

Winter 1991

Federal Criminal Sentencing: Some Measurement Issues With Application to Pre-Guideline Sentencing Disparity

William Rhodes

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

William Rhodes, Federal Criminal Sentencing: Some Measurement Issues With Application to Pre-Guideline Sentencing Disparity, 81 J. Crim. L. & Criminology 1002 (1990-1991)

This Criminology is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in Journal of Criminal Law and Criminology by an authorized editor of Northwestern University School of Law Scholarly Commons.

CRIMINOLOGY

FEDERAL CRIMINAL SENTENCING: SOME MEASUREMENT ISSUES WITH APPLICATION TO PRE-GUIDELINE SENTENCING DISPARITY*

WILLIAM RHODES**

ABSTRACT

Regression analysis is often used to study sentencing decisions. This paper presents a statistical model—a two-limit tobit model with heteroscedastic error terms—that is an improvement over techniques often used to analyze sentencing decisions. This model is applied to the sentences imposed on federal offenders convicted of bank robbery and the distribution of cocaine. Results are used to motivate a discussion of sentence disparity.

I. INTRODUCTION

The empirical analysis of sentencing practices is a frequent subject for social scientists. This predilection is understandable. Sentencing occupies a prominent role in the criminal process; furthermore, many, if not most, of the factors that affect sentencing decisions are measurable, allowing social scientists to test hypotheses that in other contexts are difficult to operationalize.

For example, analyses of sentencing patterns have been used

* Part of this work was performed while the author was a staff member of the United States Sentencing Commission. The author would like to thank Michael Block, John Lott, Candace McCoy, Luke Froeb, Cindy Alexander, Bruce Kobayashi, Scott Lyden, Liz Phillips and anonymous reviewers for their helpful comments. The views expressed in this paper are the author's and do not necessarily reflect those of the United States Sentencing Commission or Abt Associates.

** William Rhodes is Senior Scientist at Abt Associates. Dr. Rhodes' research interests include the mathematical modeling of criminal behavior and the statistical analysis of criminal justice activity. His current research focuses on drug abuse, drug law enforcement, AIDS and criminal justice modelling.

both to construct sentencing guidelines¹ and to evaluate them once they have been implemented.² Analyses of sentencing patterns have also been used to evaluate social interventions such as career criminal programs, bans on plea bargaining, and mandatory minimums for gun-law violations.³ Finally, they have been used to test hypotheses about society and law, including the behavior of complex organizations, the presence of racial and sexual discrimination, and the functionalist/conflict theories of law.⁴ Similarly, analyses of sentencing practices have been incorporated into studies of plea bargaining.⁵

This paper focuses on a related topic, sentence disparity. Sentence disparity arises when similar offenders who have committed the same crime receive different sentences, and when dissimilar offenders who have committed different crimes receive the same sentence—absent compelling justification for differences in the first instance and similarity in the second. Sentence disparity has received extended attention from researchers and critics of the judicial process.⁶

In section two, we introduce the statistical model used in this study to analyze sentence disparity. In sections three and four, we use this model to estimate the average sentences imposed on bank robbers and high-level dealers of cocaine conditional on the offense characteristics (*e.g.*, the dollar amount stolen and the amount of drugs involved), the offender's background (including his criminal record), and especially whether the offender pled guilty or was convicted after a trial. In section five, we go beyond this expectation to

¹ W. RICH, P. SUTTON, T. CLEAR & M. SAKS, *SENTENCING GUIDELINES: THEIR OPERATION AND IMPACT ON THE COURTS* (1981) [hereinafter W. RICH]; UNITED STATES SENTENCING COMMISSION, *SUPPLEMENTAL REPORT ON THE INITIAL SENTENCING GUIDELINES* (1987) [hereinafter SENTENCING COMMISSION, SUPPLEMENTAL REPORT].

² M. TONRY, *SENTENCING REFORM IMPACTS* (NIJ Issues and Practices 1987) [hereinafter M. TONRY].

³ See A. BLUMSTEIN, J. COHEN, S. MARTIN & M. TONRY, *RESEARCH ON SENTENCING: THE SEARCH FOR REFORM* (1983) [hereinafter A. BLUMSTEIN]; M. TONRY, *supra* note 2.

⁴ See A. BLUMSTEIN, *supra* note 3.

⁵ W. McDONALD, *PLEA BARGAINING: CRITICAL ISSUES AND COMMON PRACTICES* (NIJ Research Report 1985); W. RHODES, *PLEA BARGAINING: WHO GAINS? WHO LOSES* (Promise Res. Pub. No. 14, 1979); A. ROSETT & D. CRESSEY, *JUSTICE BY CONSENT: PLEA BARGAINS IN THE AMERICAN COURTHOUSE* (1976); Landes, *An Economic Analysis of the Courts*, 14 J. L. & ECON. 61 (1971).

⁶ See Albonetti, *Race and the Probability of Pleading Guilty*, 6 J. QUANTITATIVE CRIM. 315 (1990); Spiegel & Templeman, *Economics of Discriminatory Sentencing*, 5 J. QUANTITATIVE CRIM. 317 (1989); A. BLUMSTEIN, *supra* note 3. In the federal courts, the locus of this paper, we note the critique of Judge Frankel, see M. FRANKEL, *CRIMINAL SENTENCES* (1973), and the empirical work of Rhodes and Conly. See Rhodes & Conly, *Analysis of Federal Sentencing* (Fed. Just. Res. Program, May 1981).

examine more minutely the analytical meaning of the dispersion about these conditional means before tying this examination to a discussion of sentence disparity. Our analysis is based on sentences imposed prior to the introduction of the federal sentencing guidelines, and much of our discussion compares preguideline sentencing practices with the since implemented guidelines.

II. A STATISTICAL MODEL

When analyzing sentencing practices, researchers often use a statistical model based on regression analysis. A generic version of this model can be written:

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

T is the sentence imposed, measured as prison months in this application. According to this formula, a sentence is determined by measurable factors, $X_1, X_2 \dots X_n$, and an unmeasurable (or at least unmeasured) random factor, ϵ . The unmeasured factor, ϵ , is often called the "residual term." The greek symbols $\beta_0, \beta_1 \dots \beta_n$ represent how much weight each of the measurable factors receives at the time of sentencing. These weights, called "parameters," are estimated from the data.

The statistical model used in this study departs from the generic model, which assumes that all judges impose sentences using the same weights, $\beta_0, \beta_1 \dots \beta_n$, and that each judge uses the same weights every time he or she imposes a sentence. Rich *et al.*⁷ and others have argued that this assumption of homogenous parameters across both judges and across sentencing decisions is unrealistic. We agree, and in our modification of the generic model, we have assumed that the parameters $\beta_0, \beta_1 \dots \beta_n$ are random. The task is to determine the values of these parameters "on average."

One cost of adopting a random parameter model is that the statistical estimation procedure is complicated. In the generic model, the residual term ϵ is usually assumed to be normally distributed with a mean of zero and a constant standard deviation of σ . In our approach, σ is seen to vary, such that judges disagree more about the sentences to be imposed on repeat offenders who commit the most serious crimes than they do about first-time offenders who commit the least serious crimes. A mathematical expression for this assumption is written:

⁷ See W. RICH, *supra* note 1.

$$\sigma = \alpha_0 + \alpha_1(\beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n)$$

This model is closer to reality than is the generic model, it is less subject to the criticisms raised by Rich and colleagues, and—as we explain later—it supports a richer interpretation of sentence disparity.

Our approach departs from the generic model in another way. The effect of a plea bargain on the sentence imposed is usually modeled in one of two ways: either one of the X variables is used to represent occurrence of a guilty plea, or two distinct models are estimated, one for sentences imposed after a trial and one for sentences imposed after a guilty plea. Although we use both of these standard approaches, we also use a third approach, in which the sentence typically imposed after a trial is seen to be a multiple of the sentence typically imposed after a guilty plea. This aspect of the model can be represented:

$$T = (1 + \tau)\beta X + \epsilon$$

where t equals one when a trial occurs and zero otherwise, and τ is an additional parameter that requires estimation.

In a final departure from the generic model, we recognized that, in our data set, sentences have a lower limit of zero (because the dependent variable S was measured as length of a prison term) and an upper limit of 360 months (a practical limit based on preguideline parole eligibility). Thus, the estimation procedure had to account for these upper and lower limits. We used a two-limit tobit model, an estimation technique that is generally regarded as suitable for such applications.⁸

As an overview, this model is a two-limit tobit model with heteroscedastic error terms (that is, σ varies systematically) and a multiplier term to reflect the consequences of plea bargaining. Details of our statistical model are provided in a technical appendix. We now turn to using this model to describe the preguideline sentencing practices of federal judges.

III. ANALYSIS—BANK ROBBERY

Bank robbery is a unique crime among federal cases.⁹ It is not unlike that portrayed on television. Sometimes a bank robbery is committed by an individual rather than by a group. When it is a

⁸ G. MADDALA, *LIMITED-DEPENDENT AND QUALITATIVE VARIABLES IN ECONOMETRICS* (1983).

⁹ *Bank Robbery*, BUREAU JUST. STUD. BULL. (Aug. 1984).

group activity, roles are differentiated: some offenders are leaders or joint participants, others are subordinates, and still others are peripherally involved. Sometimes the offenders are armed; frequently they are not (although the victim may not know this). Rarely is anyone injured, although there are instances involving hostages, injury, and even death. What makes the crime unique, however, is its high probability of capture and its ease of conviction. These two features are to be expected, since the crime is committed before witnesses, captured on a camera, may trigger a silent alarm, and nets an exploding dye-pack along with the cash.

Its singular offense characteristics actually make bank robbery an ideal offense to study. First, conviction is so likely (93% of jury trials resulted in convictions during 1985) that evidence probably plays a secondary role during plea negotiations. The secondary role of evidence is useful given the absence from our data of variables about evidence. Second, a prison sentence is the usual punishment, so it is unnecessary to develop a prison equivalent for other sentences—probation, fines and restitution. Finally, our data include the major offense characteristics summarized in the previous paragraph.¹⁰

Variables used in the analysis are defined in Table 1. Descriptive statistics are reported in Table 2. Between January 1, 1984 and the middle of 1988, 5,168 offenders were convicted of bank robbery and sentenced under preguideline laws.¹¹ Columns 4 and 5 of Table 2 pertain to 445 offenders who were convicted by jury trial. Columns 2 and 3 pertain to the 3,697 offenders who pled guilty. A total of 400 cases had unknown dispositions. Another 74 offenders, who were convicted by bench trial, are not included in the statistics.¹²

¹⁰ Data come from the FPSSIS information system, developed and maintained by the Administrative Office of the United States Courts. The data contain over 95% of all cases that resulted in conviction between January 1, 1984, and the middle of 1988. These data are described in SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1.

¹¹ Defendants sentenced under the guidelines receive "real time" sentences, which determine time served except for a potential 15% reduction for good-time credits. Defendants sentenced prior to the guidelines received maximum sentences, and typically, only about one-third of the maximum term is served. The actual release date is determined by the Bureau of Prisons and the United States Parole Commission. Preguideline and postguideline sentences cannot be compared directly, a fact that necessitated the elimination of guideline cases. Furthermore, the guidelines have likely modified the factors entering the sentencing decision, or the weights those factors receive, or both. A single structural equation could not be used to model both pre- and postguideline sentences.

¹² Defendants who are convicted by bench trials typically are sentenced more leniently than defendants who are convicted by jury trials. See Tiffany, Yakov & Peters, *A*

Table 1
DEFINITIONS OF VARIABLES: ROBBERY

PRISON*	The maximum prison sentence imposed, coded in months. Life equals 999. "Partial months" are coded as 0.5.
AGGRAVAT	Coded 1 when the probation officer indicated that aggravating conditions existed; otherwise, coded zero.
MITIGATE	Coded 1 when the probation officer indicated that mitigating conditions existed; otherwise coded zero.
W-BRAN	A firearm or other weapon was brandished or otherwise used (that is, discharged if a firearm).
W-PRES	A firearm or other weapon was present but was not brandished or otherwise used.
MOREC	In a group crime, the offender was more culpable than other offenders.
EQUALC	In a group crime, the offender was equally culpable.
LESSC	In a group crime, the offender was less culpable.
CONVICTS	The logarithm of the number of adult and juvenile convictions not including the instant conviction for bank robbery. A one was added to all variables prior to taking logarithms.
INCLOW	The logarithm of the number of prior prison terms of between 30 days and one year. A one was added to all variables before taking logarithms.
INC1-5	The logarithm of the number of prior prison terms of between 1 and 5 years. A one was added to all variables before taking logarithms.
INC-5	The logarithm of the number of prior prison terms of more than five years. A one was added to all variables before taking logarithms.
MULTIPLE	"The current behavior was comprised of multiple acts, consisting of a number of impulse or opportunity crimes (<i>e.g.</i> robberies committed with little planning)" . . . from the coding instructions.
CONTINUE	"The current behavior was comprised of multiple recurrent acts that were planned as a long-range scheme" . . . from the coding instructions.
SDOLLAR	Logarithm of the amount of money stolen, when the crime was described as being a single event. Equal to zero when not a single event.
CDOLLAR	Logarithm of the amount of money stolen, when the crime was described as continuing. Equal to zero when not continuing.
MDOLLAR	Logarithm of the amount of money stolen, when the crime was described as multiple events. Equal to zero when not multiple events.
C1. .C11	Circuit within which the sentence was imposed. The 11th circuit is the excluded circuit.

* As explained in the text, prison terms above 360 months were treated as being censored at 360 months.

Table 2
DESCRIPTIVE STATISTICS: ROBBERY

	Guilty Pleas		Trials	
	mean	stand dev.	mean	stand dev.
PRISON	132.59	95.47	221.40	139.23
AGGRAVAT	0.20	0.40	0.22	0.41
MITIGATE	0.06	0.25	0.02	0.16
W-BRAN	0.33	0.47	0.54	0.49
W-PRES	0.17	0.38	0.15	0.36
MOREC	0.13	0.34	0.23	0.42
EQUALC	0.06	0.24	0.07	0.26
LESSC	0.08	0.27	0.08	0.28
INC-5	0.25	0.46	0.40	0.55
INC1-5	0.44	0.56	0.54	0.60
INCLow	0.50	0.64	0.47	0.60
CONVICTS	1.37	0.80	1.49	0.76
CONTINUE	0.14	0.34	0.16	0.37
MULTIPLE	0.37	0.48	0.18	0.38
SDOLLAR	3.94	4.22	5.67	4.33
CDOLLAR	1.29	3.25	1.57	3.61
MDOLLAR	3.32	4.35	1.71	3.66
C1	0.01	0.13	0.02	0.15
C2	0.04	0.21	0.02	0.16
C3	0.04	0.19	0.03	0.18
C4	0.09	0.28	0.10	0.30
C5	0.04	0.20	0.02	0.15
C6	0.06	0.25	0.10	0.30
C7	0.04	0.20	0.07	0.26
C8	0.05	0.21	0.08	0.27
C9	0.43	0.49	0.28	0.44
C10	0.03	0.18	0.05	0.22
TRIAL	0.00	0.00	1.00	0.00
N =	3697		445	

We excluded 170 cases that involved victim injury.¹³ We further excluded 337 cases that reported no dollar loss, because dollar loss was used as one measure of offense seriousness. A few other cases were dropped due to missing data.

Inspection of the descriptive statistics reported in Table 2 indicates that sentences typically were more severe following trials than following guilty pleas. One explanation is that cases that are tried tend to be more serious than those that result in guilty pleas. For example, a trial is more likely when the defendant used a weapon,

Statistical Analysis of Sentencing in Federal Courts: Defendants Convicted After Trial, 4 J. LEGAL STUDIES 369 (1975).

¹³ Victim injury is rare during a bank robbery, primarily because the offense takes place in a controlled setting and bank employees are instructed to acquiesce to the robber's demands. When overt violence occurs, the crime is an exception, and sentencing may follow a structural model that differs from the one analyzed here. We avoided the need to model these exceptions by dropping such cases from the analysis.

had an extensive criminal record, and exercised a leadership role. However, a trial is also more likely when the defendant robbed only one bank instead of several banks.¹⁴

A second explanation is that people who enter guilty pleas receive more lenient sentences holding constant the seriousness of the offense. Exploring this second explanation requires a regression model, which we discussed earlier.

Table 3 reports regression results. Variables, which were defined in Table 1, are identified in column 1. Column 2 pertains to guilty pleas only and column 3 pertains to trials only. The fourth column presents results for pleas and trials combined (with a dummy variable added for a trial). The final column presents results for the multiplier model. T-scores (parameter estimates divided by their standard errors) are reported in parentheses.

Focusing attention on the first two regressions, patterns emerge from sentences imposed both after trials and after guilty pleas. As expected, the length of the sentence increases with the severity of the crime, the offender's responsibility for that crime, and the offender's criminal record.

There is regional variation. Federal district courts are organized into circuits. The First and Second Circuits (C1 and C2) are Northeastern: for example, Massachusetts is from the First Circuit and New York Southern (Manhattan) is from the Second. The Northeastern part of the nation is often characterized as imposing relatively lenient sentences,¹⁵ which is consistent with the large negative coefficients in table 3. The Ninth Circuit (which includes California) also has a reputation for relative leniency, which is consistent with the regression results.

Judges from the South are sometimes characterized as being harsher than those in the rest of the country when sentencing. The results are consistent. Sentences tend to be especially severe in the Fourth (which extends from Maryland through the Carolinas) and Fifth (which consists of Texas, Louisiana and Mississippi) Circuits. The omitted region (which serves as a contrast) is the Eleventh Cir-

¹⁴ Perhaps this is because an acquittal is more likely when the case must be based on a single episode; therefore, the defendant has an incentive to seek a trial.

¹⁵ The circuits contain heterogeneous districts. For example, the Southern District of New York (Manhattan) has a caseload and legal culture that probably differs radically from the Northern District of New York (Albany, other upstate cities, and rural areas), although both districts are members of the second circuit. At best, use of circuit dummy variables captures rough regional effects, leaving within-circuit variation unexplained. We estimated models with districts substituted for circuits, drawing the data from the eleven districts that prosecuted 99 or more cases. The resulting regressions indicated considerable variation across the districts.

Table 3
REGRESSION RESULTS FOR PLEAS AND TRIALS: ROBBERY

	Guilty Plea	Trial	Both	Multiplier
CONSTANT	40.46	58.45	32.81	37.24
AGGRAVAT	(4.15)	(2.31)	(3.55)	(4.28)
MITIGATE	19.88	11.86	19.13	18.09
W-BRAN	(6.37)	(1.13)	(6.37)	(6.24)
W-PRES	-15.99	42.70	-13.89	-13.08
MOREC	(-3.82)	(1.72)	(-3.34)	(-3.22)
EQUALC	49.63	61.33	51.22	49.46
LESSC	(17.70)	(6.16)	(18.76)	(18.76)
INC-5	36.28	46.20	36.56	35.65
INC1-5	(11.65)	(4.06)	(12.08)	(12.18)
INCLOW	12.80	11.37	13.21	12.50
CONVICTS	(3.59)	(1.11)	(3.89)	(3.86)
CONTINUE	5.01	-26.69	2.69	2.25
MULTIPLE	(1.04)	(-1.89)	(0.58)	(0.51)
SDOLLAR	-41.72	-73.63	-44.94	-43.63
CDOLLAR	(-11.70)	(7.93)	(-13.01)	(-13.35)
MDOLLAR	37.92	44.62	38.17	36.45
C1	(12.61)	(5.10)	(13.35)	(13.26)
C2	19.47	11.82	18.58	17.67
C3	(7.80)	(1.55)	(7.74)	(7.66)
C4	-0.24	-15.83	-1.41	-1.61
C5	(-0.10)	(-2.25)	(-0.62)	(-0.74)
C6	17.81	28.27	18.84	18.18
C7	(8.25)	(4.09)	(8.98)	(9.02)
C8	17.56	-58.11	15.99	12.98
C9	(0.87)	(-0.85)	(0.84)	(0.71)
C10	-20.25	-109.19	-16.72	-19.53
α_0	(-1.45)	(-1.06)	(-1.23)	(-1.47)
α_1	3.45	5.02	4.11	3.74
γ	(3.18)	(1.81)	(4.00)	(3.88)
	6.35	17.95	7.32	7.19
	(3.13)	(2.59)	(3.81)	(3.91)
	8.40	23.31	8.84	8.80
	(6.33)	(2.06)	(6.69)	(6.74)
	-41.23	-104.05	-45.09	-45.67
	(-5.03)	(-4.62)	(-5.78)	(-6.07)
	-75.03	-78.78	-75.32	-73.15
	(-14.02)	(-4.49)	(-14.49)	(-14.58)
	-25.82	-7.55	-24.17	-23.00
	(-4.00)	(-0.32)	(-3.87)	(-3.80)
	14.99	12.87	14.88	14.43
	(2.88)	(0.86)	(3.01)	(3.04)
	15.30	56.45	17.54	18.04
	(2.27)	(1.77)	(2.67)	(2.81)
	-4.97	46.49	0.38	1.20
	(-0.91)	(2.68)	(0.07)	(0.24)
	-28.15	-5.52	-26.23	-23.49
	(-4.57)	(-0.36)	(-4.49)	(-4.29)
	-18.24	-1.47	-17.08	-15.68
	(-3.10)	(-0.09)	(-3.07)	(-2.97)
	-37.39	-37.52	-36.78	-35.09
	(-10.17)	(-3.55)	(-10.51)	(-10.45)
	-34.05	8.42	-29.64	-26.26
	(-5.23)	(0.45)	(-4.72)	(-4.42)
	42.24	27.49	43.50	42.63
	(21.29)	(3.38)	(23.18)	(22.64)
	0.21	0.25	0.20	0.21
	(13.57)	(5.64)	(14.40)	(14.70)
TRIAL			53.54	
			(12.75)	
				0.39
				(12.52)
No. Obs	3697	445	4142	4142
Log-likelihood	19591	2245	21875	21868

cuit (Alabama, Georgia and Florida). This contrast reinforces the Southern stereotype.

Another result is that σ , the standard error of the unexplained variation, increases with the average sentence (the parameter α_1 has a t-score in excess of 5 in all regressions). Furthermore, σ is about the same for pleas and trials. For sentences of ten years, σ equals 67 for pleas and 57 for trials; for sentences of fifteen years, σ is 80 for pleas and 72 for trials; and for sentences of twenty-five years, the values are 92 for pleas and 87 for trials. Although the conclusion is necessarily tentative, it appears that a guilty plea results in a reduced sentence, but otherwise does little to decrease the uncertainty of sentences.¹⁶ This finding is surprising, because "reduction of uncertainty" is often advanced to explain the preponderance of guilty pleas in American court rooms.

One more pattern is notable. The difference between the estimated trial and guilty plea structural models cannot be characterized by a shift in the intercept alone; that is, the introduction of a variable to represent a guilty plea does not fully account for the different sentences received by those who enter a guilty plea and those who are convicted by trial.¹⁷

Putting aside comparisons for the circuit dummy variables, parameter estimates derived from the regressions using only data from cases with guilty pleas appear to be a fraction of their corresponding values from the regressions using only data from cases with trials. There are exceptions to this rule, but these exceptions are for variables that were not statistically significant in the trial regression;

¹⁶ This finding is subject to interpretation. Suppose that X (the vector of exogenous variables) excludes variables that are known by the prosecution and defense to affect sentencing, or suppose that βX is an inadequate model specification, or both. Then σ_S and σ_T are not so much reflections of the uncertainty facing the prosecution and defense as they are indications of an inadequate model, and the interpretation given in the text fails. Alternatively, assume that X includes all variables known by the prosecution and defense to affect sentences, and that βX is an adequate specification of the relationship between X and the expected value of the sentence. Then the interpretation of σ_S and σ_T , as provided in the text, holds.

The model undoubtedly suffers from specification errors, and judicial decision making certainly provides illustrations of idiosyncratic sentencing practices, so σ_S and σ_T likely result from a combination of both factors. If idiosyncratic judicial decision making is a significant part of the explanation, it is remarkable that σ_S and σ_T are so close, implying—as stated in the text—that plea bargaining reduces uncertainty less than is commonly assumed. We should not attribute the size of the uncertainty completely to judicial idiosyncracies, however.

¹⁷ Regressions one and two can be treated as the "unconstrained" model. Regression three is the "constrained" model that forces all parameters except the intercept to be equated across the two regressions. We reject (at $p < .001$) the null hypothesis that the slope coefficients are the same across the equations.

thus, the exceptions might be attributed to the more precise parameter estimates for guilty pleas.¹⁸

Presuming that parameter estimates from the guilty plea regressions are a constant multiple of the parameters from the trial regressions, we estimated the fourth regression, which involved a multiplier. Results are presented as column 5 of table 3.¹⁹ The resulting model suggests that trials result in sentences that are roughly 1.39 times larger than those resulting from guilty pleas.

IV. ANALYSIS—DRUG CASES

Drug law violators are a large and growing proportion of the federal criminal case load.²⁰ Combatting the drug scourge has become a national priority. Both the Reagan and Bush administrations have increased resources for combatting drug crimes, and, in 1986, Congress passed the Anti-Drug Abuse Act, which doubled the mandatory minimum terms for most crimes of distribution, manufacturing and importation, and added enhancements to cover special conditions (such as selling to minors). The Sentencing Commission further increased the punishment for drug crimes by making the guideline sentence following a conspiracy conviction identical to the sentence following a conviction for the substantive offense.²¹ Thus, statutory provisions for sentencing drug law violators have changed twice over the last five years,²² with the expected consequence of increasing the sentences for these crimes.

Federal drug law violations are conveniently categorized as in-

¹⁸ The accuracy of the parameter estimates increases with the size of the sample. This increase in accuracy is reflected in the width of the confidence intervals for the parameter estimates. Hence, we have more "confidence" in the regressions for guilty pleas.

¹⁹ The multiplier model results in a value for minus the log-likelihood (21,868) that is between the values for the two equation model (21,836) and the dummy variable model (21,875). Although its explanatory power is less than the two-equation model, parsimony recommends the multiplier model over the two equation alternative.

We estimated two additional versions of the multiplier model. In the first, we added a constant term for trials. In the second, the multiplier term was a function of the circuit. Neither specification significantly improved the model's fit.

²⁰ In 1981, 3,700 of 30,400 felony and serious misdemeanor cases involved drug law violations as the most serious charge. As of 1985, 6,700 of 38,500 federal cases involved drug law violations. Administrative Office of the United States Courts, Annual Report of the Director, Table D-2 (1985).

²¹ The Anti-Drug Abuse Act of 1986 excluded conspiracy convictions from mandatory minimum terms. Anti-Drug Abuse Act of 1986, P.L. 99-570, 100 Stat. 3207.

²² In November 1988, President Reagan signed still another bill that increased the minimum and maximum terms for drug law violations, extended minimum and maximum terms to new offenses such as possession under specified conditions, and added collateral consequences such as loss of federal benefits. The analysis reported in this paper predates the Anti-Drug Abuse Act of 1988, P.L. 100-690, 102 Stat. 4181.

volving opiates, cocaine, marijuana, and other controlled substances. We chose to examine cases of a single drug type: cocaine. We included cases where the offense of conviction was conspiracy and distribution. We excluded cases of possession, and cases where the offense of conviction was the Continuing Criminal Enterprise,²³ the latter of which accounted for fewer than 100 cases per year.

Unlike bank robbery, cocaine-related crimes often differ from their fictional counterparts portrayed on television.²⁴ Cocaine distribution may be conducted by a hierarchical structure, but often the organization is small, and frequently, informal. We chose to examine cases where cocaine was distributed through an hierarchical structure, and further limited our examination to offenders whose roles were described as "leader" or "more culpable." This selection eliminated a large number of cases where the offender was judged to be "equally culpable," "less culpable," a "worker," or "acted alone." Prison terms are the predominant sentence for the selected group.

Some distribution chains tend to specialize by drug type; others deal with a spectrum of drugs. We selected for analysis those distribution chains where only one drug—cocaine—was mentioned.

Within this selected group, we further distinguished offenses based on the sophistication of the organization and the amount of drugs that were distributed. We measured sophistication by whether the offense was described as "ongoing and sophisticated," "multiple events, but unsophisticated," or "single act." We measured drug amount by the amount of pure cocaine distributed. After some pretesting, a log-transformation of drug amount was found to fit the data best. The probation officers who prepared presentence investigation reports for these cases recorded the sophistication and drug amount; laboratory tests typically determined the pure drug amount.

We limited the analysis to the twelve federal districts that prosecuted the largest volume of cocaine cases between January 1, 1984, and the middle of 1988. Besides reducing the amount of data to manageable size, focusing on these large districts afforded an exam-

²³ 21 U.S.C. § 848 (1988). Convictions under the Continuing Criminal Enterprise (CCE) were not included, because CCE cases are qualitatively distinct from conspiracy and distribution.

²⁴ See P. ADLER, *WHEELING AND DEALING* (1985); S. Murphy, D. Waldorf & C. Reinerman, *Drifting into Dealing: Becoming a Cocaine Seller* (NIJ Grant #7-0363-9-CA-IJ, 1989); D. Waldorf, *Final Report of the Ex-Sellers Project: An Exploratory Study of Indirect Criminal Justice Pressures on Cocaine Sellers* (NIJ Grant #89-IJ-CX-0036, 1989); M. Klein, C. Maxson & L. Cunningham, *Gang Involvement in Cocaine "Rock Trafficking"* (NIJ Grant #85-IJ-LX-0057, 1988).

ination of sentencing on the district level. Finally, for reasons discussed earlier, we excluded the records of a few hundred defendants who were sentenced under the guidelines. A total of 2,829 cases remained: 2,172 guilty pleas and 657 jury decisions.

Table 4
DEFINITIONS OF VARIABLES: DRUGS

PRISON*	The maximum prison sentence imposed, coded in months. Life terms are coded as 999. "Partial months" are coded as 0.5.
MULTIPLE	"The current behavior was comprised of multiple acts, consisting of a number of impulse or opportunity crimes (<i>e.g.</i> robberies committed with little planning)" . . . from the coding instructions.
CONTINUE	"The current behavior was comprised of multiple recurrent acts that were planned as a long-range scheme" . . . from the coding instructions.
COKE	The logarithm of the amount of drugs involved in the offense. Drugs were measured as grams of pure cocaine.
T85. . .T88	Dummy variables representing the years 1985 through 1988; 1984 is the excluded year.
CONVICT	The logarithm of the number of adult and juvenile convictions not including the instant conviction of distributing cocaine. A one was added to all variables prior to taking logarithms.
INCLW	The logarithm of the number of prior prison terms of between 30 days and one year. A one was added to all variables before taking logarithms.
INC1-5	The logarithm of the number of prior prison terms of between 1 and 5 years. A one was added to all variables before taking logarithms.
INC-5	The logarithm of the number of prior prison terms of more than five years. A one was added to all variables before taking logarithms.
PRIORS	Number of prior convictions for this type of offense.
D1. .D12	Dummy variables coded one or zero depending on the district where the defendant was convicted. District identities:
	1 New York Eastern
	2 New York Southern
	3 Pennsylvania Eastern
	4 Maryland
	5 Texas Northern
	6 Florida Middle
	7 Georgia Northern
	8 Texas Southern
	9 Michigan Eastern
	10 Illinois Northern
	11 Missouri Eastern
	12 California Central

* As explained in the text, prison terms above 360 months were treated as being censored at 360 months.

Variables are defined in table 4. Descriptive statistics appear in table 5. Regression results are reported in table 6.

Table 5—which presents descriptive statistics—has no sur-

Table 5
DESCRIPTIVE STATISTICS: DRUGS

	Guilty Pleas		Trials	
	mean	stand dev.	mean	stand dev.
PRISON	70.07	61.80	126.98	88.12
CONTINUE	0.33	0.47	0.39	0.48
MULTIPLE	0.25	0.43	0.20	0.40
COKE	6.53	2.97	7.56	2.89
T85	0.14	0.35	0.15	0.36
T86	0.28	0.45	0.23	0.42
T87	0.28	0.45	0.30	0.45
T88	0.14	0.35	0.15	0.36
CONVICT	0.49	0.65	0.44	0.62
INCLW	0.08	0.27	0.08	0.29
INC1-5	0.08	0.28	0.10	0.30
INC-5	0.02	0.16	0.04	0.18
PRIORS	0.17	0.38	0.15	0.34
D1	0.07	0.25	0.07	0.25
D2	0.13	0.33	0.08	0.27
D3	0.04	0.20	0.02	0.15
D4	0.09	0.29	0.07	0.25
D5	0.04	0.21	0.04	0.19
D6	0.08	0.27	0.08	0.28
D7	0.04	0.20	0.06	0.24
D8	0.06	0.24	0.05	0.22
D9	0.05	0.23	0.03	0.19
D10	0.04	0.21	0.03	0.18
D11	0.03	0.19	0.05	0.22
TRIAL	0.00	0.00	1.00	0.00
N=	2172	657		

prises. Prison terms are longer following convictions at trial than they are following convictions by guilty plea. This difference might be explained by the observation that more serious cases go to trial. For example, trials tend to involve more sophisticated crimes and larger quantities of drugs. A regression analysis is required to determine whether crime seriousness alone accounts for these differences.

According to Table 6, it appears that crime seriousness alone cannot account for sentence differences. The regression for guilty pleas (column 2) and the regression for trials (column 3) both indicate that sentence severity increases with the sophistication of the drug-distribution network (continuing) and with the amount of drugs involved. Both indicate that severity has increased over time (see variables T85-T88), although for trials, the trend does not seem to extend into 1988.²⁵ Both indicate large differences across

²⁵ The trend toward more severe sentences appears to predate the Anti-Drug Abuse Act, which applies only to defendants who committed their crimes after October, 1986.

Table 6
REGRESSION RESULTS FOR PLEAS AND TRIALS: DRUGS

	Guilty Plea	Trial	Both	Multiplier
CONST	4.71 (1.21)	17.40 (2.00)	-2.91 (-0.77)	0.18 (0.06)
CONTINUE	26.13 (10.44)	31.77 (5.46)	26.77 (11.13)	23.27 (11.29)
MULTIPLE	2.54 (1.23)	-4.86 (-0.91)	1.85 (0.95)	1.02 (0.56)
COKE	6.91 (18.28)	9.57 (10.22)	7.57 (21.87)	7.08 (23.34)
T85	1.03 (0.32)	6.61 (0.99)	3.47 (1.16)	2.90 (1.14)
T86	1.86 (0.65)	16.90 (2.36)	4.70 (1.75)	5.34 (2.34)
T87	8.06 (2.75)	18.64 (2.48)	12.26 (4.42)	11.49 (4.89)
T88	21.98 (5.73)	7.20 (0.86)	18.67 (5.59)	16.77 (5.79)
CONVICT	4.71 (2.25)	-2.47 (-0.42)	5.82 (2.91)	4.98 (2.77)
INCLOW	-2.18 (-0.57)	-6.39 (-0.55)	-2.20 (-0.60)	-2.44 (-0.71)
INC1-5	9.23 (2.25)	34.29 (2.75)	13.56 (3.31)	13.55 (3.64)
INC-5	14.96 (2.40)	75.23 (3.09)	25.16 (3.61)	23.91 (3.96)
PRIORS	9.32 (3.05)	15.74 (1.79)	9.65 (3.22)	9.26 (3.45)
D1	-18.06 (-5.08)	-25.64 (-3.39)	-18.94 (-5.67)	-15.67 (-5.33)
D2	-8.86 (-2.87)	-5.51 (-0.95)	-9.86 (-3.52)	-6.97 (-2.87)
D3	13.84 (2.54)	4.42 (0.31)	13.89 (2.69)	14.52 (3.04)
D4	-11.92 (-3.60)	-19.61 (-2.64)	-11.94 (-3.81)	-10.66 (-3.79)
D5	37.09 (6.05)	58.14 (2.93)	41.71 (7.06)	40.82 (7.57)
D6	3.97 (1.04)	53.02 (4.80)	11.63 (3.03)	15.12 (4.50)
D7	4.76 (0.80)	4.53 (0.45)	4.49 (0.85)	5.76 (1.27)
D8	3.88 (0.89)	17.86 (1.70)	4.97 (1.21)	8.85 (2.43)
D9	-16.13 (-3.92)	49.16 (2.92)	-10.01 (-2.34)	-10.11 (-2.69)
D10	-24.75 (-6.63)	-4.18 (-0.37)	-22.19 (-6.03)	-18.78 (-5.53)
D11	-2.29 (-0.49)	22.78 (1.72)	2.94 (0.63)	4.94 (1.21)
α_0	22.33 (10.18)	5.53 (0.75)	24.01 (13.11)	23.27 (12.86)
α_1	0.47 (13.43)	0.54 (7.82)	0.43 (17.12)	0.43 (17.09)
TRIAL			42.16 (14.52)	
γ				0.61 (14.17)
No. Obs	2172	657	2829	2829
Log-likelihood	10855	3511	14428	14408

district courts in the sentences imposed on convicted drug dealers, and while there are exceptions to the pattern, non-Southern districts tend to impose sentences that are more lenient than those imposed in Southern districts.

Using the results from the regression reported in column 4 (which includes all pleas and trials, with a dummy variable to represent the occurrence of a trial), we reject the null hypothesis that sentences are the same, except for a shift in the intercept, after trials and pleas. We estimated a multiplier model. Results are reported in column 5. Based on these results, a defendant convicted by trial appears to receive a sentence that is roughly 61% larger than a similarly situated defendant who pled guilty.²⁶

Once again, it is noteworthy that σ_S and σ_T are roughly the same. For sentences of 60 months, the standard deviation for guilty pleas is 50; for trials, it is 38. For sentences of 120 months, the values are 79 and 70; for 180 months, they are 107 and 103; for 240 months, 135 and 135. Although findings are tentative because estimates of σ cannot be attributed exclusively to judicial sentencing idiosyncracies, plea bargaining appears to do little to reduce the uncertainty of sentencing.

V. SENTENCE DISPARITY

Expanding upon our earlier working definition, a sentence is disparate when (1) two defendants who are identical in all ways that should affect the sentence nevertheless receive different sentences, or (2) two defendants who are identical except for one or more factors that should affect the sentence receive the same sentence, or (3) a sentence is imposed that is too severe or is too lenient given accepted sentencing criteria. This definition may be thus illustrated:

(1) Two first-time offenders with identical backgrounds rob banks under the identical set of circumstances. The first offender enters a guilty plea and receives a fifteen year term, while the second enters a guilty plea and receives a twenty-five year term.

(2) Two first-time offenders with identical backgrounds rob banks under identical circumstances, except that the first offender displays a firearm while the second offender is unarmed. Both offenders receive twenty year prison terms.

See supra note 21. Few of these offenders would have been convicted prior to 1987, yet sentence severity increases before that date.

²⁶ Based on a likelihood ratio test, the model with a shift parameter was rejected at $p < .001$. The multiplier model was also rejected at $p < .001$. However, parsimony seems to recommend the multiplier model, for which minus the log-likelihood (14,408) was intermediate between that recorded for the unconstrained model (14,366) and the dummy variable model (14,428).

(3) A repeat offenders robs a bank, discharges a firearm in the direction of a teller, and pistol-whips the guard. The offender receives a two year term of probation.

One notable feature of this definition is that its application requires an explicit sentencing standard. For example, illustrations (1) and (2) require *qualitative* standards for what variables should and should not be taken into account during sentencing, and illustration (3) requires a *quantitative* standard for how much each factor should affect the sentence.

To tie this working definition into the statistical analysis, it is useful to revisit the statistical model, which can be summarized as:

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

$$\sigma = \alpha_0 + \alpha_1(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

where the Xs represent variables taken into account by sentencing judges and ϵ is a random error term with no observable counterpart. We are able to estimate the β s, the quantitative effect of each explanatory variable, and σ , the amount of unexplained sentencing variation.

Which of these parameters measures disparity? It is seductive to examine the residual variance σ^2 , note that it is large relative to mean sentence imposed, and conclude that sentence disparity in the federal courts was rampant prior to sentencing guidelines (which took effect after these data were collected).²⁷ Indeed, sentence disparity may have been rampant, but for a number of reasons,²⁸ this

²⁷ The guidelines were in effect after November 1987. However, the guidelines applied only to offenders who committed their crimes after November, 1987, and somewhat more than half the district court judges found the guidelines unconstitutional. Consequently, few defendants had been sentenced under the guidelines during the time covered by this analysis.

²⁸ First, although our model is more general than models usually employed to study sentencing practices, there is no guarantee that the model's specification is correct. If the model is not linear-additive, as we have assumed, misspecification will be reflected in the residual term, and "disparity" may be nothing more than inadequate model specification.

Second, although our data provide rich descriptions of the offense and the offender, there is no guarantee that we have included all the variables considered by a judge at the time of sentencing. "Disparity" may be nothing more than omitted variables that appear as a large residual.

Third, there is a fundamental problem with equating residual variation to disparity. Assuming a sentencing standard, suppose that the X vector can be decomposed into two parts: factors that are legitimate when used during sentencing (L) and factors that are not legitimate (I). Similarly, β can be decomposed into two corresponding parts: the quantitative importance of legitimate components (β_L) and the quantitative importance of illegitimate components (β_I). Disparity can be said to exist when $\beta_I \neq 0$, because defendants, who are identical on all factors that should affect sentencing, receive different

conclusion is necessarily tentative.

Subject to appropriate caveats, however, the regression of T on X produces parameter estimates that can be used to draw inferences about sentence disparity. Before these inferences can be drawn, we must distinguish between factors that have a legitimate role (L) at the time of sentencing and those factors that do not (I).

Here is the problem for a social scientist: there exists no clearly defined standard distinguishing L from I. Consequently, there is considerable ambiguity when drawing conclusions about sentence disparity.

How does one find a standard? The regression equation itself might be said to "reveal" an implicit "standard" in the form of "typical" decision making. Unfortunately, this approach begs the question. For example, if Blacks receive longer sentences than Caucasians, should we assume that race is a legitimate factor at the time of sentencing? If repeat offenders receive extended prison terms, should we presume that the sentence is optimal from the standpoint of deterrence? incapacitation? rehabilitation? There is no compelling reason for treating a regression equation as a revealed standard.

The general search may be quixotic, but for the federal courts, Congress has mandated the Sentencing Commission to be the standard setter. This observation suggests a test for sentence disparity: Offenders who were sentenced in the past should be categorized according to variables that the Sentencing Commission has deemed appropriate for sentencing. Following this classification, descriptive statistics can be estimated for sentences within each category. The extent to which "old" sentences depart from "new standards" can be a measure of disparity.²⁹ An overview of sentencing guidelines, as they apply to bank robbery and cocaine distribution, appears in the next section.

sentences based on factors that should not influence sentencing decisions. Disparity can also be said to exist when $\beta_L=0$, because defendants who are different on a factor that should be considered at the time of sentencing nevertheless receive the same sentences. Even a non-zero estimate for β_L may indicate disparity if β_L differs from an accepted standard. Thus, sentence disparity may be reflected by variation that is accounted for by the regression but attributed to illegitimate factors (the explained variation) as well as by variation that is not accounted for by the regression (unexplained variation).

²⁹ A problem with this test is that the standard is a moving target; namely, whatever the Sentencing Commission deems to be the guideline sentence at any point in time. That the Commission has already modified its guidelines several times since 1987 points to the unavailability of anything but a rubber ruler for measuring sentence disparity.

A. SENTENCING GUIDELINES

Figure 1 provides a summary of bank robbery and drug sales sentencing guidelines, as these were promulgated in 1987.³⁰ In operation, the "levels" and "points" referenced by the guidelines are totaled separately, and a sentencing grid (with offense seriousness levels on one axis and criminal history points on the other) specifies a mandatory sentencing range. An abbreviated version of the sentencing grid, which actually has forty-three levels, appears as Figure 2.

The Crime Control Act of 1984 abolished parole,³¹ so sentences imposed pursuant to the guidelines are equivalent to time to be served, subject only to a fifteen percent reduction for good-time. The sentencing grid provides a range within which the judge can impose a sentence; any sentence within this range can be imposed. Judges can depart from this range, but only for cause, and most departures can be appealed by the defense and the government.

As an illustration, the bank robbery guidelines set a base level of 18 for a defendant who robs a single bank and steals less than \$2,500. If the defendant was unarmed, caused no injury, did not take hostages (Figure 1 does not report provisions for injury and hostages), and had no prior criminal record, then the level 18 score would result in the defendant serving between 27 and 33 months, not counting any earned good-time. If the defendant's criminal record resulted in a criminal history score of 13 points or more, he or she would serve 57-71 months, not counting good-time. Enhancements to the offense seriousness level increase time served by a compound rate of roughly 12.5% per level. A criminal history score of 2 or 3 points increases the sentence by 27%, a score of 4 to 6 points increases it by 42%, a score of 7 to 9 points increases it by 60%, a score of 10 to 12 points increases it by 80%, and a score of 13 or more increases it by 103%.

B. GUILTY PLEAS—DISPARITY OR ADMINISTRATIVE NECESSITY?

Plea bargaining, a process that is only remotely related to the

³⁰ Figure 1 excludes some guideline provisions for bank robbery and drug sales, paraphrases the provisions that are included, and omits details that are important when applying the guidelines. Nevertheless, this simplified representation of the guidelines that appears in Figure 1 is adequate for our purposes—comparing results from the statistical analysis with the standards established by the guidelines.

³¹ Comprehensive Crime Control Act of 1984, P.L. 98-473, 98 Stat. 1987, ch. II, Chapter II, entitled Sentencing Reform Act of 1984, has been codified as 18 U.S.C. § 3551 (1988).

Figure 1**SELECTED PROVISIONS OF THE FEDERAL SENTENCING GUIDELINES****Bank Robbery**

Assign a base offense level of 18

If the value of the property taken or destroyed exceeded \$2,500, increase the offense level as follows: 1 level for \$2,501-\$10,000; 2 levels for \$10,001-\$50,000; 3 levels for \$50,000-\$250,000; 4 levels for \$250,000-\$1,000,000; 5 levels for \$1,000,000-\$5,000,000; and 6 levels for more than \$5,000,001.

If a firearm was discharged, increase by 5 levels, if a firearm or a dangerous weapon was otherwise used, increase by 4 levels, if a firearm or other dangerous weapon was brandished, displayed or possessed, increase by 3 levels.

Add 2 levels if two banks were robbed; add 3 levels if three banks were robbed; add 4 levels if four to five banks were robbed; and add 5 levels if more than 5 banks were robbed.

Distribution of Cocaine

For distribution of cocaine, the offense level is determined by the amount of drugs involved: 0-25 grams, level 12; 25-49 grams, level 14; 50-99 grams, level 16; 100-199 grams, level 18, 200-299 grams, level 20; 300-399 grams, level 22; 400-499 grams, level 24; 0.5-1.9 kilograms, level 26; 2-3.4 kilograms, level 32; 15-49.9 kilograms, level 34; 50 kilograms or more, level 36.

For multiple offenses, add all drug amounts together before applying the above rules.

Guidelines Common to Both Offenses

Based on the role in the offense, increase the offense level by 4 if the defendant was an organizer or leader of a criminal activity that involved five or more participants, by 3 levels if the defendant was a manager or supervisor (but not an organizer or a leader) and the criminal activity involved five or more participants, and by 2 levels if the defendant was an organizer, leader, manager or supervisor in any other criminal activity.

Based on the role in the offense, decrease the offense level by 4 levels if the defendant was a minimal participant and by 2 levels if the defendant was a minor participant.

If the defendant clearly demonstrates a recognition and affirmative acceptance of personal responsibility for the offense of conviction, reduce the offense level by 2 levels.

Criminal History Score

To construct the criminal history score, add 3 points for each prior offense of imprisonment exceeding one year and one month, add 2 points for each prior sentence of imprisonment of at least sixty days but less than one year and one month, add 1 point for each other prior sentence (up to 4 points); add 2 points if the defendant committed the instant offense while under any criminal justice sentence; add 2 points if the defendant committed the instant offense less than two years after release from prison.

Source: October 1987 Guidelines Manual issued by the U.S. Sentencing Commission. Guidelines text was paraphrased, details were deleted, and some provisions were omitted.

Figure 2
GUIDELINES SENTENCING GRID (ABBREVIATED)

offense level	criminal history points					
	0-1	2-3	4-6	7-9	10-12	13+
18	27-33	30-37	33-41	41-51	51-63	57-71
19	30-37	33-41	37-46	46-57	57-71	63-78
20	33-41	37-46	41-51	51-63	63-78	70-87
21	41-51	46-57	51-63	57-71	70-87	77-96
22	46-57	51-63	57-71	63-78	77-96	84-105

Notes: The sentencing grid contains 43 offense levels. The lower levels provide for sentences to probation; the highest levels provide for life sentences. In this abbreviated version, the grid provides only for determinate terms of incarceration.

To determine a sentence, the offense level is computed based on the offense committed, the criminal history point total is computed based on the offender's criminal record, and the sentence range is determined by the intersection of the offense level and the criminal history points. For example, an offense level of 20 and a criminal history point score of 5 would require a prison term between 41 and 51 months.

magnitude of the crime and the defendant's culpability, is a major determinant of the sentence imposed. The guidelines contain an "acceptance of responsibility" provision, which might be expected to make sentences following conviction at trial approximately 27% longer than sentences imposed after acceptance of a guilty plea.³²

Prior to the guidelines, defendants convicted at trial could expect to receive sentences that were 61% (cocaine) to 39% (robbery) longer than sentences expected by defendants who entered guilty pleas. The guidelines, through the "acceptance of responsibility" provision, contemplate a sentence that is about 27% higher for defendants who are convicted by trial than for defendants who pled guilty. The guidelines further allow unlimited sentence concessions in exchange for substantial cooperation with the government regarding other criminal cases.

Whether such extreme sentence variation should be allowed in

³² Defendants who enter a guilty plea do not by right receive credit for acceptance of responsibility; nor are defendants who are convicted by trial precluded from receiving the 2-level adjustment. Nevertheless, Commission deliberations were consistent with an expectation that the acceptance of responsibility provision was seen as a substitute for sentence concessions for plea bargains. In addition, the Commission provided for sentencing below the guideline minimum when the defendant provided substantial assistance to the government in a criminal prosecution.

a "just" sentencing system is arguable. The guideline drafters' ambivalence about this debate is reflected in the guidelines' acceptance of responsibility provision, which provides for a sentence concession not because the defendant entered a guilty plea, but rather because he admitted and accepted responsibility for his crime. Although there is a shade of difference between "admitting responsibility" and "entering a plea of guilty," in practice the shade is not a dark one, and the acceptance of responsibility provision is principally a politically more palatable alternative to a guideline provision that would more directly reward defendants for entering guilty pleas.³³

The plea bargaining debate has many facets. Among these is the question of how large sentence concessions must be to induce a sufficient number of defendants to enter guilty pleas rather than go to trial. Our statistical analysis provides a tentative answer to this question.

Tables 3 and 6 reveal marked differences in sentences across circuits (robbery) and districts (drugs). These tables also reveal that sentences are more severe for defendants convicted by trial than for defendants who pled guilty. These two observations lead to a question: Do inter-circuit (inter-district) differences in sentence concessions affect the probability of a trial?

To answer this question, we used the regressions on guilty pleas and trials to compute the predicted sentence following a trial (ST) and following a guilty plea (SP) for every defendant, regardless of how he was actually convicted. We interpreted the resulting sta-

³³ Preliminary drafts of the guidelines used language that more directly recognized a need to exchange sentence concessions for guilty pleas. Such wording was objectionable to many outside the Commission, especially certain federal judges, who found the acceptance of responsibility provision to be more acceptable. Nevertheless, the acceptance of responsibility provision continued to be seen by the Commission as a surrogate means to induce guilty pleas, and much internal debate at the Commission was about whether the acceptance of responsibility provision would be adequate or whether some other concessions were necessary. The Commission was intentionally vague in its provisions for plea bargaining, partly because the Commissioners could not agree about the role of plea bargaining under a system of guidelines, partly because they were uncertain about how plea bargaining would actually operate under the guidelines. Once the guidelines were implemented, some Commissioners argued that the acceptance of responsibility provision was as much leverage as the government could legitimately exercise; other Commissioners held (and promoted) the position that all elements recognized under the guidelines were negotiable by the prosecution and defense. From the former view, for example, when a gun was used during a robbery, that fact must be used by the sentencing judge when imposing sentence; from the latter view, the prosecution could stipulate that no gun was used, in which case it would not be factored into the guidelines. These insights are from the author's personal observations while Director of Research at the United States Sentencing Commission. See also SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1, at 49.

tistic, $C=SP/ST$, as an indication of the concession that a defendant expects to receive for a guilty plea. We used ST and C as explanatory variables in a logit regression with conviction by trial as the dependent variable. Furthermore, we included the predicted value of the sentence following trial, ST , because of evidence that the most serious crimes were most likely to lead to trials.

For robbery, the regression results were:

$$\begin{array}{rcl} -3.88 & +0.0045S_T & +1.19C \\ (-15.37) & (7.16) & (5.00) \end{array}$$

where t-scores are reported in parentheses. For drugs, the results were:

$$\begin{array}{rcl} -2.23 & +0.0059S_T & +0.48C \\ (-10.19) & (6.21) & (2.02) \end{array}$$

These results imply two conclusions. First, the probability of a trial increases with the size of the sentence that is expected to be imposed following conviction at trial. Second, the probability of a trial decreases with the size of the concession received following a guilty plea. In quantitative terms, the derivative of the probability of a trial, when evaluated at the mean for the independent variables, equals:

$$\begin{array}{rcl} \text{robber: } \partial P/\partial ST & = & 0.0052 \\ & \partial P/\partial C & = 0.0011 \\ \text{drugs: } \partial P/\partial ST & = & 0.013 \\ & \partial P/\partial C & = 0.000086 \end{array}$$

where ST is measured as years and C as a .01 change in the indicator of a sentence concession. Although there is no absolute standard for "big" or "small" changes, these derivatives suggest that the demand for trials is inelastic with respect to the size of the plea bargain concession. Put another way, the analysis produces considerable evidence that defendants receive sentence concessions following a guilty plea, but very little evidence to indicate that a defendant's propensity to enter a guilty plea is sensitive to the size of these sentence concessions. Perhaps this insensitivity is an explanation for why reform programs that abolish plea bargaining have seldom resulted in large, permanent increases in the incidence of trials.³⁴

That the admission of guilt is so insensitive to the size of the

³⁴ M. TONRY, *supra* note 2.

sentence concession exchanged for a guilty plea raises the question: Should the sentencing guidelines provide for large sentence concessions that are unrelated to the magnitude of the crime and the defendant's criminal record without a compelling administrative necessity? Our observations are that sentence concessions prior to the guidelines were large, that sentence concessions anticipated under the guidelines are not much smaller, and that the administrative necessity of such large sentence concessions may be lacking.

Studies by the Sentencing Commission, which have been promised but are not yet forthcoming, are imperative for answering this question. For now, we conclude that under the guidelines—as was true under preguideline practices—defendants who enter guilty pleas will serve markedly shorter terms than will similarly situated defendants who are convicted at trial.

C. PREGUIDELINE SENTENCES COMPARED TO THE GUIDELINES

Beyond plea bargaining, past sentencing practices can be compared to the standard set by the guidelines. This comparison requires three steps: using results from the regression analysis to compute the sentence typically imposed on defendants who have specific characteristics, determining time to be served as stipulated by the guidelines for the same defendants, and comparing the two.

According to the regression results for bank robbery, a first-time offender who was unarmed, who acted alone, and who stole \$8,000 from a single bank could expect to receive a sentence of somewhat more than five years.³⁵ Given extant parole guidelines and good-time provisions, this offender could expect to serve 28 to 35 months.³⁶ The original sentencing guidelines required a term of between 30 and 37 months, exclusive of good-time; thus, the sentence required by the original guidelines appears to be consistent with preguideline sentences, at least for the type of offender and offense described in this paragraph.

According to the regression results, possession of a weapon increased the sentence by 21% for those convicted by trial and by 27% for those who entered a guilty plea. Brandishing or otherwise using a weapon increased the sentence by 28% to 38%. In comparison, under the guidelines, the weapon enhancement is between 27% and 54%, depending on how the weapon was used.³⁷ Regard-

³⁵ All variables other than those mentioned were set at their average values when making these calculations.

³⁶ SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1, ch. 4.

³⁷ Each level increase in the guidelines increases the time to be served by about 12.5%. We computed the sentence enhancement attributable to the guidelines as 1.125

ing sentence enhancement for possession and use of a weapon, preguideline practices are roughly consistent with the standards contemplated by the guidelines.

According to the regression results, robbers who were judged to be "more culpable" received terms that were 5% to 10% longer (following trials and guilty pleas) than those judged "equally culpable" or "acting alone." The guidelines call for a 27% sentence enhancement, so preguideline practices were more lenient. According to the regression results, robbers judged "less culpable" received sentences that were 32% to 35% shorter than those judged "equally culpable" or "acting alone." Under the guidelines, the reduction would be 27% to 60%, depending on whether the defendant was determined to be a "minimal" or a "minor" participant. Regarding the offender's role in the offense, preguideline practices were roughly consistent with the standard set by the guidelines.

According to the results from the regressions, two prior convictions that had not resulted in prison terms would be expected to increase the current sentence by 14% to 15% (trials and pleas); two prior convictions that both resulted in prison terms would be expected to increase the current sentence by 29% to 47%, depending on other assumptions. The guidelines require a sentence enhancement of about 13% for two prior convictions that did not result in prison, and between 27% and 42% for two prior convictions that both resulted in prison terms that exceeded one year. Although the guidelines again appear to be consistent with preguideline sentencing practices, the guidelines contain a career offender provision³⁸ requiring that defendants who have two prior convictions for controlled substance offenses or violent crimes must serve no fewer than 70 months and may be required to serve as many as 327 months, depending on the offenders' prior criminal records and the statute under which they are convicted. Many bank robbers will qualify for application of this provision,³⁹ so the guidelines' criminal history provisions are likely to be more severe for repeat offenders than was true of preguideline sentencing practices.

According to the regression results, a second bank robbery is likely to increase the prison term by 28% to 44% depending on

raised to the power of L, where L is the number of level adjustments stipulated by the guidelines. Other guideline effects reported in this section were computed similarly.

³⁸ The career offender provision was required by the Crime Control Act of 1984. 28 U.S.C. § 994(h) (1988). The Commission determined how this provision would be applied, however.

³⁹ Block & Rhodes, *Forecasting the Impact of the Federal Sentencing Guidelines*, 7 BEHAV. SCI. & L. 51 (1989).

whether the offenses are described as "multiple" or "continuing." Under the guidelines, the increase would be about 27%, so the guidelines might be seen as being somewhat more lenient toward multiple offenses.

For bank robbery cases, then, federal judges appear to have taken into account the same factors prior to the guidelines as they are required to take into account by the guidelines. Furthermore, the relative weights associated with each factor prior to the guidelines is roughly similar, although far from identical, to the weights assigned by the guidelines.

For drug law violations, comparisons are less useful because the minimum sentences mandated by the Anti-Drug Abuse Act of 1986 forced the Sentencing Commission to establish guideline sentences that departed from preguideline practices. According to the regression results, a first-time offender who sold 879 grams of pure cocaine (the average weight in these data and roughly a kilogram of packaged cocaine) would be sentenced to 100 months following conviction at trial. He would be eligible for parole after serving one-third of this sentence, and given prevailing release practices of the United States Parole Commission, he could expect to serve about 33 to 41 months.⁴⁰ Under the guidelines, the sentence range is 78 to 121 months exclusive of good-time, assuming that the 879 grams are not heavily diluted.⁴¹ The guidelines require prison terms that are markedly longer than preguideline sentences.⁴²

According to the regressions, an offender with two prior convictions, neither of which had resulted in a prison term, would receive no more than a marginal sentence enhancement based on his criminal record. Under the guidelines, the sentence enhancement would be about 27%. According to the regressions, an offender with two prior convictions, both of which had resulted in prison terms, would receive a sentence that is considerably longer than the sentence of a first-time offender. Under the guidelines, the difference would be about 27% to 42%.

D. DISPARITY

Although the comparisons are necessarily crude, the regression analysis seems to contain most of the factors that the Sentencing Commission has deemed "legitimate" considerations for purposes

⁴⁰ SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1, chap. 4.

⁴¹ The guidelines' sentence is based on bulk weight, rather than pure weight, but a near kilogram of cocaine is unlikely to be diluted more than about 10%.

⁴² Block & Rhodes, *supra* note 39.

of sentencing. Furthermore, for bank robbery, the quantitative importance of those factors taken into account in the past are not greatly different from the weights stipulated by the guidelines. Drug law violators will receive prison sentences that are higher under the guidelines than during the preguideline period, primarily because the Anti-Drug Abuse Act of 1986 stipulated increased mandatory minimum terms based on the amount of drugs sold and the conditions under which the drugs were sold.

With this observation made, some conclusions can be drawn. First, although specification error and omitted variables undoubtedly account for some of the unmeasured variation, the data include variables recognized by the Sentencing Commission as legitimate factors to be used at the time of sentencing, and the linearity assumption is consistent with the sentencing guidelines.

Thus, these are rich data,⁴³ and a linear-additive form is likely to capture most of the systematic variation in sentence imposed. Yet the unexplained variation—which must be attributable primarily⁴⁴ to factors that should not be taken into account during sentencing—is large. For bank robbery, σ is about eight years for robbers who as a group receive average terms of twenty-five years; for drug-law violations, σ is about ten years for dealers who as a group receive average terms of twenty years. Given the standard that is implicit in the discretionary 25% range allowed under the guidelines (a provision of the Crime Control Act of 1984 and reflected in the

⁴³ Using a smaller but more detailed data set, the Sentencing Commission, in chapter four of its SUPPLEMENTAL REPORT ON THE INITIAL SENTENCING GUIDELINES (*see supra* note 1), found other variables that affected the sentencing decision: degree of planning involved, hostages taken, degree of injury inflicted by the defendant's accomplice, and whether the defendant earned most of his income from crime. We excluded all cases with injury and all cases involving hostages from our data, so these factors cannot affect the analysis. The absence of the other two variables—degree of planning and defendant's income—appear as part of the residual, but neither had a large effect in the Sentencing Commission's statistical model. For drug law violations, the Commission reported that additional variables affect the sentence imposed: weapon present or used, cooperation with authorities, defendant was drug user, defendant imported the drug, and defendant's income came primarily from crime. Some of these variables are expected to have little importance for large cocaine distribution schemes, but others would have increased the model's explanatory power somewhat.

⁴⁴ The guidelines recognize several explicit, special conditions, such as extreme psychological injury to a victim, for which the judge may depart from the guidelines. These special conditions may have occurred during the crimes that resulted in the cases that comprised our data, where they would partly account for the unexplained residual variation. Based on qualitative and quantitative analysis, the Commission concluded that such special conditions were rare, which was why those conditions were not incorporated explicitly as elements of the guidelines. Such rare events could not explain much of the unexplained variation in these regressions.

sentencing grid), sentences imposed during the preguideline period diverged greatly from standards of equal justice.

Furthermore, the regional and district variation is large. An argument can be made that a sentencing standard should afford wide regional variation, allowing as a matter of political expression a conservative South to impose harsher sentences than a liberal North. The Sentencing Commission has rejected such a district-specific standard. Consequently, our statistics provide strong evidence of disparate sentencing across the nation's federal courts during the mid-1980s prior to the introduction of sentencing guidelines.

The unexplained residual undoubtedly contains a mixture of systematic variation that could be attributed to unwarranted factors (*e.g.*, age, sex, race) that were not included in our analysis, judicial idiosyncrasy, and pure random chance. How the unexplained residual should be divided among these factors is unknown, but we doubt that age, sex and race (factors that have much interested other researchers) explain much of the unmeasured variation. Earlier analyses conducted at the Sentencing Commission and separate analysis by Rhodes and Conly⁴⁵ generally found no strong effects attributable to race or age; gender effects were somewhat stronger, but for the two crimes examined here, men predominate. Rhodes and Conly reported that the unemployed are sentenced to longer terms than the employed, and the Sentencing Commission⁴⁶ reported that judges sentence defendants who derive their income primarily from crime more severely than those who have alternative legitimate income sources. Our data did not provide these variables, so we conclude that they are part of the residual term, but certainly not a large part of that residual.

Aside from variation in regional perspectives, we feel that sentence disparity has not resulted from biased judges imposing unacceptable sentences based on illegitimate factors; rather, disparity seems to have resulted from well-intentioned judges making individualistic decisions in a setting almost devoid of principle, structure, and regularity. Whether the federal Sentencing Guidelines Commission can further articulate the principles that should guide judges when sentencing and provide the structure for consistent sentencing is a question of pressing policy significance.

VI. SUMMARY

This paper presents a statistical model of sentencing that em-

⁴⁵ Rhodes & Conly, *supra* note 6.

⁴⁶ SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1, at 39.

phasizes the average sentences imposed after guilty pleas and trials as well as the variation about those averages. Especially notable in this statistical model is the explicit specification of a random coefficient model, which has a strong justification when analyzing sentencing data. The specification provides an improved measure of sentence dispersion, a statistic that is important to measuring sentence disparity.

The utility of this model has been demonstrated using data about bank robbery and cocaine-distribution. We used the model to draw conclusions about sentence disparity. Regarding plea bargaining, we noted that sentence concessions in exchange for a guilty plea are statistically significant and might be considered large. We uncovered only scant evidence that the frequency of guilty pleas increases or decreases with variation in the size of these concessions. Regarding sentence disparity, we concluded tentatively—subject to numerous caveats—that preguideline disparity might be considered to be a major problem in federal courts. Whether the recently introduced guidelines can reduce this disparity remains an open question.

APPENDIX: STATISTICAL METHODOLOGY

Interest centers on estimating and comparing the conditional mean and standard deviation for sentences imposed after trials and after guilty pleas. Our approach required four steps: (1) using data from trials only, we estimated a regression equation with prison time as the dependent variable; (2) next, using data from guilty pleas only, we repeated the first regression; (3) we then combined the trial and plea data, added a dummy variable to represent occurrence of a trial, and estimated a third regression; and (4) after replacing the dummy variable with a multiplier term to account for differences between sentences following trials and guilty pleas, we estimated a fourth regression.

To explain the regression technique, let:

- T prison time imposed by the court
 T* a latent variable
 X a vector of exogenous variables.
 β a parameter vector, conformable with X
 α parameters with subscripts 0 and 1.
 σ standard error
- $$T^* = \beta X + \epsilon \text{ where } \epsilon \sim N(0, \sigma)$$
- $$T = T^* \text{ if } T^* \geq 0 \text{ and } T^* \leq 360$$
- $$T = 0 \text{ if } T^* < 0$$
- $$T = 360 \text{ if } T^* > 360$$
- $$\sigma = \alpha_0 + \alpha_1 \times (\beta X)$$

The lower limit applies because prison terms cannot be less than 0.⁴⁷ The upper limit applies because sentences above 30 years have a maximum parole eligibility of ten years, and hence, sentences greater than thirty years have mostly symbolic significance.

The residual is heteroscedastic with a standard error that increases with the value of the latent variable. This specification—which is a significant feature of this model—results from an assumption that judges sentence according to a structural model $T^* = \beta X + u$, but that each judge selects B from a random vector whose elements are normally, independently, and identically distributed with

⁴⁷ Alternatively, we could have used a sequential equation model, Heckman, *Sample Selection Bias as a Specification Error*, 47 *ECONOMETRICA* 153 (1979), or a switching regression, G. MADDALA, *supra* note 8, as suggested by Klepper, Nagin & Tierney, *Discrimination in the Criminal Justice System: A Critical Appraisal of the Literature*, in *RESEARCH ON SENTENCING: THE SEARCH FOR REFORM* (A. Blumstein, J. Cohen, S. Martin & M. Tonry eds. 1983). Although the Heckman approach has a stronger theoretical justification than the approach taken here, it requires *a priori* identification conditions that could not be met.

a mean vector of β ; u is normally distributed with a mean of 0. Then ϵ is normally distributed with a standard deviation equal to $(\alpha_0 + \alpha_1 \times (\beta X))^{1/2}$. In fact, we eventually assumed that $\sigma = \alpha_0 + \alpha_1 \times (\beta X)$, which fit the data better based on the value of the likelihood function.

How to model judicial sentencing behavior is a subject that has been addressed by Rich *et al.*,⁴⁸ among others. Rich has argued that not all judges sentence according to the same structural model, and consequently, a regression equation does not truly reflect judicial decision making. Our approach, which uses slope coefficients that vary across sentencing decisions, is more general than the approach criticized by Rich. As a result, this statistical model is closer to reality⁴⁹ and less subject to the criticisms raised by Rich and colleagues.⁵⁰

Developing a model that recognizes heterogeneity in judicial decision making, and hence a heteroscedastic error term, increases confidence in the parameter estimates of the β , because limited dependent variable models are subject to bias when error terms are erroneously modeled as homoscedastic.⁵¹ Furthermore, the specification provides a more robust measure of sentence dispersion, a statistic that is crucially important to measuring sentence disparity.

⁴⁸ W. RICH, *supra* note 1.

⁴⁹ Based on our experience working with federal judges, to us it seems realistic to assume that judges disagree about the weights that should be given to different factors when sentencing convicted defendants, while these same judges agree generally about the factors that mitigate and aggravate sentences. Nevertheless, the assumption is testable by conventional significance tests on the parameter α_1 .

⁵⁰ A criticism raised by Klepper, *supra* note 47, and others is that the tobit model requires that identical factors affect both the judicial decision to sentence to prison and the length of time to be incarcerated; furthermore, parameters from the tobit model are required to have the same weights in both of these decisions. Many researchers view these assumptions as being so restrictive as to make the tobit model unsuitable for the analysis of sentencing practices.

An alternative to the tobit model is to first estimate the probability of a prison sentence and then estimate the length of time to be served after correcting for selection bias, Heckman, *supra* note 47. Although this more general model has appeal, it requires restrictive identification conditions: Variables that affect the imposition of a prison sentence cannot affect the decision about the length of prison time imposed. This restriction must be based on theory; it cannot be tested through the analysis, because a misspecified model will yield biased parameter estimates, hence invalidating the test of the null hypothesis. Theory is typically inadequate to support such *a priori* restrictions.

Nevertheless, the "bifurcated model" of sentencing (as the above approach is known) is frequently used for the analysis of sentencing decisions. Researchers who use the bifurcated approach should note that our discussion of heteroscedasticity remains valid for the bifurcated approach: parameter estimates will be biased absent a correction for heteroscedasticity. Additionally, using a multiplier to model the effects of plea bargaining can be applied when using the bifurcated approach.

⁵¹ G. MADDALA, *supra* note 8.

This model is easily extended to include a multiplier:

$$T^* = (1 + \tau)' \beta X + \epsilon$$

$$\sigma = \alpha_0 + \alpha_1 \times ((1 + \tau)' \beta X)$$

where t equals one when a trial occurs and equals zero otherwise. In effect, this model assumes that sentences following a trial are a multiple $(1 + \tau)$ of sentences following a guilty plea.

Using maximum likelihood techniques, the regressions were estimated by an iterative technique programmed in GAUSS. We performed limited pretesting. Instead, when specifying our model, we relied on earlier analysis of similar data sets,⁵² expert opinions about factors that currently enter into sentencing practices,⁵³ and an assumption that the model was additive. Of course, there is no guarantee that the resulting specification is correct.

We examined the data for collinearity. Using singular-value decomposition methods, we concluded that multicollinearity was not a serious problem.⁵⁴

Using White's test for heteroscedastic error terms, we concluded that heteroscedasticity was a problem. Our approach to overcoming this problem has been discussed.

⁵² SENTENCING COMMISSION, SUPPLEMENTAL REPORT, *supra* note 1.

⁵³ UNITED STATES SENTENCING COMMISSION, GUIDELINES MANUAL (1988).

⁵⁴ Such an approach is necessarily subjective. See D. BELSLEY, E. KUH & R. WELSCH, REGRESSION DIAGNOSTICS (1980).