NBER WORKING PAPER SERIES

SENTENCING GUIDELINES AND JUDICIAL DISCRETION:
QUASI-EXPERIMENTAL EVIDENCE FROM HUMAN CALCULATION ERRORS

Shawn D. Bushway
Emily G. Owens
Anne Morrison Piehl

Sentencing Guidelines and Judicial Discretion:Quasi-experimental Evidence from Human
Calculation Errors
Shawn D. Bushway, Emily G. Owens, and Anne Morrison Piehl
NBER Working Paper No. 16961
April 2011
JEL No. K14,K42

ABSTRACT

There is a debate about whether advisory non-binding sentencing guidelines affect the sentences outcomes
of individuals convicted in jurisdictions with this sentencing framework. Identifying the impact of
sentencing guidelines is a difficult empirical problem because court actors may have preferences for
sentencing severity that are correlated with the preferences that are outlined in the guidelines. But,
in Maryland, ten percent of the recommended sentences computed in the guideline worksheets contain
calculation errors. We use this unique source of quasi-experimental variation to quantify the extent
to which sentencing guidelines influence policy outcomes. Among drug offenses, we find that the
direct impact of the guidelines is roughly ½ the size of the overall correlation between recommendations
and outcomes. For violent offenses, we find the same ½ discount for sentence recommendations that
are higher than they should have been, but more responsiveness to recommendations that are too low.
  We find no evidence that the guidelines themselves directly affect discretion for property offenders,
perhaps because judges generally have substantial experience with property cases and therefore do
not rely on the errant information. Sentences are more sensitive to both accurate and inaccurate recommendations
for crimes that occur less frequently and have more complicated sentencing. This suggests that when
the court has more experience, the recommendations have less influence. More tentative findings
suggest that, further down the decision chain, parole boards counteract the remaining influence of
the guidelines.

Shawn D. Bushway
School of Criminal Justice
University at Albany, SUNY
135 Western Avenue
Albany, NY 12222 USA
SBushway@uamail.albany.edu

Emily G. Owens
Cornell University
Department of Policy Analysis and Management
137 MVR Hall
Ithaca, NY 14853
ego5@cornell.edu

Anne Morrison Piehl
Department of Economics
Rutgers, The State University of New Jersey
New Jersey Hall
75 Hamilton Street
New Brunswick, NJ 08901-1248
and NBER
apiehl@economics.rutgers.edu

I. Introduction

Forty years ago in the United States, almost all state criminal justice systems used indeterminate sentencing. Judges assigned sentences with large ranges and allowed parole boards to determine the actual sentence. This structure was largely created to facilitate rehabilitation and to ensure that each punishment was proportional to the severity of each particular crime. Starting in the 1970's, however, scholars and practitioners raised questions about the wide disparity in sentences such a system created (Frankel 1973, von Hirsch 1976). At about the same time, others were questioning the validity of the rehabilitation model (Martinson 1974) and calling for more definitive sentences for the purposes of deterrence (Wilson 1975). These critiques led to a wave of sentencing reforms, including the imposition of sentencing guidelines, determinate sentences, and mandatory sentences, through which legislatures attempted to exert control over punishment decision by imposing structure on judicial discretion. As of 2008, twenty-four states use some type of judicial guidelines in the criminal sentencing process.[1]

Throughout this period of reform, researchers have questioned the ability of legislatures to fundamentally affect sentencing outcomes. With few exceptions, sentencing reforms have focused on limiting judicial discretion while ignoring decisions made by other actors in the justice system, including prosecutors, defense attorneys, and parole boards (Tonry 1996).

There are competing theoretical models of the extent to which legislative actions are implemented or undone by judges, attorneys, and parole boards. Reitz (1998) formally presents a model in which – across the institutional forms of structured discretion existing in the U.S. – these actors can largely counteract legislative mandates regarding sentencing. Huber and Gordon (2007) and Jacobi and Tiller (2007) build models in which lower court judges strategically respond to restrictions placed upon them by the legislature. In contrast to these theories, many legal models are predicated on greater compliance with rules from higher (earlier) levels of government, either from a positive or a normative perspective (Schauer 1991, Sullivan 1992). Such models of legal directives predict compliance except, perhaps, when the directives take on extreme values. Several types of social norms would also predict compliance. If the sentencing recommendations produced by the legislature embody widely held social norms regarding punishment, then later actors may well comply with these even when they differ from the actor's own preferences.

The extent to which rules set by the legislature bind or influence decisions regarding sentence length is central to institutional design and to determining the practical impact of any proposed reform regarding criminal punishment. For example, it has been suggested that prosecutors gain more power relative to judges under sentencing guidelines and that this leads to sentences that are overly punitive (Alschuler 1978, Coffee and Tonry 1983, Piehl and Bushway 2007). This concern will be valid only if sentencing guidelines cause judges and other actors to deviate from their preferred level of punishment. It is true that both legislative recommendations and actual sentences have increased over time. What is unclear is whether this is due to sentencing recommendations affecting the behavior of judges or correlated changes in institutional preferences over time.

---

[1] United States Sentencing Commission 2008 Annual Report: http://www.ussc.gov/ANNRPT/2008/chap1_08.pdf. For more on variation in sentencing structures, see Pfaff (2006).

An extensive literature evaluates the impact of the adoption of structured sentencing.[2] But there is mixed evidence on the operation of discretion *within* structured sentencing regimes. Consistent with the theoretical predictions of strategic response, research on mandatory minimums has shown that they are enforced selectively, if at all (Farrell 2003, Ulmer et al. 2007). Schanzenbach and Tiller (2008) present evidence that judges manipulate charges if the cost of deviating from the federal guidelines is relatively high. On the other hand, Kessler and Piehl (1998) found that policy changes in the early 1980s which enhanced recommended punishments did raise sentence lengths for affected crimes, suggesting some limits to available judicial discretion.

The impact of guidelines on subsequent criminal justice decisions is of particular interest after *Kimbrough v. United States* and *United States v. Booker*. While sentencing guidelines no longer limit the decisions judges can make, even recommended sentences provide information to judges about what a "typical" punishment might look like. Experimental evidence on the role of information in legal decision making shows that the context can affect outcomes. Rachlinski and Jourden (2003) find strong evidence that hypothetical jurors are sensitive to "contrast effects." When subjects are asked to decide on both severe and minor criminal cases at the same time, as opposed to one case at a time, the result is lower punishments for the worse offender and higher punishments for the less serious offender. Judges are not immune to the influence of irrelevant information. In a sample of active judges, Wistrich et al. (2004) demonstrated that exposure to legally inadmissible information about either the defendant or the victim had a statistically and substantively meaningful effect on judicial decision making.

Estimating the impact of legislative action on sentencing decisions in real data is complicated by the (potentially) simultaneous determination of preferences and legislation.[3] Measures of judicial and legislative tastes for punishment are rarely comparable (Bailey and Chang 2001), and are unlikely to be constant over time (Martin and Quinn 2007). However, to the extent that legislators and judges are exposed to the same stimuli, legislative changes in recommended punishments may be positively correlated with judicial taste (Schoenfeld 2010). For example, consider the following series of events. Increased public concern about violence associated with crack cocaine leads legislatures to increase recommended sentences for crack possession. At the same time, judges observe the increased number of defendants facing crack-related charges and decide that harsher punishments are in order. Parole boards follow suit, and extend periods of incarceration yet further. In this case, simply comparing changes in sentences and time served with sentence recommendations will reveal a positive correlation, but it would not have been induced by legislative action.

The ideal test of the influence of guidelines themselves would be to hold constant judicial preferences over punishment, and observe whether sentences change when legislatively determined recommendations are random. In lieu of randomized sentence recommendations, we identify instances in the Maryland circuit court in which the case facts are not consistent with the

---

[2] For a literature review and a reporting of apparent reductions in disparity after the adoption of both voluntary and mandatory guidelines, see Pfaff (2006).

[3] Comparing sentences before and after the implementation of a guidelines system is also problematic, as even the type of sentencing data collected fundamentally changes with the reform (Mustard 2001).

final sentence recommendation. As documented later, these inconsistencies appear to be the result of human error, exogenous to the preferences of downstream actors. Inaccurate guideline calculations result in downstream actors receiving sentencing recommendations that were arbitrarily longer or shorter than the typical (appropriate) recommendation for the conviction in question. Although we never observe the preferences of any of the actors in the system, conditional on observed facts of the case, these errors will be uncorrelated with preferences. This source of variation, then, introduces random variation in the recommendations affecting sentencing outcomes, allowing us to disentangle the direct influence of the guidelines themselves from shared preferences toward criminal punishment.[4]

We find that, on average, preferences for punishment are positively correlated across the stages of the criminal justice system. On average, each month recommended by the state guidelines is associated with a 12-20 day sentence and approximately 5 days of prison time served, depending on the type of offense. The message here is mixed. Sentencing guidelines are strongly and positively correlated with the final outcomes, but the downstream actors clearly have a lot of discretion to deviate from the recommendations. It is at least possible that the guidelines themselves are irrelevant to the final outcomes in the system.

We attempt to shed light on this question by examining the effect of random variation in sentencing recommendations. In drug cases, the direct impact of guidelines is roughly ½ the size of the overall correlation between sentencing recommendations and outcomes: each month recommended by the guidelines is associated with a 12 day sentence, but each erroneously added (or subtracted) month increases (or reduces) sentences by 6 days. Arbitrarily longer recommendations are also discounted by ½ for violent offenses, but sentences are more responsive to recommendations that shorten sentences arbitrarily. For property offenses, we do not find evidence that the guidelines themselves directly influence judicial decision making. We conclude that the recommendations have limited impact when the court has sufficient experience with the type of case.

Maryland's guidelines are voluntary and the recommended sentencing ranges are wide. In this institutional setting characterized by substantial discretion of downstream actors, some may be surprised that these voluntary guidelines alter sentencing outcomes at all. For those whose prior is that public servants comply with directives, the surprise will be the extent of noncompliance. Either way, the impacts are substantial. A typical error that is "too low" has a recommendation that is off by 18 months, which implies an average sentence reduction of nearly 8 months. Positive errors tend to be larger, but the impact on outcomes is smaller. A typical error of 27 months extra in the recommendation translates into a 4-month longer imposed sentence. In more tentative results, we find that parole boards operate in the opposite direction, reducing the impact of the errant recommendations on eventual time served in prison.

Recent U.S. Supreme Court decisions have restored substantial discretion to judges in many jurisdictions, including the federal system, which were previously operating under deterministic guidelines. These results predict that a move from mandatory to voluntary guidelines will mean

---

[4] Note that we are examining the impact of the guidelines on judges and parole boards. It is outside of the scope of this paper to measure the impact of the sentencing decision on the parole board's final decision.

that sentencing outcomes increasingly, but not absolutely, reflect the preferences of downstream actors.

II. Criminal Sentencing in Maryland

The Maryland sentencing guidelines have always been advisory. That is, judges are not required to assign a sentence within the upper and lower bounds and parole boards are not compelled to ensure that the time served falls within the guidelines. In fact, there are few direct incentives for judges to respond to changing legislative preferences that are reflected in changing recommended sentences. Consistent with this, as well as with the predictions of Reitz (1998), Huber and Gordon (2007) and Jacobi and Tiller (2007), sentences fall within the guideline range less than half the time. Some states require judges to file statements justifying their decision to sentence outside the recommended range, and in some jurisdictions these statements receive court scrutiny. Under Maryland law, neither of these checks occur.

The Maryland Sentencing Guidelines consist of three matrices, one for drug offenses, one for violent offenses, and one for offenses against property. The rows of these matrices correspond to an "Offense Score" (reflecting attributes of the instant offense) and the columns to an "Offender Score" (reflecting criminal history and other offender attributes). These matrices specify the recommended upper and lower bounds for the sentence.[5] These guidelines are implemented by (roughly) four sets of actors: the Maryland State Commission on Criminal Sentencing Policy, practicing attorneys, judges, and parole boards.

The matrices are produced by the Maryland State Commission on Criminal Sentencing Policy (MSCCSP), a sentencing commission permanently established by the legislature in 1999 following several years in an advisory capacity. Sixteen of the 19 voting members are appointed to four year terms by either the governor (nine members), the chief judge of the court of appeals (three members), the President of the Senate (two members) or the Speaker of the House (two members). The Attorney General, the State Public Defender, and the Secretary of the Department of Public Safety and Correctional Services are also voting members.[6] The MSCCSP also provides worksheets to assist in the calculation of these sentence recommendations. In Figure 1, we present the 2001 version of the Maryland Sentencing Guidelines Worksheet, which was the relevant version during the time period of our analysis.

Typically, guidelines worksheets are prepared by both the state's attorney and the defense attorney, and it is commonly understood that the worksheets form the basis of the pre-trial bargaining process between the two attorneys. Completing a worksheet involves a number of steps. First, the offense in question must be located in the Maryland criminal statute. The statute lists the offense "seriousness category," which ranges from I (the most serious) to VII (the least serious). For drug or property offenses, the seriousness category is equivalent to the offense score. For violent offenses, the seriousness category is converted into "seriousness points." As evident in figure 1, additional seriousness points may be added based on the nature of the crime. To further complicate the procedure, note that the point values are highest for the lowest

---

[5] Copies of the sentencing guidelines are available at http://www.msccsp.org/guidelines/matrices.html.

[6] We think of this collective body as representing the legislature and for brevity sometimes refer to it as such.

seriousness category. For violent offenses only, the sum of all of these points is the relevant offense score. The next step in completing the worksheet is to calculate an offender score, which is the sum of four elements reflecting the juvenile and adult criminal history of the defendant.

Once an offense and offender score have been calculated, the individual completing the worksheet must locate the range of recommended sentences which corresponds to those scores. In the case of violent offenses, there are 120 cells from which to choose. The property crime matrix has 48 different cells, and the drug matrix has 56. After locating the correct cell, the upper and lower bounds of the range of recommended sentences are entered on the worksheet. After conviction, one final worksheet is submitted to the sentencing judge. Regardless of who completes the final form, both the defense attorney and the prosecutor are expected to agree to its terms.

Maryland circuit court judges, who are initially appointed by the governor then subject to non partisan elections every 15 years, are expected to observe the range of recommended sentences and issue a decision, which they record on the worksheet. The judge then signs the now complete worksheet, copies of which are then distributed to various parties. One copy is retained by the sentencing judge, one copy becomes part of the offender's criminal file which is available to the Maryland Parole Commission, one copy is sent to the MSCCSP, and copies are given to both attorneys.[7]

III. Identifying and Characterizing Inaccuracies

Through a research agreement, we obtained an extract from the MSCCSP database on 17,052 sentences of males between the ages of 18 and 35 who were sentenced between July 2001 and 2004 for a crime with a single criminal count.[8] The database contains all of the information entered on the guidelines worksheet (Figure 1), save the name of the sentencing judge. We calculated the upper and lower bounds of the recommended sentences for each single count conviction based on a) the entered offender and offense scores, and b) the crime and criminal history elements underlying the offender and offense scores. These computer-generated recommendations were then compared to what was entered on the worksheet. If either of the calculated upper and lower bounds did not match the upper and lower bounds entered on the worksheet, this sentence was flagged as a potential error.

After we identified the set of cases with potential errors, we employed two interns at the MSCCSP to locate original copies of every identified worksheet to double check the information. Some of the identified discrepancies were found to be the result of data entry errors by the interns at MSCCSP rather than mistakes in the completion of the worksheets. These cases were

---

[7] These copies are produced via contact paper; each worksheet consists of six pieces of paper, and the defense attorneys' copy is the last one. In personal communication, one public defender noted that it is often difficult for him to read his copy of the final worksheet, limiting his ability to always catch and correct errors. (email communication with Louis Brendan Curran, Esq., Baltimore City – Felony Trial Division, office of the Public Defender, 10/26/2009)

[8] The calculation of sentence recommendations for multiple charge convictions depends on whether the different offenses are considered part of the same criminal act, which is not consistently recorded in the data. We exclude these observations from our analysis.

not considered errors for the purpose of our analysis, as the inaccuracy was introduced after the sentence was determined.[9] Using this universe of verified errors, we then provided a randomly generated list of 100 of these cases to the Executive Director of the MSCCSP, Dr. David Soulé, who performed his own review to determine whether these cases were truly errors. He agreed with our classification in all but 11 cases.[10] We used his feedback to further refine our criteria for identification of errors.[11]

A Baltimore public defender that we spoke to recalled personally noticing "about one [incorrectly completed worksheet] a year."[12] We found an overall inaccuracy rate between 2001 and 2004 of 10%. Errors occur across the range of the guidelines matrices. Variation in the recommended sentence on the worksheet is substantial in both accurate and inaccurate cases, as noted in the large standard deviations. There are more high errors than low errors for drug offenses; the opposite is true for property and violent offenses.

Relative to empirical evidence on learning and accuracy rates in both the criminal justice and private spheres, we believe an accuracy rate of 90% reflects a relatively successful production process. For example, Contreras, Kim, and Tristao (2007) find that 8.2% of LASIK patients require a second surgery, and Liebman et al. (2000) point out that 41% of capital convictions between 1973 and 1975 were thrown out due to "serious error" in the original conviction. Cole, Cornyn and Gunther (1995) found that the ratings system developed by the Federal Reserve miscategorized 17% of failing banks as satisfactory, and 7% of satisfactory banks as failing. This was considered an improvement over the previous rating system.

In the set of verified mistakes, 45% of the errors were what we call *low* errors—where the recommendation from the worksheet (the midpoint of the range) was lower than it should have been if the worksheet had been filled out accurately. The remaining 55% of the errors were *high* errors, where the recommendation from the worksheet was higher than it should have been.

There were three primary sources of mistakes on the worksheet. Mistakes were most likely to occur when the person completing the worksheet had to reference other documents. In 1,588 instances, the upper and lower bound of the recommended sentence entered on the worksheet were incorrectly copied from the sentencing guidelines matrix. An additional 100 errors where made due to incorrectly copying the wrong seriousness score from the Maryland criminal statutes. Finally, on 26 occasions, the elements of the offender and offense score were summed incorrectly.

---

[9] This article only attempts to deal with inconsistencies within the sentencing worksheet itself, taking as given any possible inaccuracies underlying these data.

[10] In addition, there were 20 cases for which it was impossible to ascertain whether or not the worksheet was correctly completed, as the cases involved subsequent offenses and the correct allocation of case factors was unclear.

[11] We are confident that this process has identified cases of error according to the rules promulgated by the MSCCSP. Independent of our project, the MSCCSP has been working to implement an automated worksheet. This project has as one of its explicit goals the elimination of data entry and worksheet compilation errors (Maryland State Commission on Criminal Sentencing Policy 2005).

[12] Email communication with Louis Brendan Curran, Esq., Baltimore City – Felony Trial Division, office of the Public Defender, 10/26/2009.

Using the mean of the upper and lower guidelines recommendations as a summary measure of the sentence recommendation, the errors we identified are reasonably large in magnitude. The average correct recommendation was 45 months, downward errors were an average of 17.6 months (sd= 22.2) too low and upward errors were 27.4 months (sd=35.2) too high. The distribution of errors is predictably skewed; the median low error is 10.7 months, and the median high error is 12 months (the median correct recommendation is 24 months).

Did these human errors affect judicial sentences? We begin with graphical evidence. Figure 2 displays the distributions of the difference between the midpoint of the recommendation entered on the worksheet and the sentence imposed. When recommendations are accurate, the distribution is centered just below zero, and is bimodal. When the recommendations contain errors, the distributions of sentences are single peaked but slightly shifted either to the left or right depending on the direction of the mistake. Judges do appear to respond to the guidelines differently when the worksheet is not completed correctly. Relative to the worksheet recommendation, sentences are higher when recommendations are too low and lower when recommendations are too high. This is consistent with judges, or others, noticing the mistake, or generally ignoring the guidelines worksheet.

The distribution of judicial deviation from the worksheet recommendation suggests that, to some extent, judges counteract mistakes. However, it is not obvious that judges fully account for human error. Figure 3 displays the distribution of deviations from the midpoint of the range of recommended sentences if the worksheets had been calculated correctly -- the "accurate" midpoint. If judges identified the inaccuracies or ignored the worksheets as a general rule, then the relationship between the sentence and the accurate recommended sentence should be unrelated to whether or not the worksheet was correctly calculated.

As in Figure 2, there is a large mass of sentences between 0 and 20 months lower than the accurate midpoint. If the recommended sentences are too high, the distribution is shifted to the right, meaning that sentences tend to be higher relative to the accurate recommendation. For worksheet recommendations that are lower than they should be, there are relatively more sentences just to the left of the mass point than to the right, indicating a downward shift in the distribution of sentences. While deviations from inaccurate recommendations tend to compensate for the direction of the error, consistent with the experimental results of Rachlinski and Jourden (2003), they do not move enough to undo the error completely.

It is possible that the sentences we observe, as well as the errors themselves, are due to some case-specific factor, like the race of the defendant, or whether or not the charge was pled out. In order to evaluate this, we first look for any substantive differences in observed characteristics that are recorded on the guidelines worksheet before the judge issues a sentence. These "upstream" characteristics are listed and defined in Table 1, and include demographics (age and race), whether the case was resolved by plea bargain and the type of legal representation, the job title of the person who completed the worksheet, and the sentence recommendation itself. Table 2 provides descriptive statistics separately for the accurate worksheets, those with high errors, and those with low errors.

Table 2 also reports the value of the guidelines recommendation that corresponds with the case characteristics, the accurate midpoint. This can be thought of as a summary measure for the severity of the crime, using the severity weights of the sentencing commission. Property offenses have shorter recommended sentences than the other crime types. Among violent and property offenders, those cases with high errors have shorter accurate recommended sentences than consistent worksheets, which are in turn shorter than cases with low errors. Violent offenses have the most complicated calculations, and this result suggests that the most complicated calculations (with the longest sentences) sometimes miss some of the enhancements that are relevant for this offense type. Within drug offenses, cases with errors have shorter accurate recommended sentences than those with consistent worksheets. Because of the cross offense differences, we estimate separate models below.

Cases with errors were more likely to be resolved with a plea bargain than cases without errors. This difference in the rate of plea bargaining is statistically significant for all offense types when the recommended sentence was too high – almost 80% of these cases are resolved by plea bargain, compared with 70% of all consistent cases. One possible interpretation is that defendants were pressured to plea by the threat of the high sentence recommendation. At the same time, defendants who received a recommendation that was too low were also more likely to plea (71.7% vs. 70.5%). Both of these observations are consistent with the belief that worksheets of cases settled by plea receiving less scrutiny than others. We discuss this and other interpretations of this finding below. The nature of the representation does not vary with the accuracy of the worksheet. The same is true for the age of the offender. It appears that unconditionally, black offenders are more likely to receive high errors in violent or property cases.

In order to formally determine if there is some systematic pattern in the occurrence of errors, we estimated a logit model in which the left hand side variable takes on the value of one when the guidelines worksheet contains an error, either too high or too low.[13] We include on the right hand side dummy variables indicating whether or not a plea bargain was reached, the jurisdiction of sentence, six month fixed effects, the job title of the individual completing the worksheet, whether or not the defendant had private representation, and the age and race of the offender. We also control for the previous number of "similar" convictions occurring in that jurisdiction in that year, where "similar" is defined as a violation of the same criminal statute (e.g., "second degree assault").[14] We interpret this as a proxy for learning or familiarity – the more common the offense is, the higher the probability that the individual completing the worksheet has calculated the correct recommendation before.

Marginal effects from logit models are presented in Table 3, by offense type. We find little relationship between the seriousness of the offense, captured in the accurate midpoint, and the probability of an error. The higher error rates for plea bargains seen in Table 2 do not carry

---

[13] The use of logits in a model with so many dummy variables means that there is a substantial number of observations dropped due to perfect collinearity. This can be seen by comparing the error rate in the logit estimation sample to the error rate in the full sample. In earlier versions of the paper we estimated the same specifications using a linear probability models, which allowed us to include extensive controls without inducing sample selection. The estimates are generally robust to alternative modeling specifications (OLS, probit or logit models).

[14] Convictions in which the guidelines are incorrectly calculated are excluded in the construction of this measure.

through here; controlling for other observed features of the case, there is no statistically significant effect on the probability of an error from having a case settled by plea. There is no statistically significant relationship between type of representation and the error probability.

We find strong evidence that individuals convicted of relatively common offenses are substantially less likely to have an error in their sentence recommendations, supporting our learning hypothesis. This is particularly true for drug offenses, where a10% increase in the number of times a similar case has been resolved in that jurisdiction in that year decreases the overall probability of an error by 2.5%. There is an average of 297 similar cases for drug offenses, compared to 13 for violent cases and 10 for property cases. The hypothesis that the individual charged with completing the final worksheet learns how to complete it correctly with practice is an intuitive one.

Among drug offenses, we find corroborating evidence for this learning hypothesis from examining the job title of the person who completed the final worksheet. Parole and probation agents, the omitted group, and states' attorneys together complete over 90% of all worksheets, and these two groups are equally likely to commit an error- notice the large standard error on the state's attorney dummy in column 1. Public defenders, who complete only 21% of worksheets, and private attorneys, who complete only 2.3%, are significantly more likely to commit a mistake. Because we are controlling for the representation of the offender, this effect is not picking up differences in the quality of legal counsel, which itself may be correlated with the final outcome. Rather, we believe that this effect reflects familiarity with the guidelines and the worksheet.[15]

Inaccurate worksheet calculations that increase a recommended sentence may differ in important ways from inaccurate calculations that reduce the recommendation, as the incentives for uncovering the error differ depending upon its direction. Andreoni (1991)'s model of the judicial system predicts such asymmetry. In this model, which he calls the "reasonable doubt test," the asymmetry arises because juror or judicial demands on the evidence increase as penalties increase.

To allow for asymmetry we separately examine high and low errors in columns (4) through (9) of Table 3. Across offense types, more serious offenses (those with longer accurate sentence recommendations) are associated with fewer high errors and more low errors. Consistent with Andreoni (1991), the relationship between the number of drug cases seen in the jurisdiction and the error rate is negative and statistically significant for errors in both directions, but the magnitude is smaller for the low errors. We observe an asymmetric relationship between plea bargains and errors for drug offenses – plea bargains are strongly associated with worksheet recommendations that are "too high." We come back to the interpretation of this correlation below.

---

[15] There is no discretion intended with regard to the worksheet itself. All case facts are expected to be reported. Evidence that the facts are recorded on the worksheet comes from Farrell (2003), who studied the extent to which those who used a firearm in the crime, as represented on the sentencing worksheet, received mandatory sentence enhancements. She found that 70% of cases with firearms as a "case fact" did not receive a sentence enhancement.

Relative to parole and probation agents, states' attorneys appear more likely to calculate a recommended sentence that is too low, only for drug offenses. While there is not much precision in the estimates for state's attorney, the signs of the coefficients are inconsistent with a prosecutor trying to increase the amount of punishment an offender receives. Public defenders and private attorneys are more likely to commit both high and low errors, which is again consistent with our hypothesis that these are true errors that are less likely to occur as individuals learn how to use the guidelines worksheets. People convicted of violent offenses are about one percentage point (roughly 26%) less likely to have a worksheet recommendation that is lower than it should be.

Based on this analysis, we conclude the observed errors are not correlated with another factor that may affect plea negotiations or a judge's or parole board's decisions.[16] We do find that the more frequently a particular offense was encountered, the less likely there was to be an error. Therefore, we conclude that inconsistent guidelines recommendations were the result of simple human error. This type of human error should be uncorrelated with preferences for punishment. In the remainder of the paper, we exploit the difference between the observed recommendation and what it should have been to determine the independent effect of guidelines on sentencing and prison release decisions, disentangled from any correlated changes in preferences across institutions.

IV. Analytic Framework

To investigate how guidelines recommendations impact the length of criminal sentences imposed, we decompose the recommended sentence into three parts: the accurate recommendation (that should have been recommended based on the case facts and the guidelines), the months erroneously recommended in addition to the accurate recommendation, and the months erroneously excluded. For any observation, one or both of the latter components will be zero.

We model the impact of the recommended sentence on the outcome as:

$$(1) \quad \text{Sentence}_i = \eta^S + AM_i \theta_A^S + (WM_i - AM_i) D_H \theta_H^S + (WM_i - AM_i) D_L \theta_L^S + X_i \pi^S + v_i^S$$

where Sentence_i is the sentence (in days) passed down after conviction i, AM_i is the accurate recommended sentence (in months), modeled as the midpoint of the lower and upper ends of the recommended range, and WM_i is the recommended sentence as entered on the guidelines worksheet, again represented by the range's midpoint (in months). The dummy variable D_H takes on the value of one if the quantity $(WM_i - AM_i)$ is larger than zero, indicating that the recommended sentence given to the judge was higher than it should have been based on the information provided on the worksheet. Similarly, D_L equals one if the recommended sentence is too low, meaning that $(WM_i - AM_i)$ is negative. The matrix X_i controls for observed

---

[16] Another way to test the independence of the errors is to calculate an omnibus F-test of the differences between those with errors in their recommendations and those without. We implemented this by running a seemingly unrelated regression of a system of sentences recommendations (standardized by age and accurate midpoint) stacked with race, representation, plea, and offense type and run against the error dummy along with the number of similar offenses, jurisdiction and time dummies. The p-value on the test of significance of the error dummy is 0.22, consistent with conditional independence of the errors.

characteristics that might also affect the sentencing decision but are not contained in AM_i, including the age of the offender at sentencing, whether or not the offender had private representation, whether or not the a plea bargain was reached, the jurisdiction and date of the conviction (in six month intervals), whether or not the offender is black, the job title of the person who completed the guidelines worksheet, and our measure of offense familiarity (the number of previous convictions occurring within a jurisdiction in that calendar year). We estimate θ_A^S, θ_H^S and θ_L^S separately for each offense type.[17]

Assuming there is a general consensus between the judge and legislature on proportionality, we expect that $\theta_A^S > 0$. If the final sentence is independent of the guidelines recommendation, either because the recommendation is not relied upon during sentencing or because the inaccurate calculations are corrected during the sentencing process, we should not be able to reject the null hypothesis that $\theta_H^S = \theta_L^S = 0$; the additional or missing months in the recommendations would simply add noise to the equation.

Such a finding would indicate that the guidelines themselves do not affect the sentences imposed, and any observed correlation between changes in the guidelines and sentence length is due to simultaneously determined preferences of the judge and legislature. Alternatively, if sentencing guidelines have restricted judicial discretion, we would expect to see $\theta_H^S = \theta_L^S = \theta_A^S \neq 0$, implying that upstream actors can successfully manipulate the decisions of those downstream.[18] Estimating θ_H^S separately from θ_L^S allows a test of whether the effect of misinformation is symmetric by testing the hypothesis that $\theta_H^S = \theta_L^S$.

If observed "errors" are actually the result of strategic bargaining, we would overstate the influence of the inaccurate portion of the recommended sentence- leading us to fail the reject the hypothesis that $\theta_H^S = \theta_L^S = \theta_A^S$. For example, consider the possibility that the prosecuting and defense attorneys agree to allow the offender to be officially convicted of a lesser charge in exchange for a slightly longer sentence. The alternate situation is also possible -- in which the offender pleads guilty to a particular charge in exchange for a shorter sentence -- although we have no reason to suppose that this occurs. Because entering a plea is a predictor of having a recommended sentence that is too high (for drug offenses), but not too low, if there is such a bias we expect it is more likely to exist in our estimate of θ_H^S than our estimate of θ_L^S.

V. Results

Our estimates of θ_A^S, θ_H^S and θ_L^S are presented in the first three rows of Table 4. The first three columns report the main results by offense type; the last three restrict the sample to cases settled by plea. In addition to controls for plea bargain and whether the defendant had private

---

[17] Note that because we measure the dependant variable in days and the independent variables in months, see θ_H^S, θ_L^S, and θ_A^S, the coefficients are measured in "days per month."

[18] Note that we do not expect θ_A^S to be equal to one, since judges are observed to deviate from the recommended sentences with some frequency. The question of interest is whether or not judges respond in the same way to the component of the recommendation that is uncorrelated with their preferences.

representation, all specifications include as correlates the jurisdiction, time, and demographic controls from the previous analyses.[19]

For each additional month of incarceration recommended by the sentencing guidelines in drug case, the average sentence length increases by 11.8 days. An incorrectly long or short sentence recommendation, however, appears to be discounted by roughly 50% relative to correct recommendations. This suggests that a substantial amount of discretion remains under voluntary guidelines; judicial decisions are more highly correlated with the accurate part of the guidelines recommendation than the arbitrary part. Roughly 60% of all convictions in Maryland are for drug offenses, and recall that the drug guidelines rely on only seriousness of the offense and length of criminal history to determine recommended sentencing ranges. As a result, it is reasonable to assume that judges and attorneys are more familiar with the "typical" punishment in a drug case. P-values for F-tests (reported at the bottom of the table) show that the data reject the null hypothesis that accurate and inaccurate sentence recommendations in either direction have the same impact on sentence length, but we fail to reject the null hypothesis that judges respond in a symmetric way to errors.

In cases of violence, sentences are much closer to the recommended midpoint -- each additional recommended month increases sentences by 21.5 days. Incorrectly long recommendations add to the sentence imposed, but at 12.6 days per month. Judges appear to discount these inaccuracies by about 50%. Missing months in the recommendation are not discounted relative to the accurate midpoint – they decrease sentences by 21.7 days. Not only are violent offenses half as prevalent as drug offenses, but the guidelines calculations are much more involved. There is an additional step required to include a series of case facts in the offense score. There are also roughly twice as many total cells in the sentencing grid for violent offenses compared with drug offenses. This may make it less likely that errors are "caught."

Unlike in drug cases, we can reject the null hypothesis that the impact of errors is symmetric. The probability that $\theta_H^S = \theta_L^S$ is less than one half of one percent. One interpretation of this finding is that downstream actors are reluctant to honor recommended sentences that are "too long" and show more deference to guidelines when the recommended sentence is too low.

Turning to property offenses, we find that sentences are similarly responsive to accurate recommendations as with drug offences (13 days for each month recommended), but that inaccuracies don't have any appreciable impact. We cannot reject the null hypothesis that discretion is unaffected by sentencing recommendations - neither θ_H^S or θ_L^S is estimated to be statistically different from zero. Although there are fewer single count property offenses than drug or violent offenses, this lack of statistical significance is unlikely to be the result of type-II error as there are still a large number of cases with recommendation errors.

The estimated values of θ_H^S and θ_L^S for property offenses are also an order of magnitude smaller than θ_A^S, and smaller than the corresponding estimates among drug and violent cases.

---

[19] Note that our estimates of θ_H^S and θ_L^S are statistically robust to the inclusion of these covariates, supporting our claim that the guidelines inaccuracies approximate random human error.

Technically, we cannot reject the null hypothesis that judges respond to accurate recommendations in the same way as arbitrarily low recommendations (the probability $\theta_A^S = \theta_L^S$ is 12%), but this is due to the large standard errors on θ_L^S. The lack of responsiveness to mistakes in the guidelines is consistent with our earlier learning hypothesis. Circuit court judges in Maryland frequently have experience in lower (district) courts before their appointments to the circuit court. In the lower courts, judges will have handled numerous property crime cases, which are a substantial fraction of the docket.[20]

Across specifications, we have found that plea bargains have shorter sentence lengths, on average. This finding is consistent with defendants entering a guilty plea in exchange for a less costly punishment (Reinganum 1988), or the general idea of a trial penalty (Grossman and Katz 1983). It is also plausible that pled cases receive less attention from judges, making it plausible that they respond to sentencing recommendations in a fundamentally different way.

In specifications reported in columns (4)-(6), we restrict the sample to those cases resolved by plea bargains. We find the drug sentences to be slightly less responsive to all of the recommendations – accurate and inaccurate – in pled cases. Among the pled violent cases, we now observe symmetry in reaction to the inaccurate parts of the sentence recommendations- judges give slightly more deference to increases in recommendations, but less deference to lower recommendations. Further, note that in pled cases, we cannot reject the null hypothesis that judges respond to accurate and inaccurate increases in recommendations equally, but can reject the null hypothesis that judges completely compensate for mistakenly low sentences with 95% certainty. Conclusions for pled property offenses are the same as for the full sample, although, again, there is a weaker correlation between the accurate component of the recommendation and the judge's decision.

So far, the results indicate that guidelines do not influence sentence outcomes for property offenses. But for the other offense types, inaccuracies have an impact on sentence outcomes, by less than a literal reading would suggest, more for violent offenses than for drug convictions, and among violent offenses, more for low errors than for high.

VI. Falsification Test

The above analysis relies on the assumption that the "facts of the case" (the offender and offense data) are accurate and that any deviation of the sentence recommendation from that indicated by the facts of the case is caused by human error. This assumption would be false in a scenario in which attorneys agree on a recommended sentence and then record "facts of the case" to comport with the agreement. Because we do not have independent information on the bargaining process between the two attorneys, we cannot directly test this assumption. However, we can create a situation in which we incorrectly classify convictions as containing errors and compare our results to estimates from the falsified data. If our falsified results are identical to our main result, this would cast doubt on our conclusion that internally inconsistent worksheets indicate simple human error.

---

[20] We thank Eve Brensike Primus of the University of Michigan Law School for suggesting this interpretation.

From the sample of 15,338 cases in which the facts of the case and sentence recommendation were consistent, we randomly selected 877 of those cases to be "too low" and 875 cases to be "too high" to match the distribution of errors in the full sample. We then divided the recommended sentences into multiple components, splitting the recommendation at an arbitrary break point for cases deemed "too high," and adding an arbitrary fraction of the recommendation in cases that were "too low." As a result, in each of these cases the accurate sentence recommendation should be the sum of the "inconsistent" and "consistent" parts of the case. We then re-estimated equation (1) with this sample. Our estimated results are presented in Table 5. As expected, our estimates of the θ_A^S are very close to the previous estimates (from the first three columns in Table 4).

Contrary to our primary results, in the falsification sample we cannot reject the null hypothesis that judges defer to punitive recommendations. For all three offense types, we cannot reject (at the 5% level of statistical significance) the null hypothesis that sentences are as responsive to extra recommended months as they are to accurate or missing months. Note also that there is no evidence of a strong inframarginality problem. If the additional months added onto a worksheet for this exercise are interpreted as a spline of the recommended sentence, judges appear to respond equally to the first and last month of a recommendation.

This falsification test promotes confidence in the results of the primary empirical approach. And it argues against the plausibility of the interpretation that plea bargains lead to the errors in the guidelines values rather than the guidelines influencing the sentence outcomes.

VII. Sentencing Guidelines and Time Served

Even after the sentence is assigned, other downstream actions can influence the time actually served to fulfill that term. Because the sentencing worksheet remains part of the individual's criminal record, it is plausible that parole boards are also influenced by its contents. In order to analyze the downstream behavior, we focus our analysis on the 8,184 sentences for which we have data on time served as well as sentence length.[21]

We now estimate a model of time served that is comparable to equation (1), with number of days incarcerated replacing sentence length as the dependant variable:

$$(2) \quad \text{Time}_i = \eta_2 + AM_i \theta^T + \left(WM_i - AM_i \right) D_H \theta_H^T + \left(WM_i - AM_i \right) D_L \theta_L^T + X_i \pi^T + v_i^T$$

1,260 of the 8,184 convictions are right censored because the individuals were still incarcerated at the time of data collection. Nonetheless, these incomplete spells still contain some information about the impact of worksheet inconsistencies on time served by placing a lower bound on the dependant variable. The estimation procedure is split into two further steps - the effect of the independent variables on time served and the effect of the independent variables on whether or not the incarceration spell is completed; we are essentially treating time served as a

---

[21] Time served information was collected from state prison system and jails in Baltimore City and the counties of Prince George, Anne Arundel, and Washington. All cases that were dropped were cases with jail sentences, meaning sentences of one year or less. Thus, the sample is composed disproportionately of more serious cases that resulted in prison sentences.

latent variable. Let $Time_i^c$ be the observed time served. The relationship between the actual length and measured length is then:

$$(3) \quad Time_i = \begin{cases} Time_i^* & \text{if} \quad Time_i^* < Time_i^c \\ Time_i^c & \text{if} \quad Time_i^C \leq Time_i^* \end{cases}$$

where $Time_i = Time_i^*$ if the offender is released. Let d_i equal one if the sentence is complete, and zero otherwise. The likelihood function for equation (2) is therefore:

$$(4) \quad L(\Lambda_i) = \sum \left[d_i \ln(h(Time_i, \Lambda_i)) + (1-d_i)\ln(1 - H(Time_i^c, \Lambda_i)) \right]$$

where Λ_i represents the values of the independent variables associated with period of incarceration i, $h(\bullet)$ is the probability density function evaluated at $Time_i$, and $H(\bullet)$ is the cumulative density function evaluated at $Time_i^c$.

Due to their position as downstream actors, parole boards have more information about behavior and program participation while incarcerated in addition to the sentence recommendation and the imposed sentence. There are two mechanisms through which the guidelines can impact parole boards: a direct effect of the sentence recommendation as well as an indirect effect through the sentence itself. In order to estimate the relative importance of these two effects, we estimate equation (2) both with and without controlling for sentence length.

Table 6 reports estimates of time served.[22] The odd numbered columns exclude sentence length on the right hand side, the even numbered columns include this variable. Column (1) shows that for drug offenses each additional month suggested by the sentencing guidelines is associated with 5 additional days (s.e.=0.19) served by convicted offenders – many fewer days than suggested by looking at the sentences imposed. Each additional month erroneously recommended increased time served by 3.09 days (s.e.=0.65). Individuals receiving arbitrarily shorter recommendations were released only 1.7 days (s.e.=1.1) earlier than similar offenders who received accurate recommendations. Further, once the impact of the errors on the judge's sentence is taken into account, column (2), there is no remaining independent effect of the errors on time served. Note, however, that each accurately recommended month does increase time served by one day (se=0.18), over and above its correlation with sentence length.

In contrast to the drug offenses, for violent (columns 3 and 4) and property (columns 5 and 6) convictions we cannot reject the null hypothesis that errors in the worksheet completion process do not impact the amount of time an offender is incarcerated. Conditional on sentence length, we find no additional correlation between what the legislature recommends and what parole

---

[22] Because the time served sample is a subsample of the analysis sample for Table 4, we examined the determinants of sentence length using the new subsample and found qualitatively identical results Specifically, in this sample of drug cases, each recommended month is associated with a sentence of 11 days (se=0.4), each arbitrarily added month worth 7 days (se=1), and each arbitrarily missing month reduces sentence by 5 days (se=2). For violent offenses, the corresponding relationships are 19 days (se=1), 17 days (se=5), and 20 days (se=6). Finally, for the sub-sample of property offenses, we observe that each recommended month is associated with a 10 day sentence (se=0.8), each arbitrarily added month worth 4 days (se=3) and each arbitrarily missing month increases sentences by an imprecise 2 days (se=7).

boards do. We conclude that parole boards rely heavily on information specific to the offender and offense in question for these types of offenses.

Are parole boards "undoing" errors made prior to the sentencing decision? Note that in violent offenses, the sign of our estimates of θ_H^S and θ_L^S flip when we include sentence length. While not statistically significant, the magnitudes are non-trivial (3 days for each month) and suggests that the direct impact of arbitrarily missing or added months on time served counteracts the indirect impact. While far from conclusive, this suggests that downstream actors may, in part, be responding to the decision of upstream actors in a strategic way.

VIII. Robustness

We further investigate the nature of the relationship between guidelines and sentencing by allowing the impact of the inaccurate component of the sentence recommendation to vary across different dimensions. If large discrepancies in recommendations are more likely to be identified as errors, we would expect to see a diminishing marginal impact of the inaccuracies as they increase in absolute size. We would also expect that sentences for common offenses should be less sensitive to inaccuracies. We also address concerns that worksheet inaccuracies are the result of unobserved racial bias in the sentencing process by allowing the impact of recommendations to vary by race. We then allow the impact of the error to vary with the frequency with which the court sees similar cases. Finally, we adapt our empirical strategy to take the non-normal distribution of sentence length into account, by modeling sentence length as a count variable with excess zeros.

Tables 7a – 8c report the results from these alternate specifications. The sign of the squared values of θ_H^S is the opposite of the sign of linear parameter for drug and property offenses, implying a diminishing marginal impact of arbitrarily added months (but the effect of this is small). On the other hand, note that the estimated impact of the squared value of θ_L^S is negative, and when the quadratic effect is included we are no longer able to precisely estimate the linear relationship between missing months and sentence length. This interesting result suggests that judges become more sensitive to lower recommendations as the number of "missing" months increases.

We do not find evidence that the impact of guidelines recommendations is heterogeneous with respect to race (column 2). As expected, the interaction between similar convictions and inaccurate recommendations that are too low is opposite in sign to the first order effect of the recommendation, although not precisely estimated. There appears to be no evidence of any "learning" when recommendations are too high – the coefficient is an order of magnitude smaller than the standard error. This provides mixed support for the learning interpretation.

Finally, our general conclusions are robust to our statistical modeling assumptions. Estimating the relationship between sentence length and recommendation using a zero-inflated negative binomial also finds positive and significant effects of additional months on the worksheet and negative and statistically significant effects of inaccuracy in the opposite direction, with responsiveness to legislative reductions being larger than responses to increases for violent offenses. Note that since these coefficients are roughly partial elasticities, the size of the

coefficients are not directly comparable- a 1% increase in accurately recommended sentence is much larger than a 1% increase in a mistake.

Table 8 reports a similar set of tests using time served as the dependant variable. The substantive results are quite consistent across these specifications. We find evidence of a diminishing marginal impact of arbitrarily added or missing recommended months, and no statistically significant difference between the treatment of black and non-black males. One way in which the results vary from the previous estimates is in regard to our proxy for judicial learning. Instead of having a limiting effect on the inconsistent recommendations we find amplifying effects. While puzzling, our proxy is better suited to the sentencing decision, as it is based on the number of similar convictions occurring in the county. The analogous proxy would be the number of similar cases seen by the given parole board in that year, and we do not have access to the necessary data to construct this measure.

As with sentence length, the qualitative findings are robust to alternate statistical models, including the linear probability model and the censored negative binomial. Restricting the sample to completed cases yields a generally consistent picture.[23]

IX. Conclusion

In this paper, we take advantage a new source of identification to study how government actors interact and make decisions in the criminal sentencing process. Our use of miscalculations in recommended sentences gives a window into the complex process of criminal sentencing that has implications for how we expect the imposition of government regulations to influence policy outcomes. Under the assumption that mistakes in the guidelines recommendation occur at random the resulting variation in recommended sentences is uncorrelated with a judge's preferences for punishment in a given case.

Because we observe both the accurate and inaccurate worksheet recommendation, we are able to explicitly quantify the extent to which downstream actors have preferences that differ from those of the upstream actors. The use of mistakes in guidelines calculations, rather than changes in guidelines due to policy changes, allows us to estimate the interaction among actors without requiring the existence of any new policy or explicit change in preferences.

We find that even an advisory guideline system like the one in Maryland has a direct impact on judicial decision making in drug and violent cases. In a state where almost half of the sentences are below the recommended minimum, judges still appear eager to go along with a lesser sentence in violent offenses. Relative to the underlying correlation between legislative guidelines and judicial decision making, judges appear to reduce their sentences one for one with arbitrarily reduced recommendations. If recommended sentences are too high, judges appear to discount the mistakes, but are still influenced by what are, by definition, recommendations that are orthogonal to the case at hand.

---

[23] We also estimate models (results available on request) in which the additional months are interacted with the job title of the individual completing the worksheet, which would allow us to learn if mistakes made by certain types of individuals are more likely to be detected. We find no compelling evidence that this is the case.

This asymmetry does not occur in the simpler, more frequent, and arguably lower social cost crime types of property and drug offenses. Rather, it is among the more complicated and less frequent offenses of personal violence that guidelines recommendations are more binding and the response to inaccuracy more asymmetric. More generally, we find that experience matters. Error rates are lower for more frequently occurring offense types and lower for those court professionals who complete more of the sentencing worksheets.

The net effect of sentencing guidelines on time served appears to be small because of conflicting preferences within the criminal justice system. Parole boards appear largely unconstrained by the guidelines. In addition, they appear to prefer longer sentences and therefore counteract the errors which result in leniency. It is also possible that parole boards have different, more proximal information that is systematically less optimistic than the information available to the upstream actors.

References

Alschuler, Albert. 1978. "Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for 'Fixed' and 'Presumptive' Sentencing," 126 *University of Pennsylvania Law Review* 550-577.

Andreoni, James. 1991 "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" 22 *RAND Journal of Economics* 385-395.

Bailey, Michael and Kelly H. Chang. 2001. "Comparing Presidents, Senators, and Justices: Interinstitutional Preference Estimation," 17 *Journal of Law, Economics, & Organization* 477-506.

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

Coffee, John and Michael Tonry. 1983. "Hard Choices: Critical Trade-offs in the Implementation of Sentencing Reform through Guidelines," in M. Tonry and F. Zimring (eds.) *Reform and Punishment: Essays in Criminal Sentencing*. University of Chicago Press.

Cole, R., B. Cornyn and J. Gunther. 1995. "FIMS: A New Monitoring System for Banking Institutions," *Federal Reserve Bulletin,* January.

Contreras, J. B. Kim and I. Tristao. 2007. "Does Experience Make Better Doctors? Evidence from LASIK" *CBO working paper.*

Farrell, Jill. 2003. "Mandatory Minimum Firearm Penalties: A Source of Sentencing Disparity?" 5 *Justice Research and Policy* 95-115.

Frankel, Marvin. 1973. *Criminal Sentences: Law without Order*. NY: Hill and Wang.

Grossman, Gene M. and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare" 73 *The American Economic Review* 749-757.

Huber, Gregory A. and Sanford C. Gordon. 2007. "Directing Retribution: On the Political Control of Lower Court Judges," 23 *Journal of Law, Economics, & Organization* 386-420.

Jacobi, Tonja and Emerson H. Tiller. 2007. "Legal Doctrine and Political Control," 23 *Journal of Law, Economics, & Organization* 326-345.

Kessler, Daniel P. and Anne Morrison Piehl. 1998. "The Role of Discretion in the Criminal Justice System," 14 *Journal of Law, Economics, & Organization* 256-276.

Liebman, J., J. Fagan, V. West and J. Lloyd. 2000. "Capital Attrition: Error Rates in Capital Cases, 1973-1995," 78 *Texas Law Review* 1839-1865.

Maryland State Commission on Criminal Sentencing Policy. 2005. Annual Report. www.msccsp.org/publications/ar2005.pdf

Martin, Andrew D. and Kevin M. Quinn. 2007 "Assessing Preference Change on the US Supreme Court," 23 *Journal of Law, Economics, & Organization* 365-385.

Martinson, Robert. 1974. "What Works? Questions and Answers about Prison Reform," 35 *The Public Interest* 22-54.

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

Pfaff, John. 2006. "The Continued Vitality of Structured Sentencing Following Blakely: The Effectiveness of Voluntary Guidelines," 54 *UCLA Law Review* 235-307.

Piehl, Anne Morrison and Shawn Bushway. 2007. "Measuring and Explaining Charge Bargaining," 23 *Journal of Quantitative Criminology* 105-125.

Rachlinski, Jeffrey J. and Forest Jourden. 2003. "The Cognitive Components of Punishment," 88 *Cornell Law Review* 457-485.

Reitz, Kevin. 1998. "Modeling Discretion in American Sentencing Systems," 20 *Law and Policy* 91-428.

Reinganum, Jennifer F. 1988. "Plea Bargaining and Prosecutorial Discretion," 78 *American Economic Review* 713-728

Schanzenbach, Max M., and Emerson H. Tiller. 2007. "Strategic Judging Under the United States Sentencing Guidelines: Positive Political Theory and Evidence," 23 *Journal of Law, Economics, & Organization* 24–56.

Schauer, Frederick. 1991. *Playing by the Rules: A Philosophical Examination of Rule-based Decision-Making in Law and in Life*. Oxford: Oxford University Press.

Schoenfeld, Heather. 2010. "Mass Incarceration and the Paradox of Prison Conditions Litigation," 44 *Law and Society Review* 731-768.

Sullivan, Kathleen M. 1992. "The Justice of Rules and Standards," 106 *Harvard Law Review* 22–123.

Tonry, Michael. 1996. *Sentencing Matters*. New York: Oxford University Press.

Ulmer, J., M. Kurlychek, J. Kramer. 2007. "Prosecutorial Discretion and the Imposition of Mandatory Minimum Sentences," 44 *Journal of Research in Crime and Delinquency* 427-458.

von Hirsch, Andrew. 1976. *Doing Justice: The Choice of Punishments*. NY*:* Hill and Wang.

Wilson, James Q. 1975. *Thinking about Crime*. NY: Basic Books, Inc.

Wistrich, Andrew J., Chris Guthriet, and Jeffrey J. Rachlinski. 2004. "Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding," 153 *University of Pennsylvania Law Review* 1251-1346.

Figure 1: The Maryland Sentencing Guidelines Worksheet

MARYLAND SENTENCING GUIDELINES WORKSHEET	OFFENDER NAME - Last, First, Middle		SID#	SEX __ M__ F	BIRTHDATE		JURISDICTION

PSI __ Yes __ No	DATE OF OFFENSE	DATE OF SENTENCING	DISPOSITION TYPE		REPRESENTATION	ETHNICITY	RACE __ Unidentifiable

PSI __ Yes __ No

AT THIS SENTENCING, NUMBER OF: CONVICTED OFFENSES [ ] CRIMINAL EVENTS [ ] WORKSHEET #_____ OF CRIMINAL EVENT #_____.

DISPOSITION TYPE
__ ABA plea agreement   __ Jury trial
__ Non-ABA plea agreement   __ Reconsideration
__ Plea, no agreement   __ Review
__ Court trial   __ Probation Revocation

REPRESENTATION
__ Private
__ Public Defender
__ Court Appointed
__ Self

ETHNICITY
Hispanic/Latino Origin __ Yes __ No
INDIGENCE ESTABLISHED Yes ___ No ___

RACE
__ Black   __ Asian
__ White   __ Other
__ Native Hawaiian/Pacific Islander
__ American Indian/Alaskan Native

CONVICTED OFFENSE TITLE	I-VII	CJIS CODE	MD CODE, ART, & SECTION	STAT. MAX	MAND. MIN	CASE #/DOCKET #
1st Convicted Offense		—				
2nd Convicted Offense		—				
3rd Convicted Offense		—				

OFFENSE SCORE(S) – Offense Against a Person Only

1st Off	2nd Off	3rd Off	A. Seriousness Category
1	1	1	= V – VII
3	3	3	= IV
5	5	5	= III
8	8	8	= II
10	10	10	= I

1st Off	2nd Off	3rd Off	B. Victim Injury
0	0	0	= No Injury
1	1	1	= Injury, Non-Permanent
2	2	2	= Permanent Injury or Death

1st Off	2nd Off	3rd Off	C. Weapon Usage
0	0	0	= No Weapon
1	1	1	= Weapon Other Than Firearm
2	2	2	= Firearm or Explosive

1st Off	2nd Off	3rd Off	D. Special Victim Vulnerability
0	0	0	= No
1	1	1	= Yes

_____ _____ _____   OFFENSE SCORE(S)

OFFENDER SCORE

A. Relationship to CJS When Instant Offense Occurred
0 = None or Pending Cases
1 = Court or Other Criminal Justice Supervision

B. Juvenile Delinquency
0 = 23 years or older or crime-free for 5 years or no more than 1 finding of a delinquent act
1 = Under 23 years old and: 2 or more findings of a delinquent act or 1 commitment
2 = Under 23 years and committed 2 or more times

C. Prior Adult Criminal Record
0 = None   3 = Moderate
1 = Minor   5 = Major

D. Prior Adult Parole/Prob Violation
0 = No   1 = Yes

_____   OFFENDER SCORE

GUIDELINES RANGE

1st Con. Off.
_____
TO

2nd Con. Off.
_____
TO

3rd Con. Off.
_____
TO

Overall Guidelines Range
Multiple Counts Only
_____
TO

50% of Sentence Announced
__ Yes __ No

ACTUAL SENTENCE - Imposed, Suspended, Time Served, Probation, Restitution, Fine, Corrections Options Programs (Drug Treatment Court, Break the Cycle, Etc.)

1st Convicted Offense
For Theft, Fraud, and Related Crimes, please indicate: ☐ Economic loss $_____ ; ☐ Unknown Amount
Subsequent Offender Filed __ Yes __ No   Restitution Requested __ Yes __ No
Subsequent Offender Proven __ Yes __ No   Restitution Proven __ Yes __ No

2nd Convicted Offense
For Theft, Fraud, and Related Crimes, please indicate: ☐ Economic loss $_____ ; ☐ Unknown Amount
Subsequent Offender Filed __ Yes __ No   Restitution Requested __ Yes __ No
Subsequent Offender Proven __ Yes __ No   Restitution Proven __ Yes __ No

3rd Convicted Offense
For Theft, Fraud, and Related Crimes, please indicate: ☐ Economic loss $_____ ; ☐ Unknown Amount
Subsequent Offender Filed __ Yes __ No   Restitution Requested __ Yes __ No
Subsequent Offender Proven __ Yes __ No   Restitution Proven __ Yes __ No

Was the offender sentenced to a Corrections Options program under Commission criteria?
Drug Court __ Yes __ No Other __ Yes __ No

Institutional/Parole Recommendation or Additional Information

Parole Notification __ Yes __ No

Worksheet Completed By

Title

Sentencing Judge (Please Print)

Sentencing Judge's Signature

VICTIM INFORMATION

Victim	__ Yes	__ No
Victim Unavailable	__ Yes	__ No
NRF	__ Yes	__ No
Victim Notified Plea	__ Yes	__ No
Victim Notified Date	__ Yes	__ No
Victim Present	__ Yes	__ No
Written VIS	__ Yes	__ No
Oral VIS	__ Yes	__ No
No Contact with Victim	__ Yes	__ No
CICB Cost Imposed	__ Yes	__ No

SENTENCE DEPARTURE INFORMATION

If the actual sentence departs from the guidelines range, please indicate the Court's reason(s) using the numerical code(s) on the list of common departure factors in the manual, if applicable.

[ ] [ ] [ ]

Departure Code 9 or 18 (Please Explain):

COPIES: White – Judge; Blue – Sentencing Commission; Green – Attach to Commitment or Probation Order; Yellow – File; Pink – Prosecution; Gold – Defense

9/2007 (1.4)

Figure 2: Distribution of Sentences, 2001-2004
Relative to Worksheet Recommendation
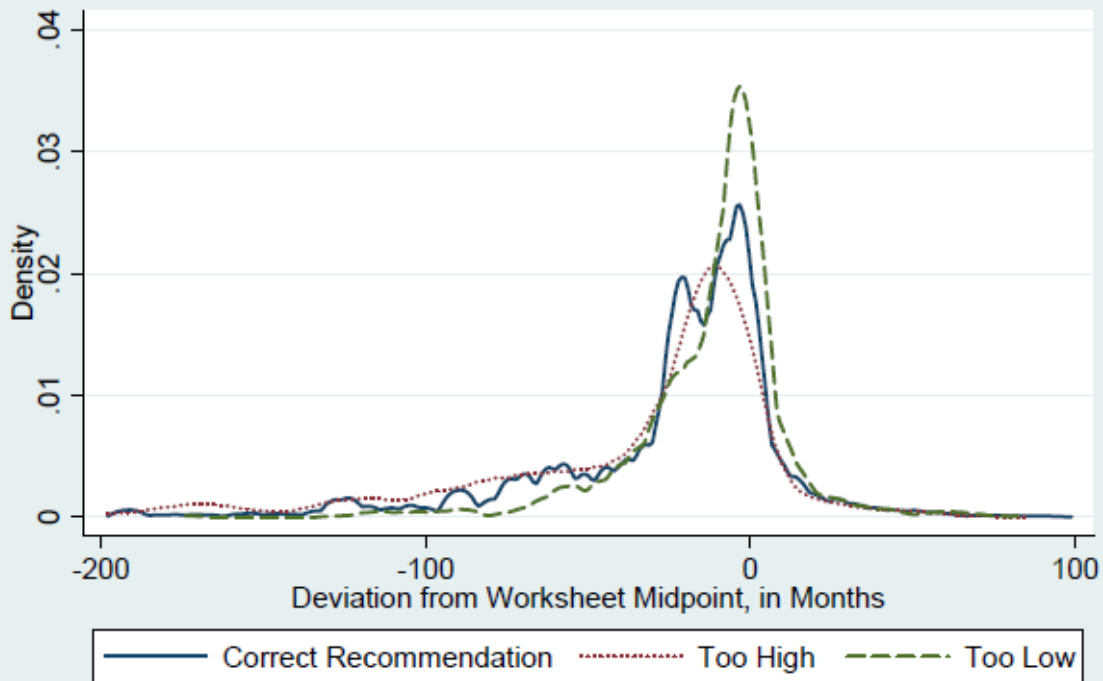
Figure 3: Distribution of Sentences, 2001-2004
Relative to Correct Recommendation
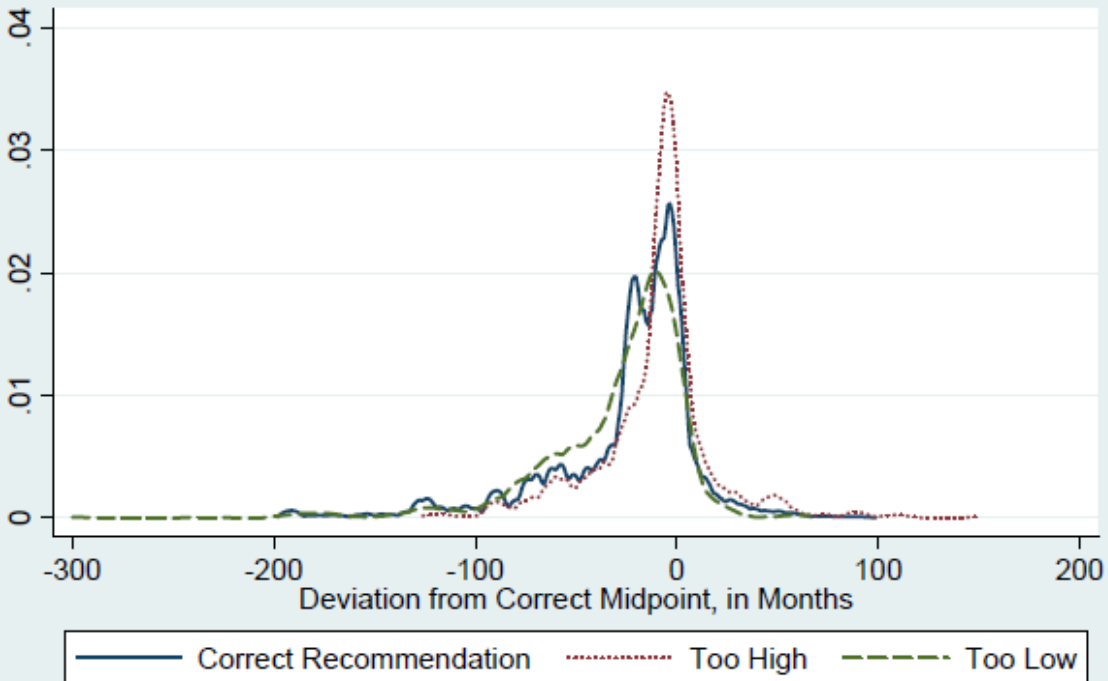
Table 1. Variable Definitions

Variable	Definition	Temporal Location
Accurate Midpoint	Mean of upper and lower range of sentencing guidelines as calculated by algorithm. (see fig. 1 "CJIS Code", "Offender Score" and "Offense Score"	"Upstream" *prior to judge and parole board decision*
Worksheet Midpoint	Mean of upper and lower range of sentencing guidelines as entered on worksheet. (see fig. 1 "Guidelines Range")	"Upstream" *prior to judge and parole board decision*
Nature of Representation	Legal representation of offender (see fig. 1 "Representation")	"Upstream" *prior to judge and parole board decision*
Plea Bargain	Outcome determined prior to trial (see fig. 1 "Disposition Type")	"Upstream" *prior to judge and parole board decision*
Race	Race of Offender (see fig. 1 "Race"	"Upstream" *prior to judge and parole board decision*
Age	Age of offender at sentencing (see fig. 1 "Birth Date", "Date of Sentencing")	"Upstream" *prior to judge and parole board decision*
Worksheet Completer	Individual responsible for filling out worksheet given to judge (see fig. 1 "Title")	"Upstream" *prior to judge and parole board decision*
Sentence	Sentence determined by Judge (see fig. 1 "Actual Sentence")	"Downstream" *prior to parole board decision*
Time Served	Number of days (if observed to be incarcerated)	"Downstream"

Table 2: "Upstream" Characteristics of Accurate and Inaccurate Worksheets

		Offense Type			
		All	Drug	Violent	Property
Sample Size					
	Accurate	15,338	8,747	4,404	2,187
	High	936	757	94	85
	Low	778	479	132	167
Worksheet Midpoint, in months					
	Accurate	45.4 (56.6)	47.2 (48.0)	50.5 (52.3)	28.2 (40.3)
	High	63.9*** (76.6)	69.0 (79.4)	52.3 (70.2)	31.6 (41.6)
	Low	30.0*** (40.4)	26.8*** (29.7)	44.2 (60.3)	27.7 (45.2)
Accurate Midpoint, in months					
	Accurate	45.4 (56.6)	47.2 (48.0)	50.5 (74.9)	28.2 (40.3)
	High	36.5** (47.9)	39.3*** (49.1)	33.7** (51.0)	14.6*** (20.0)
	Low	47.7 (53.7)	44.1 (43.9)	66.6** (73.8)	43.0*** (57.4)
% Private Representation					
	Accurate	33.6 {0.38}	32.8 {0.50}	36.8 {0.73}	30.5 {0.98}
	High	30.9* {1.51}	31.2 {1.68}	33.0 {4.88}	25.9 {4.78}
	Low	34.1 {1.70}	36.7* {2.20}	38.6 {4.25}	22.7** {3.25}
% Plea Bargain					
	Accurate	70.5 {0.37}	73.7 {0.47}	65.8 {0.72}	66.7 {1.01}
	High	79.6*** {1.32}	79.9*** {1.46}	77.6** {4.32}	78.8** {4.46}
	Low	71.7* {1.62}	71.2 {2.07}	73.5* {3.86}	71.8 {3.49}
% Black					
	Accurate	69.7 {0.37}	82.0 {0.41}	60.5 {0.74}	39.2 {1.04}
	High	77.9*** {1.36}	81.9 {1.40}	71.3** {4.69}	49.4* {5.46}
	Low	71.1 {1.63}	81.4 {1.78}	68.2* {4.07}	43.7 {3.85}
Age					
	Accurate	24.0 (4.11)	24.1 (4.06)	24.0 (4.16)	23.9 (4.22)
	High	24.4 (4.05)	24.5 (3.97)	24.1 (4.22)	24.0 (4.56)
	Low	24.3 (4.15)	24.2 (4.14)	24.0 (4.04)	24.7** (4.23)

Statistically different from consistent worksheets at * 10%, ** 5%, *** 1% significance level
Standard deviation in parentheses, standard errors in braces.

Table 3: Predicting Worksheet Errors. Marginal Effects from Logit Estimation

	DV: Worksheet is Inaccurate			DV: Worksheet is Too High			DV: Worksheet is Too Low		
	Drug	Violent	Property	Drug	Violent	Property	Drug	Violent	Property
Accurate Midpoint (divided by 1000)	-0.015 [0.076]	-0.015 [0.054]	0.294 [0.207]	-0.045 [0.078]	-0.205* [0.081]	-1.216*** [0.336]	0.052 [0.120]	0.052 [0.033]	0.548*** [0.144]
Similar Cases	-0.312*** [0.014]	-0.318 [0.671]	-2.053* [0.849]	-0.318*** [0.005]	0.173 [0.365]	-0.947 [0.854]	-0.059*** [0.008]	-0.570 [0.491]	-1.55* [0.655]
Plea Bargain	0.015 [0.011]	0.010 [0.012]	0.022 [0.015]	0.024*** [0.007]	0.002 [0.008]	0.011 [0.017]	-0.006 [0.007]	0.008 [0.011]	0.019 [0.010]
Black	-0.007 [0.006]	0.009 [0.007]	-0.020 [0.020]	-0.016* [0.007]	0.003 [0.006]	-0.014* [0.006]	0.006 [0.006]	0.008 [0.007]	-0.001 [0.015]
Private Representation	0.011 [0.008]	0.002 [0.008]	-0.020 [0.020]	0.007 [0.005]	-0.012* [0.006]	-0.012 [0.022]	0.006 [0.006]	0.009 [0.007]	-0.014 [0.020]
Worksheet Completed by State's Attorney	0.029 [0.033]	0.013 [0.026]	0.025 [0.057]	0.007 [0.024]	0.011 [0.017]	0.012 [0.022]	0.030 [0.016]	0.007 [0.022]	0.020 [0.052]
Worksheet Completed by Public Defender	0.059* [0.024]	0.014 [0.022]	0.046 [0.040]	0.037* [0.018]	0.007 [0.011]	0.018 [0.016]	0.035* [0.013]	0.007 [0.018]	0.040 [0.036]
Worksheet Completed by Private Attorney	0.100* [0.040]	0.054 [0.061]	0.055 [0.055]	0.057 [0.035]	0.056 [0.036]	0.162*** [0.038]	0.065** [0.025]	-0.003 [0.040]	-0.014 [0.034]
Pseudo R^2	0.06	0.07	0.09	0.15	0.11	0.16	0.05	0.06	0.08
N	9,979	3,272	1,842	9,427	2,486	1,314	9,153	3,080	1,712
Error Rate in Logit Sample	12.4%	6.9%	13.6%	8.03%	3.78%	6.47%	5.23%	4.28%	9.70%
Error Rate in Full Sample	12.4%	4.88%	10.3%	7.58%	2.09%	3.74%	4.80%	2.91%	7.09%

Notes: Age at sentencing, jurisdictions, six month fixed effects controls included. Standard errors in brackets allow for arbitrary correlation within jurisdiction. 56% of worksheets were completed by a State's attorney, 3.2% by a public defender, and 2.6% by a private attorney.
* = $p < 5\%$, ** = $p < 1\%$, *** = $p < 0.1\%$

Table 4: OLS Estimates, Dependant Variable: Sentence Length (in days)

	Offense Type			Pleas Only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Drug	Violent	Property	Drug	Violent	Property
Accurate Midpoint	11.8***	21.5***	13.0***	11.1***	17.7***	11.7***
(θ_A^S)	[2.46]	[1.55]	[0.91]	[2.06]	[1.80]	[1.46]
Additional months on	6.06***	12.6***	0.46	4.32***	14.2***	-0.6
Worksheet (θ_H^S)	[0.96]	[2.46]	[1.29]	[1.02]	[2.48]	[2.11]
Missing months on Worksheet	-6.71*	-21.7***	-2.1	-6.52	-13.9***	-4.13
($-\theta_L^S$)	[3.14]	[1.21]	[6.29]	[3.22]	[1.76]	[6.51]
Plea Bargain	-42.6	-190**	-63.8	--	--	--
	[26.3]	[59.2]	[38.5]			
Private Representation	-7.28	-110***	-38	-28.8	-100.4**	-62.6
	[14.2]	[23.7]	[24.8]	[24.3]	[26.7]	[42.6]
R^2	0.41	0.63	0.45	0.4	0.53	0.41
N	9983	4630	2439	7396	3070	1647
Mean Sentence Length	528.8	958.8	440.1	488.6	816.6	395.8
P-value of F-stat						
P($\theta_A^S = \theta_H^S$)	0.0038	0.02	0.0000	0.0002	0.41	0.0001
P($\theta_A^S = \theta_L^S$)	0.0022	0.90	0.12	0.0067	0.042	0.34
P($\theta_H^S = \theta_L^S$)	0.80	0.0037	0.80	0.39	0.95	0.49

Notes: Jurisdiction, six month dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of the person completing the worksheet. Standard errors in brackets allow for arbitrary correlation within jurisdiction.

* = $p < 5\%$, ** = $p < 1\%$, *** = $p < 0.1\%$

Table 5: Falsification Test: Dependent Variable: Sentence Length (in Days)

	(1) Drug	(2) Violent	(3) Property
Accurate Midpoint (θ_A^S)	11.7*** [2.51]	21.4*** [1.49]	12.7*** [1.02]
Additional months on Worksheet (θ_H^S)	13.9*** [1.37]	25.0** [8.65]	29.9 [25.9]
Missing months on Worksheet (-θ_H^S)	-8.34* [3.66]	-14.2** [5.04]	-24.2** [7.98]
Plea Bargain	-22.8 [19.7]	-205.6** [62.8]	-62.6 [41.5]
Private Representation	-4.25 [17.5]	-126.4*** [30.4]	-47.9 [28.6]
R^2	0.41	0.62	0.43
N	8747	4404	2187
P-value of F-stat			
P($\theta_A^S = \theta_H^S$)	0.23	0.66	0.52
P($\theta_A^S = \theta_L^S$)	0.14	0.17	0.14
P($\theta_H^S = \theta_H^S$)	0.078	0.30	0.84

Notes: Jurisdiction, six month, plea bargain, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Column 1 includes offense type dummy variables. Standard errors in brackets allow for arbitrary correlation within jurisdiction.

* = p < 5%, ** = p <1%, *** = p <0.1%

Table 6: MLE Estimates, Dependent Variable: Time Served (in days)

	Drug		Violent		Property	
	(1)	(2)	(3)	(4)	(5)	(6)
Accurate Midpoint (θ_A^S)	4.78*** (0.19)	1.11*** (0.18)	5.32*** (1.22)	0.89 (0.47)	5.05*** (0.55)	0.9 (0.55)
Additional months on Worksheet (θ_H^S)	3.09*** (0.65)	0.97 (0.50)	0.13 (1.82)	-3.39 (3.12)	1.29 (1.86)	-0.21 (1.17)
Missing months on Worksheet ($-\theta_L^S$)	-1.71 (1.09)	-0.11 (0.89)	-1.31 (2.69)	3.03 (2.43)	-3.52 (2.87)	-3.49 (3.06)
Plea Bargain	-32.8* (15.7)	-19.8 (12.4)	-52.9 (37.1)	32.5 (28.4)	-30.8 (36.7)	-6.85 (31.8)
Sentence Length		10.3*** (0.51)		10.2*** (0.46)		12.4*** (1.19)
R^2	0.38	0.58	0.40	0.41	0.52	0.63
N	5230	5230	2134	2134	820	820
Mean of DV	297.0	297.0	545.5	545.5	354.9	354.9
P-value of F-stat						
P($\theta_A^S = \theta_H^S$)	0.014	0.80	0.017	0.18	0.053	0.38
P($\theta_A^S = \theta_L^S$)	0.0046	0.26	0.14	0.11	0.58	0.36
P($\theta_H^S = \theta_L^S$)	0.28	0.41	0.72	0.93	0.52	0.26

Notes: Jurisdiction, six month and private representation dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of the person completing the worksheet. Robust standard errors in parentheses.

* = p < 5%, ** = p <1%, *** = p <0.1%

Table 7a: Alternate Specifications of Sentence Length (in days), Drug Offenses Only

	(1) Quadratic Effects	(2) Racial Effects	(3) Learning	(4) Any Sentence (Logit ME)	(5) Neg. Bin
Accurate Midpoint (θ_A^S)	11.9*** [2.48]	11.8*** [2.47]	11.8*** [2.45]	0.004*** [0.0005]	0.021*** [0.002]
Additional months on Worksheet (θ_H^S)	12.0* [4.54]	8.73* [3.95]	6.29*** [1.34]	0.002*** [0.00007]	0.0048* [0.0019]
(Additional months on Worksheet)^2	-0.057 [0.036]				
(Additional months on Worksheet) x Black		-2.94 [3.53]			
(Additional months on Worksheet) x Similar Cases			-1.47 [2.90]		
Missing months on Worksheet (-θ_L^S)	-4.21 [3.11]	-5.82 [7.08]	-6.36 [4.05]	-0.0001 [0.0006]	-0.0075*** [0.0022]
(Missing months on Worksheet)^2	-0.033 [0.021]				
(Missing months on Worksheet) x Black		-1.03 [5.78]			
(Missing months on Worksheet) x Similar Cases			-1.11 [2.89]		
R^2	0.41	0.41	0.41	0.19	0.40
N	9983	9983	9983	9977	5542
Mean of D.V.	528.7	528.7	528.7	0.55	951.8

Notes: Jurisdiction, six month, plea bargain, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Standard errors in brackets allow for arbitrary correlation within jurisdiction.

$* = p < 5\%, ** = p < 1\%, *** = p < 0.1\%$

Table 7b: Alternate Specifications of Sentence Length (in days), Violent Offenses Only

	(1) Quadratic Effects	(2) Racial Effects	(3) Learning	(4) Any Sentence (Logit ME)	(5) Neg. Bin
Accurate Midpoint (θ_A^S)	21.5*** [1.56]	21.5*** [1.55]	21.5*** [1.55]	0.0047*** [0.0007]	0.010*** [0.001]
Additional months on Worksheet (θ_H^S)	7.94 [7.40]	15.2 [13.7]	14.4*** [2.64]	0.004*** [0.001]	0.0015 [0.002]
(Additional months on Worksheet)^2	0.039 [0.050]				
(Additional months on Worksheet) x Black		-2.79 [15.0]			
(Additional months on Worksheet) x Similar Cases			-303.1 [193.9]		
Missing months on Worksheet ($-\theta_L^S$)	-17.5*** [3.16]	-30.2*** [3.81]	-17.8*** [2.37]	-0.004*** [0.0004]	-0.008*** [0.0013]
(Missing months on Worksheet)^2	-0.032 [0.018]				
(Missing months on Worksheet) x Black		9.40* [3.46]			
(Missing months on Worksheet) x Similar Cases			-301.6* [119.7]		
R^2	0.63	0.63	0.63	0.21	0.60
N	4630	4630	4630	4595	3217
Mean of D.V.	958.8	958.8	958.8	0.692	1,382

Notes: Jurisdiction, six month, plea bargain, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Standard errors in brackets allow for arbitrary correlation within jurisdiction.

* $= p < 5\%$, ** $= p < 1\%$, *** $= p < 0.1\%$

Table 7c: Alternate Specifications of Sentence Length (in days), Property Offenses Only

	(1) Quadratic Effects	(2) Racial Effects	(3) Learning	(4) Any Sentence (Logit ME)	(5) Neg. Bin
Accurate Midpoint (θ_A^S)	13.0*** [0.90]	13.0*** [0.91]	12.9*** [0.86]	0.006*** [0.0008]	0.015*** [0.001]
Additional months on Worksheet (θ_H^S)	1.96 [4.20]	0.53 [1.29]	0.44 [3.83]	0.001 [0.002]	0.0012 [0.002]
(Additional months on Worksheet)^2	-0.008 [0.016]				
(Additional months on Worksheet) x Black		-0.56 [2.26]			
(Additional months on Worksheet) x Similar Cases			0.6 [255.3]		
Missing months on Worksheet ($-\theta_L^S$)	-1.76 [8.29]	-1.07 [8.39]	-7.76 [4.82]	-0.002 [0.003]	-0.01* [0.0048]
(Missing months on Worksheet)^2	-0.0065 [0.26]				
(Missing months on Worksheet) x Black		-3.43 [9.13]			
(Missing months on Worksheet) x Similar Cases			548.4 [419.9]		
R^2	0.45	0.45	0.45	0.18	0.43
N	2439	2439	2439	2435	1507
Mean of D.V.	440.1	440.1	440.1	0.612	712.9

Notes: Jurisdiction, six month, plea bargain, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Standard errors in brackets allow for arbitrary correlation within jurisdiction.

* = $p < 5\%$, ** = $p < 1\%$, *** = $p < 0.1\%$

Table 8a: Alternate Specifications of Time Served (in days), Drug Offenses Only

	Quadratic Effects	Racial Effects	Learning	Any Incarceration	Censored Neg. Bin.	Completed Spells only
Accurate Midpoint (θ_A^T)	4.79*** (0.19)	4.79*** (0.19)	4.79*** (0.19)	0.005*** (0.0002)	0.0047*** (0.00050)	2.67*** (0.13)
Additional months on Worksheet (θ_H^T)	5.60*** (1.17)	5.47* (2.28)	2.63*** (0.74)	0.003*** (0.0002)	0.00095*** (0.00023)	1.65** (0.51)
(Additional months on Worksheet)^2	-0.023* (0.0091)					
(Additional months on Worksheet) x Black		-2.58 (2.38)				
(Additional months on Worksheet) x Similar Cases			3.22 (1.99)			
Missing months on Worksheet (-θ_L^T)	-1.22 (2.09)	4.3 (5.55)	-0.97 (1.51)	-0.0003 (0.001)	-0.001 (0.0011)	-0.11 (0.82)
(Missing months on Worksheet)^2	-0.006 (0.021)					
(Missing months on Worksheet) x Black		-6.42 (5.66)				
(Missing months on Worksheet) x Similar Cases			-1.94 (2.34)			
R^2	0.38	0.38	0.38	0.21	0.18	0.35
N	5230	5230	5230	4741	2931	4644
Mean of D.V.	297.0	297.0	297.0	0.515	530.0	190.1

Notes: Jurisdiction, six month, plea bargain, offense type, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Robust standard errors in parentheses.

* = $p < 5\%$, ** = $p < 1\%$, *** = $p < 0.1\%$

Table 8b: Alternate Specifications of Time Served (in days), Violent Offenses Only

	Quadratic Effects	Racial Effects	Learning	Any Incarceration	Censored Neg. Bin.	Completed Spells only
Accurate Midpoint (θ_A^T)	5.32*** (1.22)	5.32*** (1.22)	5.32*** (1.22)	0.004*** (0.0003)	0.0041*** (0.00053)	2.03*** (0.30)
Additional months on Worksheet (θ_H^T)	0.44 (3.90)	6.21 (9.01)	-0.77 (1.82)	0.004 (0.002)	-0.0022 (0.0019)	1.72 (1.87)
(Additional months on Worksheet)^2	-0.0034 (0.026)					
(Additional months on Worksheet) x Black		-6.58 (9.16)				
(Additional months on Worksheet) x Similar Cases			129.3 (196.7)			
Missing months on Worksheet (-θ_L^T)	-2.77 (5.20)	-3.56 (5.22)	1.32 (3.78)	-0.002 (0.001)	0.0075 (0.0062)	-1.9 (1.60)
(Missing months on Worksheet)^2	0.017 (0.042)					
(Missing months on Worksheet) x Black		2.5 (5.89)				
(Missing months on Worksheet) x Similar Cases			-343.2* (171.7)			
R^2	0.40	0.40	0.40	0.22	0.11	0.41
N	2134	2134	2134	1574	1624	1577
Mean of D.V.	545.5	545.5	545.5	0.676	716.8	305.2

Notes: Jurisdiction, six month, plea bargain, offense type, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Robust standard errors in parentheses.
* = $p < 5\%$, ** = $p < 1\%$, *** = $p < 0.1\%$

Table 8c: Alternate Specifications of Time Served (in days), Property Offenses Only

	Quadratic Effects	Racial Effects	Learning	Any Incarceration	Censored Neg. Bin.	Completed Spells only
Accurate Midpoint (θ_A^T)	5.04*** (0.55)	4.99*** (0.55)	5.05*** (0.54)	0.005 (0.002)	0.015*** (0.0015)	2.84*** (0.43)
Additional months on Worksheet (θ_H^T)	4.26 (5.76)	1.51 (2.17)	-2.49 (3.26)	0.003 (0.004)	0.0091 (0.0095)	1.29 (1.68)
(Additional months on Worksheet)^2	-0.046 (0.075)					
(Additional months on Worksheet) x Black		-0.098 (3.76)				
(Additional months on Worksheet) x Similar Cases			395.4 (238.0)			
Missing months on Worksheet (-θ_L^T)	-10.3 (6.41)	-5.59* (2.76)	-1.57 (3.40)	-0.001 (0.003)	-0.028*** (0.0058)	-2.34 (1.89)
(Missing months on Worksheet)^2	0.14 (0.12)					
(Missing months on Worksheet) x Black		14.2 (8.53)				
(Missing months on Worksheet) x Similar Cases			-163.7 (117.8)			
R^2	0.52	0.52	0.52	0.26	0.10	0.56
N	820	820	820	568	582	703
Mean of D.V.	354.9	354.9	354.9	0.581	500.1	226.3

Jurisdiction, six month, plea bargain, offense type, private representation and dummy variables included, as well as age at sentence, number of similar cases, race of offender, and title of individual completing the worksheet. Robust standard errors in parentheses.

* = p < 5%, ** = p <1%, *** = p <0.1%