affiliation of a judge's appointing president influences racial and gender gaps in sentencing decisions, the authors use a joint F-test to test whether there are significant differences in defendant characteristics by judge political affiliation (as well as by judge tenure and gender). Conditioning on sentencing year and district court fixed effects, they find that case characteristics are distributed across judges from each political affiliation in a way consistent with random assignment (Cohen and Yang, 2019).

Our setting is different from these studies, given that we are (1) analyzing the large, publicly available JUSTFAIR dataset of federal sentences from 2001–2018 (Ciocanel et al., 2020) and (2) interested in whether the cases in this dataset are assigned randomly to individual judges within each district (rather than to judges with different characteristics). We find that, when testing for random assignment using an F-test method similar to Cohen and Yang (2019) or using a Monte Carlo simulation method similar to Abrams, Bertrand, and Mullainathan (2012), nearly every district shows evidence of nonrandom judge assignment. Specifically, nearly every district shows statistically significant between-judge differences in at least one case characteristic, such as defendant race, defendant sex, and offense type. Due to nonrandom assignment or the process by which cases enter our dataset, some unknown component of between-judge variability in conditional racial disparities arises from differences in unobserved factors that should reasonably influence the sentence. Observed covariates probably make this component much smaller than it would be otherwise, but we cannot know how substantial the component remains after including these covariates.

2.2 Measures

The outcome of interest is the length of the sentence assigned to the defendant. We cap the sentence length at 470 months because life sentences are generally coded as 470 months. We log-transform the outcome because its distribution has a positive skew (Bushway and Piehl, 2001).[3] Accordingly, we interpret the results in terms of

[3] Because the log of zero is undefined, we add one to every sentence before log-transforming so that nonincarceration outcomes (zero-month "sentences") have a defined value in our estimation. It is challenging to address zero-month sentences in analyses where sentence length is the dependent variable, and there is no perfect solution to the issue. Adding one to all sentences is a common approach (Fischman and Schanzenbach, 2012; Light, 2022; Rehavi and Starr, 2014; Yang, 2015). This approach appreciates the fact that nonincarceration outcomes are real outcomes that can contribute to racial disparities, while also accounting for the fact that a nonincarceration outcome entails less incarceration time than any other sentence. Nevertheless, we acknowledge that this approach masks nuance in the difference between a nonincarceration outcome and a one-month sentence.

percentage differences in sentences of imprisonment rather than linear differences in months of prison time.

The case-level characteristic of principal interest is the defendant's race and ethnicity, which we measure with the categories *Hispanic*, *Non-Hispanic Black* (henceforth *Black*), *Non-Hispanic White* (henceforth *White*), and *another racial identity* (henceforth *ARI*). We collapse the groups in this way because, unfortunately, the sample sizes of the various groups contained in the final category are not large enough for us to perform statistically informative cross-judge analyses. Using these definitions of race and ethnicity, we determined that the median judge imposed a sentence in 41 cases with Hispanic defendants (mean = 97.3, SD = 210.6) and in 69 cases with Black defendants (mean = 114.1, SD = 137.7).

We include an extensive set of control variables:

- the guideline minimum sentence, with any statutory minimum sentences taken into account,[4]
- the defendant's criminal history points,
- crime type, namely *violent crime*, *drug-related crime*, *embezzlement / fraud / theft*, or *other*,
- whether the case was settled by guilty plea or trial,
- whether the government sponsored a downward departure,
- sentencing year, and
- defendant demographics, namely sex, age, U.S. citizenship status, and educational attainment (measured with a dummy variable for each level of education).

Arguably the most important of these control variables is the guideline minimum sentence, which is standardized based on factors set by the USSC. Thus, at least in a theoretical world with no racial disparities, we would expect to see roughly equal sentences imposed on two defendants who are from different racial groups but who have the same guideline minimum sentence.

While the defendant's demographics are extralegal factors that judges ought not consider in their sentencing, we nevertheless control for these demographics because they may capture variation in relevant legal factors. For example, if a judge hears cases where many defendants have a unique situation not captured by observed legal variables, and if this unique situation also is associated with age, then controlling for age accounts for some of the variation due to the unobserved prevalence of this situation.

[4] We recognize that, given the judge's role in determining guideline enhancements, using this as a control might introduce endogeneity issues. In line with Light (2022), we believe any such endogeneity should make the estimates of disparity more conservative. As such, our estimates do not capture disparity that might arise through the judicial determination of the base offense level itself or the statutory minima that might apply due to the judge–prosecutor interaction and the charges sought (see, e.g., Kim, Spohn, and Hedberg, 2015).

2.3 Analytic Strategy

To identify plausible sources of racial disparity in federal sentences, we construct a series of fixed- and random-effect models. Since sentences might vary due to case and defendant characteristics, the sentencing judge, and the sentencing district, we fit models that address three key questions:

(1) How do sentences vary based on defendant race and case characteristics? (models 1 and 2);
(2) How do sentences vary based on defendant race, district, and judge? (models 3, 4, and 5); and
(3) How do disparities in sentence based on race vary between judges (model 6)?

Model 1 predicts a sentence, Y_i, based upon a defendant's race as follows:

$$\log(Y_i + 1) = \beta_0 + \vec{\beta}_1 \vec{r}_i + \varepsilon_i,$$

where β_0 is the average log-transformed sentence in the reference category (White defendants), $\vec{\beta}_1$ is a vector of slopes relating log-transformed sentence length to the vector $\vec{r}_i$ of dummy variables for defendant race, and ε_i is the error term associated with case i.

We then estimate this same model, but with the addition of case and defendant characteristics, as described in section 2.2 (model 2). Model 3 adds in district fixed effects along with these controls.[5] We then turn to estimating the interaction of district fixed effects and defendant race in model 4. Model 5 considers sentences as a function of judge fixed effects and district fixed effects, with controls but no interactions with defendant race.

Finally, for model 6, we estimate a hierarchical linear model of log-transformed sentence lengths where cases (level 1) are nested within federal judges (level 2). In particular, we estimate the following model:

$$\begin{aligned}
\log(Y_{ijk} + 1) &= \beta_{0j} + \vec{\beta}_{100} \vec{X}_{ij} + \vec{\beta}_{1j} \vec{C}_{ij} + s_{ij} && \text{(level 1)},\\
\beta_{0j} &= \gamma_{000} + \zeta_{0j} && \text{(level 2, intercepts)},\\
\beta_{1jk} &= \delta_{100k} + \eta_{1jk} && \text{(level 2, slopes)},
\end{aligned}$$

where Y_{ijk} is the length of the sentence corresponding to case i by judge j for a defendant of race k, β_{0j} is the intercept for cases heard by judge j, $\vec{X}_{ij}$ is a vector of defendant control characteristics corresponding to case i by judge j, $\vec{\beta}_{100}$ is a vector of slopes between control characteristics and log-transformed sentence length, $\vec{C}_{ij}$ is a vector of racial group binary indicators, $\vec{\beta}_{1j}$ is a vector of random slopes between logged sentence length and each racial group indicator among cases heard by judge j, s_{ij} is an idiosyncratic error term, γ_{000} is the overall intercept,

[5] For the reference category, we use the district with the median mean sentence length, the Northern District of New York (NDNY).

ζ_{0j} is the deviation of the intercept for cases heard by judge j from the overall intercept, β_{1jk} is the random slope corresponding to racial group k represented in β_{1jk}, δ_{100k} is the overall slope between log-transformed sentence length and racial group indicator k, and η_{1jk} is the deviation of the slope for cases heard by judge j from the overall slope corresponding to racial group k. Our model assumes that s_{ij}, ζ_{0j}, and η_{1jk} each are normally distributed with mean zero.

We are primarily interested in the values of η_{1jk} for each judge. Taking the example where $k =$ Black and the reference category is White defendants, the value of $\eta_{1j\text{Black}}$ will answer the following question: how much greater is the Black–White sentencing disparity when judge j hears cases compared to when the average judge hears cases, controlling for defendant and case characteristics? This value does not necessarily reflect directly unfair treatment based on race, one reason for this being that a judge may hear cases in which Black and White defendants are unbalanced with respect to unobserved legal characteristics, even above and beyond observed control variables. The general approach of conditioning on observed factors is nevertheless common in research on racial sentencing inequalities (Feldmeyer and Ulmer, 2011; Light, 2022), in part because cases do not seem to be randomly assigned to judges even within districts. Regarding variability in $\eta_{1j\text{Black}}$ values, we note that, because our model estimation shrinks random slopes and intercepts for judges with a small number of cases, such judges do not cause random slopes and intercepts to appear unduly variable.

The model includes judge random effects instead of judge fixed effects because the former accommodates random slopes, which are the primary parameters of interest in this study. While we also fit judge fixed-effect models instead of a hierarchical model (model 5), these are mainly for the purpose of seeing how collinear these effects are with district effects – that is, to what extent district effects are explained by the sentencing practices of the judges that sit in the district rather than by collective workgroup norms. We forgo a hybrid model because it is less parsimonious and does not influence the estimated random slopes (Schunck, 2013), our primary interest. We estimated an alternative model with district fixed effects and the results did not differ in any substantial way. Thus, our preferred specification is more parsimonious and does not affect the main results.

3 Results

3.1 Aggregate Disparities

The estimates for racial disparity from models 1, 2, 3, 4, and 5 are displayed in Table 1. Without controls, Black defendants received sentences 61% greater than White defendants and Hispanic defendants received sentences that are 40% greater. All defendants, on average, receive a downward deviation from the guideline (or statutory, if higher) minimum, but this deviation is greater for White defendants (mean = −20.2) than for Black (mean = −13.8) or Hispanic (mean = −15.9)

Table 1
Fixed-Effect Model Coefficients on Defendant Race

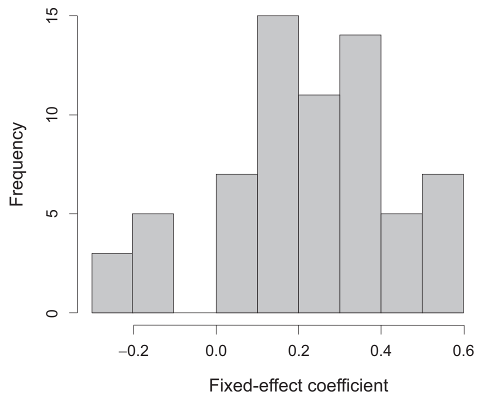
	No controls (model 1)	With controls (model 2)	With controls & district FEs (model 3)	With controls, & district FEs interacted with race (model 4)	With controls, district FEs, & judge FEs (model 5)
Black	0.482 (0.006)	0.130 (0.005)	0.101 (0.005)	0.565 (0.049) (ref = NDNY)	0.108 (0.007)
Hispanic	0.335 (0.006)	0.099 (0.006)	0.149 (0.006)	0.278 (0.066) (ref = NDNY)	0.216 (0.010)
Another racial identity	−0.143 (0.011)	−0.152 (0.009)	−0.064 (0.010)	0.331 (0.096) (ref = NDNY)	−0.136 (0.016)
Adjusted R^2	0.020	0.403	0.422	0.425	0.439

defendants. The standard deviation of this variance from the minimum is also greater for White defendants (SD = 126.3) than for Black (SD = 68.3) or Hispanic (SD = 59.6) defendants.

Adding in controls for case and defendant characteristics, these coefficients shrink dramatically; the unexplained disparity is approximately 14% and 10% for Black–White and Hispanic–White, respectively. The addition of district fixed effects further decreases the unexplained Black–White disparity but increases the unexplained Hispanic–White disparity by about 50%. This change indicates that districts might vary more in their sentencing practices for Hispanic defendants vis-à-vis White defendants, but that the Black–White disparity is explained less by disparity between judges within each district and more by the disproportionate presence of Black defendants in overall less-lenient districts.

Our aggregate results are in line with previous studies, though they perhaps point to even more aggregate disparities than found in earlier timespans with different sampling mechanisms. In an investigation of a 2000–2002 pre-*Booker* USSC dataset, Feldmeyer and Ulmer (2011) find that, after controlling for a considerable set of legal and extralegal factors, Black defendants' sentences are 6% longer and Hispanic defendants' sentences are 1% longer on average than White defendants' sentences. In an analysis of a 2006–2008 multi-agency-based dataset, Rehavi and Starr (2014) found that, in the aggregate, Black defendants receive sentences that are 9% longer than those of similarly situated White defendants who commit the same crimes. Finally, using a federal sentencing dataset similar to ours but covering cases from 2000 to 2010, Yang (2015) found that Black offenders receive 1.9 months longer sentences, and Hispanic offenders over 1.9 months longer sentences, than those of similar White offenders post-*Booker*.

Figure 1
Histogram of Significant District Fixed Effects with the Reference District as the Northern District of New York



3.2 Interdistrict Variation

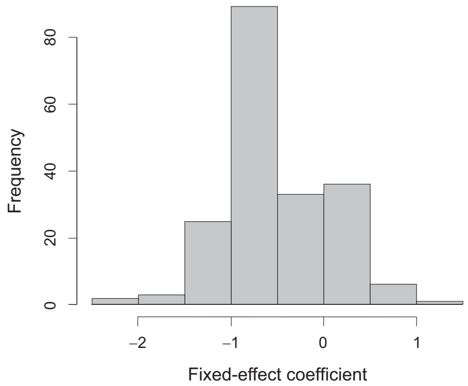
Most districts with statistically significant fixed effects have positive fixed effects in model 3; that is, their sentencing practices are harsher than the reference district conditional on case characteristics. The distribution of these fixed effects is shown in Figure 1.

To triangulate our findings on district variation, we compared the districts we identified as most and least harsh to the findings from the USSC (2020) report examining interdistrict variation. While that report looks at the mean sentence with regards to certain specific guidelines (fraud, drugs, firearms, and illegal reentry), we cross-reference our district fixed effects with the first three of those guidelines, as we have excluded immigration offenses from our sample. Seven of the eight most lenient districts[6] from our sample were amongst the most lenient for two or three of the USSC-studied guidelines. For the ten most severe districts we identified,[7] all ten were amongst the harshest for two or three of the USSC-studied guidelines. These results, for both above-average and below-average districts, show a good degree of concordance between our results and those of the USSC. For the district–race interaction in model 4, almost every district has a significant interaction term for Black defendants, and almost all are negative compared to NDNY. The interactions between districts and Hispanic defendants are more mixed, both in direction and significance.

[6] District of Vermont, District of Rhode Island, District of Arizona, District of New Mexico, Eastern District of New York, Western District of New York, and District of South Dakota.

[7] Southern District of Georgia, Western District of North Carolina, Northern District of Georgia, Southern District of Florida, Eastern District of Virginia, Northern District of Iowa, Eastern District of Tennessee, Middle District of North Carolina, Eastern District of Texas, and Central District of Illinois.

Figure 2
Histogram of Significant Judge Fixed Effects, with the Reference Judge as the Judge with the Median Mean Sentence in 38 Districts with High Match Rates



3.3 Interjudge Variation

In model 5, we seek to examine how much interdistrict variation in sentencing can best be explained as interjudge variation between districts, and how much might be due to collective workplace norms around sentencing within districts. For the purposes of this model only, we subset the data to only districts where the match rate exceeds two-thirds of all cases (38 districts out of 87 in our dataset). The reference judge is the median judge in terms of mean sentence from these districts. With judge and district fixed effects, the race main effect persists for Black defendants – about 10% greater sentences for Black defendants. This indicates that the Black–White disparity is not driven by between-judge variation in harshness but perhaps by the disparate representation of White defendants in districts that are collectively more lenient towards White defendants, as compared to Black defendants. The Hispanic–White disparity grows to about 22% when judge fixed effects are added. This indicates that significant within-judge variation (perhaps due to unobservable contextual case or defendant characteristics) might be driving this disparity. Finally, the statistically significant judge fixed effects are displayed in Figure 2. A larger proportion of judges are more lenient than the mean, which tracks with the overall initial finding that the average defendant of any race is receiving some leniency from the guidelines, but that White defendants receive more. This accords with the USSC's conclusions that both the average district (USSC, 2020) and the average judge in a subset of courthouses (USSC, 2019) sentence below the guideline minimum, but to varying degrees.

Table 2
Dispersion of Random Effects from Model 6

	Standard deviation
Random intercepts	0.35
Random Black slopes	0.21
Random Hispanic slopes	0.23
Random another racial identity slopes	0.31

Notes: Random intercepts represent judges' estimated sentencing severity toward White defendants, and random slopes represent judges' estimated racial disparities compared to the average judge's estimated racial disparities.

3.4 Interjudge Variability in Disparities

How much do judges vary in their estimated racial disparities in sentencing? The estimates in the previous sections show only the average extent of estimated racial disparities and the overall average disparities among judges. We turn now to a discussion of how much judges are spread around the average disparity based on race. Table 2 shows that the spread is considerable for each racial group, as estimated in model 7, with random slope standard deviations ranging from 0.21 to 0.31 log-transformed years. The random intercepts, estimating judges' sentencing severity against White defendants, also vary to a similar extent.

Because interpreting the foregoing standard deviations is challenging given the log-transformed dependent variable, we present a more intuitive account of the interjudge variability in Table 3. The average judge assigns Black defendants sentences that are conditionally 13% longer than White defendants' sentences. However, a judge who is one standard deviation above average in terms of estimated Black–White disparity (14.9% of judges in our data) assigns Black defendants sentences that are conditionally 39% longer than White defendants', and a judge who is two standard deviations above average (3.1% of judges in our data) assigns Black defendants sentences that are conditionally 71% longer than White defendants' sentences. Similarly, while the average judge assigns Hispanic defendants sentences that are conditionally 19% longer than White defendants' sentences, a judge who is one standard deviation above average in terms of estimated Hispanic–White disparity (15.2% of judges) assigns Hispanic defendants sentences that are 49% longer than White defendants' sentences. A judge who is two standard deviations above average (2.9% of judges) assigns Hispanic defendants sentences that are conditionally 87% longer than White defendants' sentences.

To test whether the variance in random slopes is significantly different from zero, we conduct a likelihood ratio test that compares our model to a model that excludes random slopes but is otherwise identical. The p-value resulting from this test is less than 2.2×10^{-16}, suggesting conditional racial disparities in sentencing really do vary across judges. The Bayesian Information Criterion for the model with random

Table 3
Percent Change in Sentence Length Conditionally Associated with the Defendant Being in Different Racial Groups*

	% Change (avg.)	% Change (1 SD)	% Change (2 SD)
Black	12.77	38.94	71.20
Hispanic	19.15	49.35	87.20
ARI	−10.28	22.05	66.05

Notes: * Relative to being White by the average judge (column 1), a judge one standard deviation above average (column 2), and a judge two standard deviations above average (column 3) in terms of estimated racial disparities.

slopes is 1,209,679, compared to a much less favorable value of 1,211,157 for the model without random slopes. In short, if all judges' conditional racial disparities were the same, the aforementioned values would be surprising.

Triangulating with other research helps bolster confidence in the results above. Our analysis of JUSTFAIR data found that judges from the Eastern District of Pennsylvania had the highest amounts of racial disparity and the USSC identified the Eastern District of Pennsylvania as the place with the largest deviation between judges. Furthermore, an analysis by the Transactional Records Access Clearinghouse identified Philadelphia and Tampa as two of the five courthouses with large overall interjudge disparities (TRAC, 2021); in our results, multiple judges that we identified as having the greatest conditional Black–White or Hispanic–White disparities work in these districts.

These models result in three key findings. First, accounting for district fixed effects reduced the unexplained racial disparity between Black and White defendants but not between Hispanic and White defendants (model 3). This first finding indicates that, for Black defendants, the district of sentencing explains part of the gap with White defendants. Second, the Hispanic–White gap grows and the Black–White gap remains when accounting for district and judge fixed effects (model 4). This second finding indicates that, particularly for Hispanic defendants, the disparity with White defendants is not the result of being sentenced by judges that are overall harsher towards all defendants. Finally, the wide variation in judge's individual racial disparities does not decrease when accounting for district fixed effects (model 6). This finding indicates that judges with particularly wide disparities are not grouped in certain districts due to collective sentencing norms or other factors. In sum, the Black–White disparity is likely driven by Black defendants being more likely to be sentenced in districts that are harsher towards all defendants, and not by being sentenced in districts with particularly discriminatory judges. We do not see evidence of a similar explanation for Hispanic defendants.[8] Nonetheless, both Black–White and Hispanic–White disparities vary widely between judges.

[8] Crucially, our sample excludes immigration offenses, a category that makes up 68.4% of Hispanic defendants in the federal courts (USSC, 2021).

4 Conclusion

This paper analyzes the new JUSTFAIR database of federal district court criminal sentences, which is an incomplete database of federal sentence, defendant, and judge characteristics based on merging the USSC dataset and the FJC Integrated Database, as well as matching with other publicly available data. Ciocanel et al. (2020) describe this dataset in detail. While not without its imperfections, we believe this dataset has the potential to reinvigorate the study of judicial variation within the federal system.

Using this database, we computed a series of models, beginning with district and judge fixed-effect models and building to a hierarchical linear model of log-transformed sentence length with cases nested within federal judges. We controlled for factors such as recommended sentence, crime type, whether the defendant pleaded guilty, whether the government sponsored a downward departure, sentencing year, defendant demographics, and more. From our model, we calculated the percentage by which judges' sentences given to Black and Hispanic defendants are conditionally greater than the sentences judges give to White defendants.

Consistent with prior research, we find that White defendants are sentenced more leniently than Black and Hispanic defendants. Case characteristics explain a large portion of the racial disparity for both Black and Hispanic defendants, as do district fixed effects for Black defendants, but not for Hispanic defendants. While defendants of all races receive sentences, on average, below the applicable guideline or statutory minimum, White defendants are treated most leniently and with greater variation in this leniency. This finding concords with work by Smith, Levinson, and Robinson (2014) arguing for White favoritism, rather than animus against non-White defendants, as a lens for understanding racial disparity in criminal justice outcomes. From our results, we also find that Black defendants are likelier to be sentenced in districts that are less lenient overall; this is not the case for Hispanic defendants. Adding judge effects causes mixed changes to the magnitudes of district effects, indicating some districts are harsh or lenient due to collective norms and others due to interjudge variation.

Our findings suggest that average racial disparities are sizeable and that judges vary nontrivially in terms of how racially disparate their sentencing patterns are. With respect to Black–White disparities, Black defendants on average receive 13% longer sentences than observationally equivalent White defendants receive, and a judge one standard deviation above average in Black–White disparities gives 39% longer sentences to Black defendants. With respect to Hispanic–White disparities, Hispanic defendants on average receive 19% longer sentences than observationally equivalent White defendants receive, and a judge one standard deviation above average in Hispanic–White disparities gives 49% longer sentences to Hispanic defendants. The online supplementary materials for this paper include a replication package with the data and code we used to yield these findings.

We emphasize some words of caution in interpreting our results. First, because cases are not all randomly assigned to judges, we cannot know to what extent unobserved factors that systematically differ across judges drive the interjudge differences we find. A second limitation is our inability to observe every case, or even every district. Like most social science datasets, the JUSTFAIR database contains a sample rather than the population, and as with most social science datasets, one cannot know exactly what determines whether a case in the sampling frame is missing. If sample inclusion is unrelated to our measures of interest, then lacking the full population leaves our estimates more susceptible to random sampling error than they would be otherwise. If sample inclusion is related to our measures of interest, then lacking the full population leaves our estimates more susceptible both to systematic error and random sampling error than they would be otherwise. In light of these limitations, our results should be seen as imperfect approximations for the degree of aggregate racial disparities and the interjudge variation therein. However, we find that our results are comparable to those of Cohen and Yang (2019) and Yang (2014, 2015), who use similar overall strategies of merging public datasets.

Future research should also explore how well the analysis with either the TRAC data or JUSTFAIR matches up with the between-judge analysis conducted by the Sentencing Commission (USSC, 2019). Prima facie evidence is at least supportive that JUSTFAIR does capture some real variation in sentences among judges. The interdistrict results from our model 3 match well with the Commission's 2020 interdistrict report, indicating that JUSTFAIR's sample may be like theirs in crucial ways. Similarly, key results from our model 6 about districts with unusually large variation in racial disparities between judges align with research by the USSC and TRAC. While the JUSTFAIR dataset has significant limitations, we believe this analysis is a reliable contribution to and extension of existing research on federal sentencing disparities.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan (2012), "Do Judges Vary in their Treatment of Race?" *The Journal of Legal Studies*, 41(2), 347–383.

Administrative Office of the United States Courts (2020), "FAQs: Filing a Case," http://www.uscourts.gov/faqs-filing-case#faq-How-are-judges-assigned-to-cases? accessed December 22, 2020.

Anderson, James M., Jeffrey R. Kling, and Kate Stith (1999), "Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines," *The Journal of Law & Economics*, 42(S1), 271–308.

Bushway, Shawn D., and Anne Morrison Piehl (2001), "Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing," *Law & Society Review*, 35(4), 733–764.

Ciocanel, Maria-Veronica, Chad M. Topaz, Rebecca Santorella, Shilad Sen, Christian Michael Smith, and Adam Hufstetler (2020), "JUSTFAIR: Judicial System Transparency through Federal Archive Inferred Records," *PLoS ONE*, 15(10), DOI: 10.1371/journal.pone.0241381.

Cohen, Alma, and Crystal S. Yang (2019), "Judicial Politics and Sentencing Decisions," *American Economic Journal: Economic Policy*, 11(1), 160–191.

Doerner, Jill K., and Stephen Demuth (2010), "The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in U.S. Federal Courts," *Justice Quarterly*, 27(1), 1–27.

Droske, Timothy J. (2008), "Correcting Native American Sentencing Disparity Post-*Booker*," *Marquette Law Review*, 91(3), 723–813.

Feldmeyer, Ben, and Jeffrey T. Ulmer (2011), "Racial/Ethnic Threat and Federal Sentencing," *Journal of Research in Crime and Delinquency*, 48(2), 238–270.

Fischman, Joshua B., and Max M. Schanzenbach (2012), "Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums," *Journal of Empirical Legal Studies*, 9(4), 729–764.

Frankel, Marvin E. (1973), *Criminal Sentences: Law Without Order*, Hill & Wang, New York (NY).

Franklin, Travis W. (2013), "Sentencing Native Americans in US Federal Courts: An Examination of Disparity," *Justice Quarterly*, 30(2), 310–339.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway (2018), "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment," *American Journal of Sociology*, 124(1), 49–110.

Hartley, Richard D., and Rob Tillyer (2012), "Defending the Homeland: Judicial Sentencing Practices for Federal Immigration Offenses," *Justice Quarterly*, 29(1), 76–104.

Hofer, Paul J., Kevin R. Blackwell, and R. Barry Ruback (1999), "The Effect of the Federal Sentencing Guidelines on Interjudge Sentencing Disparity," *The Journal of Criminal Law and Criminology*, 90(1), 239–322.

Humphries, John Eric, Aurelie Ouss, Megan T. Stevenson, Winnie van Dijk, and Kamelia Stavreva (2022), "Measuring Effects of Conviction and Incarceration on Recidivism Using Multi-Treatment Random Judge Designs," Working Paper presented at NBER Summer Institute 2022, July 28, 2022.

Johnson, Brian D., and Sara Betsinger (2009), "Punishing the 'Model Minority': Asian-American Criminal Sentencing Outcomes in Federal District Courts," *Criminology*, 47(4), 1045–1090.

Kim, Byungbae, Cassia Spohn, and E. C. Hedberg (2015), "Federal Sentencing as a Complex Collaborative Process: Judges, Prosecutors, Judge–Prosecutor Dyads, and Disparity in Sentencing," *Criminology*, 53(4), 597–623.

Light, Michael T. (2014), "The New Face of Legal Inequality: Noncitizens and the Long-Term Trends in Sentencing Disparities across U.S. District Courts, 1992–2009," *Law & Society Review*, 48(2), 447–478.

— (2022), "The Declining Significance of Race in Criminal Sentencing: Evidence from US Federal Courts," *Social Forces*, 100(3), 1110–1141.

—, Michael Massoglia, and Ryan D. King (2014), "Citizenship and Punishment: The Salience of National Membership in U.S. Criminal Courts," *American Sociology Review*, 79(5), 825–847.

Mustard, David B. (2001), "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts," *The Journal of Law & Economics*, 44(1), 285–314.

Payne, A. Abigail (1997), "Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts," *International Review of Law and Economics*, 17(3), 337–366.

Rachlinski, Jeffrey J., and Andrew J. Wistrich (2017), "Judging the Judiciary by the Numbers: Empirical Research on Judges," *Annual Review of Law and Social Science*, 13, 203–229.

Rehavi, M. Marit, and Sonja B. Starr (2014), "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 122(6), 1320–1354.

Schanzenbach, Max (2005), "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics," *The Journal of Legal Studies*, 34(1), 57–92.

Schanzenbach, Max M., and Emerson H. Tiller (2008), "Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform," *The University of Chicago Law Review*, 75(2), 715–760.

Schunck, Reinhard (2013), "Within and Between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models," *The Stata Journal*, 13(1), 65–76.

Smith, Robert J., Justin D. Levinson, and Zoë Robinson (2014), "Implicit White Favoritism in the Criminal Justice System," *Alabama Law Review*, 66(4), 871–923.

Starr, Sonja (2013), "Did *Booker* Increase Disparity? Why the Evidence Is Unpersuasive," *Federal Sentencing Reporter*, 25(5), 323–326.

TRAC (2021), "Equal Justice and Sentencing Practices among Federal District Court Judges," Technical Report, https://trac.syr.edu/tracreports/judge/663/, accessed September 6, 2022.

Ulmer, Jeffrey T., and Mindy S. Bradley (2018), "Punishment in Indian Country: Ironies of Federal Punishment of Native Americans," *Justice Quarterly*, 35(5), 751–781.

—, Michael T. Light, and John H. Kramer (2011), "Racial Disparity in the Wake of the *Booker/Fanfan* Decision: An Alternative Analysis to the USSC's 2010 Report," *Criminology & Public Policy*, 10(4), 1077–1118.

USSC (2010), "Demographic Differences in Federal Sentencing Practices: An Update of the *Booker Report's* Multivariate Regression Analysis," Technical Report, United States Sentencing Commission, Washington (DC).

— (2012), "Report on the Continuing Impact of *United States v. Booker* on Federal Sentencing," Technical Report, United States Sentencing Commission, Washington (DC).

— (2017), "Demographic Differences in Sentencing: An Update to the 2012 *Booker* Report," Technical Report, United States Sentencing Commission, Washington (DC).

— (2019), "Intra-City Differences in Federal Sentencing Practices: Federal District Judges in 30 Cities, 2005 – 2017," Technical Report, United States Sentencing Commission, Washington (DC).

— (2020), "Inter-District Differences in Federal Sentencing Practices: Sentencing Practices across Districts from 2005 – 2017," Technical Report, United States Sentencing Commission, Washington (DC).

— (2021), "Fiscal Year 2020: Overview of Federal Criminal Cases," Technical Report, United States Sentencing Commission, Washington (DC).

Waldfogel, Joel (1991), "Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts (D.Ct., S.D.N.Y., N.D.Cal.)," *Federal Sentencing Reporter*, 4(3), 151–154.

Yang, Crystal S. (2014), "Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from *Booker*," *New York University Law Review*, 89(4), 1268–1342.

— (2015), "Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing," *The Journal of Legal Studies*, 44(1), 75–111.

Nicholas Goldrosen
Institute of Criminology
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DA
United Kingdom
ncg36@cam.ac.uk

Christian Michael Smith
University of California, Merced
5200 North Lake Road
Merced, CA 95343
USA
csmith97@ucmerced.edu

Maria-Veronica Ciocanel
Duke University
Physics Building
120 Science Drive
Box 90320
Durham, NC 27708
USA
ciocanel@math.duke.edu

Rebecca Santorella
Brown University
Providence, RI 02912
USA
rebecca_santorella@alumni.brown.edu

Shilad Sen
Macalester College
1600 Grand Ave.
St. Paul, MN 55105
USA
ssen@macalester.edu

Shawn Bushway
RAND Corporation
14 Mountainwood Drive
Glenville, NY 12302
USA
sbushway@rand.org

Chad Topaz
Williams College
26 Hopkins Hall Dr.
Williamstown, MA 01267
USA
cmt6@williams.edu