Judicial Errors, Crime Deterrence and Appeals: Evidence from U.S. Federal Courts

Roee Sarel*

Working paper. Original version: February 29, 2016. This version: June 5, 2016

Abstract

This paper extends the empirical research on crime deterrence to judicial errors and their correction on appeal. Judicial errors have long been identified in the literature as detrimental to deterrence, but only with scarce experimental evidence to support the claim. Error correction has, thus far, been mostly overlooked. Analyzing appeal results from U.S. federal courts (1997-2013) and corresponding crime rates, I find that type I errors (wrongful convictions) decrease deterrence, as theory predicts. Error correction, however, has a more complex effect: reversals increase deterrence, but remands to a lower court decrease deterrence; which implies a need for theoretical adjustment, judicial caution and a closer look at the design of the criminal appeal system.

JEL Classification: K4, C33
Keywords: Judicial errors, crime rates, crime deterrence, appeals

1 Introduction

The body of 13 year old Tair Rada was found in the school toilet. Roman Zadorov, a construction worker, was arrested and charged with her murder. Based on a disputed confession and circumstantial evidence, he was convicted and sentenced to life in prison. Zadorov appealed his conviction. He produced expert opinions which challenged the strength of the evidence and further indicated that his confession does not match the findings at the crime scene. The appeal court decided to "remand" the case, i.e. to return it to the trial court for further hearing. Zadorov was convicted for a second time. He appealed again. The prosecution suggested a second remand, but the request was denied - the appeal court has decided to review the evidence itself. The conviction was then affirmed by the appeal court, but only

*Frankfurt School of Finance & Management, D- 60314 Frankfurt, Germany. email: r.sarel@fs.de. This paper has benefited from the helpful comments of Eberhard Feess, Ronald Wintrobe, Darwyn Deyo, Ansgar Wohlschlegel and other participants of the following conferences: Public choice society annual meeting (2016), European public choice society annual meeting (2016). An earlier version of this paper was circulated under the title "Judicial errors, crime deterrence and everything in between: evidence from U.S. federal courts".

by a 2-1 majority vote, leaving the public with a dissenting opinion of "not guilty" and doubts on whether a mistake has been made.[1] The twists and turns of such a case may breed a feeling of a possible injustice. As expressed by Blackstone's formula, it is *"better that ten guilty persons escape than that one innocent suffers"* (Blackstone 1765, pp. 352).[2] However, a less intuitive aspect is easier to overlook: judicial decisions do not only affect the parties involved, but create an externality which influences overall public incentives. Namely, the probability of conviction (possibly a wrongful one) may affect the decision of potential offenders who attempt to estimate the expected sanction. Disentangling this externality requires answering three related questions:

(1) *Do potential offenders care whether judicial errors are made?*

(2) *Does it matter whether errors are eventually corrected?*

(3) *Does the way of correction - remand or reversal - matter?*

A possible answer to the first question can be found in the law and economics literature. Following the seminal work of Becker (1968), in which potential criminals compare expected payoffs from law obedience and crime commission, a widely adopted extension (Png 1986) incorporates the effects of judicial errors on deterrence. In this Extended Beckerian Model (hereinafter also: "EBM"), judicial errors alter expected payoffs and "tip the scale" in favor of crime commission. This is argued to hold for the two possible types of errors: "type I"- i.e. wrongful convictions - and "type II" - i.e. wrongful acquittals.[3] However, the EBM has been repeatedly criticized, with some suggesting that judicial errors have a weak effect on deterrence while others suggesting the opposite - a stronger effect than the EBM implies. Therefore, theory alone does not give a clear-cut answer to the first question. Moreover, error correction has yet to be fully incorporated in the EBM and so no answer to the second or third question is available.[4]

Unfortunately, the same is true for the empirical literature. While many researchers have repeatedly tested other extensions of the Becker model, judicial

[1] Criminal Appeal 7939/10 *Zadorov v. State of Israel* (23.12.15). On February 2016, Zadorov filed a motion for an En-banc rehearing (Criminal en-banc rehearing 1329/16), which is currently pending.

[2] For a thorough discussion of the Blackstone principle and its dynamic effects see Epps (2015), responses by Johnson (2016) and Masur & Bronsteen (2015) and recent answer by (Epps 2016).

[3] While Png (1986) and Harris (1970) define type I error as releasing a guilty defendant and type II error as convicting an innocent, subsequent papers have used the opposite definitions. Since a criminal case has the null hypothesis of innocence, it seems more appropriate to define a conviction of an innocent (i.e. rejecting the null hypothesis although it was correct) as type I, and so I too shall adopt this definition.

[4] It should be noted, however, that since the EBM focuses only on the overall probability of conviction, it can perhaps be interpreted as implicitly including error correction as a factor of this probability. Additionally, one paper by Chopard et al. (2014) takes a step in the right direction by adopting the assumptions of the EBM and considering a setting with an appeal system, but focuses mainly on the effect of appeals on judicial effort.

errors have been consistently omitted. In recent years, the effect of judicial errors on deterrence has mainly been investigated experimentally, using individual decision making dilemmas in the face of varying error probabilities. However, these papers provide only partial evidence, as they use simplified models and do not take into account potential reactions of other agents in the criminal system (e.g. judges and prosecutors) to judicial errors. More importantly, the vast majority of the literature does not test error correction.[5]

In this paper I attempt to fill the gap in the literature by incorporating judicial errors and their correction into empirical research on deterrence. I use a unique panel data set gathered from U.S. federal courts, sentencing commission and the FBI's "Uniform Crime Reports" regarding a period of 17 years (1997-2013). Using appeal results as a proxy, I find a positive correlation between judicial error rates and crime rates, thus providing the first empirical support to the EBM. Namely, appeal affirmation rates (i.e. the share of appealed convictions that were affirmed on appeal), which reflect the rate in which judicial errors *do not* occur, are found to be negatively correlated with crime rates. My results also point at a possible interaction effect between conviction rates and affirmation rates, such that crime increases when both rates are high. However, the effect seems to be limited to government appeals only, which deal with punishments rather than convictions and are thus not directly related to type I errors.

I further find that when reversal rates rise, crime decreases. This suggests that error correction indeed matters and has a positive effect on deterrence. Conversely, I find that when remand rates rise, crime increases. This may seem surprising or even contradictory to theory, since intuitively, remands could be perceived as merely a "weaker" way to correct errors. Nevertheless, several competing arguments suggest that remands can in fact decrease deterrence. *A delay in the execution of the punishment*, due to the prolongment of the criminal procedure, might result in larger discounts of the sanction by impatient or myopic potential offenders. *A loss of faith in the system*, manifested by a disbelief in the error correction process, may cause the public to perceive remands as "uncorrected errors", given that the appeal court could have chosen reversal - but decided not to. Deterrence will then especially decrease, if the public does not believe that trial courts are likely to correct the error upon remand, or worse - that guilty defendants get another shot at a wrongful acquittal. *Asymmetric imposition of litigation costs* on

[5]One recent exception by Lewisch et al. (2015) considers an appeal system but does not vary judicial errors rates (see section 2 below for further details).

the innocents and the guilty in a remand process may increase crime incentives as well. Finally, I find that the two alternative paths to error treatment (reversal vs. remand) have an opposite but equally strong effect on crime deterrence, such that the *reversal-remand spread* best captures the effect.

To ensure that my findings are robust, I compare results from different methodologies that have previously been used for testing crime rates and court behavior. Furthermore, I adopt a wide identification strategy to overcome any potential endogeneity bias, stemming from omitted variables or a possible reversed causality. For example, in a reversed scenario, crime rates may affect appeal results either directly, if appeal judges adjust their behavior in response to changes in crime rates, or indirectly - through adjustments made by other agents in the criminal process (e.g. law enforcement, prosecutors or legislators), which end up affecting the rate of judicial mistakes. The wide strategy uses two complementary measures, by (1) controlling for relevant benchmarks along the legal chain, and (2) implementing an instrumental variables ("IV") approach. Instruments include, inter alia, appeal results from other fields of law (civil and administrative), which are related to both the trial court's tendency to err and the appeal courts' tendency to correct errors, but not directly related to crime. Additional instruments are created using the increasingly popular approach of Lewbel (2012), which utilizes heteroskedastic errors to form "internal instruments" from the data.

My findings hold two important implications: first, for the research community, existing theory should be adjusted to include not only the frequency of judicial errors - but also whether and how they are corrected on appeal. Empirical research should similarly incorporate appeal rates as explanatory variables when researching crime rates. Second, for judges and policy makers, my findings imply a need for caution when designing a remand mechanism and when deciding to remand cases to lower courts. Furthermore, a welfare cost-benefit analysis should be conducted in order to decide on optimal resource allocation for error prevention, scope of remand process and strengthening of public faith in the system.

The rest of the paper is organized as follows: Section 2 reviews existing literature. Section 3 presents the hypotheses development. Section 4 describes the data sample and sources. Section 5 reviews main variables and descriptive statistics. Section 6 focuses on the econometric model and methodology. Section 7 presents basic results. Section 8 reviews robustness tests, including additional data and instrumental variables. Section 9 discusses implications. Section 10 concludes.

2 Related Literature

In the original Becker (1968) model, a social planner can attempt to combat crime by using two available tools at his disposal: the probability of apprehension and conviction ("p") and the sanction imposed upon conviction ("f"). An increase in either p or f is assumed to decrease the expected payoff from crime, leading rational potential offenders to respond by opting for innocence rather than committing a crime. The model assumes that, at the margin, offenders are risk preferers, such that deterrence is more responsive to changes in p than in f. Becker's model implicitly allowed for type II errors, since p could be lower than 100%, but did not consider the possibility of a wrongful apprehension and conviction of innocent people (i.e. type I errors were not addressed).

In subsequent papers, judicial began receiving increasing attention (e.g. Harris 1970, Posner 1973) and their formal incorporation into the model is attributed to Png (1986).[6] In this extended model, both types of error (I and II) are argued to be detrimental to deterrence. The effect of type II errors (wrongful acquittals) is relatively straightforward: the more guilty defendants are wrongfully released, the more attractive crime becomes, since the expected sanction for criminals decreases. Type I errors, however, similarly increase the incentive to commit crimes, by reducing the payoff from innocence, given the possible wrongful sanction.[7]

The EBM has since been regularly incorporated in the law & economics literature (e.g. Kaplow 1994, Dari-Mattiacci & Deffains 2007, Polinsky & Shavell 1999, 2007), but opposing views challenge its core argument that type I errors reduce deterrence. Some argue that type I errors actually *increase* deterrence, for example, if they are (mistakenly) perceived as justified convictions (Ehrlich 1982) or when offenders act in a continuous action space, where over-compliance with the legal standard breeds a higher payoff (Calfee & Craswell 1984). Others argue that type I errors have *no effect* on deterrence, for example, when their frequency is too small for offenders to care (Andreoni 1991, pp. 389) or when the source of the wrongful conviction is someone else's crime ("identity error"). With respect to the latter, Lando (2006) and Mungan & Lando (2015) assume that identity errors happen with equal probability to the guilty and the innocent, leading to a zero net

[6]The extension of the model is usually associated with Png (1986), although earlier works by Harris (1970) and Posner (1973) contain some similar ideas (for a discussion on the origin of the model, see Mungan & Lando 2015).

[7]Another path in which type I errors may decrease deterrence is their effect on the stigma of a conviction, since an increase in wrongful convictions implies that it less certain that a conviction implies wrongdoing (Epps 2015, pp. 1099). My dataset does not allow to test the effects of stigma separately, but - in essence - this seems to be merely a specific case of the EBM's argument of changes in relative payoffs.

effect. Garoupa & Rizzolli (2012) argue conversely, that identity errors change the deterrence equilibrium, since each identity error is accompanied by a type II error regarding the actual perpetrator. Epps (2015, pp. 1126-1128) points out, however, that a failure to apprehend the actual offender (the type II error) stems from police behavior and might be exogenous to adjudication and wrongful convictions.[8]

A related prominent result of the EBM has also also been challenged: the argument that type I and II errors have an *equal marginal effect* on deterrence (Png 1986, Rizzolli & Saraceno 2013). Rizzolli & Stanca (2012) and Nicita & Rizzolli (2014) argue that type I errors are more detrimental to deterrence than type II errors, due to risk aversion; rank dependent utility; loss aversion *á-la* prospect theory (Tversky & Kahneman 1979, 1992); and the feeling of indignation, instigated by wrongful convictions. Contrarily, Mungan & Lando (2015) assert that the scope of the EBM's symmetrical result of error types is limited to some special cases (e.g., when the mistake is in the act rather than the legal standard).[9]

The empirical literature has, thus far, examined other extensions of the Beckerian model, but not the EBM. For example, some papers focus on unemployment (e.g. Entorf & Spengler 2000, Ihlanfeldt 2007, Lin 2008), which arguably decreases the payoff from innocence due to the potential lower value of honest work. Empirical evidence has been gathered similarly on lower wages (e.g. Gould et al. 2002), changes in police manpower (e.g. Levitt 1997, 2002, Evans & Owens 2007) or tactics (e.g. Sherman et al. 1989, Weisburd et al. 2009), arrest and conviction rates (e.g. Machin & Meghir 2004), imprisonment rates (e.g. Levitt 1997) and sanction enhancement (e.g. Bell et al. 2014). For further literature, see Khadjavi (2014*a*,*b*) and Chalfin & McCrary (2014).

The experimental literature has attempted to assess the Beckerian model as well (e.g. Baker et al. 2003, DeAngelo & Charness 2012, Friesen 2012, Schildberg-Hörisch & Strassmair 2012, Ouss & Peysakhovich 2013), but only recently included judicial errors. For example, Rizzolli & Stanca (2012) use a reversed dictator game with varying error probabilities and find that type I errors reduce deterrence (more than type II errors). Feess et al. (2015) assign subjects the roles of violators or judges and find that uncertainty regarding crime commission reduces the frequency

[8]Epps notes however, that a wrongful conviction breeds a strong incentive for the police to look for the real culprit, which then supports the argument that the two types of error are linked.

[9]The argued scope of the EBM is summarized by Mungan & Lando (2015, pp. 2) to include special cases where: *"the sanction is act-based rather than harm-based; the choice-set of the actor is bifurcated (dichotomous) rather than continuous; the likelihood of the case being brought to court is unaffected by the actors choice; the adjudicator makes a mistake in assessing the nature of the act committed and not in assessing the legality of the act; the actor does not choose whether or how much to participate in the regulated act; and the adjudicators mistake does not concern the identity of the offender."*.

of conviction, thereby decreasing deterrence (see also Baumann & Friehe 2015, for a similar setting). Only one, very recent, experimental paper (Lewisch et al. 2015) touches upon the appeal system's effect on deterrence and indicates that adding a second instance may increase deterrence. While this experiment provides preliminary evidence, dynamic marginal effects are not addressed, since the setting includes only a one-shot game where judicial error rates are not common knowledge.

Similarly, other aforementioned experiments do not fully encompass the empirical complexity of an appeal system and its effects over time, nor do they account for a potential indirect effect of crime rates on judicial behavior, through actions of agents in the legal system. This paper is therefore, to my knowledge, the first to fully incorporate judicial errors and their correction into the empirical study of crime rates and deterrence.

3 Hypotheses development

Following the EBM, judicial errors are expected to be negatively correlated with deterrence. However, translating this intuition into an empirical relationship is tricky, since judicial errors are not directly observable. In fact, there is no practical way to fully ascertain whether an error has been made.[10] Previous literature has attempted to estimate judicial error rates, for example, by comparing agreement rates of judges and juries (e.g. Spencer 2007, Kim et al. 2013) or by focusing on specific post-conviction exonerations (e.g. death sentence cases, see Gross & O'Brien 2008, Gross et al. 2014). This paper takes a different and novel approach, by utilizing instead aggregated appeal results as a proxy.

In order for this proxy to be satisfactory, however, appeal courts must be sufficiently accurate and - at the very least - more accurate than trial courts. The assumption that appeal courts are indeed more accurate is quite intuitive and may be justified on several grounds. First, this is a common assumption in the theoretical literature, based on the notion that appeal judges are either more qualified or better informed. For example, it has been argued that the appeal process harvests private information from litigants (e.g. Shavell 1995), with some papers going so far as to assume a perfectly accurate appeals court (e.g. Levy 2005, Shavell 2004). Second, if appeal courts were not more accurate, there would be little justification for their existence, since resources would be better put to use in improving the abil-

[10] For a discussion on the many challenges involved in empirical examination of judicial errors, see Gross & O'Brien (2008).

ity of trial courts instead.[11] Third, empirically, even if appeal courts sometimes err - is seems unlikely that these errors are systematic, given that the assignment of appeal judges to cases is usually random (Hettinger & Lindquist 2012, pp. 128). Thus, errors of appeal courts (insofar as they occur) can be treated as noise.

A further challenge lies in isolating judicial errors from their correction (or lack of it), namely because appeal results are not binomial: some appeals are dismissed on technical grounds and some on the merits, some reverse a decision fully and some partially, some will be the final step in litigation and some will have subsequent proceedings. The problem of categorizing appeal results is mitigated in this paper, since the U.S. federal courts already have a classification system in place, which includes four categories: *affirmed, reversed, remanded and dismissed*. The first two categories refer to the appealed decision, which can be either affirmed (i.e. the appeal court has decided that the previous decision was correct) or reversed (i.e. the appeal is accepted and the previous decision is reversed). The third category of "remand" refers to cases that are returned to the lower court for further review. The fourth category of dismissal refers to the appeal itself, when it is dismissed (i.e. rejected).

The nature of the data set ameliorates another potential problem - the separation of type I and II errors. In the U.S., only defendants are allowed to appeal their convictions while the prosecution cannot (as a rule)[12] appeal acquittals due to the "double jeopardy" rule (Stith 1990).[13] Therefore, almost solely type I errors are reviewed on appeal.[14] The isolation of type I errors entails an advantage in terms of identification, since it is clear that appeal results refer only to convictions and not acquittals. There is however a downside to this characteristic, since type II error rates remain unobservable (but may theoretically still affect deterrence). A lack of proxy for type II errors poses a challenge, but is only problematic insofar as type I and II errors are correlated (leading to an omitted variable bias). Indeed, such a correlation seems plausible, especially in light of the criminal standard of

[11] While appeal courts do not only exist to correct errors, but also - for example - to increase the coherence of legal doctrines (Hettinger & Lindquist 2012) through the creation of binding precedent, appeal courts must still be superior in order to effectively preform their other roles. Thus, higher accuracy is still a plausible assumption.

[12] See, for example, 18 U.S. Code 3731: *"In a criminal case an appeal by the United States shall lie to a court of appeals from a decision, judgment, or order of a district court dismissing an indictment except that no appeal shall lie where the double jeopardy clause of the United States Constitution prohibits further prosecution(...)"*

[13] The legal "double jeopardy rule" forbids the state from placing a citizen twice under the risk of conviction, including via a prosecutorial appeal upon acquittal. It should be noted that while the double jeopardy rule is widely accepted across countries, it is mostly interpreted much more narrowly such that an appeal is not considered to be a second jeopardy, and so this constraint is relevant mainly in the U.S.

[14] The double jeopardy rule does not preclude the government from filing an appeal with respect to the sentencing. However, the share of such appeals is generally very small (see Commission et al. 2012, part B pp. 38), and particularly in the data set used in this paper (see section 8.2).

proof, which creates a trade-off between the two error types (e.g. Andreoni 1991, Rizzolli & Saraceno 2013, Kaplow 2011). Nonetheless, given this trade-off, the correlation is expected to be negative, which can then only lead to attenuation of the effect of the type I errors. Thus, the unavailability of a proxy for type II errors would simply imply that the my estimates regarding type I errors are conservative (meaning that that true effect is even larger).[15]

Given that only type I errors are reviewed on appeal, affirmation rates should be the easiest to interpret, as an affirmation (literally) means that no type I judicial error was found by the appeal court. Consequentially I hypothesize that:

H1: Affirmation rates will be negatively associated with crime rates.

Note that this hypothesis relies on the aforementioned assumption of appeal court accuracy, such that wrongful convictions are not frequently affirmed (otherwise, affirmations might imply more - rather than less - type I errors).

Disentangling reversal rates is slightly harder, since a reversal includes two aspects: error occurrence (which should decrease deterrence) and error correction (which implies less errors overall, and should increase deterrence). However, after controlling for the frequency of errors (through affirmation) and the appeal filing frequency (see section 5.1.3 below), the marginal effect should be only that of correction, leading to:

H2: Reversal rates will be negatively associated with crime rates.

It is fairly easy to see that hypotheses 1 and 2 are actually two sides of the same coin: if the appeal court is accurate, it will reverse wrongful convictions and affirm correct convictions. Thus, hypotheses 1 and 2 rely on similar assumptions.

Remands are a somewhat different story. First note that while the effect of reversals and affirmations has a clear direction - either increasing or decreasing the probability of a conviction - remands are "neutral", in the sense that they only constitute a delegation of the decision power to the lower court. Second, remands differ from reversals in the scope of their potential effect on the components of the expected sanction, i.e. the probability of conviction (p) and sanction size (f). Namely, unlike reversals - which are related to p but presumably do not contain information on f (since appeal courts are unlikely to reverse a verdict and acquit a defendant only due to a sentencing mistake) - remands are related to both p and

[15] It should be noted, that some arguments in the literature would suggest that the correlation between type I and II errors is positive. For example, the aforementioned argument made by Garoupa & Rizzolli (2012) regarding the 1-1 correspondence between wrongful convictions and acquittals for identity errors, implies a perfect correlation for such errors. However, for this effect to be dominant, identity errors should constitute a significant share of wrongful convictions, which seems unlikely.

f. This occurs since appeal courts might opt for remands for two different reasons: (1) clarifying questions of fact, which can affect the decision on guilt (or sentence) and (2) revising a sentence following determinations made in the appeal.[16]

Intuitively, after a remand has been ordered, both the compounded probability of conviction and the severity of sentence could only *decrease* in the final verdict (assuming that the defendant has filed the appeal and managed to get "at least" a remand). Under the original Beckerian model (without type I errors), the reduction in p and f would imply that remands necessarily lower deterrence, but under the EBM, the question becomes: which p and f? i.e. for correct convictions or for wrongful convictions? If the trial court is inaccurate, the expected sanction will be lowered for correct convictions, leading to less deterrence (otherwise, if the expected sanction decreases for wrongful convictions, deterrence will increase). In other words, whether remands increase or decrease deterrence depends, first and foremost, on how accurate *trial courts* are.

It should be emphasized, however, that changes in p and f are not necessarily symmetrical in their effect on deterrence: if potential offenders are, at the margin, risk preferers (as predicted by the original Beckerian model), deterrence will be more responsive to p than to f. A higher responsiveness to p may also occur if prison sentences have decreasing marginal disutility for prisoners.[17] A small increment in the probability of a wrongful conviction (due to a remand) might then overshadow any benefit of decreased sanction for correct convictions.[18]

Since the public may not have complete information on the reasoning behind the decision to remand, a rise in remands might also serve as a signal - although it is unclear in which direction. On one hand, remands may initialize error correction (though a delayed one, as lower courts must still review the case), signaling less errors. Such a remand may even include specific instructions to lower courts, which should encourage a correction.[19] On the other hand, a remand entails the

[16] For example, 18 U.S. Code 3742(f) states explicitly that if an appeal court finds that the sentence is in violation of law, the case should be remanded.

[17] For example, in a recent paper, Masur & Bronsteen (2015) suggests that a decreasing sensitivity for prison sentences may occur for two reasons: (1) adaptation to the prison environment over time, and (2) the fact that largest disutility stems from the conviction itself (e.g. given the difficulty of convicted felons to find a job) rather than the length of the sentence.

[18] Some argue that, contrary to the predictions of the Beckerian model, a decrease in f would actually increase deterrence. However, these arguments are not relevant in my settings. For example, one argument suggests that judges who are highly averse to type I errors will prefer to acquit when the sanction is high, in order to avoid the higher cost of sending an innocent to prison for a longer time, leading to a decrease in p and infringement of deterrence (Andreoni 1991). Since this effect isn't direct but goes through p, which is controlled for in my regressions through conviction rates, it is of little concern. As another example, Smith & Vásquez (2015) argues that an increase in the sanction will decrease individual deterrence but increase overall deterrence such that the crime rates decrease. Since my analysis is in the aggregate figures, there is again little relevance. For additional arguments on how an increase in penalty might reduce deterrence, see Gneezy & Rustichini (2004), Caulkins (1993).

[19] See for example, remands in the state of Michigan (Beery 2002) as well as recent findings by Boyd (2015).

risk of a possible reenactment of the previous result, perhaps especially if the same judicial quorum reviews the case again. Therefore, the choice of remand rather than reversal may directly be perceived as a potential "uncorrected error". A third option is that the declaration of a remand holds no information at all (for example, when appeal courts remand cases purely since it is customary not to conduct evidentiary hearings on appeal).[20]

Remands also differ from reversals in the *time dimension*, since they do not entail an immediate decision but rather instigate a prolonged process. If potential offenders discount the future, a delayed sanction (even a correct one) will not have the same effect as an immediate one. Previous literature has argued that potential offenders not only discount the future, but it is precisely that fact which turns them into actual criminals - given the time difference between immediate benefit from crime and a future sanction (Davis 1988, Beraldo et al. 2013). In this context, Lee et al. (2009) develop a dynamic deterrence model and find evidence of "impatient" offenders (i.e. who heavily discount the future), which would support the argument that remands may lower deterrence by delaying sanction execution. A similar intuition can be found in models on the effects of court delays (e.g. Torre 2003), but the empirical evidence (e.g. Soares & Sviatschi 2010, Dušek 2015) on this issue are mixed. Notably, findings from Italy (Dalla Pellegrina 2008) indicate that trial court delays reduce deterrence but appeal delays do not.[21]

For remands, however, the main relevant factor is not the appeal process length but rather whether the sanction is stayed pending the final judgment. Namely, if the sanction is not stayed (e.g. a defendant appeals the verdict while sitting in prison), there should be no discount.[22] For deterrence to decrease, it is then necessary that a stay of sentence pending remand is granted to guilty defendants (stayed sentences for "impatient innocents" would increase deterrence).[23] Since defendants seeking a stay must generally show, among else, that they are likely to succeed on the merits,[24] the requirement might not hold if trial courts are accurate.

[20] For example, Rule 52(6) of the Federal Rules of Appellate Procedure states that facts cannot be set aside by the appeal court unless they are "clearly erroneous".

[21] Dalla Pellegrina (2008) attempts to explain this finding mainly by suggesting that trial delays are sufficient to crowd out appeal delay concerns. It should be noted that the appeal results in the Italian courts were not included in the analysis.

[22] It is however possible, that a delayed process can- by itself - create a loss of faith in the judicial system (Dalla Pellegrina 2008) irrespective of the sanction. Additionally, an extreme delay might cause evidence quality to deteriorate (Torre 2003, pp. 102), leading to a possible infringement of deterrence.

[23] A stay of execution depends, inter alia, on the type of sanction: death sentences are automatically stayed. Imprisonment is stayed if a request is granted. Fines are stayed if a request is granted, but a stay might include a condition of depositing money as a guarantee. See rule 38(a) of the Federal rules of criminal procedure; Rule 8 of the Federal rules of appellate procedure.

[24] Nken v. Holder, 129 S. Ct. 1749, 556 U.S. 418, 173 L. Ed. 2d 550 (2009); Leiva-Perez v. Holder, 640 F.3d 962 (9th Cir. 2011).

Whether stays are correctly granted is, however, an empirical question.

The prolongment of the proceedings following a remand (further hearings etc.) may also lower deterrence through another channel: *litigation costs*. If these costs were the same on expectation irrespective of whether the defendant is guilty or not, then deterrence would remain unaffected. However, if there is a cost asymmetry, such that remands are more costly for the innocent, then an increase in the probability of remand would lower deterrence. A cost asymmetry may arise is two ways: either (1) if remands are more frequent for innocent defendants (which seems plausible, if the apparent "guilt" of innocents raises more factual questions, leading to remands), or (2) if innocent defendants bear higher costs for proving their innocence (which is again plausible, if opportunistic guilty defendants invest less, knowing that their chances of "proving" innocence are lower).[25]

Summing up, theoretically, remands may either increase or decrease deterrence, depending on the accuracy of the trial court, signaling effects, time preferences of potential criminals and (a)symmetry of litigation costs between the guilty and the innocents. As the actual effect is an empirical question, I hypothesize that both a negative and a positive effect on deterrence may take place:

H3a: Remand rates will be negatively associated with crime rates.

H3b: Remand rates will be positively associated with crime rates.

I avoid hypothesizing about a possible effect of dismissals, which are too ambiguous, as an appeal may be dismissed for reasons that are unrelated to the decision itself (e.g. dismissals based on procedural rules, estoppels or previous legal precedent). Furthermore, some judges may be prone to dismiss an appeal when they feel that no error has been made, while others may do so when an error occurred but the result seems nonetheless just. Given this major ambiguity and since dismissals are also not a main object of interest of this paper (as they have no error correction aspect), no hypothesis is made regarding dismissals.

4 Data Sources and Sample

4.1 Federal Courts data

Quantitative data from the U.S. federal courts in 1997-2013 was hand-gathered from the "Judicial Business of the U.S. courts" reports and official published tables

[25] The argument that innocent people generally invest more in their defense is not unique to remands, and has been made regarding criminal defense in general (Johnson 2016, pp. 264).

(federal courts website). The data relates to three different levels:

- *District (trial) courts* - 94 district courts serve as a first instance for criminal cases (18 U.S. Code, §3231). Each court is located within one state or a jurisdictional territory. Appeals are filed to a circuit appeal court.
- *States and territories* - each state contains one or more district (trial) courts, but belongs to only one "circuit", which includes only one appeal court. In addition to the states, four "territories" are included in the judicial system.[26]
- *Circuit appeal courts* - 12 appeal courts review criminal appeals which originate from district courts.[27] Eleven circuit courts (numbered 1st to 11th) review appeals from multiple district courts. The 12th appeals court ("DC") reviews appeals from one district court only ("District of Columbia").

The official judicial district map is presented in Figure 2.[28] Quantitative data was gathered on each district court, namely new cases (defendants); terminated cases (defendants); convictions[29]; bench trials; jury trials and punishment type (fine or imprisonment). Criminal appeal rates were gathered from each circuit appeal court, alongside civil and administrative appeals rates for the IV approach.

4.2 Crime rates

Crime rates for the corresponding years were gathered from the FBI's Uniform Crime Reports ("UCR"), which include reported violent crime (e.g. murder, aggravated assault, forcible rape) and property crime (e.g. burglary, larceny, motor vehicle theft). State size differences are eliminated by using crime per capita (per 100,000 residents) as the crime rate measure. Data for U.S. territories was complemented from other official sources.[30] However, since these sources do not contain data for all years and lack clear definitions for some crimes (e.g. the "rape" crime is not always separated to forcible and non-forcible rape), I excluded three territories from the analysis (Virgin Islands, Northern Mariana Islands, Guam) in order to keep a reasonably balanced panel. These territories represent a small portion of the appeal supply and their exclusion should not involve a high sample error.[31]

[26] The four territories are: Puerto Rico, Virgin Islands, Guam and the Northern Mariana Islands.

[27] An additional judicial district is the "federal district", which does not review criminal appeals.

[28] Exact definitions can be found in 28 U.S.C 41, 81-131.

[29] All convictions are included, except for the Oregon district court in the year 2002, which for some reason does not appear in the reports.

[30] The FBI "Crime in the US" publication, which is the source of the UCR data tool (for Puerto Rico); statistical digests and yearbooks published by the territories (for Virgin Islands, Guam and Northern Mariana Islands); and world bank website (for Guam and Virgin Islands).

[31] Northern Mariana Islands and Guam belong to the ninth district and are about 1% of the appeals. The Virgin Islands belong to the third district and are about 5% of the appeals. The Virgin Islands are also an outlier, since the population size

4.3 Data constraints of the federal courts sample

Several inherent constraints of the data set should be mentioned. First, crime rates are published for a full calendar year (January - December) while court rates are published for October to September, thus crime rates suffer from a time lag. This is however also an advantage, as some lag should be taken into consideration, assuming that potential criminals take time to update their expectations.

Indeed, it is unclear exactly how and in which frequency potential criminals obtain their information. Some (more sophisticated) criminals may hire statisticians to collect data; some may ask their representatives to "sit in" on appeal hearings; some may only get information from reading the newspaper. It is further unclear to which extent criminals in the sample are rational, or whether pathological criminals, who are perhaps not so easily deterred (Dhami & al Nowaihi 2013), are included. It is similarly unclear whether criminals use ongoing Bayesian updating or decide according to heuristics.[32]

Moreover, under the EBM every person is in fact a potential criminal, where the deciding factor is only whether, at a given time, the benefit from crime exceeds the expected sanction. Once changes occur in either of these (among else, due to judicial errors), some non-criminals may become criminals (and vice versa), making it hard to establish an encompassing rule on how types of people receive their information. The problem is even more severe, since individuals might perceive a different probability of apprehension and conviction, depending on the information they have gathered and on past experience (Sah 1991). However, a significant effect of appeal results on crime would indicate that potential criminals are aware of appeal results. This awareness need not be due to a high profile exposure of court statistics, but can also arise due to criminals' endogenous decision to acquire the relevant information (see Dalla Pellegrina 2008, pp. 269).[33]

A second relevant constraint is the lack of ability to match appeal results to their court of origin. While the federal courts do publish data on the amount of appeals filed and terminated per trial court, appeal results are only published in the aggregate number. This, however, would only bias the sample if some courts always "get it wrong" while other always "get it right", such that appeal result

is only about 0.05%, while in other states the ratio difference is much smaller.

[32]For example, responses to news coverage of appeal results may occur due to an "availability heuristic", see Tversky & Kahneman (1973, 1974), Johnson & Payne (1986)).

[33]Dalla Pellegrina (2008) suggests that information may be clustered, such that those that are surrounded by criminal activity might even have greater access to the relevant information (including private information). While this may be the case, it is not a necessary condition for the awareness argument to hold, since my analysis relies on publicly available information.

rates are not proportional to the number of appeals filed from each court.[34] A disproportionality of this sort, however, seems unlikely, since criminal cases are generally decided by a jury, i.e. by a random group of people. Thus, there is little reason to assume that any court will systematically err more than others.

Furthermore, notwithstanding juries, judges can influence the result through different channels, such as issuing a "directed verdict" (overruling a jury); deciding whether to admit certain evidence; or through sentencing. Therefore, even if juries decide in a non-random or biased way in a certain district, professional judges can intervene and correct the errors. Indeed, one may then argue that some judges constantly err, but this is again unlikely as judges should be able to learn from appeal results and adjust their behavior to avoid future mistakes. Moreover, problematic judges may be assigned "easier" cases, or in extreme circumstances - be replaced by better judges.[35] Thus, as long as no court errs systematically on average, no bias will occur. It should be noted that although one could attempt to test the assumption of proportionality by looking further ahead at Supreme court results, it is impractical[36] since the U.S supreme court reviews less than 1% per year of all judicial decisions[37] and possibly in a non-random fashion.[38]

A third constraint lies in the fact that the appeal results in the sample are not separated into decisions on guilt, punishment or both. The distinction between errors in convictions and in punishments is of course theoretically relevant, given that the elasticity of deterrence with respect to p and f may be different (due to risk preference etc., see above). Empirically, however, the relevant question is mainly whether the probability of a wrongful conviction is correlated with errors in punishment. If there is a positive correlation, an observed effect of type I errors will be overestimated. To clarify this point, assume that wrongfully convicted offenders receive higher punishments and then consider a person on the verge of committing larceny. If he chooses innocence, both the probability that he would be wrongfully convicted and the probability that subsequently he will receive a higher punishment should affect his incentives in the same direction - towards crime. Conversely, if

[34] Suppose, for example, that two appeals are filed from two courts. A reversal rate of 50% may then mean that both appeals from one court have been reversed, while no appeals from the other court have been reversed. Alternatively, 50% may mean instead that exactly one appeal from each court has been reversed.

[35] Impeaching judges is, however, very difficult (see Pfander 2007).

[36] Additionally, it is obviously a circular argument: if one is doubtful about the usage of higher instance results (appeal courts) to measure the lower instance's accuracy (at the trial court), a similar doubt can be raised regarding the usage of Supreme court decisions to test appeal court's accuracy.

[37] See, for example, the data published by the U.S. Census, under "Table 331: U.S Supreme Court - Cases filed and Disposition: 1980 - 2010", http://www.census.gov/compendia/statab/2012/tables/12s0331.pdf

[38] For example, several papers have found that the Supreme court tends to grant review to cases it intends to reverse (Cummins et al. 2015, pp. 387).

wrongful convictions are usually accompanied by lower sentences, the effect of type I error will be underestimated. In this context, Lundberg (2016) shows that rational judges and juries engage in "compromise verdicts", where uncertainty about crime commission is compensated by lower sentences. Such behavior would lead to a negative correlation between type I errors and sanction size, and my estimates will then be conservative.

A fourth constraint emerges since crime rates are measured at the state - rather than district - level. While crime characteristics may differ from region to region within a given state[39] all offenders must, again, rely on the reversal rate of the appeal court. Thus, the effect of judicial errors on deterrence should be the same for all sub-regions, which implies that state crime rates will serve as a relatively accurate proxy for local crime rates within the state. Furthermore, this problem can be partially solved by using fixed effects at the state level to control for time-invariant attributes of each state.

A final constraint involves the appeal category of "affirm", which seems to include also appeals that were reversed in part.[40] If the portion of such appeals is small, it can also be treated as sample error. Otherwise, it may be hard to disentangle the effect of affirmation. I address the latter problem in section 8, by using additional data gathered from the U.S. sentencing commission.

5 Variable definition and descriptive statistics

5.1 Definition of main variables

5.1.1 Dependent variables

In most regressions, I use the variable $ln\_cp100_{i,t}$, which is the natural logarithm of total crime (violent and property) per capita in the state of district court i, at year t, multiplied by 100 (to make coefficient interpretation easier).[41] This log-linear model has the advantage of reducing outliers while keeping an intuitive interpretation: the effect of a marginal increase in appeal result percentage on crime rates.[42] For some regressions, I use instead the variable $tot\_crime\_s_{i,t}$, which is simply the total number of crimes.

[39] For example, crime rates are usually higher in urban regions (Glaeser & Sacerdote 1999).

[40] Starting from 2007, this is explicitly stated in the federal courts' reports.

[41] Since this is the same linear transformation used on the independent variables on the LHS of the regression equation, the same coefficient would be derived if no transformation at all would be made. However, it is more intuitive to think about changes in percentage points, and thus I have chosen to multiply both sides of the equation.

[42] Logging crime rates is common in the literature, see Osgood (2000, pp. 34-35).

5.1.2 Independent variables and treatment of multicollinearity problem

Appeal rates are calculated as shares of overall appeals decided on the merits, multiplied by 100 (one unit thus represents one percentage point).[43] The five rates used in the analysis are: (1) Affirmation, (2) Reversal, (3) Remand, (4) Dismissal and (5) Remainder share. Since these rates represent complementary shares, they are inherently linked, such that an increase in one rate always implies a respective decrease in another one. Unsurprisingly, these variables are also highly correlated (see table 13 for a correlation table). A multicollinearity problem then arises, which I address in two ways. First, since both reversal and remand rates have a significant effect when tested alone, but in opposite directions (see section 6), I aggregate the rates with opposite signs into one "*Reversal-Remand spread*" variable. Second, in later parts of the analysis I center the variables (demean by the within-court average), which has the advantage of reducing correlation without changing the interpretation of the regression coefficients (Afshartous & Preston 2011).

5.1.3 Control variables and indirect endogeneity

Most variables previously identified by the literature as affecting crime rates are unnecessary to achieve consistent estimators in my analysis, as they are presumably uncorrelated with judicial errors. For example, it is hard to think of any link between judicial errors and unemployment or wages. However, some control variables may be crucial, namely those that reflect the different stages of the criminal judicial procedure - prosecution, trial, conviction and appeal filing. Agents in these stages may respond to changes in both sides of the chain - either crime rates or appeal results. Consider, for example, a public prosecutor who is considering whether to file an indictment. In order to decide, she naturally has to consider the chances of a conviction - either due to personal career concerns (e.g. Perry 1998) or based on some guideline which limits her discretion ex-ante to cases with a high probability of conviction.[44] If reversal rates rise, such that the new expected probability of convictions is lower, the prosecutor might decide to focus only on "winner cases",[45] which in turn should lead to a higher conviction rate and perhaps less appeals (that

[43] Limiting appeal rates to those "decided on the merits" excludes appeals that have been rejected due to procedural reasons, thus allowing a more focused test of judicial errors that have actually been discussed. Using appeals decided on the merits also follows previous literature (see, for example, Scott 2006).

[44] For example, in the Israeli criminal system, prosecutors are directed to file an indictment only when there is a "reasonable chance of conviction", see Sec. 62 of the Israeli Criminal Procedure AcFt [combined], 1982 and the review of its judicial interpretation in Case 5699/07 Anonymous v. the Attorney General 62(3) 550 (26.2.2008).

[45] The implicit assumption of this argument is that prosecutors behave as self-interested agents rather than "philosopher-king" enforcers that maximize the public interest (Friedman 1999), as public choice theory would predict.

is, if guilty people file less appeals). The same is true for a police officer, when she considers which cases to pass on to the prosecutor - as she does not want to "waste her time" on arrestees who will anyway avoid prosecution later. This may cause the police to refer more guilty people to prosecutors, and lead to more convictions. In a similar but opposite chain reaction, judicial errors may decrease deterrence, thus causing crime to rise and police officers to have less free resources (Epps 2015, pp. 1097-1098). Investigations may then become less thorough, resulting in more innocent people arrested and, eventually, in less convictions, or more appeals. A different simultaneity bias can arise if the assignment of judges to courts depends on crime rates (e.g. better judges, who err less, self-select into jobs in "safer" areas where crime rates are low).[46]

The empirical literature has attempted to deal with such issues by controlling for some components which proxy expected sanctions, but usually focusing only on detection and arrests and seldom on later stages such as convictions (for a review, see Weatherburn 2012). This seems rather strange, mainly since the reactions of agents mostly boil down to changes in conviction rates and appeals filed.[47] I therefore address this by adding two main control variables: *conviction rates* and number of *appeals filed per capita*.[48] For the sake of caution, I add additional control variables: *Prosecuted defendants per capita*, to capture prosecution behavior; *Share of jury decisions*, to reflect the effect of a full trial rather than plea bargains; *Imprisonment rates* and *fines-only rates*, to capture the type and gravity of sanctions; and the number of *appeals decided on the merits*, to ensure that the results are driven by the appeal category (nominator) rather than the scope of decided appeals (denominator).

Additionally, I interact affirmation rates with conviction rates, for two reasons: first, in order to capture the full effect of the expected sanction, which depends not only on initial conviction but on a (potential) affirmation on appeal. Second, measurement issues of the affirmation category require a closer look on interdependencies, especially since full and partial affirmations may have opposite operative consequences (e.g. one affirms the conviction while the other reverses it). Table 1 summarizes the different variables and provides definitions.

[46] Additionally, judicial errors may change the incentives of crime reporters. For example, when reversal rates increase, victims may be reluctant to report a crime when they are uncertain whether a crime was committed, in order to avoid wrongful convictions. Respectively, when reversal rates drops, opportunistic false reports may be filed to purposely achieve wrongful convictions.

[47] For a criticism on neglecting to include conviction rates and focusing on arrest rates, see Mustard (2003).

[48] It should be noted, that no further lag is needed for these control variables, as the U.S. procedural rules dictate that criminal appeals will be filed shortly after conviction (usually within 14 days, see Federal Rules of Appellate procedure, Title II, Rule 4(b)).

5.2 Descriptive Statistics - Federal courts data

Descriptive statistics of the federal courts dataset are presented in Table 2. Affirmation rates are about 81% on average, suggesting that errors occur in less than 20% of all cases. Reversals rates are 6.6% on average and range from less than 1% to about 24% of all cases. Remand rates are 3.16% on average and range between 0% (i.e. no remand whatsoever) and 31% (almost a third of the decisions remanded). Figure 2(1) presents the means of appeal results by district. As can be seen, reversal rates within each court tend to be high when remand rates are low, which implies that these judicial tools are used interchangeably. Under the aforementioned assumption that trial courts do not err systematically over time, the difference can be attributed to the appeal court's characteristics, such as judicial conservativeness or degree of omission bias and reluctance to change the status-quo (Zamir & Ritov 2012). In other words, irrespective of the merits, some appeal courts prefer to reverse while others prefer to remand.

Figure 2(2) displays appeal results means by year. A conspicuous jump occurs in 2005, with lower affirmation rates and higher reversal and remand rates. This jump is most likely a result of the U.S Supreme Court's ruling in U.S v.*Booker/Fanfan*[49], which allowed appeal courts to review sentences outside the strict range of the federal sentencing guidelines (Commission et al. (2012, part B, pp. 40), Catterson (2006, p.290)). Crime rates are also heterogeneous, with an average of 3,809 offense per capita (s.d= 1124.65). Higher crime rate observations are mostly located in DC, perhaps since it mainly includes a populated city (Washington DC). Property crimes are vastly more common than violent crimes (88.5% of crimes in the sample).

6 Empirical analysis: model and methodology

Following pre-tests which indicated a multicollinearity problem, I aggregated reversal and remand rates into the aforementioned "Reversal-remand spread" variable. The hypotheses are then tested using the following model:

$$\begin{aligned} ln\_cp100_{i,t} &= \beta_0 + \beta_1 spread_{i,t} + \beta_2 aff_{i,t} + \beta_3 conrate_{i,t} \\ &+ \beta_4 (aff_{i,t} \times conrate_{i,t}) + \beta_5 X' + Y_t + C_i + \varepsilon_{i,t} \end{aligned}$$

[49] *United States v. Booker*, 543 U.S. 220 (2005). See also the previous rulings in *Apprendi v. New Jersey*, 530 U.S. 466 (2000) and *Blakely v. Washington*, 542 U.S. 296 (2004).

where: $spread_{i,t}$ is the reversal remand spread; $aff_{i,t}$ is the affirmation rate; $conrate_{i,t}$ is the conviction rate; X' is a vector of control variables; and Y_t, C_i are year and court fixed effects respectively. Estimating the model correctly requires a careful choice of methodology.[50] To avoid method bias in this paper, I use statistical tests to rule out inappropriate methodologies and compare results from different (relevant) methodologies.

For the sake of brevity, I will assume that readers are familiar with the basic concepts of panel data analysis, but provide short clarifications when appropriate. A Hausman-test (Hausman 1978),[51] suggests that a fixed-effects model is preferable to a random effects model. In a nutshell, a fixed-effects model assumes that unobserved individual-court effects are consistent, which sits well with the impression created by figure 2(1) of differences between courts.[52] It also allows for the plausible scenario in which these effects are correlated with the regression's error term. Further tests indicate the presence of heteroskedasticity,[53] and autocorrelation of the residuals.[54] Additionally, some tests point at a cross-sectional dependence ("contemporaneous correlation") in the data.[55] A possible explanation for cross-sectional dependence is judicial behavior.[56] For example, judges may compare themselves to peers in neighboring districts (or in specific courts that serve as professional benchmarks) and alter their rulings such that rates become correlated. Alternatively, "spatial correlation" (e.g. correlation between neighboring states) may occur, for example, due to regional shocks to crime incentives. To account for these issues, I compare results from several methods:

1. **Fixed-effects with two-way cluster-robust standard errors** (by court and year)[57], which is robust in the face of arbitrary correlations within courts

[50] For discussions on appropriate methodologies, see Cornwell & Trumbull (1994) (for crime rates analysis) and Scott (2006) (for U.S. federal courts analysis).

[51] Comparing a fixed and random effects regression with quadratic and cubic terms (same as in table 5, column 1), the null hypothesis that differences between panels are inconsistent iFs readily rejected ($p < 0.0001$).

[52] Hettinger & Lindquist (2012) similarly find that appeal courts differ in their tendency to reverse appeals.

[53] The null assumption of homoskedasticity was rejected in a modified Wald-test ($p < 0.0001$), implemented by the user-developed xttest3 command in Stata (Baum 2001).

[54] The null hypothesis of no-autocorrelation was rejected in a Woolridge test for autocorrelation in panel data ($p < 0.0001$), implemented by the xtserial command in Stata (Drukker et al. 2003).

[55] Since the number of panels (N) is larger than the number of periods (T), one cannot use the standard Breusch-Pagan test (xttest2 command in Stata). Instead, I use the user-developed command: xtcsd (De Hoyos & Sarafidis 2006), which allows to compare three tests for cross-sectional dependence: a Pesaran test (Pesaran 2004), a Frees test (Frees 1995) and a Friedman test (Friedman 1937). All tests share a null hypothesis of cross-sectional independence. The Pesaran test rejects the null hypothesis, almost at the 5% level (p= 0.058). The Frees test rejects the null at the 1% level (statistic 12.9 > critical value 0.3125). Conversely the Friedman test indicates no cross-sectional dependence at all (statistic 6.863, p=1.0).

[56] To ensure that the cross-sectional dependence is not a technical by-product of using the state level crime rates for all courts within that state, I test each independent variable separately, using the user-developed command xtcd (Eberhardt 2011) which implements a Pesaran test (Pesaran 2004). I then find evidence of significant cross-sectional correlation in each series of independent variables ($p < 0.01$ for all variables), indicating that the correlation is not driven by the dependent variable alone.

[57] For a recent review of clustering methods see Cameron & Miller (2015).

(across years) and within years (across courts). Since this method is relatively general in nature and does not impose strict assumptions, it will serve as the benchmark for the lion's share of the analysis.

2. **Fixed effects with panel-corrected standard errors** ("PCSE", Beck & Katz 1995), assuming either an AR1 or PSAR1 disturbance.[58] This method is robust given contemporaneous correlation, which is problematic if some - but not all - panels are correlated (Worrall & Pratt 2004).
3. **Fixed effects with Driscol-Kraay standard errors** ("SCC", Driscoll & Kraay 1998, Hoechle 2007); deals with cross-sectional and spatial correlation.
4. **Fixed-effects Poisson with robust standard errors**, which accounts for the "event-count" nature of crimes (Osgood 2000).

Table 3 compares the methods and provides examples for previous papers on either crime rates or courts that have employed the different methods.

7 Basic Results

Table 4 presents the results of the base model regressions.[59] I find a negative correlation between the reversal-remand spread and crime rates. Since the spread increases either when reversal rates rise or when remand rates fall, this implies an equally strong - but opposite - effect of reversals and remands. The effect is significant ($p < 0.05$ or $p < 0.01$) across all five regressions, with some differences in the effect size. Taking, for example, the coefficient from the two-way cluster regression, the spread coefficient can be interpreted as follows: an increase of 1 percentage point in the spread is associated with a decrease of crime rates per capita of 0.157%. Similarly, I find a negative correlation between affirmation rates and crime rates, which is again robust across methods ($p < 0.01$ in all methods but Poisson, with $p = 0.064$). The effect size of affirmations is, however, about 6-10 times(!) larger than the *spread*, implying an asymmetric effect of error occurrence and correction.[60] However, there seems to be a balancing effect by the interaction between convictions and affirmations. Specifically, while conviction rates are neg-

[58] AR1 disturbance assumes a one-period lag correlation factor which is identical for all courts. PSAR1 disturbance assumes a one-period lag correlation factor which is court-specific. For further details see Beck & Katz (1995).

[59] Note that the PCSE regressions with fixed effects do not produce an interpretable R-squared (Blackwell III 2005, pp. 250). It is however still possible to attain a Wald-statistic of the variables of interests (Blackwell III 2005, pp. 251). Testing all variables except for year and court fixed effects produce a statistic of 29.99 (p=0.0028).

[60] In order to get a complete picture, one must also consider that affirmation rates are, on average, more than 10 times higher than the spread. Therefore, the effect size parity is even greater.

atively correlated with crime rates ($p < 0.05$), the interaction term has a positive coefficient, albeit much smaller than the main effects.

To help interpret the interaction, Figure 3 presents average marginal effects of affirmation rates given different conviction rates.[61] It is then easy to see, that affirmation rates are negatively correlated with crime rates for the lower range of conviction rates - but when conviction rates become too high, affirmation rates have an opposite effect. One interpretation for this finding would be that if convictions rates are very high, the public ceases to believe that affirmations are justified, since it is "simply impossible" that errors are so infrequent. However, the interaction may just capture the effect of some omitted variable (see robustness checks below). Control variables are mostly insignificant at the 5% level,[62] with a few exceptions in some regressions:

- Prosecuted defendants per capita: positive coefficient (PCSE regressions), which can either be explained by reversed causality (more criminals = more defendants) or by increases in prosecution rates which signal more type I errors.
- Imprisonment rate: positive coefficient (SCC), perhaps due to judges attempting to (unsuccessfully) increase deterrence through more severe punishments, in response to rising crime rates. Alternatively, a higher imprisonment rate may increase crimes in two paths: (1) by diluting the stigma of imprisonment (Harel & Klement 2007) or (2) by creating a "school for crime" (Roberts & Hough 2005, pp. 297), leading to more crime.[63]
- The share of jury cases: negative coefficient (SCC), suggesting that juries are perceived as less likely to err. However, the IV approach (see below) indicates an opposite effect, i.e. that juries are perceived as more likely to err.
- Appeals decided on the merits: positive coefficient (SCC, POISSON), implying that more appeals signal more errors. However, the mere filing of an appeal alone seems to send no relevant signal (insignificant effect for appeals filed per capita). While this sits well with the argument that it is appeal results that matter, it challenges more general game-theoretic models of appeal procedure, in which filing an appeal is an informative signal (e.g. Shavell 1995, 2004, Wohlschlegel 2014). A possible explanation would be that all criminal

[61] All values in figure 3 are taken from the sample range, two-way cluster regression.

[62] Some control variables have little variation over time. Thus, their effect is absorbed by the time fixed effects. These fixed effects are, however, necessary in order to capture the full effect of all unobserved panel-invariant effect, in order to ensure proper identification of the variables of interest.

[63] Additionally, it has been argued that imprisonment can lead to more crime by reducing prisoner's future wage prospect (e.g. since employers may avoid hiring ex-prisoners) leading to irregular employment and higher crime commission incentives (see Johnson 2016, pp. 281).

appellants have little to lose from appealing, since the cost is relatively low (Ridgway 2009, pp. 159). However, some robustness checks (table 8) indicate that there may actually be an effect for appeal filings as well.

Summarizing the basic results, I find support for hypotheses 1,2 and 3b (and against hypothesis 3a). The EBM's theoretical result - a detrimental effect of type I errors on deterrence - seems therefore to hold empirically. Moreover, the correction of judicial errors matters as well for crime deterrence. The finding regarding an interaction effect between convictions and affirmations does not pose an obstacle to this argument, but quite the contrary - if high convictions alongside high affirmations are a signal of more errors, then the effect is in line with the EBM. However, as will be explained below, the interaction effect is not fully robust.

8 Robustness checks

8.1 Controlling for polynomial effects and spurious significance

Following the econometric literature (e.g. Afshartous & Preston 2011, Balli & Sørensen 2013), I test the interaction term by substituting the original term with an interaction of centered variables[64] and add quadratic terms as controls. Since quadratic terms are highly correlated with third-degree polynomials ("cubic terms"), I control for those forms as well.[65]

Results are reported in Table 5. The effect of the spread is qualitatively similar to the base results but slightly larger. Affirmation rates demonstrate two extremes - either a (much) stronger effect than the base results (two-way cluster and SCC), or no significant effect (PCSE and POISSON). When affirmation rates have a significant effect, quadratic and cubic terms are significant as well, but the main trend is the same. To help interpret the results, Figure 4 presents an updated average marginal effects of affirmation rates (with quadratic and cubic terms).[66] For low conviction rates, the effect of affirmation is more or less linear. However, when conviction rates rise, *intermediate* affirmation levels are positively associated with crime. These effects are therefore complex, but the main finding on the

[64] To center the variables, I demean the affirmation rates and conviction rates by the within-court average prior to interacting them.

[65] The need to control for cubic terms is conveniently described by Wooldridge (2012, pp. 303) in the context of conviction rates: *"The presence of the quadratics makes interpreting the model somewhat difficult. (...) We might conclude that there is little or no deterrent at all at lower values of [conviction rates]; the effect only kicks in at higher prior conviction rates. We would have to use more sophisticated functional forms than the quadratic to verify this conclusion."*.

[66] Coefficients for figure 4 are taken from column (6).

interaction term remains similar - there is a range in which the interaction of affirmation and conviction rates increases rather than decreases crime.

Unfortunately, centering and adding quadratic terms, does not fully account for endogeneity. Particularly, the aggregation of full and partial affirmations does not allow to observe which type of affirmation interacts with conviction rates. Additionally, affirmations include also government-filed appeals of sentence, which have an opposite operative result than affirmed defendant-filed appeals.

8.2 Disentangling appeal results - data on sentencing appeals

To mitigate omitted variables bias, I collected additional data from the U.S. sentencing commission's online database, which includes some important additions. First, appeal results are reported on the district court level, obviating the need to rely on an assumption of representativeness of the appeal court average. Second, appeal results involve sentencing appeals only, i.e. no appeals of "conviction-only". Third, appeals are separated into defendant and government appeals. Fourth, affirmations are separated into "full" and "partial" (i.e. affirmed in part and reversed in part).[67] Finally, the database provides a complementary proxy for the expected sanction in the form of average imprisonment length.

This database suffers, however, from two main disadvantages (1) a smaller sample which spans only 8 years, from 2006-2013, and (2) much like the federal courts database, one still cannot extract the exact appeal results rates, since conviction-only appeals are not covered. Nonetheless, sentencing appeals rates can serve as a proxy for the overall rate. Variable definitions and descriptive statistics are listed in table 1, panel C and table 6, respectively.

Sentencing appeal rates appear to be quite different than the appeal court average, especially with respect to defendant vs. government appeals. For example, recall that the affirmation rate mean was 81% in the large sample. In comparison, the defendant-appeals affirmation rate is a bit lower (76.4%), but the government-appeals rate is much lower (13.98%). Furthermore, sentencing reversal rates are higher compared to the large sample, with a defendant reversal rate of 9.4% and government rate of 26.7%(!).[68] Since government appeals constitute less than 2% of appeal results, the disparity may perhaps be treated as noise. However, in order

[67] The "remainder" category is however non-existent in the sentencing database.

[68] A higher reversal rate for government filed appeals is consistent with the findings of Hettinger & Lindquist (2012, pp. 140), who suggest that a selection effect - where the government hand-picks the cases expected to end in reversal - may be driving this result.

to ensure that my results are not driven by a "government-outlier", I also separate government and defendant rates in some further robustness checks.

A t-test comparing the aggregated sentencing affirmation rate (full and partial, mean of 79%) to the appeal court average used in the larger sample (mean of 80.9%) shows no significant difference, which supports the assumption of representativeness mentioned earlier (i.e. appeal level data is representative of each district court). However, partial affirmations constitute around 3.7% of overall affirmations, which seems substantial.[69] I thus interact conviction rates separately with full and partial affirmations. Independent variables are centered to reduce collinearity with the interaction term.[70]

Table 7 compares results from the five methods using the sentencing data. The base results regarding the spread and full affirmation hold qualitatively once again[71], but coefficients are smaller and some methods demonstrate lower significance levels. Partial affirmations have a negative coefficient, but are significant only when controlling for quadratic terms (column 6) and in POISSON regression. The interaction effects are insignificant across methods.

However, separating defendant and government rates, Table 8 presents a slightly different picture. While defendant rates coefficients are similar to the base results, government rates coefficients are mostly insignificant, with the exception of the interaction between conviction rates and full affirmations ($p < 0.01$). The latter finding implies that perhaps it is not "public faith" which drives the interaction effect, but rather *sanction size* upon sentencing. Namely, when conviction rates are high but subsequent sanctions are low - reflected by the fact that government sentencing appeals are rejected and the original sentence is affirmed - an effect similar to type II errors occurs (a guilty person receives inadequate sanction).

As another robustness check, I vary the levels of clustering and control for circuit-by-year shocks through additional fixed effects. This check allows to rule out, inter alia, any identification issue that might arise if appeal court accuracy (which is assumed to be high) was time-variant.[72] Results of this check are presented in Table 9 and suggest that the choice of clustering level doesn't matter for

[69] A t-test comparing only full affirmation rates to the appeal level average shows it is significantly different - which implies that partial affirmations matter.

[70] For example, interactions are correlated with the main terms: 0.88 for full affirmations and their interaction with conviction rates, and 0.99 for partial affirmations and their interaction with conviction rates, both correlations are significant at the 5% level. Note that in general, centering is a common method for reducing endogeneity as well, but comes with the cost of some information loss (e.g. Gardner 1998, pp. 39).

[71] In column(6), note that there is still a significant effect of full affirmation rates, expressed by the coefficient of the cubic term.

[72] Ideally, one would try to control also for more local effects through state-by-year fixed effects. Unfortunately, since the crime rates are available at the state-level only, this is not applicable.

my results. Furthermore, when controlling for circuit-by-year effects (column 5), the government spread (recall: this spread refers to sentencing, not convictions) has a positive effect, implying that even remands of sentencing appeals is positively correlated with crime rates.

Fully interpreting the latter finding goes beyond the scope of this paper, but one possibility would be that the government is more prone to appeal sentences that accompany wrongful convictions, in which increases in the sanction (as a result of a reversal or remand) decreases deterrence. Such a tendency may reflect misconduct (e.g. overly eager prosecutors seeking to ensure that framed defendant get fully punished), but may also emerge if sentences following wrongful convictions are usually lower (for example, due to "compromise verdicts" of judges and juries, as mentioned above).

8.2.1 An Instrumental variable approach

While previous checks mitigate potential endogeneity biases, reversed causality may still be in play. I therefore employ an IV approach. For those readers unfamiliar with IV, the idea is using an additional variable ("an instrument") which fulfills two conditions: (1) it is "exogenous," i.e. uncorrelated with the regressions' error term, and (2) it is "relevant", i.e. sufficiently correlated with the endogenous variable. The instrument is first used to predict the endogenous independent variable *as if* it was unrelated to the error term (in a "first stage" regression). Predicted values then substitute original values in the main regression (a "second stage" regression).

Previous papers on crime rates and deterrence have used IV for other stages of the legal chain (e.g. arrest rates (Cornwell & Trumbull 1994), police hiring (Levitt 1997, 2002) and prison population (Levitt 1995)) and indeed, the entire legal process is somewhat suspicious as endogenous.[73] However, for variables of interest in this paper, the suspicion is strong for the reversal-remand spread and less so for affirmation rates. Namely, the spread captures not only one but two rates (reversal and remand), which judges can alternate between in response to crime rates. The choice of tool may then be correlated with the error term. Furthermore, appellate judges are presumably constrained in their ability to manipulate affirmation rates

[73] IV is however only one of several identification strategies used in the literature for deterrence. For example, some papers have used discontinuities in sanction size, such as the exogenous age transition that separates between trials in juvenile and regular courts (Levitt 1998), or the score that determines to which prison a convicted defendant will be sent (Chen & Shapiro 2007). Others used exogenous shocks, such as a legislative change (Drago et al. 2009, e.g.).

by declaring that an error took place where it did not (or vice versa), among else, due to the potential for a second appeal to the Supreme court.

Moreover, by separating affirmations into full/partial and defendant/government, omitted variable bias is already weaker. Additionally, since spread and affirmation are suspicious as endogenous partially for being tightly related, instrumenting for one can "clear out" the simultaneity problem from the other. It is also worth noting, that interaction terms will be exogenous as long as one interacted variable (either affirmation or conviction rates) is exogenous (Bun et al. 2014, Nizalova & Murtazashvili 2016), given that the endogenous variable is controlled for. Since some previous studies find that higher crime rates do not necessarily lead to higher conviction rates (Jensen & Heller 2003, pp. 83), there is indeed indication that convictions (at least) do not suffer from reversed causality. Thus, I focus on finding an instrument for *spread* and complement the identification strategy through econometric tests which ensure that all other variables can be treated as exogenous.

Alas, locating a proper instrument is tricky. First, one must locate an instrument which is sufficiently correlated with the *spread* but not with the regression's error term. Second, the instrument must also be correlated with judicial errors (and their correction), such that the proxy aspect of appeal results isn't completely lost. The range of possible instruments then becomes very narrow. For example, judicial case load of trial judges should be relatively independent of crime rates and presumably correlated with judicial errors occurrence (if trial judges err due to time constraints), but is unconnected to error correction, which depends on appeal judges.[74] Similarly, there is existing evidence that appeal judges reverse less when their ideology is similar to that of the district court judge (McGuire 2012, chapter 7, pp 138). Using ideology as an instrument would then perhaps help predict error correction, but not error occurrence.[75]

Luckily, a promising source of instruments lies in appeals from other legal fields, which may capture both judicial errors and their correction - but outside the context of crime. Three instruments were eventually located for the reversal-remand spread: (1) administrative appeals reversal-remand spread, (2) civil appeals reversal-remand spread and (3) One year lagged criminal reversal-remand spread (i.e. $spread_{i,t-1}$).[76] Administrative appeals are arguably most likely to be

[74] When tested, judicial caseloads in different compositions (state/district, criminal/civil/both) were indeed found to be insufficiently correlated with the spread.

[75] Unfortunately, in my aggregated data-set, the ideology of judges is also unavailable for use. For use of judicial ideology as an instrument, see Di Tella & Schargrodsky (2013).

[76] Bankruptcy appeal rates were also considered, but found to be inadequately correlated with the spread.

exogenous, since they should not have any direct effect on crime rates nor is it likely that they are affected by crime, especially not by the types of offenses considered here.[77] Their disadvantage is, however, that they are only partially related to judicial errors made in district courts, since administrative appeals may be filed on decisions of other institutions as well.[78]

Civil appeals are also presumably not directly connected to crime, although I cannot rule out the possibility that the civil litigation involves damages caused by crimes (rendering civil appeals endogenous). A lagged criminal spread variable has the advantage of a direct link to the spread, but is also imperfect, since lagged spreads may influence the behavior of criminals, being a part of the aggregated information used for expectation updates. Nonetheless, a lagged spread refers to 15 months before the crime rate period - a time in which perhaps different judges were incumbent and circumstances were different, making any such influence weaker. Additionally, since the instruments are taken from the *appeal court level*, they are less likely to be correlated with the state-level crime rate.

It should be emphasized that these instruments are imperfect, and could - theoretically - still be correlated with some unobserved factors that impact crime rates. Of largest concern are perhaps type II errors, which may be linked to appeal results in other fields. For example, as in criminal cases, administrative appeals are a manifestation of a conflict between private persons and a public authority. An incorrect ruling in favor of an administrative petitioner then bears some resemblance to a wrongful acquittal. In light of the above, I use several econometric methods to ensure the validity of the instruments. First, assuming that at least one of the three instruments fulfills the exogeneity requirement, a test of "overidentifying restrictions" ("Hansen's J") shows that the instruments are jointly valid ($p > 0.05$, i.e. the null hypothesis of instrument validity cannot be rejected at the 5% level).

Second, an "endogeneity test"[79] of the spread variable indicates that it should not be treated as exogenous. ($p < 0.05$ for all regressions). Third, an orthogonality test (which is the flip-side of the endogeneity test, see Baum et al. 2007, pp. 15) is used to verify that all other control variables that are suspicious of endogeneity - including affirmation - can be treated as exogenous. Fourth, I check for redundancy and find that no instrument is redundant.[80] Fifth, I use various tests for identifying

[77] Recall that the crime rates includes only violent and property crime, which seem unrelated to administrative issues. This of course would be different if, for example: a crime of public officials corruption was included in the analysis.

[78] See, for example, rule 13 of the Federal Rules of Appellate Procedure, governing appeals from tax court.

[79] The endogeneity test is implemented by the Stata command xtivreg2, endog().(Schaffer 2012). P.value is smaller than 5% for all regressions

[80] The redundancy test is implemented using the Stata command xtivreg2, redundant().

weak instruments and ensuring proper inference, including:

- Underidentification LM test (Kleibergen & Paap 2006), where a rejection of the null hypothesis means that the model is not under-identified.
- Kleibergen-Paap F-statistic,(Kleibergen & Paap 2006) which is the heteroskedasticity-robust version of the Cragg-Donald statistic (Cragg & Donald 1993). Instruments are deemed relevant if this statistic exceeds the "Stock-Yogo" threshold (Stock & Yogo 2005) of a relative bias compared to OLS. [81]
- Anderson-Rubin (AR) test for significance of endogenous regressors (Anderson & Rubin 1949). A rejection of the null hypothesis means that inference can be made even if instruments are only weakly relevant.
- Fractionally re-sampled AR test (FAR), which is robust also when instruments are not fully exogenous (Berkowitz et al. 2012, Riquelme et al. 2013).

To minimize omitted variables, I use the sentencing database. The separation of government and defendant appeals remains, with the exception of the instrumented spread variable, which is in the aggregated form (as in table 7).[82] I use a two-stage GMM estimator, which is asymptotically more efficient than the traditional 2SLS estimator (Baum et al. 2007, Lee 2003). Explanatory variables and instruments are all centered.[83]

Table 10 presents the IV results. Column (1) includes the (instrumented) spread only, and each column adds further controls, until also separate government and defendant rates and quadratic/cubic terms are included. All tests indicate that the model is not underidentified and that the instruments are valid (and robust even if they would not be fully exogenous).[84] The IV results support the argument that increasing reversal-remand spreads and affirmation rates, ceteris paribus, cause crime to decrease. Moreover, this holds also for partial affirmations, providing further support for the base results (in which both kinds of affirmation were aggregated) and the argument that affirming and reversing influence crime in the same direction. In comparison to the base results, the coefficient of the spread is larger while the coefficient of affirmation is smaller. However, when taking into account

[81] For example, a threshold of 10% means that the IV estimator bias is smaller by at least 90 percentage points than the OLS bias.

[82] Since the instruments are taken from the aggregate appeal court level, they are naturally correlated with both government and defendant appeal results. To avoid leaving one category out, I use the aggregate sentencing spread.

[83] The lagged spread is centered before lagging, i.e. it is the lagged centered spread and not the centered lagged spread.

[84] As mentioned, a Hansen J test does not reject the null assumption of instrument exogeneity. Underidentification tests are all significant at the 5% level, thus the model is not underidentified. Kleibergen-Paap weak-IV test in columns (1)-(3) all have $F > 9.08$, which is the relevant Stock-Yogo cutoff value (10% relative bias). However, in columns (4) and (5), F is smaller. The Anderson-Rubin test readily rejects the null hypothesis of weak instruments. The FAR test indicates that inference is robust even if instruments are not 100% exogenous ($p < 0.05$ in all regressions).

the average rates of affirmations and the spread, affirmations still have a much larger average effect on crime rates. Interactions are insignificant, unless separated into government and defendant rates, in which case only the interaction between full affirmation and government appeals is significant. This latter finding is in line with the aforementioned argument, that a positive interaction effect on crime is driven by lower sentences and does not involve type I errors. Slightly outside the scope of discussion, it appears that dismissals have also a negative effect on crime - which means that only remands positively affect crime. The straightforward interpretation of the latter is that appeal courts are in fact more accurate than trial court or, at the very least, are perceived to be so. Then, any final decision by the appeal court implies less errors and more deterrence.

As a further robustness check, I use a second set of IV regressions to simultaneously instrument for both the spread and affirmation rates (partial and full). Since it is extremely hard to come up with multiple instruments that provide distinct sources of variation for each variable, I utilize instead the method proposed by Lewbel (2012), which can produce internal valid instruments from heteroskedastic errors.[85] This method, however, requires that at least one regressor is exogenous (see also Baum et al. 2012).[86] As mentioned, conviction rates may be exogenous but also the share of jury cases is plausibly exogenous, since it should only affect crime indirectly through judicial errors and conviction rates. Thus, for the purpose of this additional test, I assume that at least one of the following is exogenous: (1) conviction rates (2) interactions with conviction rates, or (3) share of jury cases.

Since these regressions include multiple endogenous variables, some tests reported in table 10 are no longer applicable. Namely, Stock-Yogo critical values are available up to three endogenous variables only, while the regressions include five variables. Thus, I instead use the tests developed by Angrist & Pischke (2008) and modified by Sanderson et al. (2015), which calculate separate underidentification and weak-IV statistics for each variable. These statistics can then be compared to the Stock-Yogo critical values for one endogenous variable. Additionally, I replace the FAR and AR tests by adjusted versions for multiple endogenous variables.[87]

Table 11 presents the results of the additional IV regressions. The table includes two sets of columns. In the first set (columns 1,2), appeal results are aggregated

[85] The regression is run using the user-developed command ivreg2h (Baum & Schaffer 2014).

[86] Additionally, in a recent paper, Chau (2015) argues that a specific form of heteroskedastic errors is needed for the Lewbel approach to generate valid instruments.

[87] The tests include a variation of a Wald test and of the Anderson-Rubin test, using the user-developed Stata command "weakiv" (Finlay et al. 2013).

for defendant and government appeals. In the second set (columns 3, 4) I replace the aggregate rates by defendant appeal rates (as in table 8). All variables possibly related to judicial errors are instrumented: spread, affirmation (full and partial), appeals filed and appeals decided. Within each set, the first column presents the results when only the newly generated instruments are used; and the second column adds the three "older" instruments (i.e. administrative, civil, and lagged criminal spreads) and, as suggested by Bun et al. (2014), also interactions of each "old" instrument with conviction rates. The results regarding the spread and full affirmation are similar to previous results. Interactions of full affirmations are insignificant at the 5% level, but interactions of partial affirmations are mostly significant, with a positive coefficient. If partial affirmation means reversing (only) the sentence, this again supports the argument of a punishment-driven interaction, not directly related to type I errors. Tests are reported in table 11 (at the bottom) and table 12 , and indicate that the instruments are exogenous and relevant.

To summarize, the robustness checks reinforce the base results, especially with respect to defendant appeals - which are the theoretically relevant kind for testing type I errors. An interaction effect between affirmation and conviction rates seems to be in play, but it is unclear what is driving the effect - disbelief in the system when convictions rates and affirmation rates are simultaneously high; or a side-effect of reduced sentencing when affirming government sentencing appeals or partially reversing defendant appeals.

9 Implications

My results offer several important implications, for both theory and practice. For the research community, this paper is the first to produce empirical evidence which support the key argument of the Extended Beckerian Model - that type I judicial errors have a detrimental effect on crime deterrence. However, the results also show that a theoretical adjustment is needed, such that error correction - and not only error occurrence - is incorporated in the model. Theory should further incorporate how an error is (potentially) corrected by appeal courts, as each method of correction - reversal or remand - has opposite effects on deterrence.

Since appeal results are correlated with other stages of the legal process, the results also imply that some previous evidence on crime deterrence may be somewhat inaccurate, insofar that the omitted variable bias caused by ignoring appeal results is unaccounted for. Although not at the heart of my argument, my findings also point out two interesting byproducts: (1) the importance of judicial errors with respect to sentencing errors; and (2) an asymmetric effect of these errors for appeals filed by defendants and appeals filed by the government.

For appeal judges, my results imply a need for caution when choosing how to correct an error (reversal vs. remand), such that judges who care about deterrence should prefer reversals to remands. It is important to note, however, that judges may already be aware of this effect and consider it in their decision making process, for example, by sorting high-deterrent verdicts (e.g. with a large public interest) into reversal category and vice versa (in other words, cases may "self-select" into categories). This would, however, require that judges regularly conduct ongoing statistical analysis, which seems unlikely.

A similar need for caution is implied for trial judges, such that judges who care about deterrence should be more error averse. This is especially true, since my findings suggest that preventing an error a-priori may have a much larger positive effect on deterrence than correcting an error ex-post. Trial judges who are then considering to free-ride on the appeal court's accuracy, should be aware that a retrospective remedy is not as efficient. Finally, the results are important also for policy makers, in several aspects. For judicial system planners, the effect of error occurrence and correction on deterrence should be integrated in a cost-benefit analysis, while deciding on resource allocation. From a social welfare perspective, judicial errors should only be avoided or corrected when the benefit from correction exceeds the social harm. My results are then especially important as practical tool that allows for the usage of actual numbers.

9.1 Final remarks on the effect of remands on deterrence

A remand procedure is usually justified by either (1) "institutional superiority" of the first instance to correct errors cheaper; (2) A narrow jurisdiction of appeal courts to directly correct errors; or (3) avoiding an unnecessary binding precedent (Hessick 2012). When remands accompany reversals, they may also serve as a restraining tool for appeal courts against trial court judges (Boyd 2015). The various costs of remands identified in the literature do not include deterrence effects. The above-mentioned various arguments that may explain these effect can be aggregated into three main competing arguments:[88]

[88] A fourth possible argument would be that a higher remand rate implies that trial court judges will have more work, resulting in a higher judicial load which then leads to more errors. The implicit loss of faith in the system is then larger.

1. **Delay in the sanction** - remands are perceived as a way for criminals to manipulate the justice system by prolonging their trial. Then, although remands only delay the inevitable, a higher remand rate suggests that sanctions are imposed at a later date (assuming a stay of sentence is given). Potential offenders then discount the expected sanction more heavily.
2. **Loss of faith in the system** - the public perceives remands as uncorrected errors, since the decision was neither affirmed nor reversed. The public also does not believe that it will be corrected in the future - otherwise we would expect no effect at all of remands.
3. **Cost asymmetry** - the public believes that remands are more costly for the innocents than for the guilty.

If my results are driven by the first explanation (delayed sanction), there is perhaps little to be done other than limiting the scope of remands or speeding up the legal process, which can be quite costly. Nonetheless, given that a stay of sentence is needed to delay the sanction, the problem can be mitigated by increasing the scrutiny of petitions for a stay of sentence, in order to reduce the frequency of guilty defendants whose sanctions are postponed.

If loss of faith in the system drives the effect, there is more leeway. Namely, since appeal courts do not only decide whether to remand or reverse but can also decide on *how* to remand, the problem can be attacked on several fronts. First, strengthening the appeal court's role in giving guidance to lower courts through instructions. For example, appeal courts should perhaps be given a stronger authority to provide more specific instructions to avoid ambiguity, which is an unfortunate result of generalized written opinions (Hessick 2012, pp. 5-6). The argument for using specific instructions is supported by recent empirical findings (Boyd 2015) on the effectiveness of specified instructions and published opinions on the district court's probability of changing the pre-appeal result.

Second, if the source of public disbelief is the fact that cases are remanded back to the same panel, appeal courts should perhaps be restricted from doing so. Hannibal & Worth (2012) note, that the authority to enforce reassignment to another judge[89] is reluctantly used, to avoid an implicit vote of no-confidence in the trial judge. Furthermore, some federal courts already have a formal rule in place that dictates when cases are reassigned and others have developed customary criteria (Scheinfeld & Bagley 2013). Unfortunately, my results would suggest that

[89]The authority to enforce reassignment to another judge is derived from 28 U.S.C 2106.

current checks and balances are still insufficient to prevent a decline in deterrence.

Finally, if a (perceived) cost-asymmetry explains the effect, additional compensation for wrongful convictions following a remand process may be considered. A cost reimbursement for acquitted defendant has been shown to theoretically balance the effect of judicial errors (Fon & Schäfer 2007), but has been argued to be ineffective (Kahn 2010) in the U.S.. Improving reimbursement mechanisms for post-remand acquittals may therefore provide relief to the problem. In any event, one common factor echos through all these explanations - the accuracy of trial courts. Unfortunately, since we do not know exactly why judicial errors occur, one can only speculate as to how accuracy can be improved.

There exists of course also one extreme measure that would circumvent all these problems at once: the abolishment of the remand process altogether. The full implications of such a step go far beyond the issue at hand, but even when focusing on deterrence alone, the costs of "not remanding" might be too high. Namely, restricting courts by removing the possibility to remand would force judge to alternate between affirmation, reversals and dismissal. It is unlikely that judges will switch remand-worthy cases to affirmation due to the constraints of (not) declaring an error, but some judges might then turn to reversal as the only available measure to correct errors. This might affect deterrence in two channels: first, the reversal category would become ambiguous and harder to interpret, leading to incorrect estimation of court accuracy. Second, the payoff from committing the crime may increase, since factual mistakes that would usually require remands may also happen when one is guilty. Furthermore, prosecutors may be reluctant to file indictments where questions of evidence are pivotal, given a higher chance of reversed convictions, leading a to lower prosecution rate. Conversely, incentivizing prosecutors to focus on strong cases may prevent wrongful prosecutions, thus lowering wrongful convictions. In an alternative scenario, judges who fear such ramifications may dismiss appeals that would be otherwise remanded. This might lead to a higher loss of faith in the system, as convictions will be less frequently overturned. Prosecutors will also have a higher incentive to prosecute weak cases, possibly leading to more wrongful convictions.

Summing up, there are numerous possible remedies to address the deterrence problem of remands, but a comprehensive analysis should be taken prior to the adoption of any change, in order to avoid other channels of deterrence infringement.

10 Conclusion

The empirical analysis of U.S. federal appeal results and crime rates, supports the key argument of the Extended Beckerian Model (Png 1986) - that (type I) judicial errors are detrimental to deterrence - but indicates that theory should be further extended to error correction by appeal courts. Namely, crime decreases when less type I errors occur, such that more trial decisions are affirmed. Furthermore, crime decreases when reversal rates rise but increases when remand rates rise.

Affirmation rates seem to have a stronger effect on crime than the reversal-remand spread, implying that preventing errors yields a higher deterrence gain than correcting errors on appeal. There is some evidence of an interaction effect between affirmation rates and conviction rates, but it seems to be limited to government appeals, which do not challenge convictions. These results are, however, subjected to several limitations: measurement errors may occur due to: aggregation of crime rates[90] and appeal results; usage of reported crimes (which may differ from actual crime); and the nature of the built-in classification system of appeal results, which may be too simplistic. The sample may also be non-random, since it only includes years for which data was published.

I address these concerns throughout the paper by using additional data, multiple regression models and an IV approach, but the results must still partially rely on (weak) assumptions of representativeness. Nonetheless, my results hold important implications for the research community and judicial system agents, as they imply that error correction - and not only error occurrence - affects crime deterrence. Further research should delve into this conclusion, and adjust the deterrence model to account for error correction and appeal results. Empirical research should also incorporate judicial errors and their correction into the list of deterrence variables, and perhaps attempt to also disentangle the effect of type II judicial errors in another country which allows appeal courts to review both types of errors.

[90] In addition to the aggregation on the state and year level, combining violent and property crimes may raise some objections, as these have been argued to respond differently to changes in the expected sanction. For example, some have argued that violent crimes are plausibly committed as an act of passion, while property crimes are more likely to be a result of a cost-benefit analysis (see, for example Dalla Pellegrina 2008, pp. 268). Indeed, some papers also find different elasticities for each type of crime (see Chalfin & McCrary 2014, pp. 23-25) as well as differences in time preferences (Beraldo et al. 2013). However, in my paper, separating crime rates raises a difficulty since the appeal results are aggregated, making it impossible to identify separate error rates for each type of crime. This is especially problematic, since property crime are vastly more frequent, making it unlikely to see an effect of overall rates on violent crimes. Unsurprisingly, using different specifications (not reported here, for the sake of brevity), the effects of the spread and affirmations are significant for property crimes but for violent crimes. As this may mean very little given the data constraints, I did not go into the details of this issue.

References

Afshartous, D. & Preston, R. A. (2011), 'Key results of interaction models with centering', *Journal of Statistics Education* **19**(3), 1–24.

Anderson, T. W. & Rubin, H. (1949), 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *The Annals of Mathematical Statistics* pp. 46–63.

Andreoni, J. (1991), 'Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?', *The RAND Journal of Economics* pp. 385–395.

Angrist, J. D. & Pischke, J.-S. (2008), *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Baker, T., Harel, A. & Kugler, T. (2003), 'Virtues of uncertainty in law: An experimental approach, the', *Iowa L. Rev.* **89**, 443.

Balli, H. O. & Sørensen, B. E. (2013), 'Interaction effects in econometrics', *Empirical Economics* **45**(1), 583–603.

Baum, C. (2001), 'Xttest3: Stata module to compute modified wald statistic for groupwise heteroskedasticity'.

Baum, C. F., Lewbel, A., Schaffer, M. E., Talavera, O. et al. (2012), Instrumental variables estimation using heteroskedasticity-based instruments, *in* 'UK Stata Users Group Meetings', pp. 13–14.

Baum, C. F. & Schaffer, M. E. (2014), 'Ivreg2h: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments', *Statistical Software Components* .

Baum, C. F., Schaffer, M. E. & Stillman, S. (2007), 'Enhanced routines for instrumental variables/gmm estimation and testing', *Stata Journal* **7**(4), 465–506.

Baumann, F. & Friehe, T. (2015), 'Proof beyond a reasonable doubt: Laboratory evidence', (181).

Beck, N. & Katz, J. N. (1995), 'What to do (and not to do) with time-series cross-section data', *American political science review* **89**(03), 634–647.

Becker, G. S. (1968), 'Crime and punishment: An economic approach', *The Journal of Political Economy* **76**(2), 169–217.

Beery, B. (2002), 'Be careful what you ask for: Navigating a remand after you've won the appeal', *Michigan Bar Journal, May* .

Bell, B. et al. (2014), 'Crime deterrence: Evidence from the london 2011 riots', *The Economic Journal* **124**(576), 480–506.

Beraldo, S., Caruso, R. & Turati, G. (2013), 'Life is now! time preferences and crime: Aggregate evidence from the italian regions', *The Journal of Socio-Economics* **47**, 73–81.

Berkowitz, D., Caner, M. & Fang, Y. (2012), 'The validity of instruments revisited', *Journal of Econometrics* **166**(2), 255–266.

Blackstone, W. (1765), 'Commentaries on the laws of england, 4 vols', *Oxford* **4**.

Blackwell III, J. L. (2005), 'Estimation and testing of fixed-effect panel-data systems', *Stata Journal* **5**(2), 202–207.

Boyd, C. L. (2015), 'The hierarchical influence of courts of appeals on district courts', *The Journal of Legal Studies* **44**(1), 113–141.

Bun, M. J., Harrison, T. D. et al. (2014), 'Ols and iv estimation of regression models including endogenous interaction terms', *University of Amsterdam Discussion Paper* **2**.

Calfee, J. E. & Craswell, R. (1984), 'Some effects of uncertainty on compliance with legal standards', *Virginia Law Review* pp. 965–1003.

Cameron, A. C. & Miller, D. L. (2015), 'A practitioners guide to cluster-robust inference', *Journal of Human Resources* **50**(2), 317–372.

Catterson, C. (2006), 'Changes in appellate caseload and its processing', *Ariz. L. Rev.* **48**, 287.

Caulkins, J. P. (1993), 'Zero-tolerance policies: do they inhibit or stimulate illicit drug consumption?', *Management Science* **39**(4), 458–476.

Chalfin, A. & McCrary, J. (2014), 'Criminal deterrence: A review of the literature'.

Chau, T. W. (2015), 'Identification through heteroscedasticity: What if we have the wrong form of heteroscedasticity?'.

Chen, M. K. & Shapiro, J. M. (2007), 'Do harsher prison conditions reduce recidivism? a discontinuity-based approach', *American Law and Economics Review* **9**(1), 1–29.

Chopard, B., Marion, E., Roussey, L. et al. (2014), 'Does the appeals process lower the occurrence of legal errors?'.

Commission, U. S. et al. (2012), 'Report on the continuing impact of united states v. booker on federal sentencing', *Washington, DC: US Sentencing Commission* .

Cornwell, C. & Trumbull, W. N. (1994), 'Estimating the economic model of crime with panel data', *The Review of economics and Statistics* pp. 360–366.

Correia, S. et al. (2015), 'Reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects', *Statistical Software Components* .

Cotte Poveda, A. (2011), 'Socio-economic development and violence: an empirical application for seven metropolitan areas in colombia', *Peace Economics, Peace Science and Public Policy* **17**(1).

Cragg, J. G. & Donald, S. G. (1993), 'Testing identifiability and specification in instrumental variable models', *Econometric Theory* **9**(02), 222–240.

Cummins, T., Aft, A. & Cumby, J. C. (2015), 'Appellate review iii', *Journal of Legal Metrics* .

Dalla Pellegrina, L. (2008), 'Court delays and crime deterrence', *European Journal of Law and Economics* **26**(3), 267–290.

Dari-Mattiacci, G. & Deffains, B. (2007), 'Uncertainty of law and the legal process', *Journal of Institutional and Theoretical Economics (JITE)* pp. 627–656.

Davis, M. L. (1988), 'Time and punishment: an intertemporal model of crime', *The Journal of Political Economy* pp. 383–390.

De Hoyos, R. E. & Sarafidis, V. (2006), 'Testing for cross-sectional dependence in panel-data models', *Stata Journal* **6**(4), 482.

DeAngelo, G. & Charness, G. (2012), 'Deterrence, expected cost, uncertainty and voting: Experimental evidence', *Journal of Risk and Uncertainty* **44**(1), 73–100.

Dhami, S. & al Nowaihi, A. (2013), 'An extension of the becker proposition to non-expected utility theory', *Mathematical Social Sciences* **65**(1), 10–20.

Di Tella, R. & Schargrodsky, E. (2013), 'Criminal recidivism after prison and electronic monitoring', *Journal of Political Economy* **121**(1), 28–73.

Drago, F., Galbiati, R. & Vertova, P. (2009), 'The deterrent effects of prison: Evidence from a natural experiment', *Journal of political Economy* **117**(2), 257–280.

Driscoll, J. C. & Kraay, A. C. (1998), 'Consistent covariance matrix estimation with spatially dependent panel data', *Review of economics and statistics* **80**(4), 549–560.

Drukker, D. M. et al. (2003), 'Testing for serial correlation in linear panel-data models', *Stata Journal* **3**(2), 168–177.

Dušek, L. (2015), 'Time to punishment: The effects of a shorter criminal procedure on crime rates', *International Review of Law and Economics* **43**, 134–147.

Eberhardt, M. (2011), 'Xtcd: Stata module to investigate variable/residual cross-section dependence', *Statistical Software Components* .

Ehrlich, I. (1982), 'The optimum enforcement of laws and the concept of justice: a positive analysis', *International Review of Law and Economics* **2**(1), 3–27.

Entorf, H. & Spengler, H. (2000), 'Criminality, social cohesion and economic performance', *WEP Wuerzburg Economic Papers No. 00-22* .

Epps, D. (2015), 'Consequences of error in criminal justice, the', *Harv. L. Rev.* **128**, 1065.

Epps, D. (2016), 'One last word on the blackstone principle', *Virginia Law Review Online* **102**.

Evans, W. N. & Owens, E. G. (2007), 'Cops and crime', *Journal of Public Economics* **91**(1), 181–201.

Feess, E., Schildberg-Hörisch, H., Schramm, M. & Wohlschlegel, A. (2015), 'The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory'.

Finlay, K., Magnusson, L. & Schaffer, M. (2013), 'weakiv: Weak-instrument-robust tests and confidence intervals for instrumental-variable (iv) estimation of linear, probit and tobit models', *URL: http://ideas. repec. org/c/boc/bocode/s457684. html* .

Fon, V. & Schäfer, H.-B. (2007), 'State liability for wrongful conviction: Incentive effects on crime levels', *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* pp. 269–284.

Frees, E. W. (1995), 'Assessing cross-sectional correlation in panel data', *Journal of econometrics* **69**(2), 393–414.

Friedman, D. (1999), 'Why not hang them all: The virtues of inefficient punishment', *Journal of Political Economy* **107**(S6), S259–S269.

Friedman, M. (1937), 'The use of ranks to avoid the assumption of normality implicit in the analysis of variance', *Journal of the American Statistical Association* **32**(200), 675–701.

Friesen, L. (2012), 'Certainty of punishment versus severity of punishment: An experimental investigation', *Southern Economic Journal* **79**(2), 399–421.

Gardner, R. (1998), 'Unobservable individual effects in unbalanced panel data', *Economics letters* **58**(1), 39–42.

Garoupa, N. & Rizzolli, M. (2012), 'Wrongful convictions do lower deterrence', *Journal of Institutional and Theoretical Economics (JITE)* **168**(2), 224–231.

Glaeser, E. L. & Sacerdote, B. (1999), 'Why is there more crime in cities?', *Journal of Political Economy* **107**(6 pt 2).

Gneezy, U. & Rustichini, A. (2004), 'Incentives, punishment and behavior', *Advances in behavioral economics* pp. 572–89.

Gould, E. D. et al. (2002), 'Crime rates and local labor market opportunities in the united states: 1979–1997', *Review of Economics and Statistics* **84**(1), 45–61.

Gross, S. R. & O'Brien, B. (2008), 'Frequency and predictors of false conviction: Why we know so little, and new data on capital cases', *Journal of Empirical Legal Studies* **5**(4), 927–962.

Gross, S. R., OBrien, B., Hu, C. & Kennedy, E. H. (2014), 'Rate of false conviction of criminal defendants who are sentenced to death', *Proceedings of the National Academy of Sciences* **111**(20), 7230–7235.

Hannibal, K. & Worth, J. A. (2012), 'Judicial reassignment: a proposal', *The National Law Journal* .

Harel, A. & Klement, A. (2007), 'The economics of stigma: Why more detection of crime may result in less stigmatization', *The Journal of Legal Studies* **36**(2), 355–377.

Harris, J. R. (1970), 'On the economics of law and order', *The Journal of Political Economy* pp. 165–174.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica: Journal of the Econometric Society* pp. 1251–1271.

Hessick, F. A. (2012), 'Cost of remands, the', *Ariz. St. LJ* **44**, 1025.

Hettinger, V. A. & Lindquist, S. A. (2012), 'Decision making in the us courts of appeals', *New directions in judicial politics* p. 126.

Hoechle, D. (2007), 'Robust standard errors for panel regressions with cross-sectional dependence', *Stata Journal* **7**(3), 281.

Ihlanfeldt, K. R. (2007), 'Neighborhood drug crime and young males' job accessibility', *The Review of Economics and Statistics* **89**(1), 151–164.

Jensen, E. G. & Heller, T. C. (2003), *Beyond common knowledge: empirical approaches to the rule of law*, Stanford University Press.

Johnson, E. & Payne, J. (1986), 'The decision to commit a crime: An information-processing analysis', *The reasoning criminal: Rational choice perspectives on offending* pp. 170–185.

Johnson, J. S. (2016), 'Benefits of error: A dynamic defense of the blackstone principle in criminal law', *Virginia Law Review* **102**.

Kahn, D. S. (2010), 'Presumed guilty until proven innocent: The burden of proof in wrongful conviction claims under state compensation statutes', *U. Mich. JL Reform* **44**, 123.

Kaplow, L. (1994), 'The value of accuracy in adjudication: An economic analysis', *The Journal of Legal Studies* pp. 307–401.

Kaplow, L. (2011), 'On the optimal burden of proof', *Journal of Political Economy* **119**(6), 1104–1140.

Khadjavi, M. (2014*a*), 'Deterrence works for criminals'.

Khadjavi, M. (2014*b*), 'On the interaction of deterrence and emotions'.

Kim, S., Park, J., Park, K. & Eom, J.-S. (2013), 'Judge-jury agreement in criminal cases: The first three years of the korean jury system', *Journal of Empirical Legal Studies* **10**(1), 35–53.

Kleibergen, F. & Paap, R. (2006), 'Generalized reduced rank tests using the singular value decomposition', *Journal of econometrics* **133**(1), 97–126.

Kovandzic, T. V. & Vieraitis, L. M. (2006), 'The effect of county-level prison population growth on crime rates', *Criminology and Public Policy* **5**(2), 213–244.

Lando, H. (2006), 'Does wrongful conviction lower deterrence?', *The Journal of Legal Studies* **35**(2), 327–337.

Lee, D. S., McCrary, J. et al. (2009), *The deterrence effect of prison: Dynamic theory and evidence*, Citeseer.

Lee, L.-f. (2003), 'Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances', *Econometric Reviews* **22**(4), 307–335.

Levitt, S. D. (1995), 'The effect of prison population size on crime rates: Evidence from prison overcrowding litigation'.

Levitt, S. D. (1997), 'Using electoral cycles in police hiring to estimate the effect of police on crime', *The American Economic Review* pp. 270–290.

Levitt, S. D. (1998), 'Juvenile crime and punishment', *The Journal of Political Economy* **106**(6), 1156–1185.

Levitt, S. D. (2002), 'Using electoral cycles in police hiring to estimate the effects of police on crime: Reply', *American Economic Review* pp. 1244–1250.

Levy, G. (2005), 'Careerist judges and the appeals process', *RAND Journal of Economics* pp. 275–297.

Lewbel, A. (2012), 'Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models', *Journal of Business & Economic Statistics* **30**(1), 67–80.

Lewisch, P., Ottone, S. & Ponzano, F. (2015), 'Third-party punishment under judicial review: An economic experiment on the effects of a two-tier punishment system', *Review of Law & Economics* **11**(2), 209–230.

Lin, M.-J. (2008), 'Does unemployment increase crime? evidence from us data 1974–2000', *Journal of Human Resources* **43**(2), 413–436.

Lundberg, A. (2016), 'Sentencing discretion and burdens of proof', *International Review of Law and Economics* .

Machin, S. & Meghir, C. (2004), 'Crime and economic incentives', *Journal of Human Resources* **39**(4), 958–979.

Masur, J. & Bronsteen, J. (2015), The overlooked benefits of the blackstone principle, *in* 'Harvard Law Review Forum', Vol. 128, p. 289.

McGuire, K. T. (2012), *New Directions in Judicial Politics*, Routledge.

Mungan, M. C. & Lando, H. (2015), 'The effect of type-1 error on deterrence'.

Mustard, D. B. (2003), 'Reexamining criminal behavior: the importance of omitted variable bias', *Review of Economics and Statistics* **85**(1), 205–211.

Nicita, A. & Rizzolli, M. (2014), 'In dubio pro reo. behavioral explanations of pro-defendant bias in procedures', *CESifo Economic Studies* p. ift016.

Nizalova, O. Y. & Murtazashvili, I. (2016), 'Exogenous treatment and endogenous factors: Vanishing of omitted variable bias on the interaction term', *Journal of Econometric Methods* **5**(1), 71–77.

Osgood, D. W. (2000), 'Poisson-based regression analysis of aggregate crime rates', *Journal of quantitative criminology* **16**(1), 21–43.

Ouss, A. & Peysakhovich, A. (2013), 'When punishment doesn't pay:'cold glow'and decisions to punish', *Available at SSRN 2247446* .

Perry, H. (1998), 'United states attorneys: Whom shall they serve?', *Law and Contemporary Problems* pp. 129–148.

Pesaran, M. (2004), 'General diagnostic tests for cross section dependence in panels'.

Pfander, J. E. (2007), 'Removing federal judges', *The University of Chicago Law Review* pp. 1227–1250.

Plerhoples, C. & Summit, C. F. P. (2012), 'The effect of vacant building demolitions on crime under depopulation', *Economics job market paper. Lansing: Michigan State University, Department of Agricultural, Food, and Resources* .

Png, I. P. (1986), 'Optimal subsidies and damages in the presence of judicial error', *International Review of Law and Economics* **6**(1), 101–105.

Polinsky, A. M. & Shavell, S. (1999), The economic theory of public enforcement of law, Technical report, National bureau of economic research.

Polinsky, A. M. & Shavell, S. (2007), 'The theory of public enforcement of law', *Handbook of law and economics* **1**, 403–454.

Posner, R. A. (1973), 'An economic approach to legal procedure and judicial administration', *The Journal of Legal Studies* pp. 399–458.

Ridgway, J. D. (2009), 'Why so many remands?: A comparative analysis of appellate review by the united states court of appeals for veterans claims', *A Comparative Analysis Of Appellate Review by the United States Court Of Appeals for Veterans Claims* **1**, 113.

Riquelme, A., Berkowitz, D., Caner, M. et al. (2013), 'Valid tests when instrumental variables do not perfectly satisfy the exclusion restriction', *The Stata Journal* **13**, 528–546.

Rizzolli, M. & Saraceno, M. (2013), 'Better that ten guilty persons escape: punishment costs explain the standard of evidence', *Public choice* **155**(3-4), 395–411.

Rizzolli, M. & Stanca, L. (2012), 'Judicial errors and crime deterrence: theory and experimental evidence', *Journal of Law and Economics* **55**(2), 311–338.

Roberts, J. V. & Hough, M. (2005), 'The state of the prisons: exploring public knowledge and opinion', *The Howard Journal of Criminal Justice* **44**(3), 286–306.

Sah, R. K. (1991), 'Social osmosis and patterns of crime: A dynamic economic analysis', *Journal of political Economy* **99**(6).

Sanderson, E., Windmeijer, F. et al. (2015), 'A weak instrument f-test in linear iv models with multiple endogenous variables'.

Schaffer, M. E. (2012), 'xtivreg2: Stata module to perform extended iv/2sls, gmm and ac/hac, liml and k-class regression for panel data models', *Statistical Software Components* .

Scheinfeld, R. C. & Bagley, P. H. (2013), 'Reassignment to a new judge after remand from the federal circuit', *The New York law journal* **249**(98).

Schildberg-Hörisch, H. & Strassmair, C. (2012), 'An experimental test of the deterrence hypothesis', *Journal of Law, Economics, and Organization* **28**(3), 447–459.

Scott, K. M. (2006), 'Understanding judicial hierarchy: Reversals and the behavior of intermediate appellate judges', *Law & society review* **40**(1), 163–192.

Shavell, S. (1995), 'The appeals process as a means of error correction', *The Journal of Legal Studies* **24**(2), 379–426.

Shavell, S. (2004), 'The appeals process and adjudicator incentives'.

Sherman, L. W., Gartin, P. R. & Buerger, M. E. (1989), 'Hot spots of predatory crime: Routine activities and the criminology of place*', *Criminology* **27**(1), 27–56.

Smith, L. & Vásquez, J. (2015), 'Crime and vigilance', *Available at SSRN 2629321* .

Soares, Y. & Sviatschi, M. M. (2010), 'Does court efficiency have a deterrent effect on crime? evidence for costa rica', *Unpublished manuscript, available a t http://www. inesad. edu. bo/bcde2012/papers/7.% 20Sviatschi_Crime% 20and% 20Efficiency. pdf* .

Spencer, B. D. (2007), 'Estimating the accuracy of jury verdicts', *Journal of Empirical Legal Studies* **4**(2), 305–329.

Stith, K. (1990), 'The risk of legal error in criminal cases: Some consequences of the asymmetry in the right to appeal', *The University of Chicago Law Review* pp. 1–61.

Stock, J. H. & Yogo, M. (2005), 'Testing for weak instruments in linear iv regression', *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* .

Torre, A. (2003), 'The impact of court delays on the prosecutor and the defendant: An economic analysis', *European journal of law and economics* **16**(1), 91–111.

Tversky, A. & Kahneman, D. (1973), 'Availability: A heuristic for judging frequency and probability', *Cognitive psychology* **5**(2), 207–232.

Tversky, A. & Kahneman, D. (1974), 'Judgment under uncertainty: Heuristics and biases', *science* **185**(4157), 1124–1131.

Tversky, A. & Kahneman, D. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica: Journal of the Econometric Society* pp. 263–291.

Tversky, A. & Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and uncertainty* **5**(4), 297–323.

Weatherburn, D. (2012), 'The effect of arrest and imprisonment on crime'.

Weisburd, D., Morris, N. A. & Groff, E. R. (2009), 'Hot spots of juvenile crime: a longitudinal study of arrest incidents at street segments in seattle, washington', *Journal of Quantitative Criminology* **25**(4), 443–467.

Wohlschlegel, A. (2014), 'The appeals process and incentives to settle', *Available at SSRN 2335937* .

Wooldridge, J. (2012), *Introductory econometrics: A modern approach*, Cengage Learning.

Worrall, J. L. & Pratt, T. C. (2004), 'Estimation issues associated with time-series-cross-section analysis in criminology', *W. Criminology Rev.* **5**, 35.

Zamir, E. & Ritov, I. (2012), 'Loss aversion, omission bias, and the burden of proof in civil litigation', *The Journal of Legal Studies* **41**(1), 165–207.

A Tables and Figures

Figure 1: Judicial district map (source: http://www2.fjc.gov/sites/default/files/2012/IJR00007.pdf)

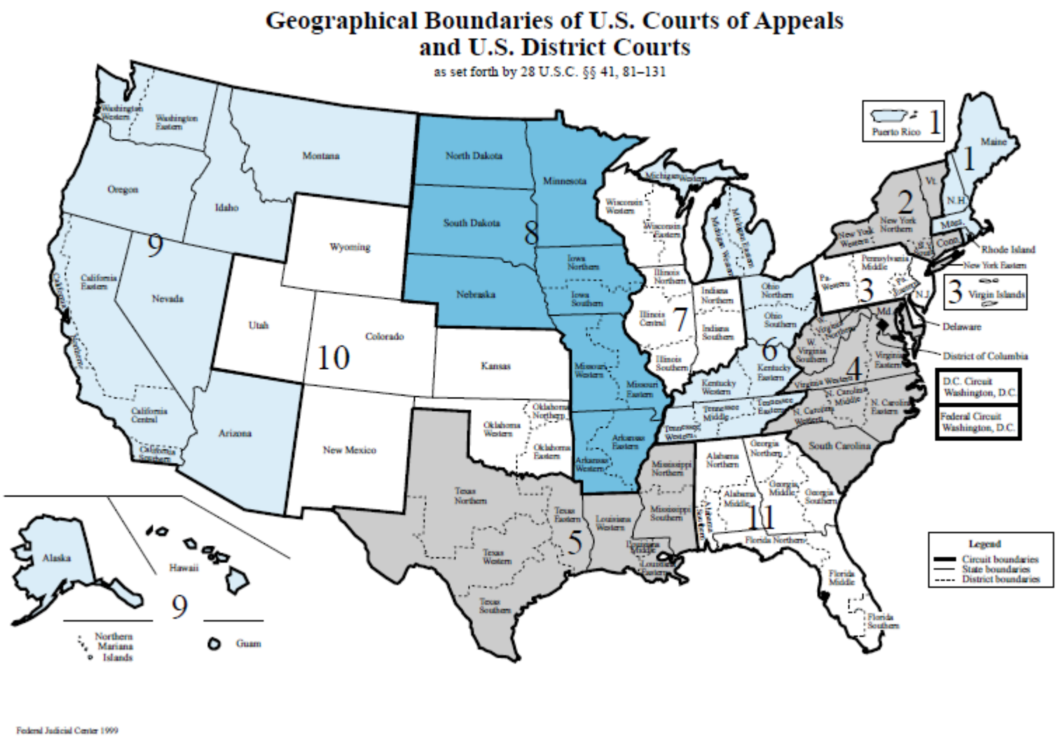
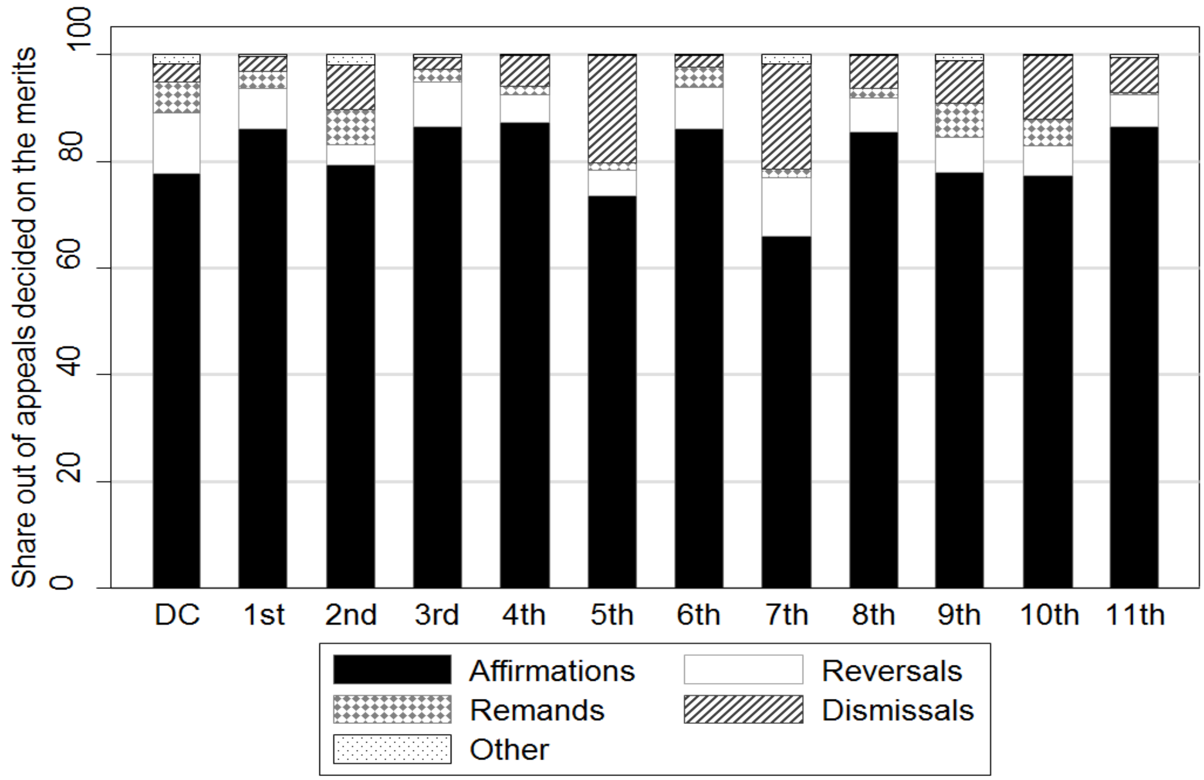


Figure 2: Appeal results - means

(1) Appeal results by district



(2) Appeal results by year

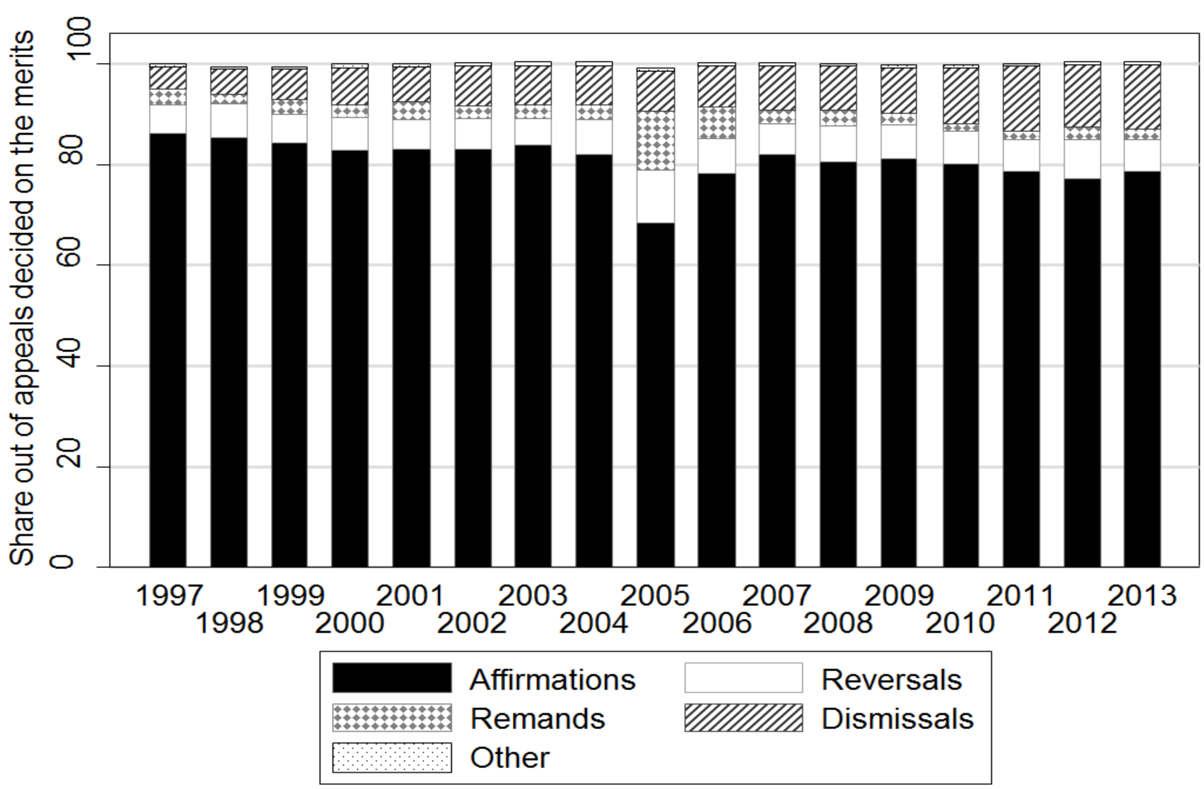


Figure 3: Average marginal effect of affirmation rates - base results

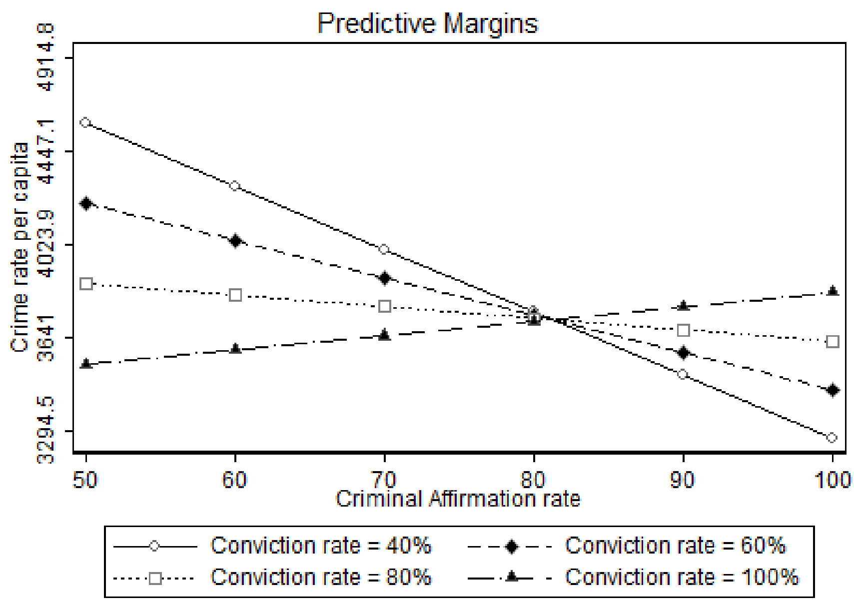


Figure 4: Average marginal effects - with quadratic and cubic terms

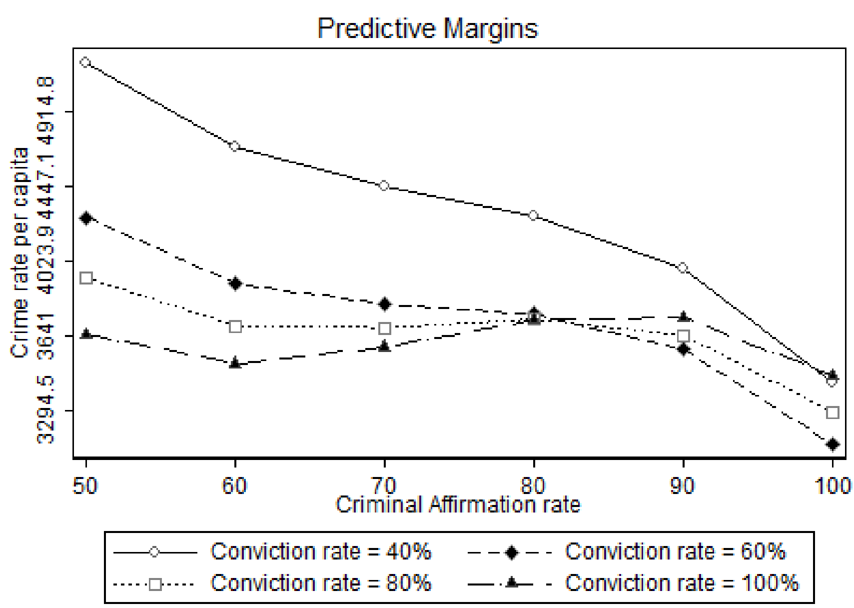


Table 1
DEFINITION OF VARIABLES

Variable	Abbreviation	Description/definition
Panel A: dependent variables		
Total crimes	$tot\_crime\_s_{i,t}$	Total number of crimes (property and violent) in the state of court i, at year t .
Logged crime per capita	$ln\_cp100_{i,t}$	100* Natural logarithm of total crime per capita (per 100,000 people in the state of court i), at year t.
Panel B: federal courts data		
Affirmation rate	$aff_{i,t}$	Affirmation rate out of appeals decided on the merits in the appeal court which district court i belongs to, at year t.
Reversal-Remand Spread	$spread_{i,t}$	Reversal rate minus remand rate out of appeals decided on the merits in the appeal court which district court i belongs to, at year t.
Prosecuted defendants per capita	$defp100_{i,t}$	Prosecuted defendants per capita in the state of court i, at year t.
Conviction rate	$conrate_{i,t}$	Conviction rate out of "terminated defendants" in district court i at year t.
Share of cases decide by a jury	$juryshare_{i,t}$	Share of cases decided by a jury, in district court i at year t.
Appeals filed per capita	$appfp100_{i,t}$	Number of criminal appeals filed per capita in the state of court i, at year t.
Appeals decided on the merits	$app\_merits_{i,t}$	Number of criminal appeals decided on the merits in the appeal court which district court i belongs to, at year t.
Imprisonment rate	$imp_{i,t}$	Share of defendants sentenced to prison out of all convicted defendants in district court i at time t.
Fine-only rate	$fine_{i,t}$	Share of defendants who received a fine-only sentence out of all convicted defendants in district court i at time t.
Dismissal rate	$dism_{i,t}$	Share of dismissed criminal appeals out of appeals decided on the merits in the appeal court which district court i belongs to, at year t.
Remainder rate	$other_{i,t}$	Remainder share of criminal appeals decided on the merits (not categorized as affirmed, reversed, remanded or dismissed)
Panel C: sentencing appeals database		
Full Affirmation rate	$AFsen_{i,t}$	Full affirmation rate in sentencing appeals decided in court i at year t.
Partial affirmation rate	$ARsen_{i,t}$	Partial affirmation rate in sentencing appeals decided in court I at year t.
Reversal-Remand spread	$SPREADsen_{i,t}$	Reversal-Remand spread of sentencing appeals in court i at year t.
Number of sentencing appeals	$apptype\_sen_{i,t}$	Number of sentencing appeals decided in court i at year t.
Average sentence length	$sen\_length_{i,t}$	Average sentence length (in months) imposed on defendants in court i at year t.
Full Affirmation rate (defendant)	$AFsendef_{i,t}$	Full affirmation rate in defendant-filed sentencing appeals in court i at year t.
Partial Affirmation rate (defendant)	$ARsendef_{i,t}$	Partial affirmation rate in defendant-filed sentencing appeals in court i at year t.
Full Affirmation rate (government)	$AFsengov_{i,t}$	Full affirmation rate in government-filed sentencing appeals that were fully affirmed in court I at year t.
Partial Affirmation rate (government)	$ARsengov_{i,t}$	Partial affirmation rate in government-filed sentencing appeals in court I at year t.
Reversal-remand spread (defendant)	$SPREADsendef_{i,t}$	Reversal-remand spread of defendant-filed sentencing appeals in court i at year t.
Reversal-remand spread (government)	$SPREADsengov_{i,t}$	Reversal-remand spread of government-filed sentencing appeals in court i at year t.
Panel D: instruments		
Civil reversal-remand spread	$civspread_{i,t}$	Share of reversed decision minus the share of remanded decisions out of civil appeals decided on merits in the appeal court which district court i belongs to, at year t.
Administrative reversal-remand spread	$adminspread_{i,t}$	Reversal rate minus remand rate in administrative appeals decided on merits in the appeal court which district court i belongs to, at year t.

Table 2

DESCRIPTIVE STATISTICS: FEDERAL COURTS DATABASE

	mean	sd	min	max
Total crimes in the state	321503.62	328709.46	13051.00	1569949.00
Total crime per capita	3848.57	1044.05	1581.71	9839.13
Logged crime rate * 100	821.84	27.49	736.63	919.41
Criminal Affirmation rate	80.93	8.91	54.75	93.68
Criminal Reversal rate	6.61	3.25	0.68	24.07
Criminal Remand rate	3.17	4.44	0.00	31.04
Criminal Dismissal rate	8.72	7.36	0.00	36.49
Criminal Remainder rate	0.57	1.21	0.00	14.29
Criminal Reversal-Remand spread	3.44	5.97	-24.83	20.37
Prosecuted defendants per capita	19.34	23.50	1.98	195.90
Share of jury cases	4.46	2.78	0.00	24.49
Conviction rate	89.33	6.62	46.55	99.01
Appeals filed per capita	2.63	2.93	0.18	29.89
Imprisonment rate	80.10	11.53	14.93	98.12
Fine-only rate	2.97	7.33	0.00	67.84
Criminal appeals decided on the merits	851.53	482.53	39.00	2784.00
Observations	1546			

NOTE.– This table presents descriptive statistics for the variables taken from the 'Judicial business of the U.S courts' over the year range 1997-2013. Appeal results are taken from the appeal court level.

Table 3

COMPARISON OF METHODS

Method	Advantages	Disadvantages	Courts/Crime paper examples
Fixed effects, two-way clustered standard errors ("Two-way cluster")	Robust for heteroskedasticity; Allows for correlation within-court and within-year; does not require strong assumptions	Weaker performance given spatial correlation and low number of clusters	Beraldo et al. (2013).
Fixed effects, Panel corrected standard errors ("PCSE")	Robust for heteroskedasticity, auto-correlation and contemporaneous correlation	Assumes very specific auto-regressive process: AR(1) assumes one-period correlation disturbance, which is the same for all panels. PSAR(1) assumes panel-specific one-period auto correlation	Kovandzic & Vieraitis (2006), Scott (2006)
Fixed effects, Driscoll-Kraay standard errors ("SCC")	Robust for heteroskedasticity, auto-correlation, spatial correlation and large N (number of panels)	Weaker performance for panels with small T (number of periods, i.e. number of years) or little spatial correlation	Cotte Poveda (2011)
Fixed effects, robust-Poisson ("POISSON")	Appropriate for count variables; robust to outliers	Does not deal with SE correlation problems	Plerhoples & Summit (2012)

Table 4
BASE RESULTS - COMPARISON OF METHODS

	(1) Two-way cluster (court,year)	(2) Panel-corrected (AR1)	(3) Panel-corrected (PSAR1)	(4) Driscol-Kraay (SCC)	(5) Poisson (robust)
main					
Criminal Reversal-Remand spread	-0.157***	-0.080**	-0.070**	-0.157***	-0.002***
	(0.001)	(0.043)	(0.046)	(0.000)	(0.000)
Criminal Affirmation rate	-1.232***	-0.627***	-0.556***	-1.232***	-0.012*
	(0.000)	(0.001)	(0.002)	(0.001)	(0.064)
Criminal Affirmation * Conviction	0.014***	0.007***	0.005***	0.014***	0.000*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.052)
Conviction rate	-1.124***	-0.521***	-0.439***	-1.124***	-0.012**
	(0.001)	(0.002)	(0.003)	(0.001)	(0.049)
Criminal Dismissal rate	0.019	-0.015	-0.042	0.019	0.003*
	(0.879)	(0.844)	(0.557)	(0.851)	(0.064)
Criminal Remainder rate	-0.175	-0.088	-0.143	-0.175	0.007**
	(0.680)	(0.609)	(0.371)	(0.680)	(0.014)
Criminal appeals decided on the merits	0.003	0.001	0.001	0.003**	0.000***
	(0.136)	(0.374)	(0.655)	(0.031)	(0.000)
Prosecuted defendants per capita	0.022	0.043*	0.057**	0.022	0.001
	(0.733)	(0.065)	(0.011)	(0.560)	(0.418)
Share of jury cases	-0.187	-0.046	-0.042	-0.187**	-0.000
	(0.210)	(0.412)	(0.424)	(0.030)	(0.862)
Fine-only rate	-0.009	-0.051	-0.040	-0.009	-0.001
	(0.893)	(0.261)	(0.283)	(0.826)	(0.163)
Imprisonment rate	0.071	0.039	0.038	0.071**	0.001
	(0.335)	(0.160)	(0.146)	(0.012)	(0.156)
Appeals filed per capita	0.061	-0.209	-0.222	0.061	0.003
	(0.859)	(0.277)	(0.188)	(0.794)	(0.505)
Dependent variable	ln_cp100	ln_cp100	ln_cp100	ln_cp100	tot_crime_s
Adjusted R-squared	0.945			0.945	
Log likelihood	-5012.463			-5012.463	-747403.159
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	1546	1546	1546	1546	1546

NOTE.– This table shows regression results of the basic model for the sample taken from the 'Judicial business of the U.S courts' over the year range 1997-2013. Appeal results are taken from the appeal court level. Each column presents the results of one of the five different methods: (1) Two-way clustered standard errors (by court and year); (2) Panel corrected standard errors, assuming an auto-regressive 1 disturbance ; (3) Panel corrected standard errors, assuming a panel-specific auto-regressive 1 disturbance; (4) Driskol-Kraay standard errors, also known as 'SCC'; (5) Poisson regression with roust standard errors. All regressions include court and year fixed effects. Constant and fixed effect coefficients are not reported. P.values are in parentheses under each coefficient. Note that the Poisson regression uses a different dependent variable than the other regressions. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 5

ROBUSTNESS CHECK: CENTERED INTERACTION WITH QUADRATIC AND CUBIC TERMS

	(1) Two-way cluster (court,year)	(2) Panel-corrected (AR1)	(3) Panel-corrected (PSAR1)	(4) Driscol-Kraay (SCC)	(5) Poisson (robust)	(6) Uncentered interaction
main						
Criminal Reversal-Remand spread	-0.215***	-0.102**	-0.090**	-0.215***	-0.002***	-0.199***
	(0.000)	(0.013)	(0.014)	(0.000)	(0.000)	(0.000)
Criminal Affirmation rate	-12.190**	-3.895	-3.352	-12.190***	0.041	-13.387**
	(0.018)	(0.223)	(0.246)	(0.001)	(0.336)	(0.011)
Criminal Affirmation squared	0.172**	0.055	0.047	0.172***	-0.000	0.172**
	(0.013)	(0.210)	(0.235)	(0.000)	(0.442)	(0.016)
Criminal affirmation cubic	-0.001**	-0.000	-0.000	-0.001***	0.000	-0.001**
	(0.010)	(0.195)	(0.217)	(0.000)	(0.548)	(0.013)
Conviction rate squared	0.068**	0.011	0.008	0.068*	0.000	0.060**
	(0.028)	(0.645)	(0.731)	(0.073)	(0.739)	(0.045)
Conviction rate cubic	-0.000**	-0.000	-0.000	-0.000*	-0.000	-0.000*
	(0.047)	(0.625)	(0.707)	(0.080)	(0.808)	(0.076)
Criminal Affirmation * Conviction (both rates centered)	0.012**	0.010*	0.008	0.012**	0.000	
	(0.049)	(0.063)	(0.105)	(0.040)	(0.191)	
Criminal Affirmation * Conviction						0.012***
						(0.001)
Dependent variable	ln_cp100	ln_cp100	ln_cp100	ln_cp100	tot_crime_s	ln_cp100
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.946			0.946		0.946
Log likelihood	-5001.451			-5001.451	-739318.799	-4997.302
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1546	1546	1546	1546	1546	1546

NOTE.– This table compares regression results of the five methods used in the basic model (see table 4), while additionally controlling for quadratic and cubic terms of affirmation rates. The first 5 columns include an interaction of centered (demeaned by court) affirmation and conviction rates, while the last column includes and uncentered interaction. Other control variables are identical to those used in table 4, i.e. dismissal rate; remainder rate; appeals decided on the merits; prosecuted defendants per capita; share of jury cases; imprisonment rate; fine-only rate; appeals filed per capita. All regressions include court and year fixed effects. Constant and fixed effect coefficients are not reported. P.values are in parentheses under each coefficient. Note that the Poisson regression uses a different dependent variable than the other regressions. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 6

Descriptive statistics - sentencing commission database

	mean	sd	min	max
Full affirmation rate (senetence)	75.47	12.58	29.73	100.00
Partial affirmation rate (sentence)	3.69	4.26	0.00	28.57
Reversal rate (sentence)	9.41	6.97	0.00	53.45
Remand rate (sentence)	2.88	4.98	0.00	45.95
Dismissal rate (sentence)	8.55	8.95	0.00	44.30
Reversal-remand spread (sentence)	6.53	8.03	-39.13	50.00
Affirmation rate (defendant appeals)	76.41	12.59	29.09	100.00
Reversal rate (defendant appeals)	8.52	6.67	0.00	56.36
Partial affirmation rate (defendant appeals)	3.58	4.05	0.00	26.32
Remand rate (defendant appeals)	2.83	4.92	0.00	45.95
Dismissal rate (defendant appeals)	2.83	4.92	0.00	45.95
Reversal-Remand spread (defendant appeals)	5.69	7.73	-39.13	52.73
Affirmation rate (government appeals)	13.98	30.30	0.00	100.00
Reversal rate (government appeals)	26.77	40.06	0.00	100.00
Partial affirmation rate (government appeals	4.41	17.55	0.00	100.00
Remand rate (government appeals)	2.47	13.30	0.00	100.00
Dismissal rate government appeals	0.44	5.87	0.00	100.00
Reversal-Remand spread (government appeals)	24.30	42.63	-100.00	100.00
Appeals decided - total	90.86	86.85	4.00	916.00
Appeals decided - sentence only	52.13	57.51	0.00	689.00
Appeals decided - sentence+conviction	17.16	17.98	0.00	207.00
Appeals decided - conviction only	21.57	18.19	1.00	129.00
Appeals decided - sentence	69.74	72.50	3.00	874.00
Share of government appeals out of all sentencing appeals	1.95	3.26	0.00	23.33
Observations	728			

NOTE.– This table presents descriptive statistics for the sample taken from the U.S sentencing commission online database over the range 2006-2013. Appeal results are measured at the trial court level (i.e. each court is represented by the respective rates of that court). Note that only appeals results regarding the sentence are included (no conviction only appeals).

Table 7

ROBUSTNESS CHECK: SENTENCING DATABASE BASE REGERESSIONS

	(1) Two-way cluster (court,year)	(2) Panel-corrected (AR1)	(3) Panel-corrected (PSAR1)	(4) Driscol-Kraay	(5) Poisson (robust)	(6) Quadratic and Cubic terms
main						
Reversal-remand spread (sentence) (centered)	-0.093***	-0.061*	-0.052*	-0.093***	-0.001***	-0.089***
	(0.003)	(0.055)	(0.082)	(0.003)	(0.000)	(0.002)
Full affirmation rate (senetence) (centered)	-0.065**	-0.051**	-0.047**	-0.065**	-0.001***	-0.024
	(0.031)	(0.039)	(0.035)	(0.038)	(0.000)	(0.576)
Full affirmation (centered) squared						0.000
						(0.968)
Full affirmation (centered) cubic						-0.000*
						(0.068)
Partial affirmation rate (sentence) (centered)	-0.051	-0.030	-0.027	-0.051	-0.002***	-0.123***
	(0.338)	(0.431)	(0.448)	(0.160)	(0.001)	(0.008)
Partial affirmation (centered) squared						-0.025**
						(0.030)
Partial affirmation (centered) cubic						0.002**
						(0.012)
Full Affirmation * conviction (both rates centered)	-0.002	-0.000	-0.001	-0.002	0.000	-0.002
	(0.654)	(0.954)	(0.795)	(0.617)	(0.229)	(0.655)
Partial Affirmation * Conviction (both rates centered)	-0.008	-0.004	-0.003	-0.008	-0.000	-0.005
	(0.252)	(0.715)	(0.680)	(0.393)	(0.846)	(0.503)
Conviction rate (centered) squared						-0.001
						(0.925)
Conviction rate (centered) cubic						-0.000
						(0.976)
Dependent variable	ln_cp100	ln_cp100	ln_cp100	ln_cp100	ln_cp100	tot_crime_s
Adjusted R-squared	0.969			0.969		0.969
Log likelihood	-2040.082			-2040.082	-114196.388	-2031.783
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	728	728	728	728	728	728

NOTE.– This table show compares regression results of the five methods used in the basic model, using the robustness database gathered from the U.S. sentencing commission for the year range 2006-2013. Appeals results are at the district court level. Quadratic and cubic terms of the appeal results are controlled for in the last column. Other control variables are: dismissal rate; sentencing appeals decided; prosecuted defendants per capita; share of jury cases; imprisonment rate; fine-only rate; share of imprisonment penalty; appeals filed per capita; average sentence length. All regressions include court and year fixed effects. Constant and fixed effect coefficients are not reported. P.values are in parantheses under each coefficient. Note that the Poisson regression uses a different dependent variable than the other regressions. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 8

ROBUSTNESS CHECK: DEFENDANT V. GOVERNMENT APPEALS

	(1) Defendant-filed appeals	(2) Government-filed appeals	(3) Defendant & Government-appeals
Reversal-Remand spread (defendant appeals) (centered)	-0.108**		-0.112***
	(0.015)		(0.007)
Reversal-Remand spread (government appeals) (centered)		0.001	0.004
		(0.896)	(0.284)
Affirmation rate (defendant appeals) (centered)	-0.070**		-0.070**
	(0.048)		(0.038)
Defendant Full affirmation * Conviction (both rates centered)	-0.002		
	(0.557)		
Affirmation rate (government appeals) (centered)		-0.002	0.003
		(0.567)	(0.605)
Government Full affirmation * Conviction (both rates centered)		0.003***	
		(0.009)	
Partial affirmation rate (defendant appeals) (centered)	-0.063		-0.056
	(0.300)		(0.330)
Defendant Partial affirmation * Conviction (both rates centered)	-0.009		
	(0.229)		
Partial affirmation rate (government appeals) (centered)		-0.006	-0.006
		(0.438)	(0.228)
Government Partial affirmation * Conviction (both rates centered)		-0.002	
		(0.253)	
Dismissal rate (defendant appeals) (centered)	-0.020		-0.019
	(0.598)		(0.626)
Dismissal rate (government appeals) (centered)		0.038***	0.034***
		(0.005)	(0.002)
R-squared	0.969	0.968	0.969
Time fixed effects	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Number of observations	728	728	728

NOTE.– This table shows results for two-way clustered (by court and year) regression, using separate variables for defendant-filed and government filed-appeals. Column (1) includes only defendant appeals. Column (2) includes only government appeals. Column (3) includes both. Appeals results are at the district court level. Control variables are: dismissal rate; sentencing appeals decided; prosecuted defendants per capita; share of jury cases; imprisonment rate; fine-only rate; appeals filed per capita; average sentence length. All regressions include court and year fixed effects. Constant and fixed effect coefficients are not reported. Dependent variable: ln_cp100. P.values are in parentheses under each coefficient. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 9

ROBUSTNESS CHECK: COMPARISON OF CLUSTERING LEVELS

	(1) Two-way, court & year	(2) Two-way state & year	(3) One-way court	(4) One-way state	(5) Circuit-by-year
Reversal-Remand spread (defendant appeals, centered)	-0.115***	-0.115***	-0.115***	-0.115***	-0.115**
	(0.007)	(0.008)	(0.003)	(0.004)	(0.012)
Reversal-Remand spread (government appeals) (centered)	0.003	0.003	0.003	0.003	0.008**
	(0.327)	(0.240)	(0.488)	(0.448)	(0.022)
Affirmation rate (defendant appeals, centered)	-0.064*	-0.063*	-0.064**	-0.064**	-0.079**
	(0.065)	(0.088)	(0.028)	(0.042)	(0.043)
Affirmation rate (government appeals) (centered)	-0.002	-0.002	-0.002	-0.002	-0.000
	(0.638)	(0.570)	(0.735)	(0.718)	(0.948)
Government Full affirmation * Conviction (both rates centered)	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Defendant Full affirmation * Conviction (both rates centered)	-0.004	-0.004*	-0.004	-0.004	0.001
	(0.203)	(0.063)	(0.241)	(0.212)	(0.630)
Partial Affirmation rate(defendant appeals, centered)	-0.048	-0.046	-0.048	-0.048	-0.005
	(0.396)	(0.404)	(0.412)	(0.420)	(0.936)
Partial affirmation rate (government appeals (centered)	-0.005	-0.005	-0.005	-0.005	-0.006
	(0.305)	(0.220)	(0.507)	(0.478)	(0.396)
Government Partial affirmation * Conviction (both rates centered)	-0.003	-0.002	-0.003	-0.003	-0.001
	(0.165)	(0.227)	(0.237)	(0.235)	(0.505)
Defendant Partial affirmation * Conviction (both rates centered)	-0.011	-0.011*	-0.011	-0.011*	-0.005
	(0.115)	(0.070)	(0.101)	(0.090)	(0.498)
Dismissal rate (defendant appeals) (centered)	-0.012	-0.009	-0.012	-0.012	-0.113
	(0.775)	(0.856)	(0.854)	(0.870)	(0.187)
Dismissal rate government appeals (centered)	0.036***	0.036***	0.036***	0.036**	0.003
	(0.005)	(0.006)	(0.008)	(0.010)	(0.715)
R-squared	0.969	0.971	0.969	0.969	0.978
Fixed effects	court,year	state,year	year, court	year, state	circuit-by-year, court, year
Cluster level	court, year	state,year	court	state	circuit-by-year, court
Control variables	Yes	Yes	Yes	Yes	Yes
Number of observations	728	728	728	728	720

NOTE.– This table shows regression results for robustness check on the level of clustering. The regression variables are the same as in table 8 column (3), but the level of clustering and fixed effects varies between columns in this table. Constant and fixed effect coefficients are not reported. For the circuit-by-year fixed effects (column 5), the user developed reghdfe (Correia et al. 2015) was used. P.values are in parantheses under each coefficient. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 10

ROBUSTNESS CHECK: INSTRUMENTAL VARIABLE APPROACH

	(1)	(2)	(3)	(4)	(5)
Reversal-remand spread (sentence) (centered)	-0.424***	-0.413***	-0.431***	-1.170**	-1.139**
	(0.001)	(0.000)	(0.000)	(0.030)	(0.012)
Full affirmation rate (senetence) (centered)	-0.135***	-0.190***	-0.213***		
	(0.001)	(0.005)	(0.002)		
Partial affirmation rate (sentence) (centered)	-0.123*	-0.179**	-0.199**		
	(0.080)	(0.027)	(0.012)		
Full Affirmation * conviction (both rates centered)		0.003	0.003		
		(0.444)	(0.482)		
Partial Affirmation * Conviction (both rates centered)		-0.007	-0.010		
		(0.473)	(0.302)		
Dismissal rate (sentence) (centered)		-0.122*	-0.140*		
		(0.097)	(0.065)		
Conviction rate (centered)		-0.091	-0.113*	-0.178	-0.136
		(0.182)	(0.099)	(0.101)	(0.204)
Prosecuted defendants per capita (centered)			0.079	0.044	0.044
			(0.303)	(0.519)	(0.478)
Share of jury cases (centered)			-0.302*	-0.196	-0.204
			(0.066)	(0.411)	(0.374)
Fine-only rate (centered)			-0.119	-0.218	-0.166
			(0.229)	(0.170)	(0.327)
Imprisonment rate (centered)			0.055	-0.039	-0.000
			(0.448)	(0.742)	(0.998)
Appeals filed per capita (centered)			-0.757**	-0.018	0.011
			(0.018)	(0.965)	(0.977)
Average imprisonment length (years) (centered)			0.049	-0.912	-0.846
			(0.896)	(0.299)	(0.293)
Appeals decided - sentence (centered)			0.007	-0.001	-0.002
			(0.166)	(0.909)	(0.761)
R-squared	0.542	0.557	0.561	0.083	0.121
Hansen J (p.value)	0.940	0.990	0.921	0.441	0.494
Kleibergen-Paap LM (underidentification)	0.000	0.000	0.000	0.083	0.030
Kleibergen-Paap Wald (weak-IV)	14.392	16.437	16.875	2.287	3.159
Anderson-Rubin chi2 weak-IV (p.value)	0.004	0.003		0.005	0.001
Stock-Wright weak-IV (p.value)	0.001	0.000		0.000	0.000
FAR	.0479	.045	.0444	.0412	.0445
Year and court Fixed effects	Yes	Yes	Yes	Yes	Yes
Defendant+Government controls	No	No	No	Yes	Yes
Defednant+Government polynomial controls	No	No	No	No	Yes
Number of observations	728	728	728	728	728

NOTE.– This table shows regression results for the IV approach. Instruments are: (1) Administrative reversal-remand spread, (2) Civil reversal-remand spread, and (3) One year lagged criminal reversal-remand spread (all centered and at the appeal court level). Standard errors are clustered at the district court level. Stata command: xtivreg2, gmm2s. Defendant+Government controls (columns 4, 5) are: Full and partial affirmation rates; their interactions with conviction rates; and dismissal rates. 'Defendant+Government polynomial controls' (column 5) are: Full and partial affirmations, squared and cubic; and conviction rate, squared and cubic. All regressions include court and year fixed effects. P.values are in parentheses under each coefficient. $^{*}p < 0.1,^{**}\ p < 0.05,^{***}\ p < 0.01$.

Table 11

INSTRUMENTAL VARIABLES: HETEROSKEDASTICITY-BASED (LEWBEL APPROACH) AND EXTERNAL INSTRUMENTS

	All appeals		Defendant-filed appeals	
	(1)	(2)	(3)	(4)
	Generated instruments	All instruments	Generated instruments	All instruments
Reversal-remand spread (sentence) (centered)	-0.113***	-0.202***		
	(0.005)	(0.000)		
Full affirmation rate (senetence) (centered)	-0.139***	-0.121***		
	(0.000)	(0.000)		
Partial affirmation rate (sentence) (centered)	-0.066	-0.052		
	(0.241)	(0.337)		
Appeals filed per capita (centered)	0.413***	0.675**	0.319**	0.632**
	(0.005)	(0.022)	(0.017)	(0.024)
Appeals decided - sentence (centered)	0.048***	0.051***	0.045***	0.049***
	(0.001)	(0.000)	(0.002)	(0.000)
Full Affirmation * Conviction (both centered)	0.008*	0.007		
	(0.084)	(0.123)		
Partial Affirmation * conviction (both centered)	0.026***	0.020**		
	(0.000)	(0.010)		
Conviction rate (centered)	-0.288***	-0.301***	-0.307***	-0.298***
	(0.000)	(0.000)	(0.000)	(0.000)
Full affirmation rate (defendant, centered)			-0.118***	-0.091***
			(0.000)	(0.000)
Reversal-Remand spread (defendant, centered)			-0.159***	-0.235***
			(0.000)	(0.000)
Partial affirmation rate (defendant, centered)			-0.060	-0.037
			(0.329)	(0.532)
Full affirmation (defendant) * conviction (both centered)			0.004	0.001
			(0.497)	(0.791)
Partial affirmation (defendant) * conviction (both centered)			0.020**	0.011
			(0.023)	(0.222)
Hansen J (p.value)	0.450	0.254	0.473	0.237
AR weak-IV (multiple endog. variables)		.006		.006
WALD weak-IV (multiple endog. variables)		1.60×10^{-15}		1.26×10^{-17}
Court fixed effect		Yes		Yes
Additional control variables		Yes		Yes
Number of observations	728	728	728	728

NOTE.– This table shows regression results for the IV Lewbel approach. Columns (1) and (3) include generated instruments only. Columns (2) and (4) include additional instrument: Administrative reversal-remand spread (centered), Civil reversal-remand spread (centered), One year lagged criminal appeals reversal-remand spread (centered), and interaction of each of these three variables with conviction rate (centered). Instruments are taken from the appeal court level. Instrumented variables (all centered): Reversal-remand spread, Full affirmation, Partial affirmations, appeals filed per capita, sentence appeals decided. Additional control variables: conviction rates, squared and cubic. (centered), Standard errors are clustered at the district court level. A two-step GMM estimator is implemented using the Stata command: xtivreg2h, gmm2s. All regressions include court fixed effects. Constant and fixed effect coefficients are not reported. P.values are in parentheses under each coefficient. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 12

Underidentification and Weak-IV for each instrumented variable

(a) Aggregated rates (government and defendant)

	Affirmation rate	Reversal-Remand spread	Parial affirmation	Appeals filed per capita	Appeals decided
Underidentificiation test(Sanderson-Meijer chi2 p.value)	0.	1.50×10^{-164}	0.	0.	2.40×10^{-322}
Weak-IV F-test (Sanderson-Windmeijer)	74.6	25.7	491	296	47.4
Stock and Yogo critical F-test value (10% relative bias)	11.26	11.26	11.26	11.26	11.26

(b) Defendant appeals only

	Affirmation rate	Reversal-Remand spread	Parial affirmation	Appeals filed per capita	Appeals decided
Underidentificiation test(Sanderson-Meijer chi2 p.value)	0.	1.20×10^{-212}	0.	0.	0.
Weak-IV F-test (Sanderson-Windmeijer)	76.2	32.3	623	294	49.3
Stock and Yogo critical F-test value (10% relative bias)	11.26	11.26	11.26	11.26	11.26

note.– This table shows the results of underidentification and weak-instruments tests used for each of the instrumented variables of table 11. Panel (a) presents the test result of the aggregated rate (table 10, column (2)) and panel (b) presents results for the defendant-filed appeals rate (table 11, column (4).

Table 13

CORRELATION TABLE: FEDERAL COURTS DATA

	Criminal Affirmation rate	Criminal Reversal rate	Criminal Remand rate	Criminal Dismissal rate	Criminal Remainder rate
Criminal Affirmation rate	1				
Criminal Reversal rate	-0.267***	1			
Criminal Remand rate	-0.422***	-0.187***	1		
Criminal Dismissal rate	-0.799***	-0.00467	-0.0482*	1	
Criminal Remainder rate	-0.239***	-0.00575	0.232***	-0.0125	1
Observations	1547				

NOTE.– This table shows correlations between the appeal result categories, for the sample taken from the 'Judicial business of the U.S courts'.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.