Do Judges Vary in Their Treatment of Race?
Author(s): David S. Abrams, Marianne Bertrand, and Sendhil Mullainathan
Source: *The Journal of Legal Studies*, Vol. 41, No. 2 (June 2012), pp. 347-383
Published by: The University of Chicago Press for The University of Chicago Law School
Stable URL: http://www.jstor.org/stable/10.1086/666006
Accessed: 28/05/2013 20:53

# Do Judges Vary in Their Treatment of Race?

**David S. Abrams, Marianne Bertrand, and Sendhil Mullainathan**

**ABSTRACT**

Are minorities treated differently by the legal system? Systematic racial differences in case characteristics, many unobservable, make this a difficult question to answer directly. In this paper, we estimate whether judges differ from each other in how they sentence minorities, avoiding potential bias from unobservable case characteristics by exploiting the random assignment of cases to judges. We measure the between-judge variation in the difference in incarceration rates and sentence lengths between African American and white defendants. We perform a Monte Carlo simulation in order to explicitly construct the appropriate counterfactual, in which race does not influence judicial sentencing. In our data set, which includes felony cases from Cook County, Illinois, we find statistically significant between-judge variation in incarceration rates, although not in sentence lengths.

## 1. INTRODUCTION

In 2008, 38 percent of sentenced inmates in the United States were African American, with African American males incarcerated at six and a half times the rate of white males (Sabol, West, and Cooper 2010). Do these differences in incarceration rates merely reflect racial differences in criminal behavior, or are they also partly an outcome of differential

DAVID S. ABRAMS is Assistant Professor of Law at the University of Pennsylvania. MARIANNE BERTRAND is Professor of Economics in the Booth School of Business at the University of Chicago. SENDHIL MULLAINATHAN is Professor of Economics at Harvard University. The authors would like to thank Josh Fischman, Chris Hansen, Max Schanzenbach, an anonymous referee, and seminar participants at the University of Chicago, Harvard University, Massachusetts Institute of Technology, and the National Bureau of Economic Research Summer Institute. Many thanks to Chief Judge Timothy C. Evans, Presiding Judge Paul P. Biebel, Jr., Malcolm Rich, and Karen Landon for providing the data and invaluable background information on the Cook County courts. Excellent research assistance was provided by Rohit Gupta, Dhruva Kothari, Jessica Pan, Kathy Qian, Tommy Wong, and especially James Wang.

[*Journal of Legal Studies*, vol. 41 (June 2012)]

prosecution or sentencing practices? A long-standing principle embedded in our system of justice is that defendants should not be treated differently because of their race. This principle is codified in the equal protection clause of the Fourteenth Amendment to the Constitution.[1] Differential sentencing or conviction rates by race are presumably a violation of this clause, which makes this an important question to answer on legal grounds. Establishing whether courts treat minority defendants differently also has important social implications: such practices might further exacerbate social inequalities and might even lead to a self-confirming equilibrium in which expectations of racial discrimination affect criminal behavior.

Numerous studies examine this question, and most encounter empirical hurdles, particularly small sample sizes and omitted-variables biases. Although almost all proceedings in U.S. courts are a matter of public record, as a practical reality it is quite challenging to obtain a statistically significant sample size. The studies using small samples of archival data have produced mixed results.[2] Of equal concern is the fact that cross-sectional studies suffer from a potentially severe omitted-variables bias. Apparently significant effects of the defendant's race may actually be due to omitted case characteristics that are correlated with race, such as criminal history or attorney quality.[3] Thus, there are two potential reasons for finding a significant coefficient on race in a cross-sectional regression: discriminatory sentencing on the part of judges or juries or unobservable characteristics that drive the sentencing gap. The central difficulty with a cross-sectional methodology is that race is not randomly assigned. Therefore, any regression and interpretation thereof is likely to suffer from an omitted-variables bias.

In this paper, we take a new approach to studying the impact of race in judicial sentencing, one that avoids some of the methodological pitfalls

1. "No state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws" (U.S. Constitution, amend. 14).

2. Given this difficulty, a number of studies (Devine et al. 2001; Sommers and Ellsworth 2000; MacCoun 1989) make use of experimental simulations of court cases, most often to understand the behavior of juries. While laboratory studies allow for the careful manipulation of the variable of interest, a defendant's race, they suffer from questionable external validity. Many studies simply involve having subjects read transcripts of cases, which removes potentially important nonverbal elements of a trial.

3. Recent research by Abrams and Yoon (2007) shows that there is substantial variation in attorney ability, although they did not find an interaction with clients' race.

just discussed and helps to shed light on the central issue.[4] We attempt to determine whether there are systematic differences across judges in the racial gap in sentencing. At the heart of our research strategy is the ability to exploit the random assignment of cases to judges. This random assignment ensures that unobservable characteristics of cases and defendants are the same across judges. It allows us to distinguish between unobservable variables pertaining to cases and defendants, on the one hand, and judicial behavior, on the other hand, as explanations for a racial gap in sentencing.

Under the unobserved-variables explanation, in which no judge is discriminatory, we may see an overall difference in sentencing by race, but we do not expect systematic variation in that difference across judges, as random assignment ensures that each judge receives the same mix of cases and defendants. Under the discriminatory-sentencing explanation, as long as there is some between-judge heterogeneity in the level of differential treatment, we have the opposite prediction; that is, some judges will systematically sentence African Americans at a higher rate, and some will sentence them at a lower rate. This logic underlies the examination in this paper of whether there is significant interjudge disparity in the racial gap in sentencing.[5]

To proceed, we use data from felony cases to compute the racial gap in the sentence length and the incarceration rate for each judge. The main empirical challenge is to identify the correct counterfactual, in which interjudge variation is due solely to sampling variability. The asymptotic $F$-distribution is inappropriate for this data set because of the small number of observations at the level at which random assignment occurs. This is a problem that occurs frequently in data sets involving randomization procedures for which data are collected over a long period of time.[6] We address this problem by employing a Monte Carlo methodology to explicitly construct the counterfactual in which

4. Ayres and Waldfogel (1994) also take a novel approach to detecting discrimination in a different legal environment—bail setting. Consistent with the presence of racial prejudice, their findings indicate that courts set bail at much higher levels for minority defendants (than for white defendants), thereby "overdeterring" them from fleeing after release on bail.

5. Several previous studies that have examined overall interjudge heterogeneity in sentencing, but none have looked at the effect of a defendant's race on this heterogeneity. See, for example, Gaudet, Harris, and St. John (1933), Anderson, Kling, and Stith (1999), Payne (1997), and Waldfogel (1991, 1998).

6. One example of a recent paper that might benefit from this technique is Cheng (2008). Fischman (2011) also employs the technique.

race has the same impact on sentencing for all judges. Besides its application to the current study, this technique could benefit a large array of empirical studies facing similar constraints without a great deal of learning costs.[7]

We find evidence of significant interjudge disparity in the racial gap in incarceration rates, which provides support for the model in which at least some judges treat defendants differently on the basis of their race. The magnitude of this effect is substantial. The gap in incarceration rates between white and African American defendants increases by 18 percentage points (compared with a mean incarceration rate of 51 percent for African Americans and 38 percent for whites) when moving from a judge at the 10th percentile to one at the 90th percentile in the racial gap distribution. The corresponding sentence-length gap increases by 10 months, but this cannot statistically be distinguished from a situation in which race plays no role in sentence length.

Although judges differ in the degree to which race influences their sentencing, we do not find evidence that observable characteristics such as judges' gender or age group significantly predict this differential treatment by race. Similarly, no systematic pattern emerges with respect to work history (such as whether the judge ever worked as a public defender). However, there is somewhat stronger evidence that the racial gap in sentencing is smaller among African American judges. Further, judges who are harsher overall (as measured by incarceration rate) are more likely to sentence African Americans than whites to jail. We also explore an important potential confounding factor: that the heterogeneity we observe in the racial sentencing gap may actually be due to heterogeneity in treatment of the type of crime. The results of this analysis indicate that there may be a difference in treatment of drug and nondrug crimes but that there is still a heterogeneous treatment of race within nondrug crimes.

One limitation to our approach is that while we can statistically establish that race matters in the courtroom, we cannot formally detect whether this is due to some judges discriminating against African Americans or some judges discriminating against whites or a mixture of both. In itself, though, the evidence we uncover on the importance of race in judicial decision making should be of direct relevance to legal policy.

The rest of the paper proceeds as follows. Section 2 provides a brief

7. The advantage of using simulations has been pointed out in other contexts, for example, by Imbens and Rosenbaum (2005) in the case of weak instruments.

overview of prior work on the role of race in judicial decisions. In Section 3, we describe the data from the courts of Cook County, Illinois. We discuss our econometric methodology, including the simulation procedure, in Section 4. In Section 5, we report our basic results, and we discuss the influence of the crime category in Section 6. Section 7 concludes.

## 2. LITERATURE REVIEW

There has been a great deal of scholarship investigating the role of race in the courtroom. Here we briefly summarize some of the previous research most relevant to this study. Many early studies were cross-sectional and frequently used data sets that were not rich enough to include controls for important case and individual characteristics, such as criminal history, crime severity, and income. Thus, it is unsurprising that an early review of the literature found a lack of consensus among these studies. Daly and Tonry (1997) note some of the shortcomings in some of the work between the 1960s and the 1980s. Kleck (1981) finds that half of the 40 studies on noncapital cases that he reviews either support a finding of discrimination in sentencing or have mixed results, while the other half do not find evidence of judicial discrimination.

Nearly 2 decades later, Spohn (2000) also reviews 40 recent studies on the role of race in sentencing but splits outcomes into incarceration and sentence length. In her survey of the literature, a majority of studies find that race impacts the incarceration decision, but fewer than one-quarter report evidence that race affects sentence length. In one of the most sophisticated critiques of work on discrimination in the criminal justice system, Klepper, Nagin, and Tierney (1983) point out numerous methodological problems, including sample selection and omitted variables. Many of their insights are still often neglected in this field of research, almost 3 decades later.

Some of the earlier papers, such as those by Thomson and Zingraff (1981) and Humphrey and Fogerty (1987), rely on relatively small data sets and are unable to distinguish a race effect from the impact of unobservable characteristics. Klein, Petersilia, and Turner (1990) use a data set from California state courts with a large number of covariates to try to minimize the concern about unobservable characteristics. They find no impact of race on either the incarceration decision or the sentencing decision and little explanatory power. Albonetti (1997) uses federal data from the U.S. Sentencing Commission (USSC) on drug offenders. She finds that African American and Hispanic defendants are more likely to be incarcerated and for longer duration. Steffensmeier and Demuth (2000) also use federal data collected by the USSC and thus have a detailed and large data set with which to work. Their cross-sectional ordinary least squares and probit regressions indicate that African Americans and Hispanics are jailed more frequently and receive longer sentences than white defendants. In another paper, the same authors find similar results using state court data from Pennsylvania (Steffensmeier and Demuth 2001). These results differ to some extent from the findings of Kramer and Steffensmeier (1993), who also use Pennsylvania state court data. In their study, they found a small impact of race on the incarceration decision but not on the length of imprisonment.

A more recent paper by Mustard (2001) improves on previous work by including additional controls in the regression analysis. Using federal data provided by the USSC, he examines the impact of race on the incarceration and sentencing decisions, as well as on departures from the sentencing guidelines. His cross-sectional regressions include controls for income as well as interaction terms for race and income, race and education, and race and criminal history. He finds that African Americans are more likely to be incarcerated and receive longer sentences, although some of this appears to be due to more extensive criminal histories and more severe offenses.

Using state data from Maryland, Bushway and Piehl (2001) estimate a Tobit model to isolate the impact of judicial discretion on sentence length. They find a greater impact of race than does most prior work. A major strength of their paper is the use of guideline recommendations to instrument for potential unobservable case characteristics. Rachlinski and coauthors (2009) approach the question from an experimental psychology perspective. In a laboratory study of judges, they find results for the implicit-association test similar to those for the general population, which has been interpreted by some as evidence of bias. In studies with explicit racial identification, however, Rachlinski and coauthors do not find race effects.

A recent contribution to the literature is Schanzenbach (2005). This study focuses on understanding the impact of judicial characteristics on case outcomes, using variation in judicial characteristics at the federal district level.[8] While he finds that female judges reduce the sex disparity

8. The study by Ashenfelter, Eisenberg, and Schwab (1995) focuses on the impact of judicial characteristics using civil rights cases. The authors find no significant impact of a judge's race, sex, or political orientation on the case outcome.

**Table 1.** Summary Statistics for Cook County and Chicago, Illinois

| | Cook County | | Chicago | | Court Data | |
|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *N* | % |
| White (non-Hispanic) | 2,558,709 | 47.6 | 907,166 | 31.3 | 120,389 | 18.0 |
| African American (non-Hispanic) | 1,390,448 | 25.9 | 1,053,739 | 36.4 | 487,732 | 73.1 |
| Other | 355,844 | 6.6 | 181,467 | 6.3 | 3,031 | .5 |
| Hispanic | 1,071,740 | 19.9 | 753,644 | 26.0 | 56,328 | 8.4 |
| Total | 5,376,741 | | 2,896,016 | | 667,480 | |

**Sources.** U.S. Census Bureau (2001); Circuit Court of Cook County felony cases, 1985–2005.

in sentencing, results for racial disparity are mixed. He also finds no main effect of judges' race on the average sentence length. Shayo and Zussman (2011) take a novel approach in an attempt to understand the impact of the ethnicities of various parties in the legal process. They exploit the random timing and location of terrorist attacks in Israel and show that there is a short-lived local difference in case outcomes that is a function of the ethnicity of the defendant, plaintiff, and judge. Price and Wolfers (2010) also find evidence for race effects in a quasi-judicial context, that of National Basketball Association referees. In this paper, we focus primarily on the effects of defendant race in one large jurisdiction.

## 3. DATA DESCRIPTION

Our data come from the cases adjudicated in the Circuit Court of Cook County of the state of Illinois. Cook County is the largest unified court system in the country, with more than 2.4 million cases processed per year in both civil and criminal courts.[9] It is also a racially mixed urban area, with a population that is 48 percent white, 26 percent African American, and 20 percent Hispanic (see Table 1). The racial breakdown in our data is 12 percent white, 72 percent African American, and 16 percent Hispanic, reflecting the substantially different rates of representation by race in the criminal justice system.

Illinois state courts are governed by sentencing guidelines, which pro-

9. For more detailed information, see Circuit Court of Cook County (http://www.cookcountycourt.org/).

**Table 2.** Summary Statistics: African American Subset

| | Mean | SD |
|---|---|---|
| African American | .86 | .35 |
| Male | .83 | .38 |
| Age | 29 | 10 |
| Cases per judge | 489 | 417 |
| Charges per case | 2.4 | 5.1 |
| Plea | .69 | .46 |
| Guilty verdict | .92 | .27 |
| Probation | .25 | .44 |
| Incarceration | .49 | .50 |
| Sentence length (months) | 20 | 36 |
| Sentence length (nonzero) | 42 | 42 |

**Note.** The results are for cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white. For judges, $N = 70$; for cases, $N = 34{,}227$.

vide suggested sentencing ranges by category of offense.[10] Previous studies, such as Anderson, Kling, and Stith (1999), have found that guidelines mitigate interjudge sentencing variation, but not substantially. Judges in Cook County courts are initially appointed or elected and subsequently subject to retention elections every 6 years.

While the original data set includes more than 600,000 felony cases tried between 1985 and 2004, we use only a subset of the data. We discuss the primary restrictions used to obtain this subset here; further details can be found in the Appendix. First, individual cases may have multiple defendants and multiple charges. In the data, the number of charges per case ranges from one to 266 (see Table 2), but the median is one. We retain one defendant and only the most severe charge for each case, since sentencing across charges for a given case will be highly correlated. Second, for the primary analysis, we restrict the data to defendants who are African American or white (excluding the 16 percent of defendants classified as Hispanic).[11] Third, we retain only cases that were initiated between 1995 and 2001. The start date is used because it was impossible to verify random assignment of cases prior to 1995. The end date is used to allow sufficient time for completion of cases

10. For a description of Illinois' sentencing guidelines, see Illinois General Assembly, Legislative Research Unit, Penalties for Crimes in Illinois (http://www.ilga.gov/commission/lru/2005PFC.pdf).

11. In the Appendix, we report the equivalent analysis on a data set including only white or Hispanic defendants and excluding African Americans.

**Table 3.** Sentencing Breakdown: African American Subset

| | Incarceration Rate | | Sentence Length | | Sentence Length Conditional on a Nonzero Sentence | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Type of charge: | | | | | | |
| Drugs | .50 | .50 | 15 | 22 | 30 | 23 |
| Violent crime | .47 | .50 | 24 | 43 | 52 | 50 |
| EFT | .56 | .50 | 23 | 31 | 41 | 31 |
| Other | .46 | .50 | 24 | 48 | 53 | 31 |
| Race: | | | | | | |
| African American | .51 | .50 | 21 | 36 | 42 | 41 |
| White | .38 | .48 | 16 | 33 | 42 | 43 |
| Total | .49 | .50 | 20 | 36 | 42 | 42 |

**Note.** The results are for cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white. Sentence length is measured in months. EFT = embezzlement, fraud, and theft. For judges, $N = 70$; for cases, $N = 34{,}227$.

initiated toward the end of the time range (since some cases can take several years to adjudicate). Fourth, murder cases were excluded from the analysis because assignment of these cases often excluded certain judges.

We further limit the data to those cases adjudicated by a subset of the judges in the Cook County Criminal Courts Building, which handles the bulk of the criminal cases in Cook County. We include judges on the basis of the following criteria: adjudicated at least 10 cases throughout the time period of study, adjudicated cases only at the central courthouse location (which insures that case randomization was performed on the same set of cases), did not preside over a special type of court (such as drug court), and did not have any unusual circumstances (such as lengthy capital trials) that would have resulted in nonrandom assignment of cases.

A summary of the data set we construct according to these criteria is provided in Tables 2 and 3. Nearly all cases (92 percent) result in a guilty finding. The majority of defendants in the sample are African American (86 percent), male (83 percent), and young (the mean age is 29 and the median age is 27). The mean length of incarceration is 20 months across all cases and 42 months conditional on incarceration. Note that the sentence length is top coded at 60 years in our data. While the median case has only one charge associated with it in the original data, the average number of charges per case is 2.4. As Table 3 shows,

**Table 4.** Judges' Characteristics

| | Mean |
|---|---|
| Male | .82 |
| White | .86 |
| Age | 49 |
| Private practice | .49 |
| Defense attorney | .27 |
| Prosecutor | .70 |

**Note.** Judges presided over cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white. *N* = 70 judges.

sentencing varies substantially by type of crime, with violent crimes receiving the most severe sentences. African American defendants receive longer sentences on average and are more than 30 percent more likely to be incarcerated than are white defendants, not controlling for any case characteristics.[12]

Table 4 reports judicial characteristics collected from a variety of sources (Law Bulletin Publishing Company 1995–2001; Chicago Council of Lawyers 1995–2001; American Bar Association 1995–2001). The judiciary included in this study is largely white and male, with an average age of 49. Approximately half of the judges have some prior experience in private practice. Experience as a prosecutor is also a very common characteristic of these judges; 70 percent have experience as prosecutors, while 27 percent had previously served as public defenders or defense attorneys.

A crucial requirement for this analysis is that the court use random assignment of cases to judges. In Section 4, we describe an econometric test for random assignment. But to establish even facial plausibility, one of the authors spent several days at the Cook County Criminal Courts Building in Chicago, in an observation arranged by Presiding Judge Paul Biebel. Every morning in the courthouse, the clerks receive files for new cases and first remove those that have charges of murder or sex crimes. The remaining case numbers are typed individually into a monochromatic green-screen computer (almost certainly around since the 1980s), which then randomly chooses one of the judges currently hearing cases.

12. Tables A1 and A2 report similar characteristics for the subset of the data containing Hispanic and white defendants.

The clerks verified that this procedure has been generally followed at least since the mid-1990s.

## 4. ECONOMETRIC METHODOLOGY

The focus of this paper is determining whether the impact of a defendant's race on sentencing varies across judges. There are two steps to testing this hypothesis. The first is to establish the random assignment of cases to judges, which ensures that sentencing outcomes can be fairly compared across judges. The second is to employ an appropriate method to evaluate whether there is excess heterogeneity in the racial gap in judicial sentencing beyond what would be expected due to sampling variability.

In theory, both steps may be accomplished with an ordinary least squares regression followed by an $F$-test. Under this approach, the random assignment of cases would be established by regressing a case characteristic, such as the defendant's age, on various controls and judge fixed effects, as in equation (1):

$$\mathrm{Age}_{ijt} = \alpha + \beta X_{ijt} + \sum \delta_j D_j + \mathrm{Mo}_t + \varepsilon_{ijt}, \tag{1}$$

where Age is the defendant's age in years, $\boldsymbol{X}$ is an array of control variables, $D_j$ are judge fixed effects, $\mathrm{Mo}_t$ are month-year dummies, $i$ is a defendant index, $j$ is a judge index, and $t$ is a time index. An $F$-test on the equality of the judge fixed effects tests the hypothesis that cases are randomly assigned (with respect to the defendant's age). Similarly, in order to test the equality of the racial sentencing gap across judges, one would regress the sentence length on a vector of control variables, defendant race, judge fixed effects, and interactions between the judge fixed effects and defendant race, as in equation (2):

$$\begin{aligned} \mathrm{Sentence}_{ijt} = \alpha + \beta X_{ijt} + \mathrm{Race}_{ijt} + \sum \delta_j D_j \\ + \sum \gamma_j D_j \times \mathrm{Race}_{ijt} + \mathrm{Mo}_t + \varepsilon_{ijt}. \end{aligned} \tag{2}$$

An $F$-test on the equality of the judge × race fixed effect coefficient, $\gamma_j$, would be a test of the equality of the racial gap in sentencing across judges.

In practice, rather than the asymptotic $F$-distribution, we rely instead on a Monte Carlo simulation to generate a correct finite-sample distribution. This methodology is analogous in spirit to the one described above, but it addresses important shortcomings of using the standard

*F*-test in this context.[13] The methodology described earlier is likely to result in overrejection of the null hypothesis (of random assignment or no excess heterogeneity) for two reasons. First, although the overall sample is large, our regressions would suffer from finite-sample bias because the sample cells are small within the short time periods that are of relevance. Indeed, it is necessary for the analysis to condition on short time periods because the random assignment of cases to judges occurs within these short periods, and there is substantial temporal variation in the judges available and the mix of case and defendant attributes. Our data structure will, therefore, not satisfy the large *N* assumption on which the distribution of the *F*-statistic relies. A second reason for not using the conventional *F*-statistic is that it will overreject the null hypothesis when the errors are not normally distributed, as is the case when the dependent variable is Bernoulli distributed with a mean substantially different from .5. This applies to several of the variables of interest here, such as race (test of random assignment) or incarceration (test of excess heterogeneity) (see Kennedy 1998, chap. 4).

The aforementioned reasons for empirically computing the finite-sample *F*-distribution are not unique to this paper; rather, they are relatively frequent occurrences. In the law and economics literature, any study that compares judge effects without very high caseloads, such as that of Cheng (2008) or Fischman (2011), is likely to suffer from the same problem. But this phenomenon is certainly not confined to studies of judges; it applies to studies of teachers, chief executive officers, and leaders (see Jones and Olken 2005) and numerous other contexts. Fortunately, the availability of cheap computing power makes the identification of the problem and the solution straightforward.

One way to test whether the small sample is a concern in this context is to simulate the *F* distribution under the null hypothesis for the given data set. Figure 1 illustrates the need for the simulation methodology in this context. In order to generate it, we ran 1,000 tests similar to those we describe later, in which, by construction, the null hypothesis should not be rejected. In theory, this should yield a uniform distribution. The data produced from the simulation methodology are nearly uniform. The data produced using the standard *F*-test methodology clearly shows

13. Methods analogous to the Bonferroni correction could also be used to address some of the shortcomings of the asymptotic *F*-test. The advantages of the simulation approach are its simplicity, transparency, and ease of interpretation.

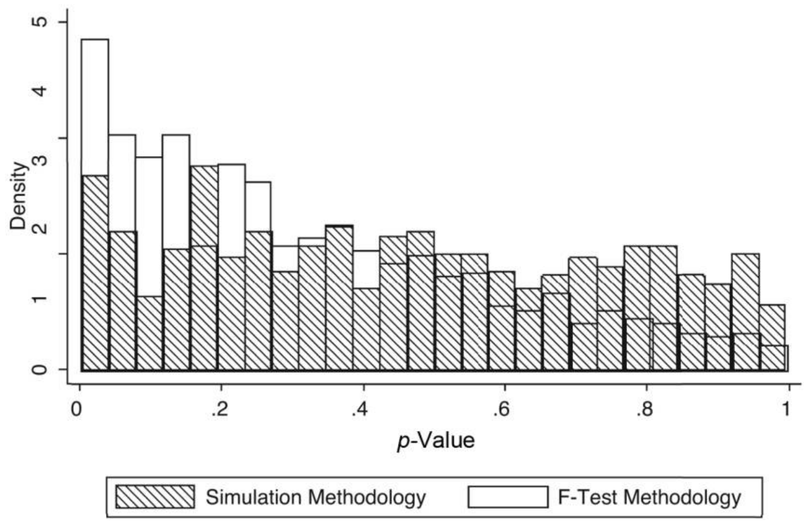


**Figure 1.** Comparison of simulation method with asymptotic *F*-test: *p*-values reported from 1,000 simulations where the null hypothesis is true by construction.

an excess of *p*-values less than .05, which would lead to an overrejection of the null hypothesis.

For these reasons, we instead use a Monte Carlo simulation methodology to both verify random assignment of cases to judges and determine whether there is excess heterogeneity in the interjudge racial gap in sentencing. Random assignment is verified by comparing the heterogeneity of the empirical distribution of case characteristics to that found in simulated data. The heterogeneity of the interjudge racial gap is tested similarly. In both cases, statistical significance is determined by the dispersion of the empirical data relative to the distribution generated by the simulations. We now describe the implementation of the simulation method, first for the random-assignment test and then for the test of excess heterogeneity across judges.

### 4.1. Testing for Random Assignment with a Monte Carlo Simulation

If cases are randomly assigned to judges, all observable case characteristics should have approximately the same moments for each judge. For example, the mean defendant age in the data set is 29 years, and therefore if cases are randomly assigned, most judges should have a set of defendants with a mean age of around 29. Likewise, since 16 percent of cases are in the violent-crime category, we expect a court that uses a random-assignment procedure to produce a distribution of cases for which most judges encounter cases of violent crimes in about 16 percent of their cases. The difficulty in determining whether a data set results from random assignment is in quantifying exactly what it means for most judges to have defendants with a mean age of around 29. The question is, how much variation would there be in a randomly assigned data set, simply due to sampling variability? A straightforward way to establish whether the Cook County data do result from a random-assignment process is by explicitly constructing a randomly assigned data set through simulation.

The procedure is as follows. Let $\boldsymbol{X}$ be a case characteristic of interest, such as defendant race, age, gender, or crime category. Denote a simulated observation by $\boldsymbol{X}_{ijcs}$ for observation $i$ of judge $j$ of simulation $s$ ($i$, $j$, $s > 0$). The term $\boldsymbol{X}_{ijc0}$ refers to the empirical data set. The data are apportioned within cells (denoted by $c$) in order to approximate the actual random-assignment procedure carried out in the courthouse.[14] Create a simulated observation, $\boldsymbol{X}_{ijcs}$, by assigning $\boldsymbol{X}_{ijcs} = \boldsymbol{X}_{\alpha\beta c0}$, where $\alpha$ is randomly chosen from the integers between 1 and $N_c$ inclusive, and $N_c$ is the number of observations in cell $c$ ($\beta$ is a function of $\alpha$). This process is iterated for all $i$ and $j$.

For each simulated data set, judge means may be computed as

$$\overline{\boldsymbol{X}}_{js} = \frac{1}{N_j}\sum_{i\in J}\boldsymbol{X}_{ijcs},$$

where $J$ is the set of cases of judge $j$ and has size $N_j$. We similarly compute a measure of interjudge disparity (such as interquartile range, $D_s^{25-75}$) for each simulated data set.[15] These measures may then be ranked across simulations, and a $p$-value may be found for the empirical distribution ($D_0^{25-75}$) on the basis of where it falls in the $D_s^{25-75}$ distribution.

Table 5 is an illustration of the simulation for the random-assignment test. For the purpose of this illustration, the outcome variable used to

14. Since cases are randomly assigned on a daily basis in the courthouse, this is the ideal cell size. Because there is unlikely to be substantial variation in case mix and judge mix within a month, we use 1 month as the cell size for computational simplicity.

15. We use three different interpercentile ranges, 25th–75th, 10th–90th, and 5th–95th. Other measures, such as standard deviation or absolute mean deviation, could be used as well. We chose interpercentile ranges because we are interested in the central tendencies of the distribution. These will not be substantially impacted by a small number of outliers.

**Table 5.** Random Assignment: Monte Carlo Race Simulation Example

| Case | Date | Real Data | Simulation 1 | Simulation . . . |
|---|---|---|---|---|
| Judge A: | | | | |
| 1001 | 01/01/00 | African American | African American | White |
| 1414 | 01/15/00 | White | African American | African American |
| Judge B: | | | | |
| 3141 | 01/05/00 | African American | African American | African American |
| 6789 | 03/12/00 | White | White | African American |
| Judge C: | | | | |
| 2718 | 01/20/00 | African American | White | African American |
| 8765 | 02/29/00 | African American | African American | White |

test random assignment is race.[16] The null hypothesis is that each judge has the same fraction of African American defendants. If the case mix and eligible judge mix were time invariant, we would not need to restrict ourselves in time. But given that there is substantial variation in both, we choose the cell size to be 1 month.

In this abridged data set, there are six cases, four of which were assigned to judges in January. Thus, the observation in simulation 1 for case 1001 will be randomly chosen from cases 1001, 1414, 3141, and 2718. Since three of the four defendants in those cases are African American, there is a 75 percent chance that the simulated data point will be African American. In fact, in simulation 1, the simulated defendant race is indeed African American.

This procedure is repeated for each observation in Table 5 to produce a full simulated data set. The process is then repeated 1,000 times to produce 1,000 simulated data sets. For each simulated data set, the mean of the race variable is then computed by judge to produce a distribution similar to the empirical distribution shown in Figure 2. We then calculate a measure of dispersion of this simulated distribution, for example, the interquartile range (IQR), which is denoted by the vertical lines in Figure 2. This measure is computed for each of the 1,000 simulations. The data are then reduced to a distribution of these simulated IQRs. We then compare the empirical IQR to the distribution produced from the simulations to obtain an estimate of how likely it is that the empirical distribution occurred due to chance. Figure 3 shows the 1,000 simulated IQRs along with the empirical IQR.

It is worth noting that the random draw in the procedure may be

16. Race is a dummy equal to zero if the defendant's race is white and one if African American.

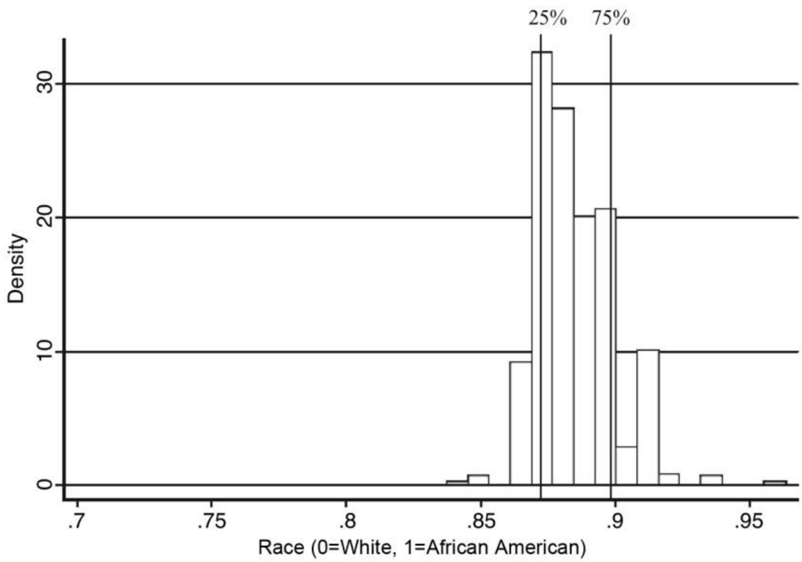


**Figure 2.** Defendant race distribution across judges

either with or without replacement (which would be akin to a permutation). Both procedures may be used but have slightly different interpretations. Drawing with replacement is correct if the data are assumed to be one manifestation of a larger universe of potential empirical realizations. The permutation approach is correct if the data are assumed to be the only relevant realization. The main results that we present were produced from random draws with replacement; however, as a check, we reproduced Figure 7 using a draw without replacement (see Figure A1). Given the size of the data set, it is unsurprising that there is no apparent difference between the two approaches.

### 4.2. Testing for Heterogeneous Sentencing by Race with a Monte Carlo Simulation

Once random case assignment has been established, we can infer that any differences in judicial decisions are due to differences across judges and not to differences in case or defendant characteristics. We may then test the hypothesis that all judges have identical sentencing propensities with respect to race through a simulation procedure similar to the one

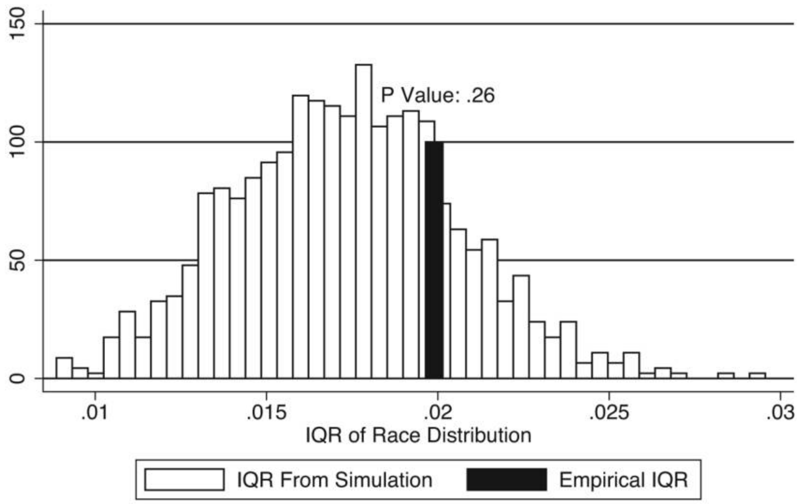


**Figure 3.** Interquartile range of race distribution by judge

already described.[17] The only difference is that we replace a case characteristic with a case outcome measure, such as incarceration rate or sentence length. The simulation procedure is as follows.

First, we compute the outcome of interest for each judge. For example, we compute the difference between the average sentence length of African American defendants and white defendants. If race has no impact on judicial decision making, this difference should be very similar across judges.[18] We can test whether there is excess interjudge disparity in this outcome by comparing the empirical dispersion with that from simulated data in which, by construction, there is no excess disparity. In order to construct the distribution under the null hypothesis of no disparity, we use the above-described process to simulate new data and replace the original case data with that from a randomly chosen case from the same cell. The only difference is that now the cells are restricted further—the simulated case must be from the same month and the defendant must have the same race as in the original case. In this way, we

17. We implicitly assume that cases do not affect each other. In particular, we assume that the racial composition of a judge's previous cases does not affect future decisions.

18. Alternatively, we would find the same result if race affected all judges' decisions in the same way.

Table 6. Sentence Length: Monte Carlo Simulation Example

| Case | Date | Race | Real Data | Simulation 1 | Simulation . . . |
|---|---|---|---|---|---|
| Judge A: | | | | | |
| 1001 | 01/01/00 | African American | 666 | 30 | 7,300 |
| 1414 | 01/15/00 | White | 0 | 365 | 60 |
| Judge B: | | | | | |
| 3141 | 01/05/00 | African American | 30 | 7,300 | 30 |
| 6789 | 03/12/00 | White | 3,650 | 0 | 730 |
| Judge C: | | | | | |
| 2718 | 01/20/00 | African American | 7,300 | 1,095 | 30 |
| 8765 | 02/29/00 | African American | 10,500 | 0 | 180 |

compute a simulated distribution of racial gaps by judge. Table 6 provides an example of simulated data for sentence length. We then calculate a measure of the interjudge dispersion of the difference in the average sentence length by race for each simulation as the test statistic. Finally, we compare the empirical measure of the test statistic to its distribution from the simulations. This comparison allows us to determine, for example, what proportion of the simulated distributions have a larger 5th–95th percentile spread than the empirical distribution. This proportion is the probability that the empirical distribution will have a dispersion of the magnitude observed or larger by chance when there is in fact no interjudge difference in the racial gap in sentencing.

This procedure has three benefits. First, it allows us to simulate the sentencing gap for each judge.[19] Second, it allows us to address the small-sample problem. The simulated data produce an unbiased distribution of the interjudge disparity measure, which is not reliant on an assumption of large *N*. Finally, this distribution allows us to compute a traditional *p*-value. Using it, we can determine the probability of observing the empirical interjudge disparity measure if cases are randomly assigned to judges and race has no impact on judicial decision making. All of these procedures focus on the racial gap but could, of course, also be used to identify the impact of any case characteristics on judicial decision making.

## 5. RESULTS

Because random case assignment is crucial for determining whether judges vary in their treatment of defendants by race, we examine it first,

19. Because judges may vary in the time periods in which they serve, the expected racial gap may be different across judges.

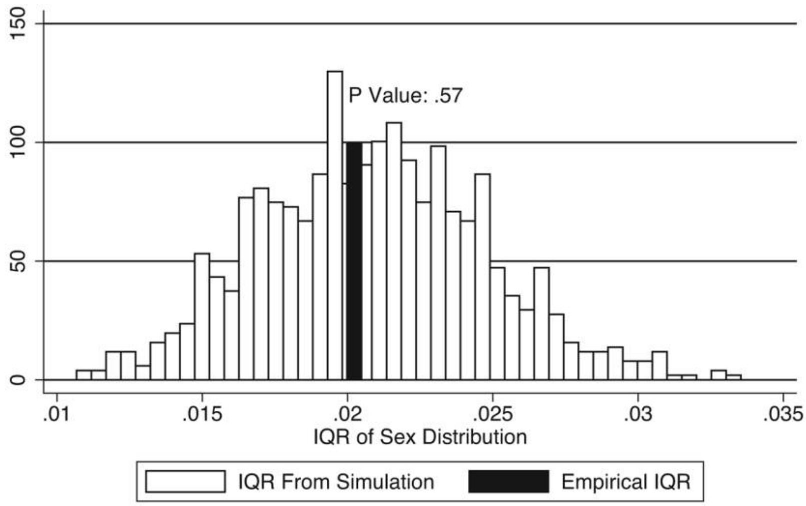


**Figure 4.** Interquartile range of sex distribution by judge

using the Monte Carlo methodology discussed in Section 4. Figure 3 displays the results of the simulation using defendants' race as a check for the random assignment of cases. Since the empirical IQR falls squarely in the middle of the simulated distribution, with a *p*-value of .26, we conclude that there is no systematic bias in the distribution of defendants by race among judges in our sample. Figure 4 reports the results of the random-assignment check using defendants' gender as the case characteristic of interest. We find a *p*-value of .57 and therefore cannot reject the null hypothesis that cases were also randomly assigned to judges with respect to the defendant's gender.

We find similar results when we perform the same Monte Carlo simulations using other specifications. In particular, we test case type and defendant's age as case characteristics, and we also test defendant characteristics by subset of case types. These test results are presented in Table 7, where we report, for each defendant characteristic or case type characteristic, the empirical IQR, the mean and standard deviation of the simulated IQR, and the associated *p*-value.

Additional measures of the spread of the distribution of observable case characteristics, including the 10th–90th percentile range and the 5th–95th percentile range, all support the basic hypothesis that cases

**Table 7.** Random-Assignment Simulation Results by Case and Defendant Characteristics

| Case | Empirical IQR | Simulation IQR Mean | Simulation IQR SD | *p*-Value |
|---|---|---|---|---|
| All ($N$ = 34,298): | | | | |
| Race | .02 | .02 | .00 | .26 |
| Age | .03 | .02 | .00 | .11 |
| Sex | .02 | .02 | .00 | .57 |
| Violent | .03 | .03 | .00 | .12 |
| Drugs | .02 | .03 | .00 | .53 |
| EFT | .02 | .02 | .00 | .53 |
| Other | .03 | .03 | .00 | .45 |
| Violent ($N$ = 5,482): | | | | |
| Race | .04 | .04 | .01 | .30 |
| Age | .06 | .06 | .01 | .60 |
| Sex | .04 | .03 | .01 | .09 |
| Drugs ($N$ = 13,322): | | | | |
| Race | .01 | .02 | .00 | .97 |
| Age | .05 | .04 | .01 | .15 |
| Sex | .03 | .03 | .01 | .37 |
| EFT ($N$ = 6,484): | | | | |
| Race | .07 | .05 | .01 | .04 |
| Age | .06 | .06 | .01 | .50 |
| Sex | .06 | .05 | .01 | .10 |
| Other ($N$ = 9,010): | | | | |
| Race | .03 | .04 | .01 | .96 |
| Age | .05 | .05 | .01 | .62 |
| Sex | .04 | .04 | .01 | .25 |

**Note.** The empirical IQR reports the interquartile range of the distribution of judge fixed effects. Simulation means and SDs are the means and SDs of the IQR distribution from 1,000 simulations. The *p*-values indicate the percentile of the simulated data to which the empirical data correspond. Simulations randomly choose an outcome for cases initiated in the same month as the original case. The data are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white. EFT = embezzlement, fraud, and theft.

were randomly assigned to judges. Based on the random assignment of all observable characteristics that we can test, we conclude that judges receive cases with the same distribution of unobservable case characteristics as well. Thus, differences between judges in their sentencing are attributable solely to their characteristics and preferences and not to differences in case types.

Having established the random assignment of cases to judges, we now examine interjudge variation in sentence lengths and incarceration rates. While not the focus of our inquiry, this is a useful baseline measure before examining differential sentencing by race. Even considered in-

**Table 8.** Dispersion of Judicial Sentencing and Incarceration Rates

| | Jail | Sentence | Sentence2 |
|---|---|---|---|
| 25th–75th Percentile: | | | |
| Empirical value | .13 | 148.28 | 257.14 |
| Simulation mean | .03 | 68.24 | 110.52 |
| Simulation SD | .00 | 13.17 | 19.25 |
| *p*-Value | <.001 | <.001 | <.001 |
| 10th–90th Percentile: | | | |
| Empirical value | .20 | 251.19 | 527.25 |
| Simulation mean | .05 | 143.69 | 231.50 |
| Simulation SD | .01 | 19.27 | 30.98 |
| *p*-Value | <.001 | <.001 | <.001 |
| 5th–95th Percentile: | | | |
| Empirical value | .25 | 390.72 | 684.25 |
| Simulation mean | .07 | 200.40 | 323.26 |
| Simulation SD | .01 | 24.50 | 41.88 |
| *p*-Value | <.001 | <.001 | <.001 |
| *N* | 34,298 | 34,298 | 16,825 |

**Note.** The results are for analogous measures of the empirical and simulated distributions of judge fixed effects. Simulation means and SDs are the means and SDs of the measure from 1,000 simulations. The *p*-values indicate the percentile of the simulated data to which the empirical data correspond. Simulations randomly choose an outcome for cases initiated in the same month as the original case. Jail is a binary variable indicating whether the defendant was incarcerated. Sentence2 is the sentence length conditional on receiving a nonzero sentence. Sentence and Sentence2 are measured in days. The data are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white.

dependent of defendant characteristics, judges in our sample demonstrate substantial heterogeneity in their sentencing decisions. In Table 8, we report results comparing actual heterogeneity to the null hypothesis of no mean differences in sentencing and incarceration rates, using the Monte Carlo methodology detailed in Section 4. All measures of dispersion are at least 20 percent lower than that in a federal district court evaluated by Waldfogel (1998). This is not a particularly concerning finding, given that federal and state courts differ in numerous ways.

In comparison with the simulated dispersion, judges' decisions show excess heterogeneity in all measures, including incarceration (Jail), average sentence length (Sentence), and average sentence length conditional on receiving a nonzero jail sentence (Sentence2). This finding holds true not only in the IQR but also in the 10th–90th percentile range and the 5th–95th percentile range. Figure 5 shows the interjudge variability in incarceration rate. We can reject the null hypothesis that the average incarceration rate does not vary across judges with a *p*-value of less than

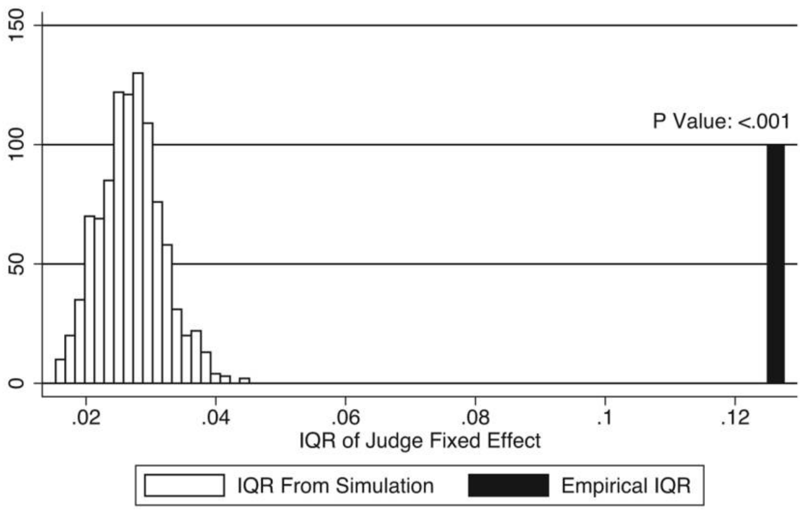


**Figure 5.** Interquartile range of judge fixed effect in incarceration rate

.001. There appears to be substantial heterogeneity in judges' sentencing in our data set. This finding of interjudge sentencing disparity is consistent with previous research focusing on other courts. In particular, Anderson, Kling, and Stith (1999) found significant interjudge sentencing variation in federal courts. They also found that this disparity was reduced only modestly by federal sentencing guidelines.

We now turn to the main objective of this paper, which is to study whether there is excess heterogeneity across judges with regard to racial differences in sentencing. Table 9 presents the results of the Monte Carlo simulations. Figure 6 shows that the IQR of the empirical distribution of the racial difference in incarceration rates is significantly larger (with a $p$-value of .01) than if judges were sentencing without regard to race. That is, we find significant judge × race interactions in the incarceration rate. This result indicates that there is variation in judicial behavior in our sample when it comes to the decision of whether to incarcerate defendants of different races.

We next examine whether there is an analogous impact of defendant race on sentence length. In Table 9 and Figure 7, we present the empirical and simulated IQRs for the racial gap in sentence length. Unlike incarceration, there is no evidence of excess interjudge variation in the racial

**Table 9.** Dispersion of the Racial Gap in Sentencing and Incarceration Rate

| Variable | Empirical IQR | Simulation IQR Mean | Simulation IQR SD | $p$-Value | $N$ |
|---|---|---|---|---|---|
| Jail | .11 | .07 | .01 | .01 | 34,298 |
| Sentence | 90.50 | 150.35 | 29.17 | .98 | 34,298 |
| Sentence2 | 238.36 | 295.21 | 53.51 | .85 | 16,825 |

**Note.** The empirical IQR reports the interquartile range of the distribution of the racial gap in the judge fixed effect. Simulation means and SDs are the means and SDs of the IQR distribution from 1,000 simulations. The $p$-values indicate the percentile of the simulated data to which the empirical data correspond. Simulations randomly choose an outcome for cases initiated in the same month and with defendants of the same race as the original case. Jail is a binary variable indicating whether the defendant was incarcerated. Sentence2 is the sentence length conditional on receiving a nonzero sentence. Sentence and Sentence2 are measured in days. The data are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white.

sentencing gap beyond what we would expect from sampling variation alone. Thus, it appears there are substantial differences in behavior across judges when it comes to the decision of whether to incarcerate defendants of different races but not when it comes to the decision of setting sentence length. The data in Table 9 also show that the lack of excess interjudge heterogeneity in the racial gap in sentence length extends to conditioning on strictly positive sentences.[20]

These findings are consistent with those in recent criminology literature describing attempts to measure the direct effect of race on sentence length. For example, Spohn (2000) notes that the evidence is more compelling for a racial impact on the incarceration decision than on the sentence length. While none of the studies reviewed avoid the difficulty of omitted-variables bias, it is worth noting that these earlier findings are consistent with those in this study. This scenario causes us to wonder why we find excess heterogeneity in the incarceration rate but not in

20. To augment the results reported in Table 9, we conduct the same analysis on a Hispanic subset of data (that is, the original data restricted to Hispanic and white defendants). We follow the same criteria in constructing this subset as we did for the African American subset (see Section 3 and the Appendix for details). The main characteristics of the Hispanic subset are reported in Tables A1 and A2. Like African American defendants, the Hispanic defendants also have higher raw incarceration rates than do white defendants. However, the difference is much smaller and not statistically significant. The main finding reported in Table A3 is that, unlike for the African American sample, we find no evidence of excess interjudge heterogeneity in the Hispanic-white gap in the incarceration rate or the sentence length.

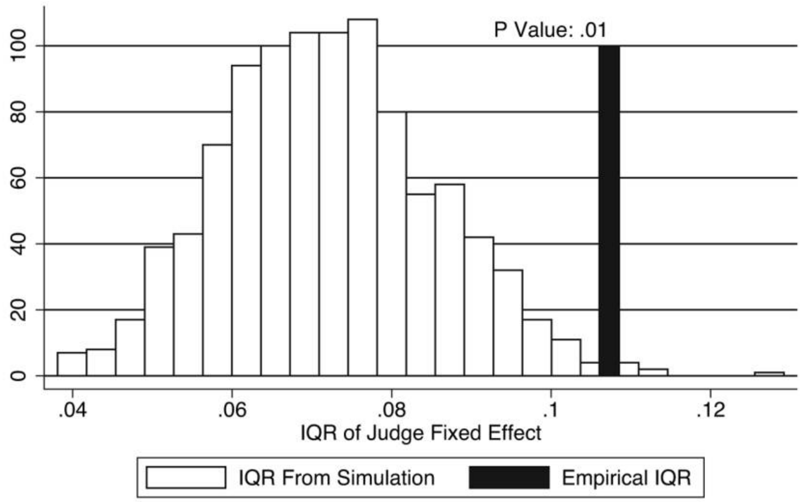


**Figure 6.** Interquartile range of the racial gap in incarceration rate

the sentence length. One possible explanation is that Illinois sentencing guidelines reduce the latitude of individual judges to tailor sentences.[21]

It is important to gain an idea of the magnitude of the interjudge racial gap in the incarceration rate. Table 10 reports the effect of a shift in sentencing from a judge at the 25th percentile of the racial sentencing gap to one at the 75th percentile. There is an increase of 11 percentage points in the probability of incarceration and nearly 3 months in the sentence length. These data compare with a mean incarceration rate of 49 percent and a racial gap of 13 percentage points and a mean sentence length of 20 months and a racial gap of 5 months. The difference between a defendant who is randomly assigned to a 10th-percentile judge and one assigned to a 90th-percentile judge is (not surprisingly) even more striking. In that comparison, the racial gap in incarceration rate rises by a full 18 percentage points, while the expected sentence length increases by 10 months. While the sentencing gap is large in magnitude, this gap cannot, as we have established, be statistically distinguished from that

21. Waldfogel (1998) shows that, under some realistic assumptions, guidelines are not an effective way to reduce interjudge sentencing disparity. Pfaff (2006) points out that Illinois guidelines are relatively broad, compared with those in other states.

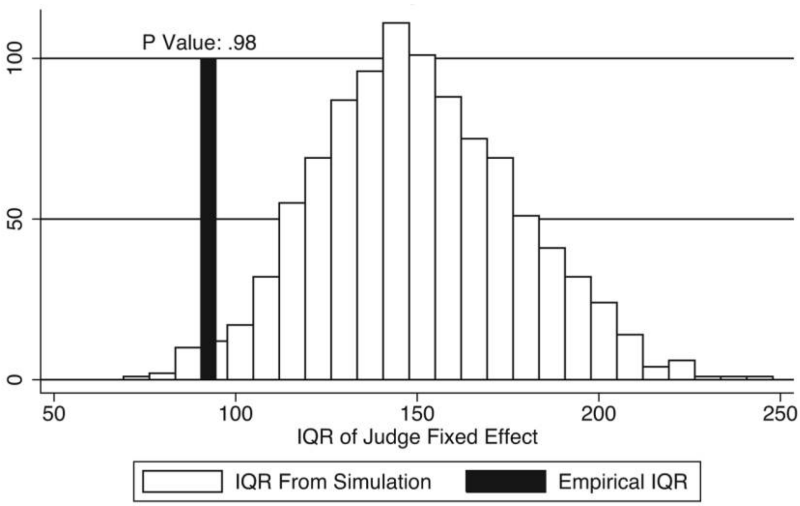


**Figure 7.** Interquartile range of the racial gap in sentencing

which would arise simply due to sampling variability (see Figure 7 and Table 9).

To make these results more concrete, consider the expectations of incarceration for two pairs of otherwise identically situated defendants who differ only by race. William L., who is white, and Bob L., who is African American, have their cases heard before Judge Lenient, who is at the 10th percentile in the racial gap in the incarceration rate. Bill H., who is African American, and Walter H., who is white, appear before Judge Harsh, whose mean racial gap in the incarceration rate puts him at the 90th percentile. Besides their race and (random) judicial assignment, all four defendants and their crimes are otherwise identical. The difference between Bill H.'s and Walter H.'s likelihood of incarceration is 18 percentage points greater than that for Bob L. and William L. So while William L. may expect a 35 percent chance of incarceration and Bob L. may expect a 45 percent likelihood, Walter H. may face a 40 percent probability of incarceration and Bill H. may face a 68 percent chance.

Given the significant heterogeneity between judges, a further question suggests itself: are any observable characteristics of judges predictive of where they fall in the empirical distribution of the racial gap in sen-

**Table 10.** Impact of Judicial Heterogeneity in Sentencing by Race: Change in African American–White Gap

| | Change in Incarceration Rate Gap | | Change in Sentencing Gap (Months) | |
|---|---|---|---|---|
| Percentile Shift | Simulation Mean | Empirical | Simulation Mean | Empirical |
| 25th to 75th | .07 (.01) | .11 | 4.85 (.94) | 2.92 |
| 10th to 90th | .14 (.02) | .18 | 9.52 (1.38) | 10.47 |

**Note.** Values indicate the impact on incarceration and sentencing of moving a defendant from a 25th (10th) percentile judge to a 75th (90th) percentile judge. The counterfactual is no interjudge variation in the racial gap, as produced by simulation. The data are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white. Standard deviations are in parentheses.

tencing? We also examine this question, and the results are presented in Table 11. To perform this analysis, we construct a data set of judge fixed effects and regress these fixed effects on judge-level characteristics such as those reported in Table 4. We estimate the judge fixed effects, $\gamma_j$, in equation (2) for both the incarceration rate and the sentence length. We use the inverse of the square of the estimated standard error to weight each observation in the judge-level regressions. For the sake of completeness, we also estimate judge fixed effects for the average incarceration rate and the average sentence length and relate those to observable characteristics of judges. We do this by estimating the judge fixed effects, $\delta_j$, in equation (1) using both the incarceration rate and the sentence length as dependent variables. Estimated standard errors are again used for weighting in the judge-level regressions.

As the data in the first two columns of Table 11 indicate, there is no systematic relationship between judges' characteristics such as race, gender, age, or experience in public defense and how harsh judges are on average. For example, while the point estimates indicate that male judges give sentences that are on average about 54 days longer and that they incarcerate about 3 percentage points more often, these differences are not statistically significant. The point estimates are of different signs for African American judges; they are associated with longer sentences on average but incarcerate at a lower rate, although again neither coefficient is statistically significant.

The data in the remaining columns of Table 11 relate judge fixed effects for the racial gap in sentencing and for the racial gap in incarceration rate to judges' characteristics. A few somewhat more robust

**Table 11.** Correlation of Fixed Effects for Racial Differences with Judge Characteristics

| | Sentence Length | Incarceration Rate | Racial Gap in Sentence Length | | Racial Gap in Incarceration Rate | |
|---|---|---|---|---|---|---|
| | | | (1) | (2) | (1) | (2) |
| African American judge | 45.03 | −.02 | −152.69 | −156.71 | −.03 | −.03 |
| | (60.20) | (.04) | (80.14) | (81.34) | (.04) | (.04) |
| Male judge | 54.02 | .03 | 61.14 | 57.6 | .02 | .02 |
| | (56.50) | (.03) | (74.22) | (75.28) | (.04) | (.04) |
| Older judge | −11.03 | −.03 | 48.80 | 48.79 | .01 | .01 |
| | (42.78) | (.03) | (57.19) | (57.59) | (.03) | (.03) |
| Judge was public defender | −.56 | .02 | 30.77 | 31.39 | −.04 | −.05 |
| | (49.19) | (.03) | (65.04) | (65.50) | (.03) | (.03) |
| Judge fixed effect in sentence length | | | | .07 | | |
| | | | | (.17) | | |
| Judge fixed effect in incarceration rate | | | | | | .3 |
| | | | | | | (.15) |
| $R^2$ | .02 | .03 | .10 | .16 | .04 | .11 |

**Note.** Standard errors are in parentheses. Each observation is weighted by the inverse of the square of the estimated standard error for the dependent variable. $N = 67$.

patterns emerge from these regressions. First, and most interesting, it appears that African American judges are associated with a smaller racial gap in sentence length. This effect is substantial (about 153 days) and statistically significant. The point estimates indicate that African American judges are also associated with smaller racial differences in the incarceration rate (about 3 percentage points), but this effect is not statistically significant. The point estimates indicate that older male judges might be associated with larger racial differences, but these effects are statistically insignificant and smaller in magnitude than those seen for African American judges. No clear pattern emerges from the data on judges with experience in public defense.

We also include additional judge fixed effects for the average sentence length and for the average incarceration rate. Both are positively correlated with the fixed effects on racial differences in sentencing. Hence, judges who are tougher on average are also relatively tougher on African Americans.

## 6. POTENTIAL CONFOUNDING FACTORS AND ANALYSIS BY CRIME CATEGORY

Our results are consistent with differential judicial treatment of African American defendants, at least with respect to the decision to incarcerate. Some judges show a much larger racial gap in incarceration rates than do other judges, even when facing the same types of defendants and cases. There are several potential concerns regarding the interpretation of these findings, which we now discuss in detail.

African Americans may commit different crimes than whites, and judges may have different sentencing policies for different crimes. For example, suppose that some judges are stricter on sentencing for violent crimes than they are for other crimes. Suppose also that African Americans commit more violent crimes. This correlation would then lead to the appearance of heterogeneity in racial gaps in sentencing even if judges were race blind. One strategy for accounting for these differences in crime categories is to look separately at different categories of crime. The difficulty with this approach is that once divided this way, each category contains a relatively small number of observations. In performing this analysis (data not shown), we find no evidence for excess heterogeneity in racial gap in any crime category. This result is almost certainly due to a lack of power.

In order to address the problem of diminishing the sample size, we run our central analysis while controlling for the category of crime com-

**Table 12.** Crime Category Analysis

| Variable | Empirical IQR | Simulation IQR Mean | Simulation IQR SD | $p$-Value | $N$ |
|---|---|---|---|---|---|
| All cases with crime controls: | | | | | |
| Jail | .090 | .069 | .012 | .046 | 34,227 |
| Sentence | 141.57 | 150.49 | 27.68 | .599 | 34,227 |
| Sentence2 | 283.06 | 279.24 | 47.91 | .457 | 16,807 |
| Drug cases: | | | | | |
| Jail | .112 | .143 | .028 | .868 | 13,317 |
| Sentence | 114.50 | 145.61 | 26.63 | .891 | 13,317 |
| Sentence2 | 175.55 | 330.76 | 66.25 | .997 | 6,588 |
| Nondrug cases: | | | | | |
| Jail | .108 | .083 | .015 | .043 | 20,910 |
| Sentence | 175.11 | 192.08 | 36.22 | .632 | 20,910 |
| Sentence2 | 350.91 | 352.24 | 71.67 | .487 | 10,219 |

**Note.** Empirical IQR indicates the interquartile range of the distribution of the racial gap in the judge fixed effect. Simulation means and SDs are the means and SDs of the IQR from 1,000 simulations. The $p$-values indicate the percentile of the simulated data to which the empirical data correspond. Simulations randomly choose an outcome for cases initiated in the same month and with the defendant of the same race as the original case. Jail is a binary variable indicating whether the defendant was incarcerated. Sentence2 is the sentence length conditional on receiving a nonzero sentence. Sentence and Sentence2 are measured in days. Cases involve felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was African American or white.

mitted. We implement this by subtracting judge-specific means by crime category for both incarceration and sentence length. The results are reported in Table 12. We find results very similar to those for the main specification. There is evidence of excess heterogeneity in the racial gap in incarceration rates but not in sentence lengths.

In order to get a firm hold on whether there is any variation in judicial decisions by case type while maintaining sufficient observations to ensure a meaningful test, we subdivide the data into drug and nondrug cases. The results from this analysis are reported in Table 12. Focusing on the incarceration racial gap, we find excess dispersion for nondrug cases ($p$ = .043) but not for drug cases ($p$ = .868). Although there are fewer drug cases than nondrug cases, the disparity is only 35 percent, and thus a lack of power is unlikely to be the cause of the difference. One plausible explanation is that the Illinois sentencing guidelines provide less judicial discretion in the incarceration decision for drug offenses than for nondrug offenses.

While correlation between race and crime type is the most obvious potential confounding factor, this is an example of a more general concern. Suppose there are unobservable (to us) features of the case that some judges care more about than others. For example, there may be details of the crime that are not captured by the statute under which the person is being charged. Alternatively, there may be details of the evidence (such as use of DNA tests) that are not in our data set. These unobservable case features could in principle generate the type of variation we observe if these unobserved features vary systematically across racial groups and judges differ in their treatment of these characteristics. This scenario would occur if DNA evidence were used more often against one racial group than another. It seems unlikely that, under this model, a characteristic such as the judge's race would systematically predict the racial gap in sentencing (as the data in Table 11 suggest). While these confounding factors are still potentially a concern, the approach in this paper advances the field in light of previous work, because now the unobservable case characteristics would have to be correlated with the defendant's race and elicit differential treatment across judges.

## 7. CONCLUSION

In this paper, we have sought to shed light on the influence of race in judicial sentencing practices. Previous research has largely made use of ordinary least squares regressions in addressing this topic. That approach may suffer from an omitted-variables problem, which could substantially bias any estimate of the influence of race on sentencing.

We make use of the random assignment of cases to judges in order to address the omitted-variables bias. With random assignment of cases, all judges receive the same distribution of case characteristics, both observed and unobserved. Thus, if all judges are unbiased, one would expect the racial gap in sentencing to be the same across judges, to within the sampling error. The core of our analysis is establishing what the gap would be for unbiased judges and comparing those figures with the actual data.

We produce these unbiased data with a Monte Carlo simulation by sampling from the actual data but mechanically breaking the judge-defendant race link. We find that there is substantial excess heterogeneity in the empirical distribution of the racial gap in the incarceration rate. The quantitative impact of this gap on sentencing disparity is of considerable magnitude. If a defendant assigned to a 10th-percentile judge was instead sentenced by a 90th-percentile judge, the racial gap in the incarceration rate would rise by a full 18 percentage points.

It is also useful to consider potential legal policy implications in light of these findings. One goal of changes in policy could be to try to reduce or eliminate the excess interjudge heterogeneity in the racial gap. This analysis can inform how big an impact that sort of policy change would make. If the excess interjudge racial gap in incarceration were eliminated, the IQR of the racial gap in incarceration would drop from .11 to .07 (Table 9). This decrease represents a 36 percent reduction in the variability of the African American–white racial gap in incarceration due just to judicial assignment. The magnitude of this potential effect would decrease one element of the randomness in the judicial process and surely would increase confidence in the fairness of the court system.

One important limitation of our work is that while we show that race appears to play a role in judicial decision making, we cannot make statements about its optimality. That is, we can say that judges vary in their treatment of race but not whether this is evidence of discrimination or reverse discrimination. It is theoretically possible that the heterogeneity in the racial gap in incarceration reflects favoritism by some judges toward African American defendants. For example, suppose that unobservable case characteristics dictated that an unbiased racial gap in sentencing would be 50 percent. In this case, heterogeneity in the race gap between 20 and 50 percent would indicate a great deal of favoritism toward African Americans, not discrimination. In future work, information on interjudge differences in the racial gap in recidivism may further guide the interpretation of our findings. In particular, one may relate the variation we observe in the racial gap in sentencing to the variation in the racial gap in recidivism. In addition, information on the success rate of appeals may provide another method of evaluating the optimality of the racial gap. The theoretical ideal would be to evaluate a social welfare function with terms that include both recidivism and appeals and all other relevant factors.

Despite this interpretational limitation, our findings nevertheless raise important legal questions. Heterogeneity across judges in sentencing by race suggests that courtroom outcomes may not be race blind. This potential lack of partiality may be one source of the substantial overrepresentation of African Americans in the prison population. Understanding the sources of variation in the criminal justice system is an important first step toward reducing disparities of various kinds.

## APPENDIX: DATA-CLEANING PROCEDURE

The data for this study come from the Circuit Court of Cook County of the state of Illinois. For each felony case that is prosecuted, a record is made of key case details, including the defendant's characteristics (race, sex, age, and so forth), case traits (crime type, assigned judge, court location), and outcomes (sentence length, plea, finding of guilt). A substantial amount of data cleaning was necessary to prepare the data for analysis, with the process detailed here.

The initial data processing removed observations with erroneous data. For example, observations for which the sentence length was inaccurate or unintelligible, such as "2 months 400 days" were excluded. Other dropped observations include those with erroneous dates (too far in the past or in the future), negative sentences, duplicate observations based on case number, and missing race.

Sentences were top coded to 60 years under the assumption that defendants were unlikely to serve longer based on the median defendant age. Life sentences were also coded as 60 years. The guilty binary indicator was set to equal guilty when sentences were nonzero and the guilty variable was missing. We dropped any observation for which the guilty and sentence variables were both nonmissing and contradicted each other (that is, defendant found not guilty but with nonzero sentence length).

Defendants with cases already pending in the courts are sometimes assigned to the same judges for their subsequent cases; thus, we keep as an observation only the first time a defendant appears in the data, because only these cases are likely to be truly random. Establishing unique defendant identities is difficult because of frequent miscoding, which we attempt to address with several procedures.

A unique defendant ID is defined by last name, race, and sex. Last name is defined as the last word in the defendant's name. The identification is further refined by a fuzzy match on the date of birth. Because of miscoding of this variable, we count two observations as having the same defendant if they match on last name, race, and sex and have at most one digit different in their dates of birth. For example, Kevin Marshall with birthday 12/42/78 (with the tens digit for the day miscoded in month/day/year format) would be the same individual as Kevin Marshall with birthday 12/02/78.

Once the data set is winnowed to a single observation per defendant, there are still a number of other data-cleaning procedures we undertake to overcome further idiosyncrasies of the data set and coding errors. Homicide cases are not allocated through the standard random-assignment method (their assignment takes into account the judicial caseload), and thus we exclude them from our sample. The variable indicating the courthouse location is often miscoded. This error poses a serious problem because cases arising in Rolling Meadows, Skokie, and other suburban courthouses have vastly different characteristics from cases in Chicago.

We use two procedures in the attempt to exclude cases actually originating from suburban locations. First, we drop all of the cases in a given year for a judge who has any cases outside the main Chicago courthouse (located at 26th and California) in that year. For example, Judge Roberts may have 100 cases at 26th and California every year from 1994 to 2003, but in 1996 he took on a case at the courthouse in Rolling Meadows. This scenario would cause us to drop all of his cases for 1996. Second, we compute a measure of the dispersion of the defendant's home zip codes for each judge. We drop all cases for a judge in a year in which this measure deviates from the mean by more than 10 percent.

For certain years in our range, the Cook County courts had judges who adjudicated only drug cases. The cases assigned to these judges were clearly nonrandom along the case-type dimension. In order to exclude them, we drop cases heard by judges for whom drug cases constitute more than 70 percent of their caseloads for the year.

After the preceding case culling, we ran the random-assignment check across multiple dimensions on the remaining data at the month level. We were unable to verify random assignment prior to 1995, so we exclude this data. We further restrict ourselves to cases begun before 2002, in order to prevent truncation bias from impacting the results, as cases can often stretch on for several years.

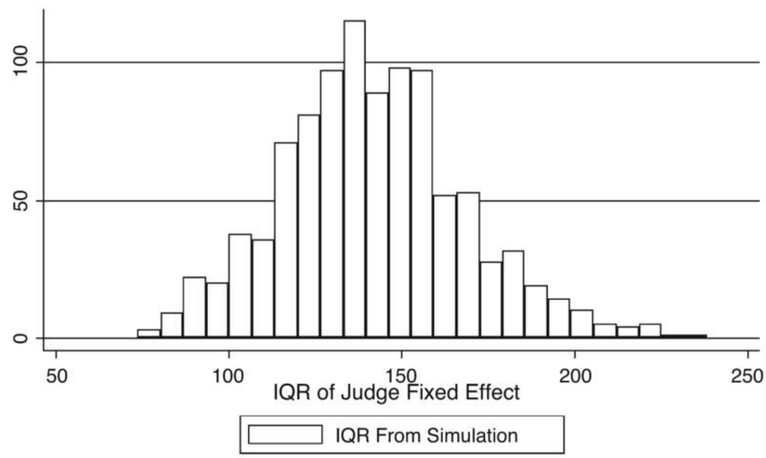


**Figure A1.** Interquartile range of the racial difference in sentencing: randomization without replacement.

**Table A1.** Summary Statistics: Hispanic Subset

| | Mean | SD |
|---|---|---|
| Hispanic | .56 | .5 |
| Male | .88 | .32 |
| Age | 29 | 10 |
| Cases per judge | 174 | 133 |
| Charges per case | 2.4 | 4.2 |
| Plea | .76 | .43 |
| Guilty verdict | .92 | .27 |
| Probation | .29 | .46 |
| Incarceration | .41 | .49 |
| Sentence length (months) | 18 | 37 |
| Sentence length (nonzero) | 43 | 46 |

**Note.** Results are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was Hispanic or white. For judges, $N = 75$; for cases, $N = 11{,}946$.

**Table A2.** Sentencing Breakdown: Hispanic Subset

| | Incarceration Rate | | Sentence Length | | Sentence Length Conditional on a Nonzero Sentence | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Type of charge: | | | | | | |
| Drugs | .34 | .48 | 7.1 | 16 | 20 | 22 |
| Violent crime | .41 | .49 | 21 | 40 | 50 | 49 |
| EFT | .48 | .5 | 19 | 29 | 40 | 30 |
| Other | .41 | .49 | 22 | 46 | 55 | 59 |
| Race: | | | | | | |
| Hispanic | .44 | .5 | 21 | 39 | 47 | 49 |
| White | .38 | .49 | 15 | 32 | 39 | 42 |
| Total | .41 | .49 | 18 | 37 | 43 | 46 |

**Note.** Results are from cases involving felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was Hispanic or white. Sentence length is measured in months. EFT = embezzlement, fraud, and theft. For judges, $N = 75$; for cases, $N = 11{,}946$.

**Table A3.** Dispersion of the Racial Gap in Sentencing and Incarceration Rates: Hispanic Subset

| Variable | Empirical IQR | Simulation IQR Mean | Simulation IQR SD | *p*-Value | *N* |
|---|---|---|---|---|---|
| Jail | .06 | .09 | .02 | .97 | 11,946 |
| Sentence | 172.58 | 193.31 | 32.52 | .75 | 11,946 |
| Sentence2 | 288.84 | 383.91 | 66.68 | .93 | 4,888 |

**Note.** Empirical IQR reports the interquartile range of the distribution of the racial gap judge fixed effect for the given variable. Simulation means report the means of the interquartile range from 1,000 simulations; SDs report the standard deviations from the simulations. The *p*-values indicate the percentile of the simulated data to which the empirical data correspond. Simulations randomly choose an outcome for cases initiated in the same month and with defendants of the same race as the original case. Jail is a binary variable indicating whether the defendant was incarcerated. Sentence2 is the sentence length conditional on receiving a nonzero sentence. Sentence and Sentence2 are measured in days. Cases involve felony offenses in the Circuit Court of Cook County initiated from 1995 to 2001 in which the defendant was Hispanic or white.

## REFERENCES

Abrams, David S., and Albert H. Yoon. 2007. The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability. *University of Chicago Law Review* 74:1145–77.

Albonetti, Celesta A. 1997. Sentencing under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991–1992. *Law and Society Review* 31: 789–822.

American Bar Association. 1995–2001. *The Directory of Minority Judges of the United States*. Chicago: American Bar Association.

Anderson, James M., Jeffrey R. Kling, and Kate Stith. 1999. Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines. *Journal of Law and Economics* 42:271–307.

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes. *Journal of Legal Studies* 24:257–81.

Ayres, Ian, and Joel Waldfogel. 1994. A Market Test for Race Discrimination in Bail Setting. *Stanford Law Review* 46:987–1047.

Bushway, Shawn D., and Anne M. Piehl. 2001. Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing. *Law and Society Review* 35:733–64.

Cheng, Edward K. 2008. The Myth of the Generalist Judge. *Stanford Law Review* 61:519–72.

Chicago Council of Lawyers. 1995–2001. *A Directory of State and Federal Judges in Chicago*. Chicago: Chicago Council of Lawyers.

Daly, Kathleen, and Michael Tonry. 1997. Gender, Race, and Sentencing. *Crime and Justice* 22:201–52.

Devine, Dennis J., Laura D. Clayton, Benjamin B. Dunford, Rasmy Seying, and Jennifer Pryce. 2001. Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups. *Psychology, Public Policy, and Law* 7:622–727.

Fischman, Joshua B. 2011. Estimating Preferences of Circuit Judges: A Model of Consensus Voting. *Journal of Law and Economics* 54:781–809.

Gaudet, Frederick J., George S. Harris, and Charles W. St. John. 1933. Individual Differences in the Sentencing Tendencies of Judges. *Journal of Criminal Law and Criminology* 23:811–18.

Humphrey, John A., and Timothy Fogarty. 1987. Race and Plea Bargained Outcomes: A Research Note. *Social Forces* 66:176–82.

Imbens, Guido W., and Paul R. Rosenbaum. 2005. Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education. *Journal of the Royal Statistical Society: Series A: Statistics in Society* 168:109–26.

Jones, Benjamin F., and Benjamin A. Olken. 2005. Do Leaders Matter? National

Leadership and Growth since World War II. *Quarterly Journal of Economics* 120:835–64.

Kennedy, Peter. 1998. *A Guide to Econometrics*. Cambridge, Mass.: MIT Press.

Kleck, Gary. 1981. Racial Discrimination in Criminal Sentencing: A Critical Evaluation of the Evidence with Additional Evidence of the Death Penalty. *American Sociological Review* 46:783–805.

Klein, Stephen, Joan Petersilia, and Susan Turner. 1990. Race and Imprisonment Decisions in California. *Science* 247:812–16.

Klepper, Steven, Daniel Nagin, and Luke-Jon Tierney. 1983. Discrimination in the Criminal Justice System: A Critical Appraisal of the Literature. Pp. 55–128 in vol. 2 of *Research on Sentencing: The Search for Reform,* edited by Alfred Blumstein, Jacqueline Cohen, Susan E. Martin, and Michael H. Tonry. Washington, D.C.: National Academy Press.

Kramer, John H., and Darrell Steffensmeier. 1993. Race and Imprisonment Decisions. *Sociological Quarterly* 342:357–76.

Law Bulletin Publishing Company. 1995–2001. *Sullivan's Judicial Profiles*. Chicago: Law Bulletin Publishing Company.

MacCoun, Robert J. 1989. Experimental Research on Jury Decision-Making. *Science* 244:1046–50.

Mustard, David B. 2001. Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *Journal of Law and Economics* 44: 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* 17:337–66.

Pfaff, John F. 2006. Continued Vitality of Structured Sentencing following Blakely: The Effectiveness of Voluntary Guidelines. *UCLA Law Review* 54:235–308.

Price, Joseph, and Justin Wolfers. 2010. Racial Discrimination among NBA Referees. *Quarterly Journal of Economics* 125:1859–87.

Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie. 2009. Does Unconscious Racial Bias Affect Trial Judges? *Notre Dame Law Review* 84:1195–1246.

Sabol, William J., Heather C. West, and Matthew Cooper. 2010. *Prisoners in 2008*. Bureau of Justice Statistics Bulletin No. NCJ 228417. Washington, D.C.: U.S. Department of Justice. April. http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1763.

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

Shayo, Moses, and Asaf Zussman. 2011. Judicial Ingroup Bias in the Shadow of Terrorism. *Quarterly Journal of Economics* 126:1447–84.

Sommers, Samuel R., and Phoebe C. Ellsworth. 2000. Race in the Courtroom:

Perceptions of Guilt and Dispositional Attributions. *Personality and Social Psychology Bulletin* 26:1367–79.

Spohn, Cassia C. 2000. Thirty Years of Sentencing Reform: The Quest for a Racially Neutral Sentencing Process. Pp. 427–501 in *Policies, Processes, and Decisions of the Criminal Justice System*, Vol. 3 of *Criminal Justice 2000*. Washington, D.C.: National Institute of Justice.

Steffensmeier, Darrell, and Stephen Demuth. 2000. Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who Is Punished More Harshly? *American Sociological Review* 65:705–29.

Steffensmeier, Darrell, and Stephen Demuth. 2001. Ethnicity and Judges' Sentencing Decisions: Hispanic-Black-White Comparisons. *Criminology* 39:154–78.

Thomson, Randall J., and Matthew T. Zingraff. 1981. Detecting Sentencing Disparity: Some Problems and Evidence. *American Journal of Sociology* 86:869–80.

U.S. Census Bureau. 2001. *Census 2000*. Washington, D.C.: U.S. Government Printing Office.

Waldfogel, Joel. 1991. Aggregate Inter-judge Disparity in Sentencing: Evidence from Three Districts. *Federal Sentencing Reporter* 4:151–54.

———. 1998. Does Inter-judge Disparity Justify Empirically Based Sentencing Guidelines? *International Review of Law and Economics* 18:293–304.