

---

## Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing

---

Shawn D. Bushway

Anne Morrison Piehl

Variation in sentencing outcomes represents the actions of a number of members of the criminal justice system. To isolate the part of the variation that is due to the discretion of the judge (or other sentencing agent, such as a prosecutor), one can model the sentencing guidelines themselves. Such a model captures any non-linearity in the sentencing grid. In practice, modeling the guidelines rather than legal factor scores (as is common in the literature) means that more of the variation that race and legal factors share in common will be attributed to the racial status of the offender. Using data from Maryland, we find that African Americans have 20% longer sentences than whites, on average, holding constant age, gender, and recommended sentence length from the guidelines. We find more judicial discretion and greater racial disparity than is generally found in the literature. Moreover, when we begin to try to explain this discretion, we find that judges tended to give longer sentences (relative to those recommended by the guidelines) to people in the part of the guidelines grid with longer recommended sentences (who are disproportionately African American) than they gave to people in the part of the grid with lower recommended sentences.

### I. Introduction

**T**here is a large literature on racial discrimination in sentencing outcomes that begins with the disproportionate representation of African Americans and other minorities in prison.<sup>1</sup>

---

The authors contributed equally to this article. Please address all correspondence to Anne Piehl at 79 JFK Street, Cambridge, MA 02138; e-mail: anne\_piehl@harvard.edu. We would like to thank Robert Brame, Gary LaFree, Raymond Paternoster, Charles Wellford, Joseph Sanders and the anonymous reviewers, as well as seminar participants at the University of Maryland for helpful comments. We appreciate the excellent research assistance provided by Andrea Cann. We would also like to say a special thank you to Claire Souryal for help with data above and beyond the call of duty. All errors remain our own.

<sup>1</sup> At year-end 1999, there were 3,408 sentenced black male inmates per 100,000 black males in the United States; 1,335 sentenced Hispanic male inmates per 100,000 Hispanic males; and 417 white male inmates per 100,000 white males (Bureau of Justice Statistics 2000). Therefore, the rate of incarceration of black males is 8.17 times the incarceration rate of white males and the incarceration rate of Hispanic males is 3.20 times the rate of white males.

But black offenders tend to have other characteristics, such as longer criminal histories, that are considered by most observers of the sentencing process to be legitimate reasons for harsher outcomes. Researchers have dealt with this correlation between race and legally relevant factors by dividing racial disparity into two parts: warranted and unwarranted disparity. Warranted disparity is the variation in sentence outcomes due to legally relevant factors, such as criminal history, crime type, and crime severity. Unwarranted disparity (Stolzenberg & D'Alessio 1994) is the variation in sentencing outcomes that can be reasonably identified as being the sole result of race or other extra-legal factors (e.g., gender), after all legally mandated sentencing factors are taken into account.<sup>2</sup> Partialing out warranted and unwarranted disparity is generally done using standard regression techniques.

In general, the body of research on unwarranted disparity in sentencing outcomes has shown that legal factors have large effects on sentencing outcomes. Furthermore, there is little evidence of direct racial discrimination once these legal factors are included in the statistical models. This evidence led Sampson and Lauritsen (1997:362) to conclude that "there is little evidence that racial disparities reflect systematic, overt bias on the part of criminal justice decision makers (as a whole)." In response, some researchers have argued for an interactive, rather than additive, model to examine whether racial discrimination occurs indirectly through certain court contexts (e.g., plea vs. jury trial) or individual characteristics (e.g., employed vs. unemployed). This can be done by including interaction terms in multivariate models (Miethe & Moore 1986), testing for differences among age and gender subgroups of each race group (Steffensmeier et al. 1998), or estimating separate models for each racial group (Albonetti 1997).<sup>3</sup>

This type of research has emphasized greater statistical rigor and model specification over the development of a theoretical framework for understanding the sentencing process (Albonetti 1991). But the statistical parsing of unwarranted and warranted disparity may conflict with the desire to understand the key factors driving the sentencing process. For example, Dixon (1995) tested her organizational context perspective on sentencing using the same basic "racial disparity" statistical model (described previously) as Albonetti (1991) used to test her bounded rationality model of judicial decisionmaking.<sup>4</sup> Yet, Dixon explicitly con-

---

<sup>2</sup> Spohn (2000) refers to direct discrimination rather than unwarranted disparity. The definitions are similar. Spohn (2000) defines direct discrimination as the situation in which race and/or ethnicity significantly affect sentence severity after all legally relevant case and offender characteristics are taken into consideration.

<sup>3</sup> For the origins of this work, see Miethe and Moore (1986), Myers and Talarico (1987). For a review of the work using interactive models, see Spohn (2000).

<sup>4</sup> We are not arguing that the two researchers use the same exact model. Dixon (1995) used data from Minnesota, which is a guideline state, while Albonetti (1991) used

sidered the behavior of the system, including prosecutors and judges, while Albonetti focused on judicial discretion. Clearly, the same model cannot simultaneously isolate system-wide behavior and judicial behavior.

This conflict illustrates a foundational point of research on the sentencing process: The sentence outcome is the result of the decision of multiple actors interacting in a system. Therefore, the researcher seeking to understand the role of a given factor on the sentencing process must first accurately model the behavior of individual actors (Klepper et al. 1983). The warranted versus the unwarranted disparity model does not attempt to do that however. Instead, it focuses on system-wide disparity, a point recognized by Sampson and Lauritsen (1997) and Dixon (1995).

To be sure, there is a body of research that focuses on actions of particular types of actors involved in the sentencing process (Kramer & Ulmer 1996). For example, there are six studies of judicial departures in sentencing outcomes in guideline jurisdictions (Kramer & Ulmer 1996, Langan 1996, Miethe & Moore 1986, Maxfield & Kramer 1998, Ulmer 1997, U.S. Sentencing Commission 1995). Because departures represent deviation from the presumptive sentence provided by the guidelines, these departures can reasonably be assumed to reflect judicial discretion. (Throughout this article, we follow convention in the literature and use the term “judicial discretion” to refer to discretion in criminal sentencing regardless of whether a judge or a prosecutor is responsible for setting a given sentence and regardless of whether the sentence results from a trial or a guilty plea.)

Five of these studies show some evidence for racial discrimination by the judges. However, this evidence is not necessarily inconsistent with the finding of no overall discrimination in some of these jurisdictions, since the judges may be compensating for the actions of prosecutors or reacting to the guidelines. The net effect may be no racial disparity even though race was a factor in the judge’s decisionmaking process. For example, Miethe and Moore (1986:269) used a study of judicial dispositional and durational departures from Minnesota’s guidelines to conclude that there is “a process of sentence adjustments (by judges) designed to bring the actual sentence more in line with what judges and other criminal justice officials may consider an appropriate sanction for the crime or person involved or both.”

Each of these six studies focuses on departures from sentence guidelines, but judges have additional discretion within the ranges provided by the guideline grids. In this article, we present

---

data from Washington, DC, which is not a guideline jurisdiction. Both, however, include measures for legal factors and extra-legal factors in an incarceration and in a sentence length model. Although there are minor differences between them (for example, Albonetti includes crime type while Dixon does not), both sets of models have similar frameworks for structuring discretion in the system.

a model, built upon work by Griswold (1987) and Engen and Gainey (2000a), that attempts to capture the full amount of judicial variation from the presumptive sentence recommended by the guidelines. This variation includes both that which falls inside the guideline range and that which is outside the range. In presenting our model, we make use of statistical innovations, such as Tobit models (Tobin 1958) and the natural logarithm of sentence length, which have been introduced in the larger literature looking at unwarranted disparity in sentence outcomes. However, our interpretation of the results is different from the bulk of the literature because of our focus on judicial discretion.

Focusing on judicial discretion directly can provide an alternative way to examine the question of racial disparity in sentencing outcomes. In the usual modeling approach, only the variation in sentence outcomes that is uncorrelated with legal factors can be considered *unwarranted* disparity. Yet, it is not hard to think of cases in which racial discrimination could be masked by correlation with legal factors. Suppose that African Americans tend to be disproportionately represented among those with long criminal history records and individuals with long criminal history records tend to get disproportionately long sentences. Although it is possible that race has a direct causal impact on the decisions of the actors in the system, high correlation between race and the legal factors leads to imprecise estimates of the role of race in the sentencing outcome.<sup>5</sup> Because of the concerns about ignoring legitimate variation due to legal factors, and in the absence of alternative methods, researchers using regression techniques have been forced to focus only on the variation in race that is uncorrelated with legal outcomes.

Suppose, however, that we could distinguish between the use of legal factors by the guidelines and the use of legal factors by the judges. The literature contains a range of opinions on this matter. Social theorists such as Savelsberg (1992) suggested that guidelines represent a neoclassical return to formal rationality in sentencing where the goal of sentencing is simply to “be just” (Savelsberg 1992). Miethe and Moore (1986) argued that crime severity and criminal history should *not* play an important role in judicial discretion outside of the guidelines. From the perspective of highly structured guideline systems, departures are primarily intended as a way of dealing with highly case-specific attributes, for the very reason that the entity setting the guidelines already specified roles for crime severity and criminal history. Using Weber’s writings on substantive rationalization, Savelsberg (1992) suggested that actors such as judges and prosecutors will work to countermand the intent of sentencing guidelines and

---

<sup>5</sup> Multicollinearity deals with the fact that high (but imperfect) correlation between two or more independent variables yields imprecise estimates of the individual effect of each of these variables on the outcome.

will continue to exercise discretion through legal and extra-legal factors. Theorists such as Ulmer (2000) and Kramer and Ulmer (1996) in fact argue that legal factors such as criminal history and crime severity are used informally by the judges as a way of assessing blameworthiness and protecting the community, two of their three “focal points” of judicial decisionmaking. Legal factors can also be used by judges in Albonetti’s related theory of bounded rationality (1991). In a somewhat different formulation, Kessler and Piehl (1998) concluded that prosecutors and judges exercise discretion within the class of factually similar crimes, allowing laws to spill over to defendants with related legal factors. Such a finding is consistent with the notion of informal use of legal factors in sentencing decisions.

If there *is* informal use of legal factors by the judges, then it seems reasonable to ask if this use is appropriate. For example, evidence exists that judges use their discretion to give departures to those with less-serious criminal histories and crime severity (Kramer & Ulmer 1996, Miethe & Moore 1986), and these offenders tend to be white. While not necessarily discrimination, this disparity would automatically be considered warranted in the traditional approach. If we draw a distinction between what the guidelines tell judges to do and what the judges actually do, we no longer have to conclude that all disparity that can be attributed to both race and legal factors should be termed *warranted*. Identifying the behavior of a particular class of actors is also useful from a policy perspective because it makes intervention possible. Knowing whether racial disparity is reduced or increased as the result of judicial discretion helps move one toward action (relative to knowing simply the level of disparity of the criminal justice system as a whole). Also, it may be interesting to compare how the exercise of judicial discretion varies across jurisdictions with different institutional environments. This exercise is only possible once judicial discretion has been isolated.

We implement our approach using data on sentenced offenders from the state of Maryland. We find evidence of substantially more discretion on the part of judges (and/or prosecutors) than was previously believed to be the case. Furthermore, we find evidence of substantial racial disparity in sentencing outcomes, which is attributable to judicial discretion. This disparity is on the order of 20% higher sentences for African Americans than for whites. The difference between our results and those in the rest of the literature is, in some cases, the result of a different modeling approach. In other cases, modeling is unlikely to explain the divergence, and we speculate that differences in the institutional structure across jurisdictions may be responsible. Before getting to the model and results, however, we review the current literature on disparity in sentence outcomes.

## II. Literature Review

The study of unwarranted disparity in sentence outcomes was first implemented in jurisdictions with indeterminate sentences and later in jurisdictions that had adopted determinate sentencing structures, primarily sentencing guidelines. (See Spohn 2000, Sampson & Lauritsen 1997, and Albonetti 1998 for reviews of the literature.) Although it is risky to generalize over such a large and disparate literature, these models always include measures of offense severity and criminal history of the defendant, and are usually estimated in two stages (the in/out decision and sentence length). Models in jurisdictions with determinate sentences apply the measures of criminal history and crime severity used in the guideline scoring systems. The two models (in/out and sentence length) taken together tend to have substantial explanatory power because of the legal factors, as expected. After the most recent round of research, a consensus has emerged that there is little evidence of direct racial discrimination. The racial discrimination that is observed tends to occur in the in/out decision, and these findings are of modest size. Discrimination is generally estimated at less than a 10% increase in the probability of prison for blacks relative to whites (Spohn 2000, Sampson & Lauritsen 1997).

Over the past 20 years, people have refined the approach to this question. This evolution was driven by the definition of unwarranted disparity, which implies that the validity of research findings depends on the adequacy of the methods and measures used to control for the impact of legal factors on sentence outcomes. If these controls are inadequate, or the statistical model is inappropriate, then the findings (or absence of findings) of racial discrimination might be the result of specification bias.<sup>6</sup> Because of the importance of specification, much of the research in this area has focused on improving the explanatory fit of the models by improving the measures or the statistical models. Peterson and Hagan (1984) raised the issue that the lack of discrimination in the sentence-length model was the result of the inclusion of only those people who received a prison sentence. To solve this problem, they proposed a two-stage sample-selection technique that attempts to control for sample-selection bias. They reported larger race effects with the sample-selection technique, and this technique and variations of this technique are now used in virtually all models in the field. Albonetti (1997, 1998) has moved to a one-stage model incorporating probation and prison outcomes, using the Tobit to correct for censoring in the same spirit as the two-stage sample-selection corrections. As

---

<sup>6</sup> This is an argument first made by Kleck (1981) and Wilbanks (1987). See also Zatz (1987).

another example of specification improvements, Wheeler et al. (1982) and Albonetti (1998) logged the sentence-length variable in order to deal with the positively skewed sentence-length data, another empirical artifact that could bias results.

The introduction, in many states, of sentencing guidelines changes the statistical problem researchers face. Guidelines tell the judges exactly which legal factors should be used, and how. In particular, individuals are assigned a score for criminal history and a score for crime severity, and these scores are then used to locate a unique cell on the sentencing grid. Each cell in the guidelines grid specifies a minimum and maximum sentence length.

Mustard (2001) pointed out that the federal sentencing guideline grid is not linear, although the measures of criminal history and crime severity are typically entered in a linear way into statistical models. He modeled the scores in a nonlinear way by including a dummy variable for each cell in the grid, which improved the fit of his model. This improvement in fit was taken as a sign that the controls for legal factors are improved and therefore the resulting estimate of the race effect is more valid. Engen and Gainey (2000a) made the same observation about non-linearity of the grid in Washington state, but recommended the use of the presumptive sentence prescribed by the guidelines (in their case, the midpoint of the cell). When controlling for legal factors the standard linear way, by linearly entering the scores on which the grid is based, Engen and Gainey (2000a) have an adjusted  $R^2$  of 0.512. The adjusted  $R^2$  improves to 0.726 when they use dummy variables rather than the linear scores,<sup>7</sup> and improves even more to 0.804 in their preferred model using the presumptive sentence. Using federal data, Albonetti (1998) also showed the inclusion of the recommended sentence (maximum sentence in her case) increased model fit. Ulmer (2000) also showed an increased fit in Pennsylvania data when the guideline minimum was added to the sentence-length model in place of the linear controls for crime severity and criminal history. In each case, the estimates for the effects of extra-legal factors decline as the model fit improves.

Based on these types of models, some reviewers have concluded that judges have only a modest ability to discriminate, since legal factors account for so much of the variation in sentence outcomes. According to this line of thought, major discrimination, if it exists, must occur before the actions of the judge (Sampson & Lauritsen 1997). Researchers skeptical of this conclusion have begun to look for indirect discrimination by

---

<sup>7</sup> This model is not exactly the model recommended by Mustard (2001), which includes a dummy for each cell, not one for each value of each of two legal factor scores. The Engen and Gainey dummy variable model has 22 dummy variables; using the Mustard approach would have added approximately another 100 variables.

judges by estimating interactive models that allow for the impact of legal factors to vary by race. (See Miethe & Moore 1986, Myers & Talarico 1987, Chiricos & Crawford 1995, Zatz 1987, and Spohn 2000.)

The problem with concluding that judges (and/or prosecutors) have a limited ability to discriminate is that the basic approach to identifying discrimination was not created to isolate and explain the decisions with regard to sentencing. Instead, the statistical models were developed to estimate the system-wide impact of extra-legal factors on sentence outcomes over and above that of legal factors. While this latter line of research is valuable, it necessarily limits the conclusions one can reach about the actions of particular actors in the criminal justice system. The finding of a high  $R^2$  in such a model could be the result of judicial discretion, but it could also be the result of the actions of other actors in the criminal justice system.

This important observation has been made before by researchers such as Klepper et al. (1983) and more recently by Kramer and Ulmer (1996). The sentence outcome for any particular offender is the result of a process that involves many actors, starting with the arrest and ending with the judge's sentencing decision or the plea bargain agreed to by the prosecutor. Just as failure to adequately model legal factors could lead to misattribution of explanatory power to extra-legal factors such as race, failure to adequately model the actions of a particular actor could lead to misattribution of racial bias to other actors (Klepper et al. 1983).

As a result, if researchers want to study the variation in sentence outcomes that is the result of a particular actor, they must first isolate the variation that can be attributed to that actor. In the case at hand, the first step therefore should be to divide the variation in sentence outcomes into that due to the judge or prosecutor and that due to other actors. At that point, the researcher can begin to explain that variation.

Researchers studying guideline departures use this approach (Kramer & Ulmer 1996, Langan 1996, Miethe & Moore 1986, Maxfield & Kramer 1998, Ulmer 1997). Guideline departures must be the result of judicial discretion because everything that happens prior to sentencing places the offender in a particular cell on the grid.<sup>8</sup> The fact that the offender received a sentence departing from the range associated with that cell can only be attributed to the judge (either directly or through agreeing to a plea bargain offered by the prosecutor). These models tend to

---

<sup>8</sup> We do not mean to minimize the complicated (and perhaps strategic) behavior that takes place prior to sentencing, including, e.g., bargaining over charges. Instead, we abstract from this process to emphasize the sentencing stage. Such an abstraction may be more appropriate in jurisdictions with lower incentives to conclude all bargaining well before confronting the guidelines grid.



have fairly low  $R^2$  (usually no more than 20%), which is consistent with the idea that departures depend on highly case-specific factors. Studies of departures also almost always find evidence of racial bias, even in data sets that generally find small or nonexistent race effects on sentence length, such as in Pennsylvania (Kramer & Steffensmeier 1993). This finding suggests that the actions of one set of actors can be hidden by the actions of the court as a whole.

Moreover, these papers provide evidence that race can exert an indirect influence through legal factors. For example, Miethe and Moore's (1986) Table 5 shows that race and other extra-legal factors can predict criminal history and crime severity. The effect of race is then attenuated when criminal history and crime severity are entered into the model of guideline departures. Although attenuation is expected in a general model of sentencing outcomes, it is less expected in a model of departures, since legal factors already drive the guidelines themselves.

In their qualitative work on court decisionmaking, Kramer and Ulmer (1996) provide an example of how race can affect outcomes through criminal history and crime severity. One reason that judges give for a downward departure is that the offense or prior record is less serious than the guideline scores indicate. Judges use this reason 14% less often for African Americans than for whites. This result is consistent with the Kramer and Ulmer (1996) empirical result that those with less-severe criminal histories and crime severity are more likely to be granted a downward departure from the guidelines. The net effect is that judges counteract the guidelines at a greater rate for whites than for African Americans. This example demonstrates that the use of criminal history and crime severity by the judges is not necessarily innocuous. It is possible that this use of sentencing departure, which is entirely discretionary, is being used to disproportionately sentence African Americans to longer sentences. In this example, what is "legal" action may also be discriminatory behavior. Our approach of focusing on judicial discretion opens up this use of particular case characteristics to legitimate scrutiny.

Of course, guideline departures are not the only source of judicial discretion. The guidelines provide a minimum and maximum for each cell, and a judge can select a sentence anywhere within that range without departing from the guideline range. This variation within the guideline ranges is part of judicial discretion. Griswold (1987) recognized this in his study of judicial discretion under Florida's guidelines and reflected it when defining his dependent variable as the difference between the sentence outcome and the recommended sentence,<sup>9</sup> as in equation (1).

---

<sup>9</sup> Griswold (1987) standardized the dependent variable by dividing by the recommended sentence length. We omit this step in equation (1) for conceptual clarity but

$$\text{Actual Sentence Length}_i - \text{Recommended Sentence Length}_i = \alpha + \gamma X + \varepsilon_i \quad (1)$$

Here,  $X$  is a vector including both legal and extra-legal factors that might predict judicial discretion.

In Florida, the guidelines actually suggest a recommended sentence length, roughly the midpoint of the cell. In other jurisdictions, the midpoint can be thought of as a reasonable estimate of the recommended sentence (Engen & Gainey 2000a). As with the models of departures from guidelines, Griswold is only able to explain a small amount of this variation. But, Griswold also showed that offenders with less-serious criminal histories and crime severity tended to get lighter sentences. If these types of offenders tend to be white, then judicial discretion is acting to provide lesser sentences to white offenders, against the recommendations of the guidelines. Finally, although he was unable to include race in the model,<sup>10</sup> he did show that extra-legal factors accounted for some of the explained variation, even when legal factors were included in the model. For example, women convicted of robberies get a 43% reduction in sentence length below the recommended sentence, relative to men.

Engen and Gainey (2000a) also used the guideline midpoint to capture the recommendation of the sentencing commission. However, their model differs in an important way from Griswold's. As shown in equation (2), Engen and Gainey included the recommended sentence on the right-hand side, with a coefficient that is allowed to vary.

$$\text{Actual Sentence Length}_i = \alpha + \beta \text{Recommended Sentence Length}_i + \gamma X + \varepsilon_i \quad (2)$$

This makes sense in the context of the original problem addressed in their paper, that of the non-linearity of the guideline grid. But in their response to a critique of their paper by Ulmer (2000), they began to argue for a different interpretation:

Therefore, we argue that the presumptive sentence in fact has a qualitative meaning apart from the actual sentence. It represents the sentence that the guidelines would have the court impose in most normal circumstances. In other words, *the presumptive sentence represents the formal structure*. It is impossible to observe the substantive process operating to determine sentences "above and beyond" the formal structure of the guidelines unless we control for what the structure prescribes. (Engen & Gainey 2000b:1250–51)

This interpretation is consistent with our understanding of the Griswold model. But a close examination of the Engen and

---

return to it in the next section. Also, for reasons that are not completely clear, Griswold subtracted the actual sentence from the recommended sentence, rendering a downward departure as a positive change and an upward departure as a negative change. We reverse the two for the sake of clarity.

<sup>10</sup> Griswold notes that "even though it was found that race was related to sanctions in the experimental study, no information was collected on race after the guidelines were introduced because of the potential political ramifications" (p. 328).

Gainey model reveals that the coefficient  $\beta$  captures how the sentence outcomes vary with the recommendation of the guideline. In equation (1),  $\beta$  is set equal to 1, by definition, since the recommended sentence is simply the starting point for judicial discretion. By allowing the coefficient  $\beta$  on the recommended sentence to vary, Engen and Gainey are capturing the judges' response to the recommendation. In their model (Engen & Gainey 2000a:1218, model 4), the coefficient on the recommended sentence is greater than 1, indicating that as the recommended sentence increases, the judges use their discretion to punish offenders yet more harshly.<sup>11</sup> They suggest that the legal score variables included in addition to the recommended sentence in the model should capture whether the legal factors influence the use of judicial discretion, but by allowing the coefficient on the recommended sentence to vary, they have not isolated judicial discretion in the same manner as Griswold.

To be fair to Engen and Gainey (2000a), they were trying to focus on the distinction between the guideline's use of legal factors and judges' use of these same factors. This is a difficult distinction to make in a literature so long focused on improving the fit of the models to avoid specification bias. This tension is evident in Ulmer's response to Engen and Gainey's new ideas.

On one hand, Ulmer and his colleagues are responsible for the main theoretical argument supporting the idea of judicial use of legal factors above and beyond the guidelines. For example, Ulmer (2000:1240) states:

As my colleagues and I have always argued, sentencing is an interpretive decision making process even under guidelines and one that is highly variable according to local court community contexts. Offense severity and prior record have a very real qualitative meaning to court community actors—above and beyond the role they play in establishing presumptive sentence recommendations in a guidelines grid.

Yet, in interpreting Engen and Gainey's model, Ulmer reverts to the question of fit that drives the literature on warranted versus unwarranted disparity. He does not evaluate the inclusion of criminal history and crime severity above and beyond that of the recommended sentence in terms of their ability to capture judicial discretion, but in terms of their ability to improve model fit ( $R^2$ ). As Griswold makes clear, one should only include these legal factors in the model if one wants to understand judicial discretion. The inability to improve fit is irrelevant. A zero coefficient on the legal factor scores means only that judges do not make use of criminal history and crime severity above and beyond the recommendation of the guidelines. Intuitively, the presumptive sentence and the legal-factor scores serve two very dif-

---

<sup>11</sup> Ulmer's (2000) estimate of the coefficient on the recommended sentence is also greater than one, although the difference is negligible (1.03).

ferent functions. The first serves to identify judicial discretion, the second serves to explain it. Having said that, the findings in the literature on sentencing departures (showing that judges use criminal history and crime severity over and above how the guidelines recommend they be used) raise the very real possibility that this behavior is discriminatory.

In the next section, we present a model that builds on Griswold's model of judicial discretion. While we are not interested in measuring unwarranted disparity as described in the larger literature previously reviewed, we repeatedly appeal to lessons learned in the larger literature.

### III. A Model of Judicial Discretion

The key empirical task is to identify an approach that allows us to model the guidelines themselves. Once this is done, we can assess the variation in sentencing outcomes accounted for by the guidelines separately from the variation in sentencing outcomes that is not accounted for by the guidelines. We start from the simple model provided by Griswold (equation 1), which can be rewritten as equation (3).

$$\text{Actual Sentence Length}_i = \alpha + \text{Recommended Sentence Length}_i + \gamma X + \varepsilon_i \quad (3)$$

Stated directly, this model says that the sentence-length decision starts from the recommended sentence length from the guidelines, but there may be additional observed variation that can be explained by a vector of explanatory factors  $X$ , including race. Removing all of the variation due to sentencing guidelines allows us to isolate the discretion of the judge.

This understanding of sentencing implies that the decision process occurs in one stage. After receiving input from the guidelines, the judge makes a decision about sentence length, a decision that can include a prison sentence of zero (meaning probation of some unobserved length). Although practice varies by jurisdiction, most guideline grids do not allow probation for most of the cells. For those cells in which a prison sentence is the minimum sentence, probation only occurs as a departure from the guidelines. Although judges can and do go below the recommended minimum to give a sentence of probation, it makes sense to view this as a deviation from the guidelines on a continuous dimension rather than as the first step in the sentencing decision. For those cells in which probation was allowed, it makes sense to think of the probation decision as an unobserved, censored outcome.

The literature is divided on whether to estimate in one or two stages. Most researchers argue that it is imperative to model the sentencing decision in two stages and that failure to do so is incorrect and misleading (Spohn 2000, Steffensmeier et al. 1993).

A small minority, as will be discussed later, have been estimating a single-stage process for some time. Most studies cite Wheeler et al. (1982) as the basis for the conceptual distinction between probation and prison sentences. Interviews with federal judges in white-collar crime cases revealed that the first and most important decision these judges made was the decision to incarcerate.

The Wheeler et al. study, however, was conducted on federal data before the onset of structured sentencing. At that time judges had virtually unfettered discretion with respect to sentencing, and parole boards had wide discretion in determining release from prison. Under this indeterminate sentencing, "parole boards generally control the precise duration of prison terms" (Reitz 1998:390). Thus, modeling sentencing in two stages was an appropriate structure (though given the power of the parole board, it is unclear whether anything meaningful about actual experience could have been inferred from the results of the second stage). Once guidelines were introduced, judicial discretion was curbed, by design, and in many jurisdictions the discretion of the parole board was eliminated. For many defendants, incarceration is the only option unless the judge decides to deviate from the guidelines. The conceptual argument for modeling the sentencing decision as two stages is considerably weaker under determinate sentencing than it was when it first became standard in the literature, before sentencing reform.

Of course, if we decide to model the judicial decision as one stage in response to the presumptive sentence of the guideline, it is still important to take into account the fact that the data on probation is censored, as recommended first by Peterson and Hagan (1984). To deal with censoring, Albonetti (1997, 1998), Rhodes (1991), and Steffensmeier and Demuth (2001) modeled the sentence outcome in one stage using a Tobit regression. Censoring occurs when the dependent variable is not observed because it falls below (or above) a certain threshold. In the case of sentence length, we have two censoring points. All individuals who are sentenced to probation are censored at the cutoff value of zero, and all individuals who are sentenced to life are censored above a certain number of years. For example, in the data we use, the longest sentence for someone with less than a life sentence is 40 years (1 person), while another 4 people have sentences of 35 years. Therefore, any standard greater than 40 years would distinguish a life sentence from one with a defined term, so we will use 40 years as the cutoff in the Maryland data set.

Regardless of the cutoff, there is clearly variation in the severity of the sentence among individuals who are sentenced to probation, but we are not able to observe this variation. Another way to say this is that there is a latent variable  $y^*$  which is only ob-

served when its value is between the threshold values. This can be written formally as:

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ y_i^* & \text{if } 0 \leq y_i^* \leq 40 \\ 40 & \text{if } y_i^* \geq 40 \end{cases} \quad (4)$$

The Tobit model (Tobin 1958) solves this problem by entering all of the information we have into a likelihood function based on the assumption that  $y^*$  would be distributed normally if it was fully observed.<sup>12</sup> We can then look at the separate contribution of each part of equation (4) to the likelihood function, as in equation (5):

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 & \rightarrow & F_N(0 | \mu_i, \sigma) \\ y_i^* & \text{if } 0 \leq y_i^* \leq 40 & \rightarrow & f_N(y_i | \mu_i, \sigma) \\ 40 & \text{if } y_i^* \geq 40 & \rightarrow & 1 - F_N(40 | \mu_i, \sigma) \end{cases} \quad (5)$$

In effect, this likelihood function is the product of two probit regressions and one linear regression model.

As with all statistical models, there are some limitations to the Tobit. The underlying distribution may not in fact be normal, and there is no direct way to test this assumption. In addition, the coefficients in the probit equations are assumed to be proportionate (based on the estimate of the standard error) to the coefficients in the truncated linear regression model. As noted by Steffensmeier and Demuth (2001), this can be a restrictive assumption if the same causal factors are not responsible for the in/out decision and the sentence-length decision. From our viewpoint, however, this is not a methodological issue of general versus restrictive models, but rather a theoretical issue. We believe that judicial decisionmaking in the presence of guidelines is a one-stage process. Therefore, by definition, the causal factors must be the same across the truncation boundary. Because of our theoretical approach, the Tobit becomes the model of choice because it deals with the truncation boundary but does not force us to treat the truncation boundary as a separate decision stage with its own causal structure.<sup>13</sup> Ultimately, the main benefit of the Tobit model lies in its ability to deal with the truncation problem evident in sentencing data while implementing a model that reflects our theory of judicial decisionmaking.

<sup>12</sup> Since we only observe  $y$ , the observed distribution does not need to be distributed normally.

<sup>13</sup> It is worth noting that selection correction mechanisms, as they are usually implemented in this literature, are not a solution. Unless exclusion restrictions are specified, these models rely on a parametric assumption to identify the selection correction. To make these models less reliant on functional form, one must specify exclusion restrictions, or factors that determine the probability of incarceration but do not influence sentence length for those who are sentenced to prison terms. There are no good examples of such factors in the literature. Econometricians have strongly criticized the use of selection correction models without exclusion restrictions. See Johnston and DiNardo (1997) section 13.12.2, for "some cautionary remarks about selectivity bias" in selection models without appropriate exclusion restrictions.

The fact that the dependent variable is censored is not the only problem with the dependent variable. In levels, the data on sentence length have substantial positive skew. The outliers could lead linear regressions of sentence length to have curvilinear error terms, leading to inefficient standard errors and potential bias. As first observed by Wheeler et al. (1982), the problem of outliers can be dealt with by taking the natural log of the sentence length. This approach has the added advantage of giving the coefficients of the model very straightforward interpretations as the percentage change in  $y^*$  caused by changes in  $X$ . In what follows we deal with the natural log of all sentence lengths.

It is worth noting that Griswold (1987) standardized the dependent variable by dividing the gap between actual and recommended sentence by the recommended sentence. Engen and Gainey (2000a) did not standardize the dependent variable, leaving the units in months. Not standardizing the dependent variable means that the statistical model estimates relationships in levels. Standardizing by taking logs (or using Griswold's approach) means that any discriminatory effect is estimated in (approximately) constant percentage terms. The importance of standardizing becomes clear when thinking about individuals at different points on the guideline grid. A departure of ten months when the recommendation is only twelve months is very different from a departure of ten months when the recommendation is 144 months, yet regressions in levels forces these two departures to have identical meaning. Using logs means not only that the positive skew mentioned previously is better accommodated but also that the model more accurately compares departures from any recommended sentence length. In practice, this means that estimates of racial disparity using a standardized sentence length will reflect short recommended sentence lengths more than using an unstandardized measure will.

The last key challenge is to develop a measure of the recommended sentence. We agree with Ulmer (2000) and Engen and Gainey (2000a) that this measure will be dependent on the jurisdiction in question. Engen and Gainey (2000a) and Griswold (1987) chose the midpoint, but Ulmer (2000) chose the standard range minimum. All three of these choices were legitimate, because they were based on an understanding of the guidelines they were intended to reflect. As we discuss in more detail later, our preferred model for the Maryland data will use the guideline midpoint. With these choices, theoretical equation (3) becomes model (6):

$$\text{Actual Sentence Length}_i = \alpha + \text{Recommended Sentence Length}_i + \gamma X + \varepsilon_i \quad (3)$$

$$\ln(\text{Actual Sentence Length}_i) = \alpha + \ln(\text{Guideline Midpoint}_i) + \gamma X + \varepsilon_i \quad (6)$$

In summary, our model of judicial discretion begins with the traditional desire to control directly for all previous actors in the system prior to sentencing. We contend that the sentence recommendation from the sentencing commission (based on the guidelines) represents the collective input of these actors. Working from this recommendation, the judge assigns a final sentence length (or accepts the sentence recommendation of the prosecutor).

Our approach has three concrete implications for statistical models of racial disparity in guideline jurisdictions. First, the role of “legal factors” as determined in the guidelines should be controlled for in the model by including the recommended sentence length in the model, not the crime severity and criminal history scores used by the guidelines. Second, the coefficient on the recommended sentence length should be fixed so that the discrimination model attempts to explain only the difference between the recommendation and the outcome. Estimating a coefficient on the recommended sentence length would combine the discretion of the judge with the actions of the prior actors, directly defeating our attempt to separately estimate the actions of the judge from the actions of the other actors in the criminal justice system. Finally, any statistical model should estimate probation and sentence length as the outcome of one decision, the decision of the judge in the presence of the recommended sentence. This decision to incorporate probation sentences in the same model with prison sentences requires that the statistical model deal with censored data using an approach such as the Tobit model. After describing sentencing in Maryland in the next section, we operationalize each of these recommendations in the context of the Maryland sentencing guidelines and provide empirical estimates of racial disparity in sentencing.

#### **IV. Sentencing in Maryland**

In May 1983, Maryland’s Judicial Conference, a statewide body of all Maryland judges, voted to apply sentencing guidelines statewide beginning July 1, 1983.<sup>14</sup> This action was undertaken voluntarily by the judges, without legislative mandate. These guidelines were administered by a sentencing guidelines board of 14 justices representing each of Maryland’s judicial circuits. The guidelines had four explicit goals: (1) increasing equity in sentencing, (2) articulating an explicit sentencing policy with a regular basis for review and change, (3) providing information

---

<sup>14</sup> This section draws heavily upon Administrative Office of the Courts (1993).



for new judges, and (4) promoting increased visibility and understanding of the sentencing process.

The guidelines apply to circuit courts only, with separate guideline grids for crimes against the person, crimes against property, and crimes involving controlled substances. The sentences are based on the seriousness of the primary offense and the offender’s prior criminal history record. The guidelines board placed all crimes into five seriousness categories. Offense seriousness for person offenses also takes victim injury, weapon usage, and special vulnerability of the victim into account. The “prior criminal history” score is based on prior adult criminal record, prior adult parole and probation violations, prior juvenile record (for offenders age 25 or younger), and relationship to criminal justice system at the time of the offense. Table 1 presents the sentencing grid for person offenses. The axis across the top of the grid represents prior criminal history, while the vertical axis represents the severity of the current crime. Seriousness of prior record increases from left to right; seriousness of the current offense increases as one moves down the table. (See the Appendix for details on how the scores are calculated.)

**Table 1.** Maryland Sentencing Grid for Person Offenses Lower End of Guideline Range/Upper End of Guideline Range

Offense Category	Criminal History Score							
	0	1	2	3	4	5	6	7 or more
1	P/P	P/P	P/3M	3M/1Y	3M/18M	3M/2Y	6M/2Y	1Y/3Y
2	P/6M	P/1Y	P/18M	3M/2Y	6M/3Y	1Y/5Y	18M/5Y	3Y/8Y
3	P/2Y	P/2Y	6M/3Y	1Y/5Y	2Y/5Y	3Y/7Y	4Y/8Y	5Y/10Y
4	P/3Y	6M/4Y	1Y/5Y	2Y/5Y	3Y/7Y	4Y/8Y	5Y/10Y	5Y/12Y
5	3M/4Y	6M/5Y	1Y/6Y	2Y/7Y	3Y/8Y	4Y/10Y	6Y/12Y	8Y/15Y
6	1Y/6Y	2Y/7Y	3Y/8Y	4Y/9Y	5Y/10Y	7Y/12Y	8Y/13Y	10Y/20Y
7	3Y/8Y	4Y/9Y	5Y/10Y	6Y/12Y	7Y/13Y	9Y/14Y	10Y/15Y	12Y/20Y
8	4Y/9Y	5Y/10Y	5Y/12Y	7Y/13Y	8Y/15Y	10Y/18Y	12Y/20Y	15Y/25Y
9	5Y/10Y	7Y/13Y	8Y/15Y	10Y/15Y	12Y/18Y	15Y/25Y	18Y/30Y	20Y/30Y
10	10Y/18Y	10Y/21Y	12Y/25Y	15Y/25Y	15Y/30Y	18Y/30Y	20Y/35Y	20Y/L
11	12Y/20Y	15Y/25Y	18Y/25Y	20Y/30Y	20Y/30Y	25Y/35Y	25Y/40Y	25Y/L
12	15Y/25Y	18Y/25Y	18Y/30Y	20Y/35Y	20Y/35Y	25Y/40Y	25Y/L	25Y/L
13	20Y/30Y	25Y/35Y	25Y/40Y	25Y/L	25Y/L	30Y/L	L/L	L/L
14	20Y/L	25Y/L	29Y/L	30Y/L	L/L	L/L	L/L	L/L
15	25Y/L	30Y/L	35Y/L	L/L	L/L	L/L	L/L	L/L

NOTES: P = probation; M = months; Y = years; L = life.

The sentencing board considered but ultimately rejected the formal use of mitigating factors. Mitigating factors such as “providing substantial assistance” are to be taken into account by the judge with no formal assistance from the guidelines board. No official list of proscribed factors (that cannot be taken into account) is provided. The guidelines themselves are voluntary, with no penalty for sentencing outside of the range provided by the guidelines. The sentence lengths attached to the grid are deliberately descriptive, rather than prescriptive.

There was a revision of the grids in July 1987, with some changes in the assigned penalties.<sup>15</sup> For eight years, there were no changes in the structure of the guidelines. More recently (and subsequent to the end of our data set), Maryland has become a legislatively mandated guideline state, although the guidelines remain descriptive and the penalty for departure is still weak. There are also several mandatory maximums (and fewer mandatory minimums) prescribed by the legislature that override the guidelines. In almost all cases, the mandatory maximums are below the maximum of the guidelines range. We incorporate these laws into the analysis below.

According to Tonry (1996), voluntary guidelines have not led to measurable declines in unwarranted disparity, leading some to doubt whether judges follow voluntary guidelines. Nevertheless, it is possible for the guidelines to have no net effect on outcomes yet still be used by judges as the starting point for their decisions. This is especially true in Maryland, since the guidelines were descriptive and therefore not intended to change the average sentencing practices of judges.

We attempted to validate in a number of different ways our claim that judges are in fact using the guidelines as a point of departure. First, according to records, guideline worksheets are filled out for the vast majority of all cases. The judges are required to sign the forms, indicating that they are aware of the scores awarded. Second, personal communication with the current director of the Sentencing Commission, along with an article written by one of the founding judges, suggest that judges do in fact use the sentencing guidelines as a starting point for the sentencing (Levin 1984). Apparently, although not bound by the law, judges appreciate the political cover provided by a sentence that falls within the guideline ranges. It is harder to determine, absent a formal survey, whether the judges or prosecutors offering plea bargains focus on the midpoint of the sentence range. We test the robustness of this assumption later.

## **V. Empirical Results**

To estimate racial disparity under our theory of sentencing, we use data collected by the Maryland Administrative Office of the Courts from the worksheets for each convicted offender filled out by court clerks at each circuit court. The database contains demographic information about the offender, criminal history measures, indicators of the characteristics of the crime(s), and case-processing variables.<sup>16</sup> Given that there are separate

---

<sup>15</sup> There was also a change in the method of calculating the adult record in the case of multiple charges, several new offenses were added, and the offense seriousness category was changed for 9 offenses (3 downgrades, 6 upgrades).

<sup>16</sup> For details about data collection, see Souryal and Wellford (1997).

sentencing grids for person offenses, property offenses, and drug offenses, the data must be analyzed by crime type. We have chosen to restrict our attention to person offenses.<sup>17</sup> We further restrict our attention to those convicted of a single count.<sup>18</sup> We limit our analysis to offenses taking place in the period after the set of sentencing reforms went into effect on July 1, 1987, and running for eight years. Less than 10% of the observations were dropped due to missing demographic or inconsistent sentencing information. After these exclusions, the resulting sample is 14,633 people.

Table 2 reports means and standard errors of the key variables for this sample. The first column contains descriptive statistics for the full sample, while other columns include the values for subsets of the sample by race and ethnicity (whites, African Americans, and Hispanics). For the full sample, 73% were sentenced to prison. Among those with a prison sentence, the average length was 45 months. Averaging the prison sentence across the full sample yields a mean sentence of nearly 33 months. Nine percent of the sample was female, and the average age at time of offense was 28 years. The sample was 65% African American and 1.4% Hispanic. The small size of the Hispanic population will limit the extent to which statistically valid statements can be made about that group. The two dimensions of the sentencing grid are summarized in the average scores. The average score for the criminal history dimension was approximately 2, while the average score for the current offense was 4.3. Ninety-two percent of the cases were settled by a guilty plea.

As whites and blacks together account for 98% of the sample, the subsequent analysis of potential discrimination will necessarily emphasize those two groups. Therefore, it is instructive to compare the descriptive statistics in columns (2) and (3). Compared to whites, African Americans are more likely to be sentenced to prison and, if so, they are more likely to have a longer sentence. Putting these two factors together, the overall average sentence for African Americans is more than 35 months and the average sentence for whites is almost 28 months, a gap of 28%. A somewhat higher proportion of African Americans are female. African Americans are somewhat younger than whites. The higher levels of punishment are likely related to the higher crimi-

---

<sup>17</sup> Person offenses include murder, rape and sexual assault, kidnapping, robbery, assault, and similar offenses. The most common crimes in the data set are robbery and assault.

<sup>18</sup> Due to the complexity of multiple-count data, most analysts restrict themselves to single-count data, which is the approach we adopt here. This means that certain situations are omitted from the analysis, including single-event convictions in which a group of offenses was committed during one continuous criminal act (for example, robbery with a deadly weapon and use of a handgun in the commission of a felony); single events with more than one person offense with a seriousness category I or II (the least serious offenses); and multiple criminal events (e.g., a defendant who has been arrested on different dates for drug distribution).

**Table 2.** Descriptive Statistics: Maryland Offenders Convicted of Person Offenses

	(1) All	(2) Whites	(3) Blacks	(4) Hispanics
Any prison term	0.729 (0.004)	0.655 (0.007)	0.769* (0.004)	0.696 (0.032)
Prison term (in months) (if > 0)	44.67 (0.668)	42.34 (1.205)	46.12* (0.820)	28.76* (3.810)
Prison term (in months)	32.57 (0.514)	27.72 (0.840)	35.48* (0.662)	20.02 (2.800)
Female	0.087 (0.002)	0.076 (0.004)	0.093* (0.003)	0.033* (0.012)
Age at offense	28.14 (0.079)	30.33 (0.148)	27.03* (0.090)	27.39* (0.556)
African American	0.646 (0.004)	0	1	0
Hispanic	0.014 (0.001)	0	0	1
Other race/ethnicity	0.006 (0.001)	0	0	0
Criminal history score	1.967 (0.018)	1.854 (0.031)	2.069* (0.021)	0.636* (0.084)
Offense seriousness score	4.326 (0.020)	3.867 (0.034)	4.567* (0.026)	4.200* (0.177)
Guilty plea	0.917 (0.002)	0.909 (0.004)	0.922* (0.003)	0.897 (0.021)
<i>N</i>	14,633	4,877	9,448	214

NOTE: Means with standard errors are in parentheses. An asterisk indicates that the mean or proportion is statistically different from the mean for whites at the 5% level of significance.

For the second row, only those with prison terms are included. *N*'s are, respectively, 10,669; 3,198; 7,268; and 149.

nal history and current offense scores of African Americans. Finally, African Americans are somewhat more likely to have settled their cases through a plea.

Table 3 begins with specifications of the form generally estimated in the literature assessing discrimination in sentencing. Column (1) reports a probit of the "in/out" decision, finding that African Americans are 5.3 percentage points more likely than whites to be sentenced to prison, while Hispanics are about 9 percentage points more likely. Guilty pleas are associated with a lower likelihood of prison. As is to be expected, higher offense and history scores are associated with a higher likelihood of prison. All coefficients have the expected signs.

Column (2) reports results from OLS of sentence imposed (in months) for those sentenced to prison. Here there are no significant differences among the racial/ethnic categories. In contrast, women continue to receive less punishment than do men. Pleas continue to be associated with reductions in sentence length. These findings are quite consistent with the general picture in the literature: Little or no difference is found among the

**Table 3.** Estimates of Racial Disparity: Two Stages  
Maryland Person Offenses

	(1) Probit: In/Out	(2) OLS: Sentence Length
African American	0.053* (0.008)	-1.207 (1.161)
Hispanic	0.087* (0.020)	-6.724 (4.440)
Other race/ethnicity	0.022 (0.040)	-11.200 (6.927)
Female	-0.173* (0.015)	-10.226* (2.181)
Age at offense	-0.006* (0.002)	0.496 (0.273)
Age squared (*100)	0.003 (0.002)	-0.500 (0.394)
Guilty plea	-0.079* (0.011)	-25.311* (1.766)
History score	0.074* (0.002)	8.888* (0.248)
Offense score	0.040* (0.002)	13.436* (0.230)
Sexual assault	-0.231* (0.027)	-28.54* (2.424)
Assault	-0.238* (0.023)	-44.63* (2.236)
Robbery	-0.109* (0.022)	-41.93* (2.169)
Weapons offenses	-0.332* (0.038)	-39.82* (4.378)
Kidnapping	-0.227* (0.064)	-42.22* (6.279)
Other	-0.158* (0.025)	-52.21* (2.428)
-Log likelihood	6,771.3009	—
R-squared	—	0.4191
N	14,633	10,669

NOTE: A constant and a linear time trend are included in all specifications. Column (1) reports coefficients transformed into changes in probability. Column (2) includes only those observations with prison terms. The excluded crime category is homicide.

racess when considering prison-term length, while a statistically significant difference is found for the probation/prison distinction. With confidence that our data conform to the literature when traditional techniques are used, we now begin to implement our theoretical model.

The primary empirical results for our statistical model are reported in Table 4. Column (1) reports a preliminary model. Control variables in this and all other models are Hispanic, other race or ethnicity, gender, age, and age squared (to allow the effect of age to be nonlinear), guilty plea, and a time trend. In column (1), we estimated the coefficient on the logged midpoint as 1.1, which means that a 10% increase in the midpoint is associ-

ated with an approximately 11% increase in sentencing outcome. The estimated coefficient on African American is 17%, which is substantially higher than values generally reported in the literature. The estimates on the other racial and ethnic groups continue to be poorly estimated due to small samples. The coefficients on the other variables are of the expected direction. Together, the two age variables indicate that there is a concave relationship between age and sentence length. That is, the positive first-order effect (age) shows that sentences increase with age, and the negative second-order effect (age squared) shows that those increases get smaller as age increases.

**Table 4.** Modeling Sentencing Outcomes: Tobits of Ln (Sentence Length) Maryland Person Offenses

	(1)	(2)
African American	0.167* (0.037)	0.203* (0.036)
Hispanic	0.013 (0.145)	-0.014 (0.143)
Other race/ethnicity	-0.288 (0.224)	-0.292 (0.221)
Female	-1.289* (0.064)	-1.301* (0.063)
Age at offense	0.012 (0.008)	0.014* (0.008)
Age squared (*100)	-0.038* (0.011)	-0.042* (0.011)
Guilty plea	-0.750* (0.060)	-0.769* (0.059)
Ln (guidelines midpoint)	1.111* (0.014)	1
Standard error	1.9165	1.9020
-Log likelihood	24,205.944	24,237.281
<i>N</i>	14,633	14,633

NOTE: A constant and a linear time trend are included in all specifications.

Allowing the guidelines to enter unconstrained, as in column (1), means that we controlled for the guidelines and their implementation, rather than only the former. The second column of Table 4 estimates equation (6) by fixing the effect of the Recommended Sentence-Length variable to be proportionate. Fixing the guidelines in this manner means that the demographic variables are only permitted to explain the variation in sentencing outcomes that remains once the variation in the sentencing recommendation has been removed. This model is our preferred specification, as we argued for in section III of this article. The effect of constraining this coefficient is a small increase on the estimated gap between African Americans and whites, to 20%.

With this estimate, we have isolated the variation that is associated with judges, whether at trial or by approving plea bargains. The remaining variation should be highly case-specific, as argued

by Griswold. Thus, as in the departure literature, it should not be surprising (nor of concern) that explanatory power is low. Among the variation in sentences that is the result of judicial discretion, a substantial part of it is correlated with race.

This estimate of racial disparity may be biased due to measurement error if our particular choice of the midpoint does not accurately represent the sentence length recommended by the commission. If this is the case, the error term will capture some of the variation in the recommended sentence. This in turn will be positively correlated with the race variable, since African Americans tend to have higher recommended sentences. The net result is that measurement error in the recommended sentence might lead to positive bias in the race coefficient. Although we believe that a race effect of 20% is of reasonable magnitude given the evidence (recall from Table 2 that African Americans have 28% longer sentence lengths than whites when probation is viewed as a sentence of zero years), we are concerned about the possibility of overstating the size of the race effect.

To explore this issue, we estimated the model with several other measures for the recommended sentence length. Replacing the midpoint of the guidelines range with the minimum (in recognition that many sentences fall in the lower end of the range) did not change the substantive results, in either the unconstrained or the constrained version of the model. We also tried using two variables, rather than one, to represent the guidelines. Specifically, we used the lower endpoint and the range. When we ran this model, we could not constrain the coefficients to fix the impact of the guidelines because a proportionate constraint is not well defined. However, from the previous results it is clear that the constraint on the guideline does not change the substantive results—large disparities of similar magnitudes are estimated in both cases. The unconstrained version of the model with two controls for the sentencing guidelines yielded results little changed from those reported in column (2) of Table 4. These results give support to our approach because, although they each are subject to the claims of measurement error, the error, and therefore the bias caused by that error, will be very different in each case. The fact that the race effect did not change in any substantive way when various measures were used suggests that bias due to measurement error is not an important part of the estimate. This specification check leads us to conclude that our estimate of a 20% difference in the sentence length between African Americans and whites as a result of judicial discretion is reasonable and defensible as a first attempt at estimating our theoretical model.

We performed a similar set of specification checks to explore sensitivity of the results to various methods of incorporating mandatory maximum sentencing laws that override the guide-

lines. In our base model, we reflected mandatory sentences in the same manner as a guideline value. That is, if a mandatory sentence level were lower than the upper end of the guideline range, the maximum was used instead of the commission's recommendation. We felt this was the most reasonable approach, as this is the way the sentencing options are presented on the scoring sheets used by judges. However, to check for robustness, we re-estimated all models, excluding those subject to mandatory sentencing laws. The results were not affected. Finally, given that guilty pleas are more frequent among African Americans than among whites (in contrast to the usual findings in the literature), we estimated our models for only pled cases. Again, the results were unaffected.

We now turn to explaining the difference between our estimate and the much smaller estimates generally found in the literature. We first consider how our results compare to the traditional approach that includes the criminal history and offense-severity measures as controls. To do this, we disaggregate the differences between the model in Table 3 and our preferred model in column (2) of Table 4 by estimating the model in stages, which we report in Table 5. The first two columns of Table 5 isolate the effect of the Tobit model. Both specifications estimate all sentencing outcomes simultaneously, using the dependent variable in logs. In the first specification, OLS is run. In the second, the Tobit is used. Switching from the OLS to the Tobit leads to a larger and now statistically significant estimate for racial disparity (from 4% longer sentences for African Americans in column (1) to 11% longer in column (2)). The substantive interpretations of the other variables in the models are little affected by the change in model specification. This result is consistent with the standard econometric prediction that censoring and truncation can lead to downward bias on coefficients of interest. More substantively, it means that moving to the Tobit accounted for approximately half of the 17 percentage point difference between column (2) of Table 4 and column (1) of Table 5.

The last two columns of Table 5 illustrate the importance of the primary innovation of this article, that of how to control for sentencing guidelines when identifying the role of race. Column (3) contains results from our preferred specification (column (2) of Table 4) augmented with the linear controls for criminal history and current offense. This exercise allows for a direct comparison of the various variables used to control for how the guidelines punish the legally relevant factors. Even though the sentencing guidelines are already controlled for through the constrained coefficient on the guidelines midpoint, these legal-factor scores have explanatory power. Here, both scores help to explain the deviation in sentencing outcomes from the recommended sentence. The inclusion of these factors absorbs a fair



**Table 5.** Modeling Sentencing Outcomes: Additional Specifications—Maryland Person Offenses

	(1) OLS: Ln (Sentence Length)	(2) Tobit Ln (Sentence Length)	(3) Tobit Ln (Sentence Length)	(4) Tobit Ln (Sentence Length)
African American	0.036 (0.024)	0.112* (0.035)	0.141* (0.037)	0.141* (0.037)
Hispanic	0.116 (0.092)	0.233 (0.135)	0.173 (0.145)	0.175 (0.144)
Other race/ethnicity	-0.117 (0.137)	-0.233 (0.207)	-0.194 (0.223)	-0.174 (0.222)
Female	-0.549* (0.040)	-0.923* (0.061)	-1.127* (0.064)	-1.121* (0.064)
Age at offense	0.002 (0.005)	-0.003 (0.008)	-0.015 (0.008)	-0.017 (0.008)
Age squared (*100)	-0.012 (0.007)	-0.014 (0.011)	-0.002 (0.012)	-0.001 (0.012)
Guilty plea	-0.517* (0.040)	-0.637* (0.056)	-0.697* (0.060)	-0.693* (0.059)
History score	0.319* (0.005)	0.419* (0.008)	0.147* (0.008)	0.085* (0.016)
Offense score	0.312* (0.005)	0.390* (0.007)	0.052* (0.007)	0.024* (0.009)
Offense score *	—	—	—	0.014* (0.003)
History score				
Crime categories (7)	Yes	Yes	No	No
Ln (guidelines midpoint)	—	—	1	1
-Log likelihood	—	23,602.967	24,049.079	24,039.436
R-squared	0.4691	—	—	—
Standard error	—	1.8497	1.9053	1.8993
N	14,633	14,633	14,633	14,633

NOTES: A constant and a linear time trend are included in all specifications.

Crime categories, controlled for using a set of indicator variables, are homicide, sexual assault, assault, robbery, weapons offenses, kidnapping, and other.

amount of variation from the African American variable, reflecting the multicollinearity apparent in the descriptive statistics. African Americans tend to have higher history and severity scores. But column (3) says that judges tend to give people with higher scores, who tend to be African American, longer sentences than those recommended by the guidelines. Moreover, there still appears to be a substantial direct difference in sentencing outcomes across whites and African Americans, on the order of 14%.<sup>19</sup> This result implies that the “legal factor” scores in the

<sup>19</sup> The 3 percentage point difference between column (3) and column (2) can be attributed to the lack of crime type in our preferred specifications. Variation due to crime type was specified by the Maryland Sentencing Commission in two ways: drawing the broad distinction drawn among person offenses, drug offenses, and property offenses when they created the three separate grids, and building crime type into the offense seriousness dimension through the crime-severity measure. Use of offense type (on top of the guidelines recommendation) by the judges in their decisions represents the explicit exercise of their discretion outside of the guidelines and is therefore open for scrutiny by the researcher under two conditions: (a) that this variation represents the difference be-

model captured roughly 6 percentage points of the race effect, telling us that judges use legal factors above and beyond the guidelines. In the traditional approach this variation would have been ascribed to the guidelines themselves.

The final column of Table 5 presents an attempt to start modeling judicial discretion more formally. The model adds one more variable to those included in the previous column: an interaction between the two linear measures of legal factors in order to capture more of the non-linearity in the sentencing grid. Recall that the sentencing grid itself is not linear. In our preferred model, equation (6), the non-linearity of the grid is accounted for by the explicit inclusion of the recommended sentence length. Yet, when we begin to study the impact of legal factors above and beyond the impact of the guidelines, we again need to address the issue of non-linearity, since we are now attempting to observe how the judges use legal factors over and above how the guidelines applied them. A simple way to reflect this nonlinear use of legal factors is with the interaction term, which allows the effect of the criminal history score to depend on the value of the offense severity score (and vice versa). The positive coefficient on this term indicates that sentences are particularly likely to deviate positively from the guidelines when both the criminal history and current offense scores are high. This effect occurs over and above the non-linearity in the guidelines themselves.

These final two columns indicate that a substantial part of the difference between our results and those in the literature is due to the way legal factors have been controlled. More precisely, when legal factors are included as controls for the guidelines with no restriction on their coefficients, as in the traditional approach, the estimated coefficients will reflect not only what the guidelines told the judge to do based on the legal factors but also the response of the judge to those legal factors. This approach fails to provide us with a way to distinguish between the actions of the other actors in the criminal justice system and judicial behavior. Moreover, given the strong correlation between legal factors and race, failure to differentiate between the guidelines and legal factors means that there is no way to determine if any of the common variation between race and legal factors might in fact be questionable. Column (4) of Table 5 provides at least the starting point for an exploration into the potentially problematic use of legal factors above and beyond what is mandated by law. In future research, we recommend entering legal factors in even less restrictive ways to more fully understand the behavior of the judges.

---

tween what the law said to do and what the judge did and (b) that this variation is correlated with race.

In much of this section, we have focused on explaining how our specification differs from the traditional focus on warranted versus unwarranted disparity. But as noted in the introduction, Engen and Gainey (2000a) and Ulmer (2000) have utilized approaches similar to ours. (Recall that Griswold [1987] did not have race in his model.) One finding common across all three studies is that the unconstrained inclusion of the presumptive sentence gains a coefficient greater than one, meaning that those with longer prescribed sentences are punished relatively more harshly. This suggests to us that if our modeling approach constraining the coefficient on the presumptive sentence to be one were taken in these other jurisdictions, they too would find more racial disparity after constraining the coefficient on the presumptive sentence, although this increase would likely be small, as it was in our model.<sup>20</sup> It should be kept in mind, however, that there are several additional differences in specification in this article, most notably logging the dependent variable and the use of the Tobit. (The Tobit both substitutes for the use of a selection correction to deal with probation sentences and corrects for censoring for life sentences.)

Although we do not have space here to provide a detailed comparison of the alternative specifications, we believe that the differences among this set of results are partially explained by differences in the guideline structures in the three jurisdictions (Washington, Pennsylvania, and Maryland). Washington has guidelines mandated by the legislature, with very narrow ranges; Pennsylvania is a legislatively mandated guideline state with formal mitigating factors but with voluntary compliance; and Maryland is a judicially approved guideline state with wide ranges, limited structure, and voluntary compliance. A priori, we expect to find more judicial discretion as we move from Washington to Maryland, and this should drive the findings of the competing models.

For example, in the Washington data, Engen and Gainey found that the inclusion of presumptive sentence leads to increased explanatory power. We, like Ulmer, find that inclusion of the presumptive sentence leads to lower explanatory power than the use of the criminal history and current offense scores. While, unlike Ulmer, we are unconcerned about the reduction in model fit, this result does suggest in part that judicial discretion represents a smaller amount of the overall variation in the sentence outcomes in Washington than in the other two states. This is the same interpretation suggested by Ulmer in comparing Pennsylvania to Washington. Moreover, in the traditional specification

---

<sup>20</sup> The fact that the unconstrained coefficient is only slightly above 1 implies that if judges are using legal factors above and beyond the guideline recommendations they are doing so in a nonlinear, non-continuous way, perhaps focusing on certain types of histories and crime types.

the legal factor scores capture both the recommendation of the guidelines and judicial discretion. Therefore, a drop in the fit after a respecification that captures only the recommendation of the guidelines suggests that the judges in the states with the less-strict guidelines are using the legal factors in a manner not approved by the guidelines commission. Both our analysis and Ulmer's (2000) demonstrate this to be true by showing that legal factors remain significant after the presumptive sentence is included in the model.

To a limited extent, we can also compare our findings to those in Albonetti (1998). She analyzed federal white-collar offenders, a far different population than those in any of the studies previously mentioned. Because she does not report specifications with only the guidelines included, it is hard to make the same direct comparisons made earlier. Her primary specification included the guidelines recommendation (as the maximum of the range), the criminal history score, the offense seriousness score, and the offense category. In this model, the offense type and legal factor score variables had explanatory power over and above the guidelines, echoing the result we found in Table 5. However, the guideline maximum has a coefficient of only  $-0.04$ . If this coefficient were constrained to equal one, as in our model, the coefficients on the legal factor scores would surely fall (given that they are highly positively correlated with the recommended sentence). In her model, Albonetti found statistically significantly longer sentences for African Americans. However, it is hard to predict how changing the specification to isolate judicial deviations from the guidelines would affect the estimate of racial disparity because, unlike in our data set, race has a complex set of interactions with the legal factor variables: Being African American is negatively correlated with offense seriousness and positively correlated with criminal history.

## VI. Discussion

The interpretation of quantitative estimates of race coefficients depends heavily on the theoretical model of decisionmaking by the actors in the criminal justice system. Our approach explicitly considers the role of the sentencing commission in setting the backdrop for a judge's (or plea bargainer's) final decision as to the sentencing outcome. Given our view that the recommended sentence in the guidelines represents the actions of all of the previous actors in the system, any difference between this recommendation and the sentence outcome reflects judicial discretion.

This starting point leads to several implications for empirical estimation of possible racial disparity in sentencing. First, careful modeling of the sentencing guidelines allows the researcher to

identify the part of the variation in sentencing outcomes that is due to the discretion of the judge (or other sentencing agent). Second, such a model provides no reason for arbitrarily separating the sentencing decision into multiple stages, a position also suggested by other researchers (e.g., Albonetti 1998). Third, modeling the guidelines means that any non-linearity in the sentencing grid, a problem identified by Mustard (2001) and Engen and Gainey (2000a) can be easily captured. Finally, in practice, modeling the guidelines rather than legal factor scores means that more of the variation that race and legal factors share in common will be attributed to the racial status of the offender.

In fact, using data from Maryland, we find that African Americans have 20% longer sentences than whites, on average, holding age, gender, and recommended sentence length from the guidelines constant. This result is 17 percentage points larger than what we found when we allowed legal factors to control for the guidelines and judicial discretion simultaneously. Using the logged guideline midpoint to isolate the recommendation of the guidelines reveals that considerable racial disparity can be attributed to judicial discretion. Moreover, when we begin to try to explain this discretion, we find that some of the racial disparity can be accounted for with the use of legal factors over and above the recommendations of the guidelines. In a result that finds support in the literature on guideline departures, we found that judges in Maryland tended to sentence people in the part of the guidelines grid with longer recommended sentences, who tend to be African American, to longer sentences relative to the guidelines recommendations (relative to offenders situated in other parts of the grid). While this exercise of judicial discretion is not necessarily discrimination, we believe such a result warrants more exploration. This exploration would not have been possible if we had not first drawn a clear distinction between what the guidelines tell the judge to do and what the judge actually does. In effect, this distinction can replace the distinction between warranted and unwarranted disparity that is prominent in the existing literature.

As noted previously, the fact that legal factors such as criminal history are correlated with extra-legal factors such as race provides challenges for empirically estimating racial disparity. Researchers have been forced to identify disparity as only that part of the variation in sentencing outcomes that is correlated with race and not correlated with legal factors. Yet, this is a conservative standard forced on us by the limits of statistical power rather than by theory. There is no reason to assume that variation in the sentence outcomes that is correlated with both race and legal factors is not discrimination.

With the exception of work on guideline departures, researchers have not been able to identify any useful structure to

separate what may be legitimate and what may be illegitimate, so standard practice errs on the side of caution and concludes that the disputed disparity is legitimate. Focusing specifically on judicial discretion provides an alternative approach that we hope will create a new discussion about the ways in which racial disparity might occur in the sentencing process. In this article, our main effort has been directed at clarifying what we see are some important conceptual problems in the debate concerning how best to model racial disparity in sentencing. As such, we make no claim to having fully explained racial discrimination in the criminal justice system in Maryland. Nevertheless, we hope that others might develop other models that foster a richer understanding of how judges react to various legal imperatives and structures. Research on this question might use judge interviews, ethnographic observation, or experimental methods. Whatever the method, however, we believe that a renewed focus on the actions of individual actors will provide a useful new focus on research into racial disparity.

## References

- Administrative Office of the Courts (1993) *Review of the Maryland Sentencing Guidelines: Concepts and 1987 Revisions*. Annapolis, MD.
- Albonetti, Celesta A. (1991) "An Integration of Theories to Explain Judicial Discretion," 38 *Social Problems* 247–66.
- (1997) "Sentencing under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991–1992," 31 *Law & Society Rev.* 789–824.
- (1998) "The Role of Gender and Departures in the Sentencing of Defendants Convicted of a White-Collar Offense under the Federal Sentencing Guidelines," 1 *Sociology of Crime, Law, & Deviance* 3–48.
- Amemiya, Takeshi (1985) *Advanced Econometrics*. Cambridge: Harvard Univ. Press.
- Chiricos, Theodore G., & Charles Crawford (1995) "Race and Imprisonment: A Contextual Assessment of the Evidence," in D. Hawkins, ed. *Ethnicity, Race and Crime*. Albany: State Univ. of New York Press.
- Dixon, Jo (1995) "The Organizational Context of Criminal Sentencing," 100 *American J. of Sociology* 1157–98.
- Engen, Rodney L., & Randy R. Gainey (2000a) "Modeling the Effects of Legally Relevant and Extralegal Factors under Sentencing Guidelines: The Rules Have Changed," 38 *Criminology* 1207–29.
- (2000b) "Conceptualizing Legally Relevant Factors under Guidelines: A Reply to Ulmer," 38 *Criminology* 1245–52.
- Griswold, David B. (1987) "Deviation from Sentencing Guidelines: The Issue of Unwarranted Disparity," 15 *J. of Criminal Justice* 317–29.
- Johnston, Jack, & John DiNardo (1997) *Econometric Methods*, 4th ed. New York: McGraw Hill.
- Kessler, Daniel P., & Anne Morrison Piehl (1998) "The Role of Discretion in the Criminal Justice System," 14 *J. of Law, Economics, & Organization* 256–76.
- Kleck, Gary (1981) "Racial Discrimination in Sentencing: A Critical Evaluation of the Evidence with Additional Evidence on the Death Penalty," 43 *American Sociological Rev.* 783–805.

- Klepper, Steven, Daniel Nagin, & Luke-Jon Tierney (1983) "Discrimination in the Criminal Justice System: A Critical Appraisal of the Literature," in A. Blumstein, J. Cohen, S. E. Martin, & Michael H. Tonry, eds. *Research on Sentencing: The Search for Reform*, vol. 2. Washington, DC: National Academy Press.
- Kramer, John, & Darrell Steffensmeier (1993) "Race and Imprisonment Decisions," 34 *Sociological Q.* 357–76.
- Kramer, John, & Jeffrey J. Ulmer (1996) "Sentencing Disparity and Departure from the Guidelines," 13 *Justice Q.* 81–106.
- Langan, Patrick A. (1996) "Sentence Reductions for Drug Traffickers for Assisting Federal Prosecutors," unpublished manuscript (U.S. Department of Justice).
- Levin, Marshall A. (1984) "Maryland's Sentencing Guidelines—A System by and for Judges," 68 *Judicature* 173–80.
- Maxfield, Linda, & John Kramer (1998) *Substantial Assistance: An Empirical Yardstick Gauging Equity in Current Federal Policy and Practice*. Washington, DC: U.S. Sentencing Commission.
- Miethe, Terence D., & Charles A. Moore (1985) "Socioeconomic Disparities under Determinate Sentencing Systems: A Comparison of Preguideline and Postguideline Practices in Minnesota," 23 *Criminology* 337–63.
- (1986) "Racial Differences in Criminal Processing: The Consequences of Model Selection on Conclusions about Differential Treatment," 27 *Sociological Q.* 217–37.
- Mustard, David B. (2001) "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts," 44 *J. of Law & Economics* 285–314.
- Myers, Martha A., & Susette M. Talarico (1987) *The Social Contexts of Criminal Sentencing*. New York: Springer-Verlag.
- Peterson, Ruth D., & John Hagan (1984) "Changing Conceptions of Race: Towards an Account of Anomalous Findings of Sentencing Research," 49 *American Sociological Rev.* 56–70.
- Reitz, Kevin R. (1998) "Modeling Discretion in American Sentencing Systems," 20 *Law & Policy* 390–428.
- Rhodes, William (1991) "Federal Criminal Sentencing: Some Measurement Issues with Application to Pre-guideline Sentencing Disparity," 81 *J. of Criminal Law & Criminology* 1002–33.
- Sampson, Robert J., & Janet L. Lauritsen (1997) "Racial and Ethnic Disparities in Crime and Criminal Justice in the United States," in M. Tonry, ed. *Ethnicity, Crime and Immigration* (vol. 21 of *Crime and Justice*). Chicago: Univ. of Chicago Press.
- Savelsberg, Joachim J. (1992) "Law That Does Not Fit Society: Sentencing Guidelines as a Neoclassical Reaction to the Dilemmas of Substantivized Law," 97 *American J. of Sociology* 1346–81.
- Souryal, Claire, & Charles Wellford (1997) *An Examination of Unwarranted Sentencing Disparity under Maryland's Voluntary Sentencing Guidelines*. (November 21) [www.gov.state.md.us/sentencing/reports/disparity.html](http://www.gov.state.md.us/sentencing/reports/disparity.html)
- Spohn, Cassia C. (2000) "Thirty Years of Sentencing Reform: The Quest for a Racially Neutral Sentencing Process," in *Policies, Processes, and Decisions of the Criminal Justice System*, vol. 3. Washington, DC: Department of Justice.
- Steffensmeier, Darrell, & Stephen Demuth (2000) "Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who Is Punished More Harshly—White, Black, White-Hispanic, or Black-Hispanic Defendants?" 65 *American Sociological Rev.* 705–29.
- (2001) "Ethnicity and Judges' Sentencing Decisions: Hispanic-Black-White Comparisons," 39 *Criminology* 145–78.
- Steffensmeier, Darrell, John Kramer, & Cathy Streifel (1993) "Gender and Imprisonment Decisions," 31 *Criminology* 411–46.

- Steffensmeier, Darrell, Jeffrey Ulmer, & John Kramer (1998) "The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male," 36 *Criminology* 763–98.
- Stolzenberg, Lisa, & Stewart D'Alessio (1994) "Sentencing and Unwarranted Disparity: An Empirical Assessment of the Long-term Impact of Sentencing Guidelines in Minnesota," 32 *Criminology* 301–10.
- Tobin, J. (1958) "Estimation of Relationships for Limited Dependent Variables," 26 *Econometrica* 24–36.
- Tonry, Michael (1996) *Sentencing Matters*. New York, NY: Oxford Univ. Press.
- Ulmer, Jeffrey T. (1997) *Social Worlds of Sentencing: Court Communities under Sentencing Guidelines*. Albany: State Univ. of New York Press.
- (2000) "The Rules Have Changed—So Proceed with Caution: A Comment on Engen and Gainey's Method for Modeling Sentencing Outcomes under Guidelines," 38 *Criminology* 1231–43.
- U.S. Department of Justice (2000) *Prisoners in 1999*. NCJ 183476. Washington, DC: Bureau of Justice Statistics.
- U.S. Sentencing Commission (1995) *Substantial Assistance Departures in the United States Courts*. Draft Final Report. Washington, DC.
- Wheeler, Stanton, David Weisburd, & Nancy Bode (1982) "Sentencing the White-Collar Offender: Rhetoric and Reality," 47 *American Sociological Rev.* 641–59.
- Wilbanks, William (1987) *The Myth of a Racist Criminal Justice System*. Monterey, CA: Brookes/Cole.
- Zatz, Marjorie S. (1987) "The Changing Forms of Racial/Ethnic Biases in Sentencing," 24 *J. of Research in Crime & Delinquency* 69–92.

## **Appendix: Explanation of Maryland Guideline Scoring**

Criminal history scores range from zero to nine. One point each is given for minor juvenile history (fewer than two commitments to juvenile facilities), minor adult history, being under criminal justice supervision at the time of the offense, and having prior probation violations. One additional point is given for having a more-serious juvenile history (a point in addition to the "minor" one) or a longer adult history (two more points for a moderate adult history, and another two more for the most serious). Juvenile history is expunged at age 26.

The offense score ranges from one to 15. Offenses are grouped into five "seriousness categories," which receive one, three, five, eight, or ten points. Then elements of the crime may add to this score. One point is added for injury to a victim, another point if the victim is permanently injured or killed, one point for use of a weapon, an additional point if that weapon is a firearm, and finally, a point if the victim is classified as "vulnerable." That is, one would receive an offense score of 15 if one committed the most serious of crimes in which a vulnerable victim was killed with a firearm.

A separate guideline sets the minimum sentence for anyone convicted of first-degree murder at life imprisonment, regardless of any other current offense or criminal history attributes. In order to avoid complications due to this supplementary law, we have excluded those convicted of first-degree murder from our sample.