Accountability and Coercion:

Is Justice Blind when It Runs for Office?

October 7, 2002
Version 3.0

Gregory A. Huber
Department of Political Science
Yale University
gregory.huber@yale.edu

Sanford C. Gordon
Department of Politics
New York University
sanford.gordon@nyu.edu

Author order is random. We thank Brandon Bartels, Kevin Eirich, Shaun Holness, and Paul Sciarra for research assistance. We would also like also to recognize the helpful suggestions of Rachel Barkow, Paul Brace, Larry Baum, Jose Cheibub, Alan Gerber, Don Green, Laura Langer, Todd Lochner, Nolan McCarty, Deborah Schildkraut, Tasos Kalandrakis, James Vreeland, and seminar participants at Yale University. Earlier versions of this paper were presented at the 2002 Midwest Political Science Association and American Political Science Association annual meetings. The staff of the Pennsylvania Commission on Sentencing, the Honorable James J. Fitzgerald of the Philadelphia Court of Common Pleas, and Joe Cairone of the Philadelphia Criminal Trial Division were exceptionally helpful in assembling and interpreting the data used herein. Gordon also gratefully acknowledges the support of the Political Science Department at the Ohio State University where he was an assistant professor for the initial phase of this project. Portions of this research were funded by generous grants from the Institution for Social and Policy Studies at Yale University and the Criminal Justice Research Center at Ohio State.
Abstract

Through their power to sentence, trial judges exercise enormous authority in the criminal justice system. In 39 American states, these judges stand periodically for reelection. Do elections degrade their impartiality? We develop a dynamic theory of sentencing and electoral control. Judges discount the future value of retaining office relative to implementing their preferred sentences, and voters are only selectively attentive to judicial behavior. Further, lacking specific information about most trials, citizens tend to assume the worst about criminal defendants and believe judges are too lenient. Our theory predicts that elected judges will consequently become more punitive as standing for reelection approaches. Using sentencing data from 22,095 Pennsylvania criminal cases in the 1990s, we find strong evidence for this effect, which is most pronounced for first term judges from liberal districts. For the cases we examine, we attribute at least 2,777 years of additional incarceration to the electoral dynamic.
I. Introduction

In the United States, only courts can authorize the incarceration or killing of individual citizens. Because appellate courts review only a tiny fraction of criminal cases, trial court judges occupy an enormously significant role in the administration of criminal justice. There are nearly 5,000 state trial court judges. In 1998, they sentenced almost one million convicted felons to a total of more than two million years of confinement in state jails and prisons (Maguire and Pastore 2002, Tables 1.81 and 5.40-5.44). Prosecutors charge and juries convict, but it is trial judges whose sentencing decisions ultimately affect the way in which this coercive element of state power is brought to bear on individual defendants.¹

In 39 American states, trial judges must periodically stand for reelection.² Do periodic elections, which exist ostensibly to assure accountability to citizen preferences, degrade the impartiality of these officials? The near-consensus among legal scholars is that this tradition – particularly in the form of partisan, competitive contests for office – is politically unassailable but insidious in its potential for degrading judicial independence (for example, Grodin 1988; ABA 1997, 2000). Indeed, judicial impartiality is a core value of Anglo-American jurisprudence. These normative concerns have received much scholarly attention, but work on the subject has yielded neither comprehensive theories concerning the relationship between trial judge behavior and the method of their selection, nor systematic empirical analysis of this relationship. Political scientists have examined the effects of the electoral connection on judicial performance, but have typically confined their attention to behavior in appellate courts.

Studying the performance of elected trial judges fuses two dominant concerns for political scientists: the coercive function of the state and the representation of citizen preferences by elected officials. This paper develops and tests a theory that specifies conditions under which trial judges will alter their sentencing behavior to improve their electoral prospects. We begin with a simple question: Why don’t elected trial judges always assign the statutory maximum sentence? Individually, judges bear little if any of the cost of excessive punishment. At the same time, there is widespread public support for more punitive sentencing, and sentencing is the most visible discretionary activity of lower court judges. Consequently, it would seem natural for a judge who desires to retain office to “throw the book” at each
convicted felon appearing before her. Doing so would seem to provide the judge with the ideal record to advertise to punitive voters in her quest for reelection. In practice, however, judges do not behave this way, and voters do not punish them for failing to do so. Sentences for the same statutory offenses vary considerably across cases and judges (Kramer and Ulmer 1996), and trial court judges retain office at a high rate (Baum 1983; Aspin 1998, 1999).

Previous research suggests that elected state Supreme Court justices take public preferences into account when making decisions (Gryski 1986; Hall 1987, 1992, 1995; Brace and Hall 1995, 1997). We seek to understand whether the same dynamics hold for less-visible trial court judges, of whose decisions a higher court will ever review only a tiny fraction. Our empirical analysis focuses on Pennsylvania, where trial court judges are first elected in partisan competitive elections and subsequently face voters in non-competitive retention elections every ten years thereafter. Unlike existing work on the effects of public preferences on sentencing behavior (Kuklinski and Stanga 1979), we examine the sentences of individual judges and model variation in responsiveness over the course of a judge’s term in office.

Because voters are at best only selectively attentive to sentencing behavior and because elected officials discount the future value of retaining office, we hypothesize that trial judges will become more punitive as their terms proceed. Further, we anticipate that the effect of electoral proximity will be diminished when judges have already succeeded in retaining office, are more conservative, and are either very easy or very hard for voters to monitor. Examining over 22,000 Pennsylvania trial court sentences for aggravated assault, rape, and robbery convictions in the 1990s, we confirm that sentences for these crimes are significantly longer the closer the sentencing judge is to standing for reelection. This finding is robust across numerous statistical model specifications and to several alternative explanations. We also find support for the first two of our three secondary hypotheses. Lacking comparative analysis of alternative systems, we withhold judgment on the relative desirability of Pennsylvania's method of judicial selection. However, we can impute a baseline estimate of the aggregate increase in prison time that occurs as a consequence of electoral incentives for the cases we examine. We can attribute at least 2,777 years of additional prison time to this electoral dynamic.
II. Judicial Performance and the Electoral Connection: Why Don’t Elected Trial Judges Always Impose the Statutory Maximum Sentence?

The method of choosing trial court judges is a matter of substantial controversy. There are obvious democratic concerns associated with removing voters from the choice of officials who will exercise enormous state authority. Judicial nominating commissions negotiating in secret might produce the worst kind of elite-level politics. Similarly, delegating the selection of trial court judges to other elected officials such as governors and legislators may undercut the very separation of powers that is supposed to render judges independent from the other branches of government. Lastly, the erosion of accountability that accompanies lifetime judicial terms may enable judges to render arbitrary or capricious decisions.

On the other hand, electing trial court judges may compromise other aspects of judicial integrity. There are four main criticisms of using elections to choose judges. First, citizens are relatively ignorant of the judicial process and individual candidates. As such, they are unlikely to pick qualified jurists. The “common man,” in other words, may choose too common a man to fill the bench, with deleterious consequences. Second, because financing judicial campaigns is costly, judges must rely on the backing of wealthy donors, creating a conflict of interest if a donor comes before the judge in a suit (Beechen 1974; Abramson 2000; Barnhizer 2001). Third, in order to secure office, judges may feel compelled to announce policy positions, thereby “politicizing” judicial elections and degrading public trust in the judiciary (De Muniz 2002). Nine states imposed restrictions on judges’ speech in judicial elections, but in June 2002 the U.S. Supreme Court ruled that these gag orders violated the first amendment.

Finally, elections may tie judges too closely to the whims of public opinion (Croley 1995). Elected judges presiding over controversial cases may base their decisions on the potential political effects of those decisions instead of legal precepts or an unbiased reading of the facts of a case. Bryce raised such concerns in *The American Commonwealth*, although he conceded (1995, 456) that examples of abuse were difficult to come by. Earlier, a concern for the potential deleterious effects of public opinion motivated Hamilton’s defense in *Federalist* 78 of the autonomy the U.S. Constitution granted federal judges. Replacing the lynch mob—and its attendant prejudices—with a judge elected by that mob
may not improve the prospects of impartial justice for parties appearing in court. Likewise, a judge looking over her shoulder when determining how a given sentence will play in the court of public opinion may offer little in the way of true judgment.

This last concern leads naturally to a puzzle. Studies of public attitudes about crime indicate that large majorities of citizens would prefer more punitive measures be taken against convicted felons. If elected trial judges value staying in office, and do not directly bear the cost of excessive sentencing, why don’t they always impose the statutory maximum? The most obvious answer is that the judges do not want to, either because of their own personal beliefs or because of some negative electoral consequence of doing so.

Even in the face of strong public pressure to sentence harshly, judges may have reason to depart from a maximally punitive model of sentencing. They may simply be more liberal than the electorate at large, either by training or, in the eyes of conservative critics, because of a proclivity for “coddling criminals.” Given that voters and conservative governors often put these judges in office, however, it is more likely that judges feel compelled by training and professional obligation to judge; that is, to make distinctions among individual defendants based on a perceived degree of culpability, remorse, or likelihood of recidivism. Alternatively, judges may make distinctions based on prejudice or stereotype. A well-developed debate in sociology and criminology concerns whether defendant characteristics such as race, gender, class, age, education, and community attachment affect judicial sentencing behavior, or whether they are merely correlated with variables a judge may justifiably take into consideration (Albonetti 1997; Zingraff and Thompson 1984; Miethe and Moore 1985; Bushway and Piehl 2001).

Lastly, proponents of the “organizational process” model of judicial decision-making note that the majority of criminal cases conclude via plea bargain. Increases in caseload, according to this view, create an incentive for overworked judges to sanction negotiated pleas more lenient than those that would be arrived at following a jury’s conviction (Landes 1971; Adelstein 1978). Also, a complex set of cooperative arrangements among police, prosecutors, defense attorneys, and judges may generate norms about appropriate punishments (e.g. Heumann 1977; Flemming, Nardulli, and Eisenstein 1992).
Voters may also recognize that there is a benefit to giving judges latitude to make decisions. Indeed, even in states where legislatures have curtailed judicial discretion by imposing mandatory minimum sentences, in no place has judicial discretion been totally eliminated. Experimental evidence suggests that citizens understand that there is variation in the level of criminal culpability and that extenuating or aggravating circumstances warrant variation in imposed sentences (Roberts and Edwards 1989). Moreover, a voter who observed a judge uniformly doling out the same sentence time after time might have reason to suspect she was shirking her responsibilities.

III. A Theory of Elections and Sentencing

To our knowledge, no systematic empirical research exists addressing the question of how standing for election shapes the behavior of criminal court trial judges. Our research builds on a larger body of theoretical and empirical work concerning the operation of the criminal justice system and electoral control of public officials generally. Scholars have long been concerned with the capacity for citizens to control government agents via the mechanism of regular elections. In political science research on the non-judicial branches, this linkage between citizen preferences and officials’ behavior has been deemed the “electoral connection” (Mayhew 1974). Building on earlier work by Downs (1957), Miller and Stokes (1963), Ferejohn (1986), and others, the core question for scholars working in this tradition is how the prospect of standing for reelection alters the behavior of elected officials.

Simply put, the literature on electoral control contends that if officials place a sufficiently high value on holding office relative to implementing their own policy preferences, they will alter their behavior to reflect the wishes of their constituencies when failing to do so would substantially decrease their chances of retaining office. Officials are presumed to care about retaining office because it provides them with policy-making authority and the perquisites of holding office. A great deal of attention has thus been devoted both to identifying the preferences of elected officials and to determining the extent and conditions of electoral control by citizens. If citizens overcome the adverse selection problem by initially electing skilled, hard-working individuals who share their preferences, then the incentive effects of subsequent elections are irrelevant (so long as citizen preferences do not change). If not, those effects
increase in importance, and citizens’ difficulties in monitoring the performance of officials may impair effective democratic control.

The electoral connection will only motivate changes in official behavior if voters can form judgments based on relevant appraisals of incumbent performance. Because information is costly to obtain, verify, and analyze, however, there may be numerous opportunities for elected officials to take actions contrary to the wishes of their constituents without being punished (Downs 1957). In light of these difficulties, numerous explanations have been offered for how citizens make voting decisions. They may, for example, simply abstain, rely on retrospective evaluations of incumbent performance (Fiorina 1981), or use partisan cues or endorsements to evaluate candidate claims and make vote choices (Lupia and McCubbins 1998). Also, challengers in competitive elections can serve as auditors of incumbent actions and claims.

With respect to citizen oversight of trial judges, it is likely that voters are even less informed of the most basic aspects of these officials’ behavior and responsibilities (Mathias 1990). Exacerbating this paucity of voter information in certain institutional settings is the lack of contextual cues like party labels (DuBois 1984). Further, until recently, many states imposed restrictions on the positions candidates could take during campaigns. Lastly, because many trial court judges run in non-competitive retention elections, voters cannot rely on challengers to provide information about candidate performance (Aspin 1988; Volcansek 1981).

Likewise, there is little research on the relationship between voter preferences and trial judge behavior. Gibson (1980, 365-367) demonstrates that elected trial judges who have experienced defeat at the polls are more responsive to the sociopolitical characteristics of their districts than undefeated ones. Kuklinski and Stanga (1979) examine the effect on aggregate judicial sentencing in different counties of variation in voter support for a California referendum endorsing less stringent penalties for certain drug crimes. They find that sentences diminished in areas where voters were more supportive of the referendum. Their analysis, however, cannot distinguish whether observed changes in sentences accompanied changes over time in public preferences, incumbent judges’ preferences, or the composition
of the bench.

In general, static models of electoral accountability that relate voter preferences to the performance of elected officials face two challenges. First, the models suggest that if elected officials represent their constituencies, it is because they fear losing office. Unfortunately, testing this claim is difficult because we almost always lack independent measures of voters’ and officials’ preferences. Consequently, perfect ex post control is indistinguishable from success in choosing likeminded officials ex ante. Second, static models suggest that when elected officials do deviate from the wishes of their constituents, it is because either the value to officials of retaining office is insufficient or voter review of their behavior is imperfect. As such, they offer an incomplete view of the nature of voter and incumbent interaction, failing to account for the fact that both the weight incumbents place on winning reelection and voter attentiveness to officials’ performance are likely to vary over time. Fortunately, by employing a dynamic model of representation, one can simultaneously address this second problem while providing a (partial) solution to the first.

In a dynamic model of representation, elected officials do not make a one-time choice about how much to heed the wishes of their constituents. Rather, over the course of their terms, these officials continuously reevaluate the balance between their own preferences and electoral concerns. There are two reasons elected trial judges would do this. First, they may discount the future value of retaining office. At the beginning of their terms, when the need to secure reelection and retain office is a far-off prospect, judges will place relatively high value on imposing their most preferred sentences given their own ideologies and the circumstances of individual cases. (In fact, at this stage of their terms they may not yet have decided whether they will run again.) Toward the end of a term, however, retaining office in the not-so-distant future becomes a paramount concern. Later in their terms, therefore, judges will give more weight to voter preferences when assigning sentences.

Second, because elections are only periodic, voter evaluation of candidate performance is likely to be temporally proximate to each election (Popkin 1991). As such, judges may ignore citizen preferences when voters are inattentive. If a challenger is attentive to an incumbent’s performance over
the course of her entire term, then (ignoring the judge’s discounting of the future) competitive elections may be sufficient to induce uniform electoral control over the course of a judge’s tenure. In many states (including Pennsylvania), however, judges preserve office via retention election. Because there are no challengers to audit incumbent performance, the problem of selective attention is likely to magnify variation over the electoral cycle.10

Recent empirical research along these lines offers suggestive results. For instance, there is evidence that Senators’ votes become more representative of constituency preferences over the course of their 6-year terms (Thomas 1984; Wright and Berkman 1986; Bernstein 1991; Ahuja 1994; Treier 2000). Likewise, the decisions of elected state Supreme Court justices appear more reflective of voter preferences when the justices’ terms are shorter (Brace and Hall 1997).

What are voter preferences about the sentencing behavior of trial court judges? As mentioned earlier, research on public opinion regarding criminal justice issues consistently reports large majorities of both liberal and conservative citizens favoring harsher penalties for convicted felons. One interpretation of these findings is that voters are more punitive than judges. An experiment by Roberts and Edwards (1989), however, demonstrates that voters’ punitive tendencies may also be a consequence of their informational environment. When randomly selected respondents were shown a newspaper account of an assigned sentence, they almost uniformly preferred a more punitive sentence. In contrast, among respondents asked to read a more detailed account of the courtroom proceedings in the case, a much smaller proportion believed the assigned sentence was too lenient. Similar results are reported by Roberts and Doob (1990).

One possible source of this bias emerges from citizen access to information about criminal justice mistakes.11 A perfectly informed voter might conclude that some sentences are overly punitive, and others overly lenient. Real voters, however, are asymmetrically informed about over- and underpunishment. Nearly all convicts claim their punishment is too severe, and newsworthy cases of wrongful convictions by definition come to light only years after their occurrence. Underpunishment is more easily observed. News accounts of recidivism make voters aware of convicts who committed additional crimes after
(seemingly) brief periods of incarceration. An asymmetry of this sort could motivate even moderate or liberal voters to assume the worst about defendants and judges.

If incumbent judges continuously reevaluate the balance between the value of office and implementing their own preferences (either because they discount the future or because voters are selectively attentive), the balance will shift increasingly toward satisfying constituents as election approaches. Further, if satisfying voters entails more punitive sentencing (because voters are innately more punitive, lack trial-specific information about extenuating circumstances, or learn only about overly lenient sentencing), then judges will “moderate” their sentencing behavior by becoming more punitive the closer they are to standing for reelection. This is the primary hypothesis we test below.

IV. Trial Courts in Pennsylvania

In order to test our hypotheses concerning the effects of the electoral connection on sentencing behavior, we linked information about sentencing in criminal courts, state elections, and judges’ backgrounds in Pennsylvania. Conventional wisdom concerning different methods of electing trial judges suggests that the pernicious effects of the electoral connection will be minimized when judges serve long terms and are retained via noncompetitive, nonpartisan elections. Pennsylvania fits these criteria, and thus constitutes a difficult case for our theory of electoral control. If elected judges do take voters into account when sentencing, they are least likely to do so in this setting. (We discuss the effects of the selection mechanisms used by other states in the last section of the paper.)

Selection to the Pennsylvania Trial Courts

The general jurisdiction trial courts in the state of Pennsylvania are the Courts of Common Pleas. When a judgeship vacates, replacements are selected via a partisan competitive election. (Mid-term vacancies are filled via gubernatorial appointment; the seat is considered open in the subsequent election.) In the primary election, judges compete for one or (usually) both of the major party nominations. In the general election, the top vote getter(s) will fill the one or more open seats in a particular judicial district. Judicial districts correspond to counties, although in sparsely populated areas some judicial districts encompass two counties.
Once elected, judges stand for reelection every ten years on the basis of a non-competitive retention vote. Retention elections are staggered and occur in November of odd years. As in other states, the vast majority of judges who stand for retention succeed. In fact, of the 223 Common Pleas retention elections from 1991-1999, there were only five cases of judges failing to secure greater than 50 percent of the vote. By itself, however, the high rate of incumbency among Common Pleas judges is evidence neither of judicial autonomy stemming from voter ignorance nor of faithful agency on the part of judges. Members of the House of Representatives also win reelection at high rates. Further, voters may know little about the day-to-day proceedings of trial courts in their counties, but this may itself be a consequence of judicial compliance with voter preferences.

Sentencing in the Pennsylvania Trial Courts

As in all states, the manner in which criminal cases wind their way through the judicial system in Pennsylvania is enormously complex. Consequently, a full accounting of the intricacies of the criminal justice system and the interrelated and strategic behavior of each actor in this system is impractical in the current context. (We point out some implications of the strategic behavior of prosecutors and defendants in the paper’s final section.) Here, we discuss the range of options available to judges once a defendant pleads guilty or a jury (or judge in a bench trial) finds a defendant guilty of a misdemeanor or felony.

Despite the existence of a sentencing commission that offers sentencing guidelines for particular crimes, Common Pleas judges generally exercise enormous discretion in imposing sentences. There are several exceptions to this general pattern. First, by statute, certain crimes carry mandatory minimum sentences. For instance, in second-degree murder cases, the mandatory minimum sentence is life imprisonment. In first-degree murder cases, the law requires a jury, in a special sentencing hearing, to consider whether the circumstances of the case warrant a life sentence or death by lethal injection. Other crimes with mandatory minima include certain drug infractions, criminal acts resulting from drug transactions, driving under the influence, and the use of illegal bullets in committing or attempting violent crimes (Commission on Sentencing 1997). Second, the judge is constrained by statutory maximum sentences. Respectively, the sentences for most first-time level one, two, and three felonies are capped at
twenty, ten, and seven years of imprisonment, although these maxima increase for repeat violent offenders (42 Pa.C.S. § 9714).

In addition to these statutory provisions, the Pennsylvania Commission on Sentencing (PCS) offers voluntary sentencing guidelines for most felonies and misdemeanors. The PCS is an agency of the state legislature composed of four legislators, four judges, and three gubernatorial appointees. Judges are obliged to take account of PCS instructions, but are not required to abide by them. The Pennsylvania guidelines work as they do in many other states: PCS classifies crimes by offense gravity and defendants by prior record. Given those two variables, a judge can determine the recommended sentencing range by referring to a sentencing matrix.

Suppose, for example, a defendant is found guilty of a single count of aggravated assault resulting in significant bodily injury. This offense has a “gravity score” of eleven points. Presume also that the defendant has a prior record score of four. (The prior record score weights serious offenses more heavily than misdemeanors.) According to the guidelines released in 1997 and currently in use, the prescribed sentencing range is then 60 to 78 months of incarceration. The guidelines expand the recommended penalty range upward or downward by twelve months in the presence of aggravating or mitigating factors. Additional matrices exist for separate sentence enhancements. For example, if the same crime were committed using a deadly weapon, the recommended sentencing range would jump to between 78 and 96 months in prison.

The stated purpose of the sentencing guidelines is consistency and proportionality of punishment. Nonetheless, even cursory examination of contemporary Pennsylvania sentencing data reveals numerous departures from the recommended PCS guidelines. There are a number of reasons for this pattern. First, defendants are often convicted of multiple crimes. If an individual is convicted of four counts of theft, for example, the presiding judge chooses whether to impose the sentences for each crime concurrently or consecutively. If the sentences will run consecutively, the judge can lower each individual sentence. Alternatively, the judge can simply issue a sentence for the most serious count of the conviction that incorporates the punishment for the lesser counts. Second, as in all state courts, most cases are settled via
plea bargain. Prosecutors and defendants in such instances negotiate a settlement, *subject to the approval of the presiding judge*, whereupon the defendant pleads guilty to reduced charges or in exchange for a recommendation to the judge by the prosecutor of a reduced sentence. Third, the judge may simply believe the guidelines are inappropriate given the particular circumstances of the case.

For a given conviction, sentencing judges in Pennsylvania hand down both a minimum and maximum sentence. In cases involving incarceration, the defendant is obliged to spend at least the minimum term in prison before becoming eligible for parole. After this term, a state parole board may or may not grant the defendant parole up to the release time specified by the judge as the maximum sentence. (According to personnel at PCS, defendants rarely go before the parole board at the immediate conclusion of their minimum terms.) Overall, the manner in which Pennsylvania incarcerates and releases defendants falls somewhere between the extreme case of fully indeterminate sentencing (in states where a parole board is granted large discretion to reduce judicially imposed sentences) and fully determinate sentencing (in states where parole has been abolished.) Thus, a convicted defendant can expect to serve a sentence somewhere between the minimum and maximum imposed.\(^\text{14}\)

V. Data and Method

The details of the Pennsylvania criminal justice system suggest a need to avoid several pitfalls in our analysis. First, we must account for a judge’s discretion in a given case. We restrict attention to a class of felonies for which judges both always have some discretion in sentencing and usually assign prison time. There are a number of such felonies, including rape, sexual assault, arson, robbery, theft, and possession with intent to distribute Schedule I and II narcotics. To keep the analysis as simple as possible, we focus our attention on all convictions in which the highest count was some form of aggravated (felony) assault, robbery, or rape. These encompass nearly all cases with high offense gravity scores under the Pennsylvania guidelines.\(^\text{15}\) We have 22,095 observations for discretionary sentences imposed from 1990 to 1999 according to guidelines issued in 1988, 1994, and 1997.\(^\text{16}\)

As noted above, judges assign two sentences for each case. The dependent variable in our analysis is the smaller of these two quantities, measured in months of incarceration. (Summary statistics
for model variables appear in Table 1.) This represents the determinate portion of the judge’s discretion over sentencing, as defendants must spend at least the smaller sentence behind bars before becoming eligible for parole. We do not consider the larger sentence the judge imposes. We interpret this quantity not so much as punitiveness on the part of the judge, but as a delegation of authority by the judge to the state parole board.\(^{17}\)

**Table 1 About Here**

By statute, the smaller sentence imposed by the judge cannot exceed one-half the larger sentence, which itself cannot be greater than the statutory maximum. Additionally, for certain crimes, statutes mandate a minimum prison sentence. Together, these rules place upper and lower boundaries on the range of a judge’s sentencing options, creating a censoring problem. When a judge’s smaller assigned sentence is the statutory minimum, the dependent variable is left-censored. She may have preferred an even lower sentence, but was forbidden by law from imposing it. Similarly, when a judge’s smaller sentence is one-half the statutory maximum, the dependent variable is right-censored; the law prevents her from being more punitive. OLS regression produces biased coefficient estimates in the presence of censoring. In order to compensate for censoring problems while retaining the OLS assumption of normally distributed errors, we employ a two-limit tobit model with observation-specific left and right censoring points (Tobin 1958; Maddala 1983, 160-162).\(^{18}\) Employing this model also allows us to address a second problem created by the 16% of cases in which no prison time was imposed. In these cases, defendants were placed on probation, forced to pay a fine, or given some other form of limited restrictive punishment. We treat these cases as left-censored, assuming they represent punishment less than the minimum jail time.\(^{19}\)

The third issue we confront is that factors other than electoral proximity and statutory limits may explain assigned sentences.\(^{20}\) Crimes, defendants, and cases vary independently in ways that will affect judges’ use of their discretion. The Pennsylvania Sentencing Commission’s recommended minimum and maximum sentences provide the best measure of the severity of the offense committed and the defendant’s prior criminal record. These reflect consensus within the state about appropriate punishments and the latitude judges should enjoy in particular cases.\(^{21}\) Additionally, we employ dummy variables for
the applicable sentence guideline regime (1988, 1994, or 1997).\textsuperscript{22} As supplementary controls for the nature of particular crimes, we employ indicator variables that distinguish the type of crime (rape and robbery – the baseline category is aggravated assault) and whether it involved the possession or use of a deadly weapon. Finally, we account for the possibility that judges distinguish among defendants based on demographic characteristics. These include age and age squared (because judges may treat young and old defendants more leniently than others), race, and sex.

We also control for variation in the disposition of cases. In 51.5\% of the cases in our sample, the defendant was convicted on more than one count. (In only 14\% of cases was the defendant convicted on more than three counts.) As stated above, judges can decide whether to impose sentences consecutively or concurrently, and which counts to issue sentences on. To circumvent this issue, we examine only the sentences associated with the most severe count on which the defendant is convicted, and control for the number of counts. For many cases, the most severe count is the only one accompanied by a sentence, and the most severe count virtually always has an associated sentence (we omit from the sample the handful of cases in which the defendant was given prison time only for less severe counts).\textsuperscript{23} Finally, we include indicator variables for negotiated and non-negotiated guilty pleas (the baseline category is conviction at trial). We remain agnostic about the expected effect of case disposition on assigned sentences, although we discuss in greater detail below what we learn from its observed effect.

Our primary hypothesis concerns the effect of electoral proximity. We code proximity as the number of days elapsed in the judge’s term at the date of sentencing divided by 3,653. The measure is thus scaled from zero to one, with zero representing ten years until the next election, and one an imminent retention vote (Election Day).\textsuperscript{24} We expect a positive coefficient on this measure: As proximity increases, so should assigned sentences. Three other variables, whose relevance we explain below, appear in the summary statistics. One is a measure of district political conservatism on criminal justice issues. Lacking a perfect measure, we employ the district Republican share of the two-party vote in the previous statewide attorney general race. The second is the number of terms the judge has served (including the current one). The last is the number of active judges in the district in a given year.
VI. Results

The Effect of Electoral Proximity

First, we are interested in the independent effect of electoral proximity on sentencing. Table 2 displays the results of the first round of tobit estimates. The coefficient estimates in column (1) come from a regression that includes the electoral proximity variable and the controls discussed above. The estimates in column (2) are from a regression that includes all the variables from the first specification, in addition to a vector of judge- and year-specific fixed effects whose coefficient estimates we do not report. This second specification provides a test of robustness, permitting us to control for all static characteristics of the sentencing judge (e.g. ideology) as well as global changes in sentencing practices over time (e.g. those caused by uniform responses to changing state conditions).  

Before proceeding to our test of the primary hypothesis, we examine the parameter estimates for the control variables. The results are without exception consistent with our expectations and prior quantitative research on criminal sentencing cited above, and suggest the validity of our specification and coding. Increases in the guideline sentences lead to increases in actual sentences. A non-monotonic relationship exists between age and length of incarceration (Per the estimates from the first specification, 21 year-old defendants can expect to see the most prison time.) Men receive longer sentences than women, and the specification in column (1) suggests that all else equal, judges hand down sentences 11 days longer for nonwhite defendants than for their white counterparts. One of the strongest predictors of additional punishment is possession of a deadly weapon. Interestingly, once one controls for possession of a weapon, using it does not significantly add to the sentence.

We also point out an interesting finding concerning the disposition of cases. The negative coefficient on the negotiated plea variable provides information about the way that plea-bargaining typically occurs (Taha 2001). A positive coefficient would suggest that the dominant form of negotiation between prosecutors and defendants concerns charge reduction: A defendant might plead guilty to a lesser charge and receive a penalty that, while stiff for the reduced charge, is nonetheless lighter than the penalty
for the higher count. The negative coefficient that we observe across specifications suggests that negotiation typically concerns the length of the sentence that the prosecutor recommends to the judge given a particular charge, more often than on the charge itself.

Next we consider the effect of electoral proximity. It is important to think carefully about how one should interpret the null hypothesis for this variable. One might argue that a failure to detect dynamic moderation suggests judicial autonomy. Actually, autonomy is one of two possible explanations consistent with a null finding. Judges may behave in a manner totally independent of the preferences of their constituents. Alternatively, a null finding could indicate judicial subservience: Judges may prioritize the desires of their constituents throughout the electoral cycle.

We need not confront this indeterminacy, however. In both specifications, the parameter estimate for electoral proximity is positive and highly statistically significant (one can reject the null hypothesis at above the 0.001 level in both specifications). All else equal, the sentence imposed by a judge whose election is imminent is likely to be about three to 4 ½ months longer than if the judge were recently elected or retained (depending on specification). A standard deviation shift in electoral proximity raises an assigned sentence by about 25 to 37 days. At first glance, this would seem a trifling difference. But, if one considers that the median sentence in the sample is only about 12 months, the magnitude of the result is substantial. A standard deviation increase in electoral proximity produces an increase of 7 to 10% from the median sentence, while a change from zero to one produces an increase of 24 to 37%.

An even more useful measure of the impact of electoral proximity can be derived by imputing an estimate of the aggregate increase in prison time stemming from judges’ desire to secure reelection. Assume that on the first days of their terms, judges feel completely unconstrained by the electoral consequences of their sentencing decisions. For each sentencing decision in our dataset, we can calculate an estimate of the sentence the judge would have imposed had she just been elected or retained, and compare that with an unconstrained prediction. Adjusting for statutory constraints on sentencing, the coefficient estimates in column (1) suggest that the proximity effect augmented sentences for the cases in our dataset by 2,777 years (+/- 1,328), or 6.1% of total prison time. For two reasons, this is a
conservative estimate. First, even a judge serving her first day in office may not feel completely unconstrained from the future electoral consequences of her actions. Given lifetime tenure, she might sentence even less punitively. Second, these figures correspond only to the cases in our Pennsylvania dataset, which comprise only a fraction of total convictions in the state.

We wish to be clear about what these numbers do and do not represent. They do represent inconsistency in sentencing, as similar defendants will be subjected to different sentencing criteria for electoral reasons having nothing to do with the technically relevant circumstances of their cases. It is tempting to view these figures as evidence that judges, in an effort to pander to voters, sentence unduly harshly toward the ends of their terms. It is also possible, however, that judges starting new terms are sentencing too leniently. Absent further information or a model of the social costs and benefits of incarceration, it is impossible to distinguish between these two potential failings.

*When Does Electoral Proximity Matter More?*

While the proximity effect is the core prediction of our dynamic electoral control model, it is unlikely that the magnitude of this effect will be the same in all circumstances. Fortunately, the panel nature of our dataset (which includes observations across many judges and districts over ten years) allows us to test several hypotheses about the conditions under which proximity is more or less likely to matter. We estimate several statistical models to test these hypotheses. Parameter estimates appear in Table 3.

**Table 3 About Here**

First, it may not be the case that all judges facing imminent retention votes react by ratcheting up sentences in the same way. There are three reasons to suspect that judges will respond to electoral proximity less after they have been through the process at least once. The first is the length of the judicial term in Pennsylvania. Our biographical data on 494 judges (assembled from multiple volumes of the *Pennsylvania Manual*) demonstrate that the median judge achieves her position at age 45. In Pennsylvania, terms last ten years. Half of all judges, then, run for retention a second time at age 65 or older. Since retirement is a likely option, the value of retaining the office will surely have declined by that point. Second, judges learn over time what they must do to be retained. A comfortable retention margin
the first time around may prompt a decrease in concern about appearing soft on crime. Third, over multiple
elections, a judge may develop a reputation with voters, such that the marginal impact of a very experienced judge’s sentences on voters’ evaluations of her is relatively small.

To test this hypothesis we reproduced the specifications in Table 2, this time including a measure of how many terms the judge has served and an interaction of the proximity variable and a dummy variable representing whether a judge was in her first term. The coefficients from this estimation, which includes judge and year fixed effects, are shown in column (1) of Table 3 and confirm our hypothesis. The estimates suggest that electoral proximity has a positive effect on sentencing, but that this effect is only statistically significant in the judge’s first term. A judge whose first retention election is imminent imposes sentences about 3.1 months longer than a judge who was just elected for the first time. This coefficient estimate is only slightly larger than the one reported in column (2) of Table 2, suggesting that a great deal of the aggregate effect of electoral proximity is associated with first-term judges. Once a judge has been retained for the first time, however, the effect of electoral proximity is tiny and indistinguishable from zero. We also note that the number of terms a judge serves plays an insignificant role in explaining the sentences she imposes.

Our second hypothesis is that electoral proximity is unlikely to play a large role in judicial sentencing when judges are punitive to begin with. If we assume that on average more conservative counties tend to select more conservative judges, then jurists there are already inclined to sentence punitively over the entire course of their terms. (They may also find upper statutory limits on sentencing binding more frequently.) Consequently, we anticipate a negative interaction between county conservativism and proximity. Column (2) of Table 3 provides support for this hypothesis. Our measure of county ideology on criminal justice issues is the Republican share of the two-party vote in the most recent statewide attorney general race. The coefficient on electoral proximity remains positive in this specification, while the coefficient on its interaction with Republican vote share is negative, suggesting that the proximity effect decreases in more conservative counties.

The average Republican share across counties is 58.1%. In a hypothetical average county, moving
from the day after election to imminent election actually reduces the expected sentence, but only by a statistically insignificant eleven days. In a county one standard deviation more liberal than average, such a move increases the sentence the judge would impose by 2.66 months (±1.09). In a county one standard deviation more conservative, the effect falls to -3.37 months (±1.85). This suggests that in the most conservative counties, judges actually reduce their sentences over the course of the electoral cycle.  

We now turn to our final supplementary hypothesis. Suppose electorally-induced sentencing moderation is attributable primarily to selective attention on the part of voters and not variation in the present value to the judge of retaining office. If this is the case, we would anticipate that electoral proximity would matter least when judges are either very easy or very difficult to monitor. If very easy, then electoral incentives will bind the judges as fully faithful agents over the entire course of their terms. If very difficult, then judges will be fully autonomous over the entire electoral cycle. Only when judges are somewhat difficult to monitor will the proximity dynamic we describe play itself out. 

Of the many factors that influence ease of monitoring, one is surely the number of judges whose behavior the voter must consider. In areas with a single judge, local news coverage of sentencing is easily attributed to a single candidate on a ballot. When judges are numerous, however, voters would have to be spectacularly well informed to discern the behavior of an individual judge. Because our hypothesis suggests a non-monotonic relationship between proximity and ease of monitoring, we employ a quadratic specification of the interaction. Specifically, we create two new variables, one in which proximity is interacted with the number of judges in the county and one in which it is interacted with the number of judges in a county squared. A negative coefficient on the second interaction would suggest a concave downward relationship in which the effect of electoral proximity was maximized at some intermediate number of judges, thus lending support for our hypothesis. 

Parameter estimates appear in column (3) of Table 3. Contrary to our expectations, the estimated coefficient for the interaction between proximity and number of judges squared is positive, which would indicate that the effect of electoral proximity is concave upward in the number of judges in the county. A judge serving alone in a county assigns almost the same sentence when election is imminent
as when just elected. As the number of judges increases toward eight, an impending election is associated with a sentence that is almost 0.5 months shorter than in the first year of the judge’s term. (The minimum effect of proximity is observed when there are eight judges in the county.) In this specification, only when the number of judges in a county reaches 16 is an imminent election associated with a longer sentence than at the beginning of a judge’s term.

There are three possible interpretations of this unexpected finding. The first is that many factors influence how easy or hard it is to monitor a judge. Our model does not capture all of these. Further, there is little reason to suspect that the presence of only one or two judges in a county reflects “easy” monitoring. Second, it is conceivable (though unlikely) that voters actually pay attention to judicial behavior over the entire course of their terms in small counties. Because the combined effect of proximity is very small until the number of judges in a county exceeds about 20, our findings might indicate that judges are bound relatively tightly over the course of their entire terms in smaller counties. In larger counties, where comprehensive auditing of sentencing behavior over an entire term is impossible, judges need only moderate as election nears. Third, voters may be selectively attentive in all cases, but may only have the media resources necessary to monitor judicial behavior in larger counties like Philadelphia and Allegheny (Pittsburgh). If this is the case, judges in small counties can “shirk” over the course of their entire terms, while those in these more urban counties are compelled to moderate over time.

One issue of concern in these last two models is the extremely high negative correlation (-0.88) between the number of judges in a county and the Republican vote share measure. This association is due almost entirely to the fact that 42% of our cases come from a single county (Philadelphia). To distinguish between the effects of these two variables, we combined the specifications in columns (2) and (3) of Table 3 in a single model, whose parameter estimates are displayed in column (4). This model simultaneously controls for the number of judges and district conservatism, and includes year and county fixed effects. The statistical significance and signs on the relevant coefficients remain unchanged, although the magnitudes differ somewhat. This specification suggests that electoral proximity would matter least in a county with about 13 judges. Further, the change in the effect of electoral proximity in
moving from liberal to conservative counties is diminished.

Assessing Alternative Explanations

Three plausible objections may be raised against our findings. First, perhaps judges simply “learn” to become more punitive as they grow older and more experienced. To the extent that such an effect might manifest itself over the entire career of the judge, we offer two responses. One is that this alternative story cannot explain our second-order finding concerning the dominance of the proximity effect among first term judges. The second response is to re-estimate the models from Table 2, controlling for judge age and experience. In both specifications, the magnitude of the electoral proximity effect is slightly diminished, but remains positive and statistically significant (3.54 months in the non-fixed effects model and 2.88 in the fixed-effect specification).

To rule out the possibility that learning is strictly a “new judge” phenomenon, we can distinguish learning early in a judge’s tenure from the proximity effect. We created a new variable equal to the electoral proximity variable for the first five years of a new judge’s term, and zero afterwards. Adding this “early first term learning” measure to the models reported in Table 2, the coefficients on the pure proximity variable remain positive and statistically significant (4.36 in the non-fixed effects specification and 2.99 in the fixed-effect model). In other words, despite the fact that judges may be becoming more punitive early in their first term in office, this effect does not drive our result.

Second, any non-random assignment of cases to judges may bias our results. In particular, it may be that less experienced judges hear less serious cases. Before considering case assignment methods directly, it is important to remember that our basic statistical model accounts for variation in case seriousness and judicial discretion by controlling for the guideline sentences. Thus, even if seriousness were correlated with electoral proximity, this would not by itself produce the results we observe.

We also investigated how cases are actually assigned. In small counties with only one or two judges, this is not a concern. However, selection may play a role in large counties such as Philadelphia or Allegheny. We contacted the Administrative Office of the Philadelphia Courts to inquire about case assignment methodology. In Philadelphia, cases are divided into three pools: homicides, other major cases
(which encompass those that we consider), and minor (“list room”) cases. During their first two to three years on the bench, judges are generally restricted to list room cases. After that period, judges are randomly selected to trials, with slightly reduced selection probabilities afforded to historically slow judges with large open caseloads. Further, the test conducted above allows us to account for potential easy-case bias early in a judge’s tenure. Even if we disaggregate the proximity effect into the first few years in office and overall proximity, our results still hold.

Third, perhaps additional strategic considerations on the part of defense attorneys and prosecutors influence our results. Suppose one side or the other anticipated the judge’s electoral concerns and altered its strategy accordingly. Defense attorneys might seek to delay sentencing until after an election, whereas prosecutors might seek to accelerate it. If defense tactics in this regard dominate the sequence, this would bias against our findings, as judges would have fewer cases with which to demonstrate punitiveness toward the ends of their terms. If prosecutor tactics dominate, this would contribute to our finding, but not reject our basic story. Such anticipative action would still constitute evidence that judges take electoral considerations into account.

VII. Conclusion

The foregoing analysis provides insight into two important areas of concern for political scientists: the state’s use of coercion and the nature of representation. Judges assign sentences to convicted criminals, thereby determining in part how governments use their authority to deny liberty. We show that judges’ sentencing behavior varies over the course of their terms, and that judges become significantly more punitive the closer they are to standing for retention. In Pennsylvania, for the ten years and 22,095 cases we analyze, we can attribute about 2,777 years of additional incarceration to impending election. This may suggest that judges sentence too harshly near elections in an effort to pander to voters, or that they sentence too leniently early in their terms with resultant social costs. In either case, the finding suggests a troubling downside to the electoral control of judges. Because judges’ power to incarcerate is applied to individuals on a case-by-case basis, we can attribute substantial inconsistency in the exercise of this power to this electoral connection.
Critics of judicial election might be tempted to seize on our results to justify removing judges from direct citizen review. It is not clear, however, that the same phenomena would not occur with any other form of periodic review. Legislative or gubernatorial reappointment might also lead judges to alter their decisions in order to retain office. Moreover, as we discuss in the paper’s second section, there may be larger costs associated with leaving the selection of judges to others. The electoral connection may have pernicious effects on consistency, but for some, this may be an acceptable side effect to ensuring that judges’ decisions are at least partially representative of citizen preferences.

Our research also provides insight into the nature of the relationship between citizen preferences and the behavior of elected officials. Had we found no effect of electoral proximity on sentencing behavior, it would have been difficult to discern whether judges already agreed with their constituents, were so tightly bound by the electoral connection that they dared not shirk, or were free to sentence as they saw fit without any consequence. Instead, we demonstrate that elected judges in Pennsylvania appear to be bound by the (weak) threat of losing office, and alter their behavior in an effort to appeal to voters. This effect is strongest for first-term judges from more liberal districts. (The effect of ease of monitoring is uncertain.) We can thus say with near certainty that elections are not merely a method for picking “good types.” The sentences handed down by the elected judges we study appear to depart from those they might assign under different selection and retention systems. From the perspective of voters, elections also provide a method to create incentives for incumbent officials to change their behavior, even though few judges lose office.

Our analysis is a first empirical cut at what is in reality a very complicated problem. One characteristic of criminal justice data analysis that is at once challenging and exciting is the enormous complexity of the system generating the observed outputs. A primary source of this complexity is the opportunity for strategic behavior on the part of all actors involved. Here, we discuss some of the issues surrounding our analysis in light of this strategic complexity. First, our disposition-related control variables are endogenous to a bargaining game between prosecutors and defendants that is reviewed by the presiding judge in a criminal case. In other words, whether a case is disposed of via a plea bargain,
non-negotiated plea, or jury trial is not randomly assigned and fixed in repeated samples, but rather
determined in part by unobservable features of the particular case that affect the likely sentence as well. A
more comprehensive model would examine sentences as equilibrium outcomes to this bargaining game.
More generally, it may be the case that negotiated settlements do not “cause” reduced sentences in a
meaningful way, but rather, that the prosecutor’s poor case, which would eventually manifest itself in a
reduced sentence, “causes” negotiated settlements. Insofar as the factors that influence bargaining (apart
from electoral proximity) are unlikely to vary systematically over the course of a judge’s term, they
should not upset our basic results.  

Second, we referred earlier to the role that judges play in the state of Pennsylvania through the
maximum sentence they are authorized to impose. The judge’s discretion in imposing the maximum
sentence is not absolute, as it must fall within statutory limits and be no less than twice the size of the
minimum. However, given the role of the state parole board, judges have effectively been granted the
authority to set its discretion. This represents a fascinating problem of delegation in a hierarchical setting:
the state legislature effectively delegates discretion to the trial judge to delegate discretion to the parole
board. How the composition of the board in relation to the preferences of the judge influences the latitude
the board is granted is a complicated question deserving further attention.

While judges grant discretion to the parole board, they also make decisions in an institutional
environment where limits have been placed on their behavior. Both mandatory statutory prescriptions and
voluntary sentencing guidelines constrain the sentences judges may assign. In states like Pennsylvania
where judges are elected locally, state officials must account for likely local variation in the preferences
of judges chosen by very different constituencies. Thus, it is important to recognize that state statutes are
most likely related endogenously to the particular institution of judicial selection. This provides an
important caveat for those intending to conduct cross-institutional research on the behavior of judges.
Any system that brings with it possibilities of judicial deviation from the preferences of legislators and
governors may lead these officials to impose compensating ex ante restrictions on judicial discretion.  
Further, recent reform of judicial selection mechanisms suggests that those methods are malleable as well.
This research also sets the stage for more extensive inquiry into the comparative politics of judicial selection. Even in the low information setting created by non-partisan retention elections, and despite the ten year terms that afford judges significant distance from electoral review, Pennsylvania judges appear to respond to the potential electoral consequences of sentencing leniently by becoming more punitive as reelection approaches. At the very least, we can conclude that the retention method does not remove politics from the sentencing process.

No system of selection is perfect, but at present we know little about the tradeoffs associated with mechanisms other than the one studied here. Future work should build on this project to compare across term length, informational environment (competitive versus non-competitive, partisan versus non-partisan), and immediate political principals (voters, governors, legislatures), variation in the types of judges selected, and the incentive effects of subsequent review. For example, how does running for reelection in a partisan race alter the incentives of judges? One might initially believe that the proximity effect would be more pronounced under competitive races, as securing reelection is more difficult. On the other hand, if the challenger plays the role of auditor, reviewing the incumbent’s entire sentencing portfolio, then the judge has reason to moderate her sentencing behavior upward over her entire term. Parsing out these effects requires additional theoretical and empirical work.

This, of course, is not an easy task. Gathering data about the behavior of trial court judges is time consuming and expensive. Understanding the restrictions (formal and informal) placed on judicial discretion is similarly complicated. Comparative analysis of electoral systems is made possible, however, by the enormous institutional variation afforded by the American states, which serve in this regard as institutional “laboratories,” to borrow Louis Brandeis’ well-known metaphor. This variation permits us to test theoretically derived claims about responsiveness, representation, and fairness in different settings while holding fixed the broad contours of the legal system. Further, differences in the methods employed to select trial judges will contribute to our understanding of how the techniques used to choose and subsequently reevaluate other elected officials influence their behavior.

Endnotes
For several reasons, trial judges retain broad discretion over sentencing in nearly all criminal matters. First, defendants can appeal sentences only on very narrow grounds of law. Second, even though most cases are resolved via plea bargain, the presiding judge must approve the proposed sentence. The main exception to judge's sentencing authority concerns the administration of the death penalty. In June 2002, the U.S. Supreme Court mandated, in *Ring vs. Arizona* (No. 01-488), jury input in sentencing for capital crimes. (Jury input had already been required in the vast majority of states prior to the decision.) This paper does not consider civil cases, where, depending on context, judges or juries have the power to assign damage awards (See Tabarrok and Helland 1999).

There is substantial institutional variation in this regard. Incumbent trial judges run in competitive partisan elections in eight states, in competitive non-partisan elections in twenty-one states, and in non-competitive retention elections in ten states. In a retention election, the incumbent judge’s name appears on the ballot with no opponent listed. She must receive more “yes” than “no” votes in order to keep her position. In seven other states, the legislature, governor, or a judicial nominating commission periodically evaluates incumbent judges for retention.


Averaged across years, in the 1990s roughly 85% of American respondents to the General Social Survey answered “not harshly enough” when asked whether they thought local courts dealt too harshly or not harshly enough with criminals. Conservatives are only slightly more likely than liberals to take this position (88% versus 80%).

A substantial and growing literature addresses the behavior of elected state Supreme Court justices (e.g. Brace and Hall 1997.) Additionally, the financing of races for trial and appellate court positions has garnered recent attention (See Thielemann 1993, Kaplan and Davidson 1998, and Hansen 1998).

The theoretical literature on electoral control is vast. See, for example, the seminal article by Ferejohn 1986; as well as, more recently, Austen-Smith and Banks 1993; Banks and Sundaram 1993; and Landa
and Hafer 2001.

8 Of course, this is necessary but not sufficient to induce electoral control. Elections must also be fairly administered, and incumbents must believe viable replacements exist.

9 For an able exception, see Gerber and Lewis 2002. For efforts to discern the difference between pure and induced ideology in the study of Congress, see Cox and Poole 2002.

10 In their article on presidential pandering, Canes-Wrone, Herron, and Shotts (2001) offer a third source of dynamic variation in the behavior of elected officials. Later in a term, the likelihood that a policy choice is revealed prior to an upcoming election to have been correct or incorrect is reduced, thus altering the incentives of a president to implement popular policies contrary to the public interest.

11 For a more detailed discussion of citizen attitudes about these mistakes, see Gordon and Huber (2002).

12 Three of these demonstrate the extreme circumstances surrounding instances of rejection. Cambria county judge Joseph O’Kicki garnered less than ten percent of the retention vote in 1991 after having been convicted on corruption charges in 1989 (Seelye 1991). Voters in Cameron and Elk Counties dismissed Gordon J. Daghir after he was sanctioned for judicial misconduct just months earlier. Daghir had accepted Penn State football tickets from a litigant in a divorce case he knew would come before him (Pennsylvania Court of Judicial Discipline 1995). Finally, Frank Eagen of Lackawanna County failed to secure retention in 1997 after being the target of a grand jury investigation concerning whether he had looted the estates of incapacitated persons (Flanagan 1997).

13 42 Pa.C.S. § 9781 sets conditions for the appeal of an assigned sentence. If a sentence is within the guideline range, the state or defendant can appeal only for clerical mistakes or if the “application of the guidelines would be clearly unreasonable.” If outside the guideline range, the standard for appeal is reasonableness.

14 Unfortunately, we lack data on the relationship between imposed and executed sentences.

15 Also, for these three crimes, punishment is politically uncontroversial. In more liberal areas, by contrast, citizens may oppose incarcerating nonviolent drug offenders.
The 1988 guidelines were revised in 1991. We have accounted for these changes as well as alterations in the criminal code during the period under study. In a given case, the authoritative guideline is the one in place when the crime was committed. For example, if a crime was committed in 1993 and the trial occurred in 1995, the judge would use the 1991 guidelines, not the 1994 ones. We also discarded cases for which we were missing the information necessary to calculate a judge’s place in the electoral cycle.

Our results are nearly identical if we examine the larger sentence imposed.

The tobit model assumes the existence of a latent variable representing, in this case, the sentence the judge would prefer to assign. It is fully observed only when the judge’s sentence falls between the censoring boundaries, in this case the statutory minimum sentence and one half the statutory maximum. Coefficients from this model reflect the effect of a unit shift in the independent variable on this imperfectly observed latent variable.

This need not imply that non-incarcerative punishments are non-punishment, only that they are, from the perspective of judges and their constituents, less punishment than any prison time.

In theory, because (as we discuss below) cases are assigned to judges randomly within districts, we could forego controlling for case characteristics, controlling instead for time-varying and static district features. Employing case level controls, however, is a conservative strategy. If case assignment is at all non-random, the controls minimize the threat of omitted variables bias. If it is random, they improve predictive efficiency.

Alternatively, one may employ as controls the defendant’s prior record score and the offense gravity score. Because these are ordinal scales that change between guideline regimes and map non-linearly to the guideline sentences, we strongly prefer using the guidelines themselves. Substituting the former for the latter, however, does not alter our substantive findings.

The 1994 and 1997 guidelines did not result in unambiguous increases (or decreases) in punitiveness for the cases we study. Rather, they altered the relative classification of case severity across crimes (e.g. aggravated assault versus robbery) and within crimes (e.g. aggravated assault with significant bodily
injury versus simple aggravated assault). Similarly, they revised the scaling of previous offense history.  

23 As a practical matter, the results are not substantially different if we confine our analysis to cases with only a single count.

24 For midterm appointments, we exclude all cases a judge hears before her initial partisan election.

25 We believe this fixed effect approach is asymptotically consistent, since the number of convictions will approach infinity faster than the number of judges (which is fixed) and years (See Heckman and MaCurdy 1980). In the current setting, consistency appears to be only a theoretical issue. Fixed effects OLS estimates (whether restricted to uncensored observations or for all cases) produce nearly identical results.

26 Throughout, we report the effect of changes in independent variables on the latent rather than on the observed dependent variable. Because it reflects the sentence a judge would hand down in the absence of statutory constraints, this quantity is the more easily interpretable of the two. To calculate for a particular observation the estimated effect of a change in an independent variable on the observed dependent variable, multiply its coefficient by the probability the observation is in the censoring region (Greene 2000, 909). Evaluated at the data sample means and based on the current set of estimates, that probability is approximately 0.68.

27 The effect is larger and significant in the fixed-effect specification. We consider this finding to be disturbing evidence of sentencing disparity, a critically important topic but one beyond the scope of this paper.

28 Parameter estimates in column (2) suggest a total sentence augmentation of 1,817 years (+/- 842). Simulated 95% confidence intervals are in parentheses.

29 Unlike our earlier models, we employ county and year dummies here instead of judge and year dummies. This is because we are most interested in controlling for unmeasured county variation that could bias our results. (A model including year, county, and judge dummies did not converge.)

30 The negative effect of proximity in very Republican districts may simply be an artifact caused by the fact that there are many cases from one very liberal county: Philadelphia. This raises the additional
possibility that the ideology variable interacted with proximity simply captures features of Philadelphia other than its liberalism. To determine whether this was indeed the case, we introduced a second interaction effect, between electoral proximity and a Philadelphia county dummy, into the model. The estimated coefficient on this effect was positive and significant, implying that the proximity effect is larger in Philadelphia than in other districts. The inclusion of this interaction did not, however, substantially alter the statistically significant, negative coefficient on the interaction between proximity and our ideology measure. Thus, our finding holds in the presence of this alternative explanation.

31 Note that we do not include a non-interacted measure of the number of judges because it is nearly constant within counties. Including such a measure does not alter the reported results.

32 The early first term learning variable is also positive and statistically significant. Manipulating the definition of “early first term” by increasing or decreasing the learning cutoff from five years does not upset this result until we extend it to ten years. (In that case, the model is identical to the one reported in the first column of Table 3.)

33 OLS regressions of seriousness on proximity, controlling for county fixed-effects, reveal no statistically significant association after the first two years in office.

34 This raises the possibility that “speedy” judges dominate our sample because of these slight perturbations in the selection weights. Our models with judge fixed-effects account for any systematic bias across judges. Additionally, the Administrative Judge we interviewed maintained that efficiency in the courtroom was not likely correlated with either judicial experience or punitiveness. Finally, even if one presumes that the selection criteria cause more punitive judges to dominate the field, this effect would bias against our findings. As we discuss above, a punitive judge would hand down harsher sentences over the entire course of her term, rather than increase them as election approached.

35 Additionally, these scheduling moves are likely to be reflected in the disposition of both plea bargained cases and those cases that go before a jury. When we disaggregated the proximity effect into plea bargained and non-plea bargained cases to see if cases settled out of court were somehow different than
those resolved through trial, we found that the proximity effect in plea bargained cases was statistically indistinguishable from the proximity effect in all other cases.

36 Moreover, by employing judge specific fixed-effects in our statistical estimation, we can (partially) account for factors specific to individual judges that may explain case disposition. If prosecutors are strategic, they can demand more punitive sentences in negotiating plea bargains after a conservative judge has been assigned to a case. In this circumstance, the judge-specific dummy will account for this conservatism both in plea-bargained and jury-conviction cases. Similarly, if (as we argue) judges become more conservative as election approaches and plea bargains become more punitive as a result, our proximity variable will capture this effect. Indeed, if we eliminate the disposition dummies from the analysis reported in Table 2, column (2), the results do not change appreciably. See also footnote 35.

37 Indeed, many of these restrictions on discretion were enacted, at least in part, to address perceived inconsistency in judicial sentencing (Greenberg and Humphries 1980; Tonry 1996; Free 1997). One potential source of this inconsistency may be differences in local electorates.
References


Table 1: Summary Statistics for Model Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned &quot;Smaller&quot; Sentence (months)</td>
<td>24.825</td>
<td>28.730</td>
<td>0</td>
<td>240</td>
</tr>
<tr>
<td>Guideline Minimum</td>
<td>20.068</td>
<td>19.218</td>
<td>0</td>
<td>120</td>
</tr>
<tr>
<td>Guideline Maximum</td>
<td>32.546</td>
<td>24.265</td>
<td>6</td>
<td>120</td>
</tr>
<tr>
<td>1988 Guidelines</td>
<td>0.541</td>
<td>0.498</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1994 Guidelines</td>
<td>0.263</td>
<td>0.440</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Male</td>
<td>0.913</td>
<td>0.281</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Non-white</td>
<td>0.656</td>
<td>0.475</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age (years)</td>
<td>28.566</td>
<td>9.056</td>
<td>14.73</td>
<td>88.42</td>
</tr>
<tr>
<td>Age squared</td>
<td>898.036</td>
<td>648.399</td>
<td>216.96</td>
<td>7817.42</td>
</tr>
<tr>
<td>Non-negotiated guilty plea</td>
<td>0.202</td>
<td>0.402</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Negotiated guilty plea</td>
<td>0.446</td>
<td>0.497</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Deadly Weapon Enhancement</td>
<td>0.108</td>
<td>0.310</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Deadly Weapon Use</td>
<td>0.020</td>
<td>0.140</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Counts in Conviction</td>
<td>2.171</td>
<td>2.287</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>Rape</td>
<td>0.063</td>
<td>0.243</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Robbery</td>
<td>0.521</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Electoral Proximity</td>
<td>0.435</td>
<td>0.277</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Judge's Term</td>
<td>1.417</td>
<td>0.676</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Republican Percentage of Vote for Attorney General</td>
<td>0.432</td>
<td>0.148</td>
<td>0.198</td>
<td>0.740</td>
</tr>
<tr>
<td>Number of Judges in County</td>
<td>22.157</td>
<td>16.439</td>
<td>1</td>
<td>47</td>
</tr>
</tbody>
</table>

N=22095
<table>
<thead>
<tr>
<th></th>
<th>(1) No judge, year effects</th>
<th>(2) Judge, year effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline Minimum</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>(11.13)</td>
<td>(16.20)</td>
</tr>
<tr>
<td>Guideline Maximum</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(5.21)</td>
<td>(6.36)</td>
</tr>
<tr>
<td>1988 Guideline</td>
<td>-2.71</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>(6.87)</td>
<td>(2.35)</td>
</tr>
<tr>
<td>1994 Guideline</td>
<td>0.47</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(3.32)</td>
</tr>
<tr>
<td>Male</td>
<td>8.70</td>
<td>8.18</td>
</tr>
<tr>
<td></td>
<td>(16.68)</td>
<td>(14.80)</td>
</tr>
<tr>
<td>Non-white</td>
<td>0.36</td>
<td>2.71</td>
</tr>
<tr>
<td></td>
<td>(1.12)</td>
<td>(7.72)</td>
</tr>
<tr>
<td>Age</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(1.74)</td>
<td>(1.83)</td>
</tr>
<tr>
<td>Age squared</td>
<td>-0.0036</td>
<td>-0.0034</td>
</tr>
<tr>
<td></td>
<td>(2.95)</td>
<td>(3.13)</td>
</tr>
<tr>
<td>Non-negotiated guilty plea</td>
<td>-6.60</td>
<td>-5.14</td>
</tr>
<tr>
<td></td>
<td>(14.68)</td>
<td>(11.30)</td>
</tr>
<tr>
<td>Negotiated guilty plea</td>
<td>-6.84</td>
<td>-7.50</td>
</tr>
<tr>
<td></td>
<td>(19.19)</td>
<td>(19.61)</td>
</tr>
<tr>
<td>Deadly Weapon Enhancement</td>
<td>19.02</td>
<td>17.89</td>
</tr>
<tr>
<td></td>
<td>(25.15)</td>
<td>(32.35)</td>
</tr>
<tr>
<td>Deadly Weapon Use</td>
<td>-1.12</td>
<td>-1.17</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.95)</td>
</tr>
<tr>
<td>Counts in Conviction</td>
<td>1.70</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>(9.82)</td>
<td>(25.39)</td>
</tr>
<tr>
<td>Rape</td>
<td>14.33</td>
<td>12.81</td>
</tr>
<tr>
<td></td>
<td>(15.60)</td>
<td>(18.69)</td>
</tr>
<tr>
<td>Robbery</td>
<td>6.43</td>
<td>6.90</td>
</tr>
<tr>
<td></td>
<td>(19.82)</td>
<td>(21.19)</td>
</tr>
<tr>
<td>Electoral Proximity</td>
<td>4.42</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>(8.13)</td>
<td>(4.23)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-14.65</td>
<td>-50.90</td>
</tr>
<tr>
<td></td>
<td>(8.86)</td>
<td>(3.80)</td>
</tr>
<tr>
<td>Standard error</td>
<td>21.46</td>
<td>20.55</td>
</tr>
<tr>
<td></td>
<td>(82.34)</td>
<td>(179.17)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-79940.74</td>
<td>-79034.04</td>
</tr>
</tbody>
</table>

**Notes:** Dependent variable is the minimum (smaller) sentence assigned by the judge. Absolute values of parameter t-ratios are in parentheses. Judge and year dummies omitted from column (2). N=22,095 (4,527 left-censored cases, 690 right-censored). Robust standard errors are employed to estimate t-ratios in in column (1).
Table 3: The Conditional Impact of Electoral Proximity: Two-Limit Tobit Models

<table>
<thead>
<tr>
<th></th>
<th>(1) Judge, year effects</th>
<th>(2) County, year effects</th>
<th>(3) County, year effects</th>
<th>(4) County, year effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline Minimum</td>
<td>0.67</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>(16.18)</td>
<td>(11.16)</td>
<td>(11.25)</td>
<td>(11.21)</td>
</tr>
<tr>
<td>Guideline Maximum</td>
<td>0.21</td>
<td>0.24</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(6.38)</td>
<td>(5.42)</td>
<td>(5.34)</td>
<td>(5.36)</td>
</tr>
<tr>
<td>1988 Guideline</td>
<td>2.74</td>
<td>3.63</td>
<td>3.64</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(2.41)</td>
<td>(2.43)</td>
<td>(2.38)</td>
</tr>
<tr>
<td>1994 Guideline</td>
<td>2.47</td>
<td>2.84</td>
<td>2.93</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>(3.26)</td>
<td>(3.21)</td>
<td>(3.31)</td>
<td>(3.25)</td>
</tr>
<tr>
<td>Male</td>
<td>8.19</td>
<td>8.30</td>
<td>8.26</td>
<td>8.28</td>
</tr>
<tr>
<td>Non-white</td>
<td>2.69</td>
<td>2.86</td>
<td>2.86</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>(7.69)</td>
<td>(8.43)</td>
<td>(8.44)</td>
<td>(8.44)</td>
</tr>
<tr>
<td>Age</td>
<td>0.14</td>
<td>0.13</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(1.80)</td>
<td>(1.53)</td>
<td>(1.59)</td>
<td>(1.61)</td>
</tr>
<tr>
<td>Age squared</td>
<td>-0.0034</td>
<td>-0.0034</td>
<td>-0.0035</td>
<td>-0.0035</td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(2.80)</td>
<td>(2.85)</td>
<td>(2.86)</td>
</tr>
<tr>
<td>Non-negotiated guilty plea</td>
<td>-5.13</td>
<td>-5.34</td>
<td>-5.25</td>
<td>-5.26</td>
</tr>
<tr>
<td></td>
<td>(11.27)</td>
<td>(11.40)</td>
<td>(11.21)</td>
<td>(11.22)</td>
</tr>
<tr>
<td>Negotiated guilty plea</td>
<td>-7.50</td>
<td>-7.61</td>
<td>-7.52</td>
<td>-7.53</td>
</tr>
<tr>
<td></td>
<td>(19.62)</td>
<td>(20.11)</td>
<td>(19.87)</td>
<td>(19.90)</td>
</tr>
<tr>
<td>Deadly Weapon Enhancement</td>
<td>17.90</td>
<td>17.94</td>
<td>17.96</td>
<td>17.92</td>
</tr>
<tr>
<td></td>
<td>(32.36)</td>
<td>(23.82)</td>
<td>(23.85)</td>
<td>(23.80)</td>
</tr>
<tr>
<td>Deadly Weapon Use</td>
<td>-1.17</td>
<td>-0.93</td>
<td>-0.81</td>
<td>-0.84</td>
</tr>
<tr>
<td></td>
<td>(0.95)</td>
<td>(0.56)</td>
<td>(0.48)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Counts in Conviction</td>
<td>1.72</td>
<td>1.75</td>
<td>1.74</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>(25.34)</td>
<td>(9.73)</td>
<td>(9.70)</td>
<td>(9.72)</td>
</tr>
<tr>
<td>Rape</td>
<td>12.78</td>
<td>13.02</td>
<td>13.06</td>
<td>13.07</td>
</tr>
<tr>
<td></td>
<td>(18.64)</td>
<td>(14.37)</td>
<td>(14.42)</td>
<td>(14.43)</td>
</tr>
<tr>
<td>Robbery</td>
<td>6.91</td>
<td>6.89</td>
<td>6.91</td>
<td>6.91</td>
</tr>
<tr>
<td>Electoral Proximity</td>
<td>0.43</td>
<td>17.03</td>
<td>0.04</td>
<td>7.56</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(9.96)</td>
<td>(0.03)</td>
<td>(2.06)</td>
</tr>
<tr>
<td>Term</td>
<td>-0.34</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>First Term Electoral Proximity</td>
<td>2.70</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Republican Percentage of Vote for Attorney General x Proximity</td>
<td>—</td>
<td>-29.90</td>
<td>—</td>
<td>-12.15</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(8.56)</td>
<td>—</td>
<td>(2.12)</td>
</tr>
<tr>
<td>Number of Judges x Proximity</td>
<td>—</td>
<td>—</td>
<td>-0.1538</td>
<td>-0.2728</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(0.92)</td>
<td>(1.60)</td>
</tr>
<tr>
<td>Number of Judges Squared x Proximity</td>
<td>—</td>
<td>—</td>
<td>0.0101</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(2.85)</td>
<td>(2.98)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-48.27</td>
<td>10.72</td>
<td>13.43</td>
<td>9.19</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td>(1.30)</td>
<td>(1.49)</td>
<td>(1.04)</td>
</tr>
<tr>
<td>Standard error</td>
<td>20.55</td>
<td>20.97</td>
<td>20.96</td>
<td>20.96</td>
</tr>
<tr>
<td></td>
<td>(179.18)</td>
<td>(81.13)</td>
<td>(81.21)</td>
<td>(81.21)</td>
</tr>
</tbody>
</table>

Log-likelihood: -79030.78, -79419.54, -79414.65, -79407.15

Notes: Dependent variable is the minimum (smaller) sentence assigned by the judge. Absolute values of parameter t-ratios are in parentheses. Judge and year dummies omitted from column (1). County and year dummies omitted from columns (2) through (4). N=22,095 (4,527 left-censored cases, 690 right-censored). Robust standard errors are employed to estimate t-ratios in columns (2) through (4).