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

Roee Sarel *

Frankfurt School of Finance & Management, 60314 Frankfurt am Main, Germany

December 19, 2016

Abstract

This paper seeks to empirically examine how the accuracy of a multi-tier adjudication system affects crime deterrence. An ongoing scholarly debate regarding the effects of judicial errors on deterrence provides mixed arguments, but the role of a multi-tier system - where errors can be corrected on appeal - has been mostly overlooked. Analyzing appeal results from U.S federal courts and corresponding crime rates, I find that error occurrence, reflected by affirmance rates, decreases deterrence. Error correction, conversely, entails a complex effect: reversals increase deterrence, but remands decrease deterrence; which implies a need for theoretical adjustment and judicial caution.

JEL Classification: K

Keywords: Judicial errors, Adjudication errors, Crime rates, Crime deterrence, Appeals, Courts

1. Introduction

Adjudication systems are imperfect and prone to errors. In criminal cases, these errors amount to wrongful convictions (type I errors) and wrongful acquittals (type II errors).[1] Since such errors entail a welfare loss, among else due to potential infringement of crime deterrence, social planners typically try to avoid and correct them, using multi-tier court systems. Appellate courts are thus established to review cases and intervene where errors have been detected, by either exercising direct authority (reversing a decision) or delegating the decision power back to lower courts (remands). Whether these actions, taken be appellate courts, indeed result in higher deterrence is, however, an open question.

Traditional deterrence models *á la* Becker (1968) assume that the criminal legal process can deter criminals through adjustments in expected sanctions, i.e. by changing the probability of apprehension and conviction (”p”) or the penalty size (”f”). However, such models seldom account for how p and f are actually reflected in the appellate process. It is however crucial to identify the marginal effects of appeals, if one is to design an optimal deterrence regime.

*Frankfurt School of Finance & Management, Sonnemannstraße 9-11, 60314 Frankfurt am Main, Germany. Email: roeesarel@gmail.com.

[1] The definition of wrongful convictions (acquittals) as type I (II) errors has been somewhat inconsistent in the literature. Earlier papers (Harris, 1970; Png, 1986) have used the opposite definitions, where wrongful acquittals are referred to as ”type I errors” and wrongful convictions are referred to as a ”type II errors”. Since a criminal trial usually has the null hypothesis of an innocent defendant, it seems more appropriate to define a wrongful conviction (i.e. rejecting the null hypothesis although it was correct) as type I, and so I too shall adopt this definition.

Optimizing deterrence subject to an appellate process requires answering three questions:

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

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

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

The first question proves difficult to answer, empirically as well as theoretically. While the effect of type II errors is clear – a decrease in deterrence, when guilty defendants may (wrongfully) escape punishment – the effect of type I errors is largely debated. The mainstream view (e.g. Png, 1986), which I shall refer to as the "Extended Becker Model" (or "EBM"), suggests that type I errors also decrease deterrence. Namely, by introducing a (wrongful) sanction to those who abstain from crime and remain innocent, type I errors decrease the 'opportunity cost' of crime and may "tip the scale" towards crime commission. Conversely, critics of the EBM claim that type I errors can in fact be beneficial (e.g. Craswell and Calfee, 1986) or inconsequential (e.g. Lando, 2006) to deterrence.

Since the effect of judicial errors on deterrence is already theoretically ambiguous, predicting how the process of correcting errors will affect deterrence, i.e. answering the second question, becomes equally challenging. Intuitively, correcting an error should reverse the effect of its occurrence, at least in part. However, the handling of errors in a hierarchical environment introduces further complications in terms of signaling, public perceptions and time preferences. Thus, even when an error is eventually corrected, the 'damage' to deterrence caused by the direct influence on the Beckerian "p" and "f" may be partially irreparable. Moreover, the damage may spillover to larger systematic effects, such the public willingness to obey the law, as facilitated by faith in the legal system. The extent to which such spillover effects are pivotal for crime deterrence requires answering the third question, i.e. to identify how the appellate courts' choice between reversing and remanding erroneous decisions is (1) reflected in the components of the expected sanction and (2) perceived as an informational signal on the reliability of the legal system. Existing theory offers little guidance, unfortunately, as to how these questions should be answered.[2]

While the empirical literature on expected sanctions is vast, evidence on judicial errors and their correction has been scarce. Instead, judicial errors have been almost solely researched experimentally, with evidence largely supporting the EBM (see section 2). However, experiments seldom include error correction and lack the empirical complexity of a full-scale legal system. In particular, potential responses of different players - such as prosecutors and trial judges - to the appeal system are usually omitted in order to keep the models stylized.

In this paper, I attempt to fill the gap in the literature by incorporating judicial errors and their correction by a multi-tier adjudicative system into the empirical research on deterrence. I use a panel data set for the period 1997-2013, which captures the different stages of the legal process in U.S. federal courts: detection/arrest, prosecution, trial, conviction, sentencing, appeals and Supreme court review. To overcome the empirical challenge of identifying judicial

[2] 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. One paper (Chopard et al., 2014) takes a step in the right direction by adopting the EBM's assumptions and considering a setting with an appeal system, but focuses mainly on judicial effort rather than deterrence.

errors (see section 3), I utilize affirmance rates (i.e. the rate in which appellate courts affirm trial court decisions), which reflect the rate in which judicial errors *do not* occur. I find that affirmance rates are generally negatively correlated with crime rates, implying that judicial errors decrease deterrence. However, the effect seems to also depend on conviction rates, such that when conviction rates are high, affirmance rates are positively - rather than negatively - correlated with crime. This dependency can be explained by the aforementioned spillover effects: when the public observes that (almost) all defendant are convicted and (almost) all convictions are affirmed, it ceases to believe in the reliability of the court system.

The rates of reversals and remands are then used to asses how the handling of errors that *did occur,* by appellate courts, affects crime. Although both reversals and remands arguably aim at error correction, their effects on deterrence might be very different. Namely, while reversals are a clear statement that an error was found and corrected, remands might be interpreted as a statement that an error was found but *not* corrected. If trial courts are perceived as unlikely to fairly reconsider the case upon remand and refrain from repeating the error, then remands are congruent with more errors overall. The EBM would then suggest that reversals should increase deterrence but remands should decrease deterrence. However, there are additional reasons to suspect that remands may be detrimental to deterrence. First, even if the error is eventually corrected by the trial court (or on a subsequent appeal), the time that innocent defendants spend in custody prior to their release constitutes a form of wrongful sanction by itself. Second, for guilty defendants whose case is remanded and who succeed in obtaining a 'stay of sentence', the delay caused by the prolongment of the legal process might result in discounts of the sanction, especially for myopic criminals. Third, spillover effects may again be in play, where appellate courts' decision to remand rather than reverse is interpreted as a signal of low judicial reliability. Finally, asymmetric litigation costs for the innocents and the guilty in a remand process may increase crime incentives as well.

Since reversals and remands operate through the same channel– the expected sanction conditional on the appellate court's belief that a judicial error occurred at trial – but arguably in opposite directions, including both rates as independent variables is problematic, particularly in light of collinearity. I therefore use instead the the Reversal-Remand spread (i.e. reversal rate minus remand rate) in order to capture the *net effect* on "p" and "f". I then find that this spread is negatively correlated with crime, which implies that indeed reversals increase deterrence but remands decrease deterrence.

In order to establish these effects as causal, it is necessary to ensure that the results are not driven by endogeneity problems, such as omitted variable bias or reversed causality. For example, in a reversed causality scenario, appellate judges with deterrence concerns may observe rising crime rates and respond by reversing less. Alternatively, other agents in the criminal process (e.g. prosecutors or legislators) may respond to crime by adjusting their behavior in a manner which ends up indirectly affecting the judicial error rate. To overcome such concerns, I use a two-fold strategy, by (1) controlling for relevant benchmarks along the legal chain, and (2) implementing an instrumental variables ("IV") approach.

As an instrument for the Reversal-Remand spread, I use the ideology of judges in the appellate and Supreme courts, which are known predictors of the tendencies to reverse (Hettinger and Lindquist, 2012; Iaryczower et al., 2013) and to remand (Boyd, 2015; Borochoff, 2008). Judicial ideologies have been previously used as an instrument for decisions in criminal cases (Di Tella and Schargrodsky, 2013) and indeed serve as effective instruments, since they reflect stable preferences that are unlikely to suddenly change in response to contemporaneous changes in crime rates. For robustness, I show that my results hold also when (1) instrumenting for affirmance rates and (2) using alternative instruments, such as heteroskedasticity-based generated instruments (Lewbel, 2012) and reversal-remand spreads in non-criminal cases (administrative and civil).

My analysis seems to be the first to provide observational empirical evidence for the effects of appeal results on crime deterrence. I show that the evidence is in line with the EBM, which has thus far been in theoretical dispute. The paper also extends the emerging empirical literature on remands (Boyd, 2015; Borochoff, 2008, e.g.). On the methodological side, the paper compares different panel methods, thus demonstrating the importance of avoiding methodological bias in the study of court behavior. Finally, my findings hold two important implications: for researchers, theory should be adjusted to include not only the frequency of judicial errors - but also whether and how they are corrected on appeal. For judges and policy makers, my findings imply that appeal results must be considered when optimizing deterrence regimes. In particular, the mechanism of remands may need adjustment, to eliminate negative effects on deterrence. However, 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 briefly presents related literature. Hypotheses are developed in Section 3. Section 4 describes the data and sources. Section 5 reviews variables and descriptive statistics. Section 6 discusses the methodology. Section 7 presents the 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 influence deterrence through the expected sanction $p \times f$. Since p could be lower than 100%, type II errors could occur, but type I errors were not considered. Judicial errors received attention in subsequent papers (e.g. Harris, 1970; Posner, 1973) and were integrated by Png (1986) to form the EBM.[3] In this extended model, errors of both types (I and II) are argued to be detrimental to deterrence: type II errors decrease the expected sanction conditional on committing the crime and type I errors increase the expected sanction conditional on abstaining. In other words, type I errors

[3]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 model's origin, see Mungan and Lando, 2015).

reduce the payoff from "innocence", thus decreasing the opportunity cost of crime.[4] The EBM then posits that deterrence hinges on whether E(sanction | commit) > E(sanction | abstain). A type II error decreases the LHS, while a type I error increases the RHS - both reducing deterrence.

The EBM has been widely adopted (e.g. Kaplow, 1994; Dari-Mattiacci and Deffains, 2007; Polinsky and Shavell, 2000, 2001, 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 space, where over-compliance breeds a higher payoff (Craswell and Calfee, 1986). Others argue that type I errors have *no effect* on deterrence, for example, when their frequency is too small (Andreoni, 1991, pp. 389) or when the source of the wrongful conviction is someone else's crime ("identity errors"). The latter argument is, by itself, highly debated. Lando (2006) and Mungan and Lando (2015) assume that identity errors happen with equal probability to the guilty and the innocent, leading to a zero net effect. Garoupa and 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) notes that such 'type II errors' stem from police behavior (failure to apprehend the actual perpetrator) and might be exogenous to adjudication.

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 and Saraceno, 2013). Rizzolli and Stanca (2012) and Nicita and 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 and Kahneman, 1979, 1992); and indignation, instigated by wrongful convictions. Contrarily, Mungan and 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).[5]

The empirical literature has, thus far, refrained from examining the EBM and instead focused on other extensions, such as the impact of unemployment (e.g. Entorf and Spengler, 2000; Ihlanfeldt, 2007; Lin, 2008), lower wages (e.g. Gould et al., 2002), police manpower (e.g. Levitt, 1997, 2002; Evans and Owens, 2007) police tactics (e.g. Thaler, 1977; Weisburd et al., 2009), arrest rates (e.g. Machin and Meghir, 2004) and imprisonment rates (e.g. Levitt, 1997). For further literature, see Khadjavi (2014, 2015) and Chalfin and McCrary (2014).

Experimental research on deterrence has been equally extensive (e.g. Baker et al., 2003; DeAngelo and Charness, 2012; Friesen, 2012; Schildberg-Hörisch and Strassmair, 2012; Ouss and Peysakhovich, 2013; Montag and Tremewan, 2016), but only recently began including judicial errors. Experiments typically assign subjects the role of thieves who can steal money

[4] 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.

[5] The argued scope of the EBM is summarized by Mungan and 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."*.

and then vary the error probabilities either explicitly (e.g. Rizzolli and Stanca, 2012; Rizzolli and Tremewan, 2016) or implicitly, using other subjects in the role of judges (e.g Feess et al., 2015; Baumann and Friehe, 2015). Findings are fairly consistent and largely support the EBM. One recent experiment (Lewisch et al., 2015) also 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, its stylized settings does not allow to disentangle the marginal effect of appeal results on deterrence, since judicial error rates are not common knowledge. Similarly, other stylized experiments cannot fully encompass the complexity of a full-scale appeal system and its effects over time, nor account for potential indirect effects of crime rates on judicial behavior, through actions of other agents in the legal system.

In this paper, I overcome these difficulties by controlling for the stages of the criminal process, thus accounting for the actions of the various agents.

3. Hypotheses development

3.1. Using appeal results as a proxy for judicial errors

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 occurred.[6] 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), using expert panels (e.g. Gould et al., 2014), focusing on specific post-conviction exonerations (e.g. death sentence cases, Gross and O'Brien, 2008; Gross et al., 2014) or - more recently - utilizing Michigan Law School's "National registry of exonerations" (e.g. Gross and Shaffer, 2012; Olney and Bonn, 2014).

While it is tempting to restrict attention to exonerations, where it has been factually proven that the defendant was wrongfully convicted, it would be problematic for testing the EBM. First, exonerations typically take many years to be proven. Thus, even if one is exonerated, it is unclear how this should be translated into a contemporaneous probability of type I errors, which might be very different than 'p' at the time of conviction. Second, exonerations presumably reflect those rare cases where even the appeal process did not yield a correct result. It will therefore fail to capture the more common process of error correction, where wrongful convictions are overturned on appeal. Third, wrongful convictions due to a legal error (rather than factual errors) would not be represented in exonerations. Incorporating legal errors is however necessary in order to settle the theoretical dispute, which partially leans on distinction of error types (see section 2 above). Fourth, exonerations may suffer from a severe selection bias, by looking only at specific types of cases. For example, most exonerations rely on DNA evidence, which might correspond mainly to identity errors, where the actual culprit left DNA at the scene. Similarly, exonerations might include only defendants with sufficient resources

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

for access to representation (Olney and Bonn, 2014).

This paper takes a different approach, by utilizing aggregated appeal results as a proxy for judicial errors and their correction. Using appeal results allows to overcome many of the shortcomings of exonerations: decisions are made in a (relatively) short time and all types of cases and mistakes are represented. Nonetheless, in order for appeal results to serve as a satisfactory proxy, appellate courts must be sufficiently accurate, particularly more accurate than trial courts. The assumption that appellate courts are indeed more accurate is intuitive and may be justified on several grounds. First, this is a common assumption in the theoretical literature, based on the notion that appellate 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 appellate court (e.g. Levy, 2005; Shavell, 2006). Second, if appellate courts were not more accurate, there would be little justification for their existence, since resources would be better spent improving trial court ability. Third, even if appellate courts sometimes err - is seems unlikely that these errors are systematic, given that the assignment of appellate judges to cases is generally random (Hettinger and Lindquist, 2012, pp. 128). Thus, any (infrequent) errors of appellate courts can be treated as noise.

Superior ability alone is, however, insufficient to ensure that appellate courts actually reach more accurate decisions. It is rather the target-function of appellate courts - i.e. their goal - which will determine their accuracy in practice. For example, appellate courts might be interested not only in reaching factually correct decisions, but also in other goals, such as promoting legal coherence through precedent creation, preserving constitutional rights of defendants, or simply minimizing effort cost. While these goals might work against accuracy maximization, most economic models assume that appellate judges either inherently aim at reaching correct decisions or are incentivized to do so given a fear of reversal by the Supreme court.[7] Furthermore, the key question is not whether appellate courts are actually more accurate than trial courts but whether they are *perceived* by potential criminals as such. Thus, as long as criminals anticipate higher accuracy on appeal - the proxy should be relatively strong.

A different 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 appellate 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.

[7]The exact utility function of judges is a subject of ongoing debate. For a review of the literature on this issue, see Epstein et al. (2013, chapter 1) and Feess and Sarel (2016).

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 whereas the prosecution cannot (as a rule)appeal acquittals due to the "double jeopardy" rule.[8] Therefore, only type I errors are reviewed on appeal.[9] The restriction to type I errors is advantageous for identification, since appeals clearly correspond only to convictions and not acquittals. The omission of type II errors can be addressed by controlling for conviction rates and using an IV approach, but any omission bias is likely mild to begin with. Namely, since type II errors are presumably *negatively* correlated with type I errors,[10] due to the trade-off created by the burden of proof (e.g. Andreoni, 1991; Rizzolli and Saraceno, 2013; Kaplow, 2011), the attenuation bias of omitting type II errors only imply that the estimations of type I errors are conservative, i.e. that the true effect can can only be larger.

3.2. Interpreting appeal result categories

Affirmance rates should be the easiest to interpret, as an affirmance (literally) means that no type I judicial error was found by the appellate court. However, an increase in the affirmance rate - conditional on conviction rates - not only implies less type I errors but also less type II errors. Namely, a higher affirmances rate means that some convictions are moved from the pool of incorrect convictions and inserted into the pool of correct convictions, simultaneously reducing both error types.[11] While this simultaneity does not allow to fully disentangle which reduction influences deterrence, it is clear that an increase in affirmance should unequivocally increase deterrence. Consequentially I hypothesize that:

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

Disentangling reversal rates is slightly harder, since reversals includes two aspects: error occurrence (which should decrease deterrence) and error correction (which implies less errors overall and increase deterrence). However, conditional on the frequency of errors (proxied by affirmance) and the appeal filing rate (used as control variable), the marginal effect should be that of correction, leading to:

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

Hypotheses H1 and H2 are actually two sides of the same coin: if the appellate court is accurate, it will reverse wrongful convictions and affirm correct convictions.[12]

[8] 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 (see, for example, Stith, 1990). Note that while the double jeopardy rule is widely accepted, some countries interpret it more narrowly, such that governmental appeals are allowed.

[9] 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.

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

[11] To illustrate this point, suppose that 20 defendants have been convicted, where 10 are innocent and 10 are guilty. Suppose further that 7 guilty defendants are convicted (reflecting 30% probability for a type II error) and 3 innocent defendants are convicted (reflecting 30% probability for a type I error). If the appellate court is fully accurate, affirmance rate would be 70%. Conditional on conviction rate, an increase in affirmance implies that more guilty are convicted and less innocents are convicted. For example, an increase to 80% affirmance implies that 8 guilty defendants and 2 innocents were convicted.

[12] Reversals may also increase deterrence through another channel - by increasing public faith in the system through the enforcement of due process rights. While such an enforcement might actually lead to a lower expected sanction (e.g. guilty defendants whose rights have been violated manage to reverse their conviction based on procedural violations), it seems plausible that such an effect would be weaker.

Remands are a somewhat different story. First note that while the effect of reversals and affirmances 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 contain little information on f[13] - remands concern both p and f. This occurs since a reversal typically implies that the conviction is overturned, such that no sanction is imposed, but remands may entail a different final sanction. Notably, this is captured by the two common motivations for remands: (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.[14] Thus, remands may affect either only p or only f or both.

How remands affect deterrence through p and f is, however, far from straightforward. Since a defendant whose case is remanded was already convicted, a remand can only imply a decrease in p (the defendant has some chance of being acquitted). However, the question then becomes – p conditional on what? i.e. p for wrongful or correct convictions? decreases in p for correct convictions is detrimental to deterrence, but a decrease in p for wrongful convictions will increase deterrence. Therefore, the effect of remands depends on (1) the accuracy of trial courts - i.e. how many wrongful convictions there are - and (2) the appellate court's ability (and strategy) when deciding which cases to remand. Since the public typically does not have complete information on the motivation to remand, an increase in remand rates might serve as a signal on p. Such a signal can contain information for both guilty and innocent defendants, since both defendant types may constitute a share of remanded cases. Wrongful convictions may constitute a larger share of those cases remanded to address factual questions, since such questions seem likely to arise when the defendant is in fact innocent.[15] Correct convictions may constitute a larger share of those cases remanded for resentencing only, since guilty defendants might be less likely to challenge their conviction and focus on the punishment only, leading to that sort of remands. Alas, even when looking at wrongful convictions alone, the signal on p may be ambiguous. On one hand, remands may initialize error correction (though a delayed one, as lower courts must still review the case), signaling less errors. Such remands may even include instructions to lower courts, which encourage a correction.[16] On the other hand, a remand entails the 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 be perceived as a potential "uncorrected error". A third option may be that the declaration of a remand holds no information at all (for example, when appellate courts remand cases purely since it is customary not to conduct evidentiary hearings on appeal).[17] .

The effect of remands on f is equally equivocal, mainly in light of the *time dimension*,

[13] Appellate courts are presumably unlikely to reverse a verdict and acquit a defendant only due to a sentencing mistake.

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

[15] The ration of wrongful to correct conviction depends, of course, also on the base rate for each conviction type.

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

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

since remands instigate a prolonged process. How this prolongment affects f depends on two main factors: First, again, the frequency of innocents vs. guilty within the remand pool, since the effect should be the opposite for each type of defendant. Second, whether the sanction is "stayed" pending the remand. A stay of sentence implies the sanction has not been executed yet, and is delayed. Intuitively, future sanctions are deemed less detrimental than present sanctions due to time preferences. However, the effect on the guilty and innocent may be asymmetric, since guilty defendants may posses different time preferences. Namely, guilty criminals may commit the crime precisely because they heavily discount any sanction, in the near or far future (Davis, 1988; Beraldo et al., 2013; Mastrobuoni and Rivers, 2016). In this context, Lee and McCrary (2009) find evidence of "impatient" offenders, which discount the future hyperbolically. A similar intuition can be found in models of court delays (e.g. Torre, 2003), but the empirical evidence (e.g. Soares and 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.[18] When sanctions are not stayed, (e.g. a defendant appeals the verdict while sitting in prison), there should be no discount.

Deterrence can then decrease if one of two combinations occurs: either the defendant is innocent and the sentence is not stayed, such that the time spent in custody pending remand constitutes an additional reduction in the payoff from innocence (see Domenech and Puchades, 2014, for a similar argument regarding wrongful arrests), OR the defendant is guilty and the sentence is stayed.[19] Since defendants seeking a stay must generally show, among else, that they are likely to succeed on the merits,[20] the sentence of guilty defendants will infrequently be stayed if trial courts are accurate. Inaccuracy of trial courts, conversely, would make it more likely that either deterrence-reducing combination would occur.

It is however possible, that a delayed process can – by itself – have a spillover effect and 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), reducing the probability of error correction.

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 (see discussion on the share within remands above), or (2) if innocent defendants bear higher costs for proving their innocence (which is plausible, if opportunistic guilty defendants invest less, knowing that their chances of "proving" innocence are lower (Johnson, 2016, pp. 264)).

It should be emphasized, however, that changes in p and f are not necessarily symmetrical

[18] 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.

[19] 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.

[20] 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).

in their effect on deterrence. Namely, if potential offenders are, at the margin, risk preferers (as is assumed in the original Becker (1968) 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.[21] A small increment in the overall probability of a wrongful conviction might then overshadow any benefit of decreased sanction for correct convictions.[22]

Summing up, theoretically, remands may either increase or decrease deterrence, depending on the trial courts' accuracy, signaling effect, time preferences of criminals and (a)symmetry of litigation costs between guilty and innocent defendants. 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, as an appeal may be dismissed for reasons unrelated to the merits (e.g. dismissals based on procedural rules or estoppels). Furthermore, some judges may be prone to dismiss an appeal when they feel that no error has occurred, 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

This section describes the data collection process (see also annex A for citations).

4.1. Courts and criminal procedure statistics

Quantitative data on proceedings in U.S. federal courts (1997-2013) was hand-gathered from the "Judicial Business of the U.S. courts" reports and official published tables. The data relates to three different levels:[23]

- *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 appellate court.
- *States and territories* - each state contains one or more district (trial) courts, but belongs to only one "circuit", which includes only one appellate court. In addition to the states, four "territories" are included in the judicial system.[24]

[21] For example,Masur and Bronsteen (2015) suggest 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.

[22] Some argue that 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 the conviction rate, which is controlled for in my regressions, it is of little concern. As another example, Smith and 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 aggregate figures, there is again little relevance.

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

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

- *Circuit appellate courts* - 12 appellate courts review criminal appeals which originate from district courts.[25] Eleven courts (numbered 1st to 11th) review appeals from multiple district courts. The 12th court ("DC") reviews appeals from one district court only.

The data contains information on proceedings in each district court: prosecuted defendants, conviction rates[26]; bench/jury trials; and punishment type (fine/imprisonment). Criminal appeal rates were similarly gathered regarding each appellate court, alongside civil and administrative appeals rates for the IV approach.

The database was complemented using various sources. Data on sentencing cases was extracted from U.S. Sentencing Commission (”USSC”) databases, as disseminated by the Inter-university Consortium for Political and Social Research (”ICPSR”). Appeal results of the U.S. Supreme court were extracted from the ”Supreme Court database” (Spaeth et al., 2016). Arrest rates and prison populations were gathered using online data-tools of the ”Bureau of Justice Statistics”. Data on U.S. territories was complemented from other official sources.[27] 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).[28]

4.2. Crime rates

U.S crime rates are usually measured using two complementary sources: (1) the FBI's Uniform Crime Reports (”UCR”) and (2) the National Crime Victimization Survey (”NCVS”). The former uses index crimes that are traditionally reported to the authorities, namely violent crime (e.g. murder, aggravated assault, forcible rape) and property crime (e.g. burglary, larceny, motor vehicle theft). The second source - NCVS - is derived from a survey of crime victims, reflecting also unreported crime. The distinction between reported and unreported crime may be of special importance to the study of judicial errors, since the incentive to report a crime might be affected by such errors. 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.[29] Respectively, when reversal rates drop, opportunistic false reports may be filed to purposefully achieve wrongful convictions. Unfortunately, NCVS rates are generally unavailable at the state-by-year level, rendering them problematic for the analysis at hand. Therefore, for the lion's share of the analysis, I use UCR rates. Nonetheless, recent work by Diallo et al. (2015) allows to approximate a 3-year average of NCVS rates at the state-level for a sub-sample (2006-2013). Thus, as a robustness check, I utilize these approximations to compare the effects with those found for the UCR.

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

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

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

[28] These territories represent a small portion of the appeal supply and their exclusion should not entail a high sample error. 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 is only about 0.05%, while in other states the ratio difference is much smaller.

[29] Iyengar (2009) finds that increases in the expected sanction for domestic violence lead to fewer reports, due to victims' reluctance to expose their attacker to larger sanctions. Such an effect may be even stronger for wrongful convictions, where victim's are unsure whether they are accusing an innocent.

4.3. Data constraints

Several inherent constraints of the data set should be mentioned. First, UCR crime rates are published for a full calendar year (January - December) while court rates are published for October to September, thus appeal 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 and al Nowaihi, 2013), are included. It is similarly unclear whether criminals use ongoing Bayesian updating or decide according to heuristics.[30] 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. Furthermore, individuals might have different estimations, depending on the information they have gathered and on past experience (Sah, 1991). However, a significant effect 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 information (Dalla Pellegrina, 2008, pp. 269) [31] or to network effects (Sparrow, 1991).

A second constraint concerns granularity. While the federal courts do publish data on the *amount* of appeals filed and terminated per district court, appeal results are only published in aggregate numbers, such that one can only extract the overall rate of an appellate court but not individual rates for each district court. This aggregation is of little concerns as long as no trial court within a given circuit errs systematically more than others.[32] A stable disparity in trial court accuracy also 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) or deciding whether to admit certain evidence. Therefore, even if juries decide in a non-random fashion in a certain district, professional judges can intervene and correct the errors. Indeed, one may then argue that some judges systematically err, but this is again unlikely as judges should be able to learn from appeal results and adjust their behavior to avoid future mistakes.

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

[31] 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.

[32] 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.

Moreover, problematic judges may be assigned "easier" cases, or in extreme circumstances - be replaced by better judges.[33] Thus, as long as no court errs systematically on average, no bias will occur.[34] Nonetheless, I use a second measurement for appeal results for robustness: the rates of sentencing appeals, which are disaggregated and do not raise granularity concerns (see section 8 below).

A third constraint arises since the published data also aggregates decisions on guilt and punishment. Errors in guilt (convictions) and errors in punishment may have different effects on deterrence, since the elasticity with respect to p and f may be different (due to risk preference, see above). Empirically, however, the question is mainly whether errors in conviction and punishment are correlated. A positive correlation would lead to overestimation of the effect of type I errors. To clarify this point, assume that wrongfully convicted defendants receive higher punishments and consider a person on the verge of committing larceny. If he chooses innocence, both the probability of a wrongful conviction and the probability of a subsequent higher punishment should affect his incentives in the same direction - towards crime. Conversely, if wrongful convictions are more likely to yield lower sentences, the effect of type I errors will be underestimated. Lundberg (2016) shows that judges and juries may engage in "compromise verdicts", where uncertainty about crime commission is compensated by lower sentences. Such behavior would imply a negative correlation between errors in guilt and punishment. Using senetencing appeals for robustness checks allows to overcome this issue as well.

A fourth constraint emerges since crime rates are measured at the state – rather than district – level. While crime may differ from region to region within a state (e.g. between urban and rural areas, see Glaeser and Sacerdote, 1999), all offenders must rely on the same appeal rates. Thus, the effect on deterrence should be the same for all sub-regions, implying that state crime rates are an accurate proxy for local crimes. Furthermore, this problem can be partially solved by using fixed effects at the court level.

A final constraint involves the appeal category of "affirm", which seems to include also appeals that were reversed in part.[35] 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 affirmance. I address the latter problem in section 8.

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

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

[34] Note that although one could attempt to test the assumption of proportionality by looking further ahead at Supreme court results, it will lead to a circular argument: if one is doubtful about the usage of higher instance results (appellate 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 appellate court's accuracy.

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

100.[36] 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.[37] For some regressions, I use instead $tot\_crime\_s_{i,t}$, which is simply the total number of crimes.

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 represents one percentage point).[38] As independent variables, I include the Affirmance Rate and the aforementioned Reversal-Remand spread. I also control for the Dismissal Rate and the rate of unclassified ("other") appeal results.

The justification for using the reversal-remand spread rather than each rate individually is two-fold. Theoretically, both rates operate through the expected sanction conditional on the appellate court's belief that some error occurred. Thus, reversals and remands may capture one concept - the appellate court's choice between reversing and remanding. However, practical reasons dictate a form of aggregation as well. Namely, since appeal results represent complementary shares of the same pie, they are inherently linked, such that an increase in one rate always implies a respective decrease in another. Unsurprisingly, these variables are then also highly correlated.[39] Multicollinearity then requires aggregation. Since several pre-tests indicated that both reversal and remand rates have a significant effect when tested alone, but in opposite directions, the Reversal-Remand Spread allows to use the appropriate mathematical signs for each category. In later parts of the analysis I also center the variables to further overcome multicollinearity, which has the advantage of reducing correlation without changing the interpretation of the regression coefficients (Afshartous and Preston, 2011).

5.1.3. Control variables and indirect endogeneity

Most variables previously identified as affecting crime rates are unnecessary to achieve consistent estimators in my analysis, as they are presumably uncorrelated with judicial errors. For example, judicial errors are unlikely to be correlated with unemployment or wages. However, some control variables may be crucial, namely those that reflect the different stages of the criminal legal process - prosecution, trial, conviction, sentencing 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 to cases with high probability of conviction. If reversal rates rise, such that the overall conviction probability decreases, the prosecutor might decide to focus only on "winner

[36] 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.

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

[38] 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 (see, for example, Scott, 2006).

[39] For example, affirmance rates and dismissal rates have a correlation of -0.79, significant at the 1% level. Reversal rate and remand rate have a correlation of -.18, significant at the 1% level as well. Both reversals and remands are highly collinear with affirmance rates, with a correlation of -0.26 and -0.42 respectively

cases", leading to a higher conviction rate and perhaps less appeals (that is, if guilt defendants appeal less). 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, leading to more convictions. In an opposite chain reaction, judicial errors may decrease deterrence, 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).

The existing literature deals with such issues by controlling for components which proxy expected sanctions, but usually focusing on detection/arrests and seldom on 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.[40] I address this issue by adding two main control variables: *conviction rates* and number of *appeals filed per capita*.[41] For the sake of caution, I add (many) additional control variables:

- *Arrest rate per capita*, which capture the probability of apprehension.
- *Prosecution rate (share of prosecuted arrestees)*, which captures both prosecutorial behavior and the probability of a "type II-like" error as a result of failure to prosecute guilty defendants.[42] Prosecution rates are often neglected in the literate and are assumed to reflect a concept similar to arrest rates, but are naturally important given the aforementioned incentives of prosecutors.
- *Share of UCR index crimes in sentences:* Since defendants accused of UCR index crimes can also be prosecuted in state-courts (rather than federal courts), I control for the share of sentences relating to UCR crimes. Thus, both the general probability and specific probability of prosecution is accounted for.
- *Share of jury decisions* (rather than bench trials or plea bargains) allows to control for divergence in conviction probabilities (or judicial errors), depending on the decision maker's identity. A difference in conviction rates may rise either due to different accuracies of judges and juries or to self-selection of defendants into jury and bench trials. For example, Gay et al. (1989) develop a model where, in equilibrium, innocent defendants opt for bench trials while the guilty choose jury trials (hoping for a "noisy jury" to wrongfully acquit them). Following a similar logic, type I errors may be less common in jury trials.
- *Imprisonment rates, Fines-only rates* and *Average sentence length*: these variables capture the type and gravity of sanctions, which may be correlated with type I errors, (e.g. due to copromise verdicts, where the sanction size becomes a function of the probability of guilt). Imprisonment rates serve as a control for the possible crime-reducing effects

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

[41] 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)).

[42] Unfortunately, I cannot differentiate between "mixes" of prosecution (e.g. weak v. strong cases).

of incapacitation. The literature has struggled in the past with isolating the effects of incapacitation and deterrence (see Chalfin and McCrary, 2014), usually in the context of identifying sentence severity. Namely, since higher prison sentences can mechanically reduce crimes committed by recidivists by incapacitating them, it becomes hard to distinguish between these effects. However, the problem is less severe when looking directly at imprisonment rates, rather than using prison population as a single proxy.

- To further separate incapacitation from deterrence I control also for *prison populations*, both in state-prisons and federal prisons.[43]
- *Shares of conviction appeals:* share of "conviction-only" and "sentence+conviction" appeals are added to disentangle the effects of punishment from convictions.
- *Share of black, Hispanic and female defendants*: this variable is used to control for changes in the pool of defendant, since previous papers have shown that minorities are more likely to be wrongfully convicted (Olney and Bonn, 2014, e.g.).
- *Appeal-filing rate*: this variable accounts for the possibility that changes in the expected appeal results lead to a change in the mix of appealed cases.
- *Number of appeals decided on the merits*: this variable ensures that results are driven by the appeal category (nominator) rather than the scope of decided appeals (denominator).
- *Supreme court ("SC") appeal rates:* contemporaneous SC appeal result rates are important for two reasons: first, criminals may be concerned with their prospect of overturning convictions using the Supreme court. Second, appellate judges might be concerned with reputation effects, leading them to opt for a more certain outcome (e.g. reverse the conviction, in order to avoid a defendant-appeal to the Supreme court). Given the scarcity of criminal appeal decisions per court and year, it is impractical to estimate an exact contemporaneous rate for each court[44] Therefore, I use instead the rates for all cases originating from each circuit court in a given year (criminal and other). Presumably, this may capture the SC's specific judicial temper towards each circuit court.[45] Since the SC traditionally does not remand cases without an accompanying reversal or affirmance, remanded cases are categorized as reversal, affirmance or partial affirmance.

Finally, I interact affirmance rates with conviction rates, for several reasons: first, in order to capture the full effect of the expected sanction, which depends not only on initial conviction but on a (potential) affirmance on appeal. Second, measurement issues of the affirmance category require a closer look on interdependencies, especially since full and partial affirmances may have opposite operative consequences. Third, measuring spillover effects require checking whether the effect of affirmance rates depends on conviction rates.

[43]Given that I only check for contemporaneous effects on deterrence, I cannot, however, rule out the possibility of dynamic effects brought by changes in incapacitation (e.g. prisoners that are released in t=1 commit crime in t=2).

[44]The Supreme court reviews only around 1% of all court decisions. See, for example, U.S. Census "Table 331" for the year 1980-2010, http://www.census.gov/compendia/statab/2012/tables/12s0331.pdf

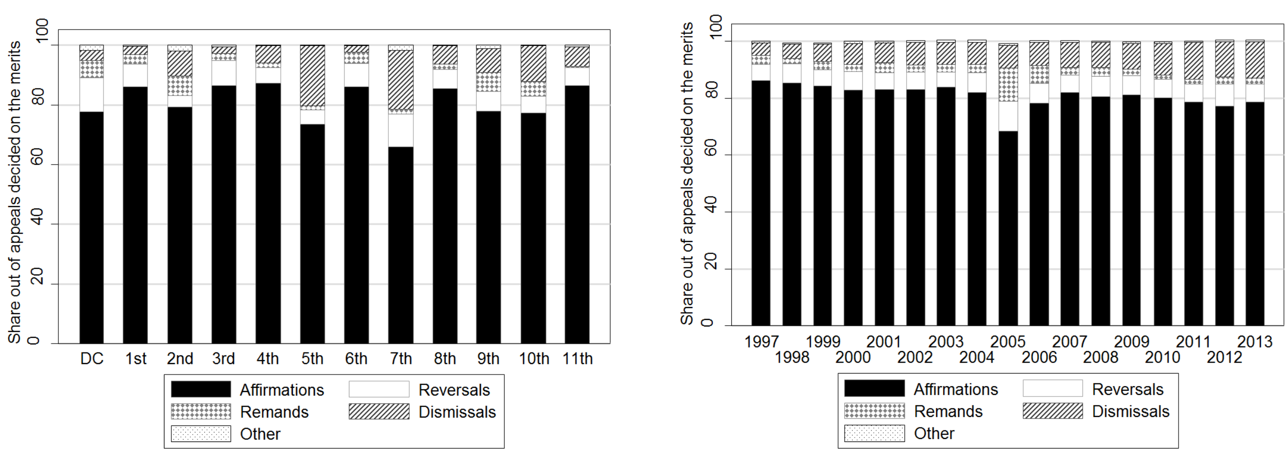
[45]Note, however, that the SC tends to to grant review to cases it intends to reverse (Cummins et al., 2015, pp. 387). Thus, these rates do not reflect the probability of review by the SC, only the probability of a certain outcome (conditional on review).

5.2. Descriptive Statistics - Federal courts data

Table 1 presents descriptive statistics. Affirmance rates are fairly high on average (80.6%), indicating a low error rate. Reversals and remands spans a wide range (0-1% to 24-31%), mainly due to inter-court variation. Figure 1.1 compares means across courts. Reversal rates tend to be high when remand rates are low, implying these are used interchangeably. If indeed no appellate court systematically reviews more erroneous decisions, the variation can be attributed to judicial preferences. In other words, irrespective of the merits, some appellate courts prefer to reverse while others prefer to remand. Disparate attitudes may stem, for example, from different degrees of omission bias (Zamir and Ritov, 2012) - where remands are preferred to reversals, in order to avoid explicit changes of the status quo. Figure 1.2 compares means by year. A conspicuous jump occurs in 2005, with lower affirmance rates and high reversal and remand rates. This jump is likely a result of the "Booker Case" (*United States v. Booker*, 543 U.S. 220 (2005))[46], which allowed judges to deviate from the strict federal sentencing guidelines and allowed to reopen cases on appeal (Commission et al. (2012, part B, pp. 40), Catterson (2006, pp. 290)). Crime rates per capita are heterogeneous, with a mean of about 3,600 for the UCR and 2,600 for the NCVS.[47] Higher crime rate observations are mostly located in DC. Property crimes are vastly more common than violent crimes.

Figure 1: Appeal results - means

(1) Appeal results by district (2) Appeal results by year



6. Empirical analysis: model and methodology

The hypotheses are tested using the following model:

$$ln\_cp100_{i,t} = \beta_0 + \beta_1 spread_{i,t} + \beta_2 aff_{i,t} + \beta_3 conrate_{i,t} + \beta_4(aff_{i,t} \times conrate_{i,t}) + \beta_5 X' + \beta_6 Y_t + \beta_7 C_i + \varepsilon_{i,t}$$

where: $spread_{i,t}$ is the reversal-remand spread;[48] $aff_{i,t}$ is the affirmance 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.

[46]See also the previous rulings in *Apprendi v. New Jersey*, 530 U.S. 466 (2000) and *Blakely v. Washington*, 542 U.S. 296 (2004).

[47]A lower measure of NCVS rates may seem surprising, given their inclusion of non-reported crimes. However, the difference in measurement methods and the possibility of excessive reporting in response to judicial errors, make lower measured rates equally plausible.

[48]Formally, the spread is calculated as: $spread_{i,t} = rev_{i,t} - rem_{i,t}$, where rev is the reversal rate and rem the remand rate.

Table 1
Descriptive statistics: Federal courts database

Panel A: crime rates	UCR				NCVS			
	mean	sd	min	max	mean	sd	min	max
Total crimes in the state	323535.42	326424.18	13051.00	1420637.00				
Total crime per capita	3783.37	946.23	1946.23	8835.56	2630.25	608.35	1153.15	5982.37
Property crime per capita	3349.76	831.97	1724.34	7117.02	1924.45	442.32	896.75	4220.61
Violent crime per capita	433.61	172.20	66.91	1718.55	705.80	200.24	256.40	1761.76
Observations	1424				630			

Panel B: appeal rates and control variables	mean	sd	min	max
Criminal Reversal-Remand spread	3.43	6.10	-24.83	20.37
Criminal Affirmance rate	80.56	9.07	54.75	93.68
Criminal Reversal rate	6.60	3.28	0.68	24.07
Criminal Remand rate	3.17	4.54	0.00	31.04
Criminal Dismissal rate	9.11	7.48	0.00	36.49
Criminal Remainder rate	0.56	1.18	0.00	9.95
Conviction rate	89.50	6.64	46.55	99.01
Share of jury cases	4.25	2.59	0.00	18.34
Criminal appeals decided on the merits	874.51	480.77	58.00	2784.00
Imprisonment rate	80.39	11.45	14.93	98.12
Fine-only rate	3.00	7.23	0.00	67.84
Jury trials - conviction rate	86.70	11.63	0.00	100.00
Appeal-filing rate	19.35	8.48	1.95	51.41
Arrests per capita	26.33	47.83	2.70	567.01
Prosecution rate	0.84	0.24	0.18	3.49
State-prisons population	39885.34	44093.27	915.00	175512.00
Federal-prisons population	182931.34	29534.25	123041.00	217815.00
Female defendants share	14.89	4.13	2.68	32.05
Black defendants share	30.84	18.20	0.00	78.03
Hispanic defendnats share	28.20	21.78	0.68	96.25
UCR Violent crime sentences share	4.68	3.71	0.00	33.11
UCR Property crime sentences share	4.25	3.59	0.00	42.86
(mean) senttot	78.04	53.22	12.25	551.69
Conviction-only share	28.40	13.14	0.00	83.33
Sentence+Conviction share	20.31	9.56	0.00	68.18
SC full affirmance rate	30.81	28.09	0.00	100.00
SC reversal rate	64.52	28.74	0.00	100.00
SC partial affirmance rate	2.85	9.15	0.00	100.00
SC dismissal rate	1.41	4.86	0.00	25.00
Observations	1424			

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.

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. Estimating the model requires a careful choice of methodology (For discussions on appropriate methodologies, see Cornwell and Trumbull (1994) and Scott (2006)). I therefore use statistical tests to rule out inappropriate methods. These tests indicate that (1) fixed effects are preferable to random effects, (2) heteroskedasticity is present and (3) the residuals are auto-correlated.[49] Additionally, there is indication of cross-sectional dependence ("contemporaneous correlation").[50].

[49] A Hausman-test (Hausman, 1978), rejects the null hypothesis of random effects, $p < 0.0001$). The null hypothesis of homoskedasticity was rejected in a modified Wald-test ($p < 0.0001$), implemented by the user-developed xttest3 command in Stata (Baum, 2001).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).

[50] Since the number of panels (N) is larger than the number of periods (T), one cannot use the standard Breusch-Pagan test (xttest2 command

Cross-sectional dependence possibly results from judicial behavior. For example, judges may compare themselves to peers in neighboring districts (or in 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, e.g. due to regional shocks to crime incentives. I address these issues, by using different methods:

1. **Fixed-effects with two-way cluster-robust standard errors** (by court and year), which is robust in the face of arbitrary correlations within courts and within years. Since this method is general in nature and does not impose strict assumptions (Cameron and Miller, 2015), it will serve as the benchmark for the lion's share of the analysis.
2. **Fixed effects with panel-corrected standard errors** ("PCSE", Beck and Katz, 1995), assuming either an AR1 or PSAR1 disturbance. This method is robust given contemporaneous correlation, which is problematic if some - but not all - panels are correlated (Worrall and Pratt, 2004).
3. **Fixed effects with Driscol-Kraay standard errors** ("SCC", Driscoll and 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 2 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 3 presents regressions results,[51], which reflect three key findings:

Finding 1 *The Reversal-Remand spread is negatively correlated with crime*

Since reversals and remands are equally weighted within the spread, this implies a symmetric - but opposite - effect of each rate. The effect is significant ($p < 0.05$) across the five methods, with some differences in the effect size. Taking, for example, the benchmark two-way cluster regression, the 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 0.179%.

Finding 2 *Affirmance rates are negatively correlated with crime*

Affirmance rates are similarly negatively correlated with crime ($p < 0.05$ across methods), but the effect size is about 6-11 times larger than the spread, implying an asymmetric effect of error occurrence and correction. That is, error occurrence seems to have a much larger effect than error correction. Taking elasticity into account (i.e. that an increase of 1 pp. in affirmance rates is much smaller than a 1 pp. increase in the spread), the parity is even greater.

in Stata). Instead, I run the regression and test the residuals using the user-developed command xtcd2, which is an adjustment of xtcd command (Eberhardt, 2011). A Pesaran test (Pesaran, 2004) then indicates cross-sectional dependence, with a p.value marginally above the 10% level. . 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 also test each independent variable separately and 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

[51] 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)

Table 2
COMPARISON OF METHODS

Method	Advantages	Disadvantages	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); Gathmann (2008).
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 and 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 and Summit (2012)

NOTE.– This table compares the five different methods used in the base model regressions. Columns (2) and (3) briefly describe the advantages and disadvantages of each methods. The last column list examples for papers that have used the method. The Stata commands used in each methods are as follows: (1) FE two-way cluster: ivreg2 or reghdfe (Correia et al., 2015); (2) PCSE: xtpcse, c(ar1) *or* c(psar1). (3) ivreg2, dkraay(2) (Baum et al., 2007a,b), which produces the same result as the xtscc command (Hoechle, 2007). Bandwidth of 2 was chosen by the rule of thumb $T^{0.25}$; (4) POISSON: xtpqml, fe (Simcoe, 2008).

Finding 3 *The effect of affirmance rates on crime depends on conviction rates*

While both affirmance and conviction rates are negatively correlated with crime, a balancing effect occurs through their interaction (albeit much smaller in size than the main effects), which has a positive coefficient ($p < 0.05$). To help interpret this finding, Figure 2 presents marginal effects of affirmance rates given different conviction rates. Notably, the main effects dominate for moderate ranges, but when conviction rates become high – affirmance rates have an opposite effect. One interpretation for this finding would be that if virtually every defendant is convicted, the public ceases to believe that affirmances are justified, since it is "simply impossible" that errors are so infrequent.

Control variables are mostly insignificant at the 5% level, with a few relevant exceptions:

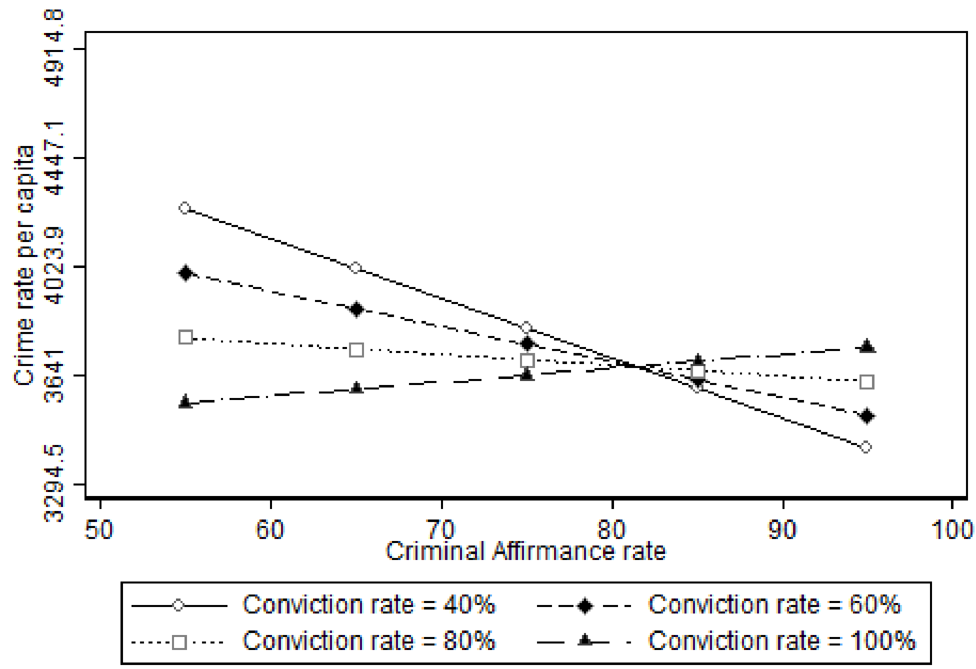
- Arrest rates: a negative coefficient, implying that increases in the probability of apprehension decreases crime (in line with theory and previous literature).
- Prosecution rates: opposite signs in different regressions. A negative coefficient is in line with theory (an increase in p), but a positive coefficient can also be explained by two possibilities: (1) reversed causality - prosecutors who want to appear "tough on crime" prosecute a larger portion of arrestees when crime rises and (2) prosecutors are perceived as instigators of wrongful convictions, i.e. if the base population of arrestees did not change, a higher prosecution rate means that more innocents are prosecuted.
- Federal prison population: this variable is panel-invariant, thus much of its effect is captured by the year fixed effects. However, some small coefficients are significant, with different signs across regressions.

Table 3
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	-0.179***	-0.081***	-0.061**	-0.179***	-0.001***
	(0.01)	(0.00)	(0.03)	(0.00)	(0.01)
Criminal Affirmance rate	-1.006**	-0.753**	-0.673**	-1.006**	-0.012**
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Criminal affirmance * Conviction	0.011**	0.008***	0.007**	0.011***	0.000**
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)
Conviction rate	-0.921**	-0.593**	-0.505**	-0.921***	-0.011**
	(0.01)	(0.01)	(0.03)	(0.00)	(0.02)
Criminal Dismissal rate	0.070	-0.011	-0.042	0.070	0.001
	(0.60)	(0.87)	(0.49)	(0.47)	(0.25)
Criminal Remainder rate	-0.363	-0.101	-0.159***	-0.363	0.003
	(0.46)	(0.27)	(0.01)	(0.41)	(0.15)
Arrests per capita	-0.034**	-0.020***	-0.019***	-0.034***	0.000
	(0.04)	(0.00)	(0.00)	(0.00)	(0.79)
Prosecution rate	-0.014	0.010	0.018***	-0.014	-0.000
	(0.62)	(0.14)	(0.00)	(0.39)	(0.62)
Share of jury cases	-0.028	0.009	0.032	-0.028	0.003*
	(0.87)	(0.86)	(0.54)	(0.85)	(0.05)
Fine-only rate	-0.004	-0.049	-0.064	-0.004	-0.001
	(0.97)	(0.38)	(0.20)	(0.95)	(0.19)
Imprisonment rate	0.056	0.035	0.023	0.056	0.001
	(0.42)	(0.22)	(0.41)	(0.14)	(0.30)
Appeal-filing rate	-0.058	-0.021	-0.023	-0.058	-0.001
	(0.34)	(0.38)	(0.37)	(0.25)	(0.27)
State-prisons population	-0.000	0.000	0.000	-0.000	0.000
	(0.98)	(0.86)	(0.12)	(0.95)	(0.13)
Federal-prisons population	0.000	0.007***	0.007***	-0.000***	-0.000***
	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female defendants share	0.008	-0.008	-0.012	0.008	0.001
	(0.93)	(0.81)	(0.73)	(0.88)	(0.65)
Black defendants share	0.008	0.000	-0.004	0.008	-0.001
	(0.86)	(0.99)	(0.77)	(0.71)	(0.19)
Hispanic defendnats share	-0.039	-0.010	-0.011	-0.039	0.001
	(0.48)	(0.61)	(0.57)	(0.25)	(0.20)
UCR Violent crime sentences share	0.004	0.008	-0.026	0.004	-0.002
	(0.98)	(0.89)	(0.67)	(0.97)	(0.10)
UCR Property crime sentences share	-0.375**	-0.105**	-0.105**	-0.375***	-0.003*
	(0.03)	(0.02)	(0.02)	(0.00)	(0.06)
Conviction-only share	0.017	0.007	0.007	0.017	-0.001*
	(0.59)	(0.24)	(0.26)	(0.50)	(0.09)
Sentence+Conviction share	0.020	0.016**	0.013	0.020	-0.000
	(0.57)	(0.03)	(0.16)	(0.41)	(0.97)
Average imprisonment length (years)	0.308***	0.145***	0.143***	0.308***	0.003*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)
Criminal appeals decided on the merits	0.003	0.001***	0.000	0.003*	0.000***
	(0.18)	(0.00)	(0.38)	(0.05)	(0.00)
SC full affirmance rate	-0.124***	-0.050***	-0.048**	-0.124***	-0.001***
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
SC reversal rate	-0.101***	-0.049***	-0.048**	-0.101***	-0.001***
	(0.01)	(0.00)	(0.02)	(0.00)	(0.00)
SC partial affirmance rate	-0.151***	-0.064***	-0.058***	-0.151***	-0.002***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
SC dismissal rate	-0.101**	-0.057**	-0.058**	-0.101***	-0.002***
	(0.03)	(0.01)	(0.05)	(0.01)	(0.00)
Dependent variable	ln_cp100	ln_cp100	ln_cp100	ln_cp100	tot_crime_s
Adjusted R-squared	0.944			0.944	
Log likelihood	-4506.079			-4506.079	-567885.887
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	1424.000	1424.000	1424.000	1424.000	1424.000

NOTE.– This table compares regression results of the basic model using 5 different econometric 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) Poission 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. Control variables whose coefficients are not reported are: arrests per capita rate; prosecution rate; share of jury cases; remainder rate; appeals decided on the merits; prosecuted defendants per capita; share of jury cases; imprisonment rate; fine-only rate; appeal rate; state prison population; federal prison population; The following shares of sentenced defendant: black, female, hispanic, UCR violent crimes, UCR property crimes, conviction only appeals, conviction and sentence appeals; average sentence length, appeals decided on the merits; court and year fixed effects. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Figure 2: Marginal effect of affirmance rates - base results



- Average imprisonment length: positive coefficients, which seem counter-intuitive, as more years in prison imply a higher f. However, this might again be a result of endogeneity, where judges handing out higher sentences to try and increase deterrence. Alternatively, longer imprisonments may facilitate a"school for crime" (Roberts and Hough, 2005, pp. 297) within prisons, leading to more crime. Imprisonment may increase crime also by reducing prisoners' future wage prospect, leading to irregular employment and higher crime commission incentives (Johnson, 2016, pp. 281). For additional arguments suggesting that higher punishments may decrease deterrence, see also Gneezy and Rustichini (2004); Caulkins (1993); Kahan (1996); Dickens (1986); Livernois and McKenna (1999).
- Appeals decided: extremely small coefficients, which seem economically insignificant.
- SC rates: significant negative coefficients, slightly smaller than those of appellate courts. While in line with my hypotheses, it is nonetheless surprising that the effect is not much smaller, given the low likelihood of SC review. A possible explanation would be that the salience of Supreme Court decisions leads to behavioral overweighting of their importance (see Korobkin and Ulen, 2000, for a discussion on criminals' response to salient outcomes).

Summarizing the basic results, I find support for hypotheses 1,2 and 3b (and against hypothesis 3a). The EBM's argument that judicial errors lower deterrence seems therefore to hold empirically, but error correction matters for deterrence as well. Note that finding 3, regarding an interaction between convictions and affirmances, does not pose an obstacle to this argument, but quite the contrary - if high convictions alongside high affirmances are a signal of more errors, the effect is also in line with the EBM.

8. Robustness checks

8.1. *Controlling for polynomial effects and spurious significance*

Following the econometric literature (e.g. Afshartous and Preston, 2011; Balli and Sørensen, 2013), I test the interaction by substituting the original term with an interaction of centered (demeaned) variables 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.[52] Results are reported in table 4. The findings are qualitatively similar to the base results (note that in some regressions, it is the quadratic terms that are significant). Figure 3 presents updated marginal effects of affirmance rates.[53] The marginal effect of affirmance on crime remains virtually linear, except for extremely high conviction rates (100%), where intermediate affirmance levels are positively associated with crime. Since no court actually reaches 100% conviction rates, the latter seems to have little economic significance in practice.

Table 4
ROBUSTNESS CHECK - CENTERED INTERACTION WITH QUADRATIC AND CUBIC TERMS

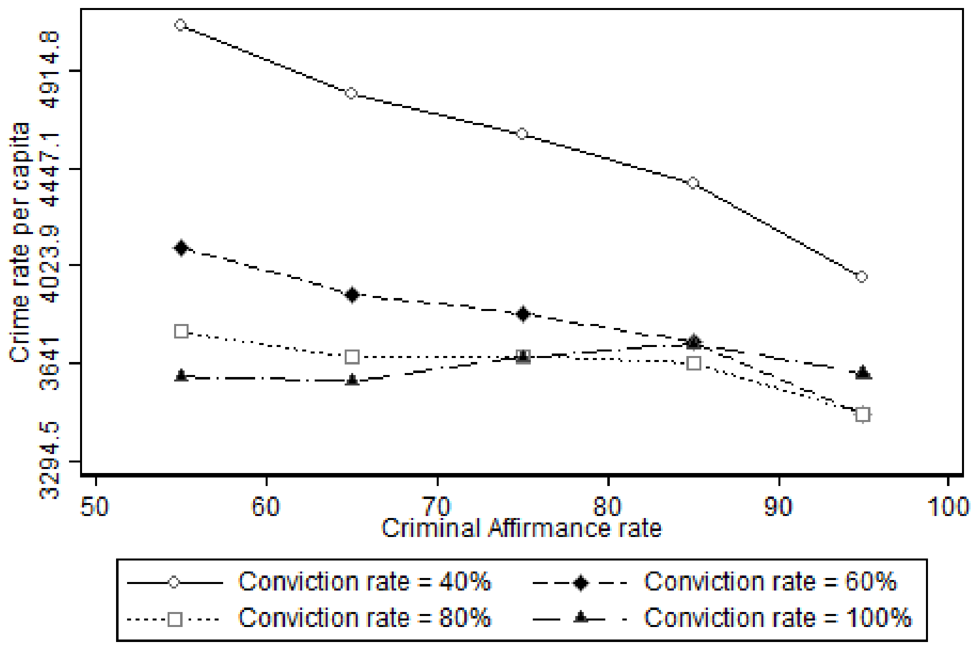
	(1) Two-way cluster (court,year)	(2) Panel-corrected (AR1)	(3) Panel-corrected (PSAR1)	(4) Driscol-Kraay (SCC)	(5) Poisson (robust)	(6) Uncentered interaction
main						
Criminal Reversal-Remand spread	-0.216***	-0.098***	-0.077***	-0.216***	-0.001**	-0.216***
	(0.004)	(0.000)	(0.004)	(0.000)	(0.017)	(0.004)
Criminal Affirmance rate	0.072	-0.037	-0.084	0.072	-0.003***	-0.902***
	(0.592)	(0.583)	(0.162)	(0.504)	(0.000)	(0.008)
Criminal affirmance rate (centered) squared	-0.015*	-0.008***	-0.007***	-0.015***	-0.000	-0.015
	(0.065)	(0.002)	(0.001)	(0.000)	(0.957)	(0.105)
Criminal affirmance rate (centered) cubic	-0.001	-0.000***	-0.000**	-0.001***	0.000**	-0.001
	(0.103)	(0.007)	(0.034)	(0.000)	(0.013)	(0.125)
Criminal affirmance * Conviction (both rates centered)	0.011**	0.008**	0.006**	0.011***	0.000**	
	(0.043)	(0.018)	(0.030)	(0.002)	(0.042)	
Criminal affirmance * Conviction						0.011***
						(0.009)
Conviction rate (centered) squared	-0.004	0.000	0.001	-0.004	0.000	-0.004
	(0.641)	(0.914)	(0.746)	(0.535)	(0.920)	(0.648)
Conviction rate (centered) cubic	-0.000	-0.000	0.000	-0.000	-0.000	-0.000
	(0.243)	(0.856)	(0.948)	(0.164)	(0.596)	(0.255)
Criminal Dismissal rate	0.126	0.018	-0.018	0.126	0.000	0.126
	(0.357)	(0.833)	(0.813)	(0.179)	(0.731)	(0.359)
Criminal Remainder rate	-0.463	-0.093	-0.121*	-0.463	0.004*	-0.463
	(0.372)	(0.276)	(0.067)	(0.300)	(0.063)	(0.372)
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	0.945			0.945		0.945
Log likelihood	-4494.604			-4494.604	-551871.139	-4494.604
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Court fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1424.000	1424.000	1424.000	1424.000	1424.000	1424.000

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 affirmance rates. The first 5 columns include an interaction of centered (demeaned by court) affirmance 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[52]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."*.

[53]Coefficients for figure 3 are taken from an uncentered interaction with quadratic and cubic terms.

Figure 3: Marginal effects - with quadratic and cubic terms



8.2. Comparing UCR and NCVS, property and violent crimes

In the basic results, appeal results cannot be separated by crime type. Controlling for the shares of violent/property crimes in sentences already (partially) addresses this issue, but still does not allow for heterogeneous effects on crimes types. A different approach can be taken by running separate regressions for each crime type, but these are only reliable if appeal rates are representative for both types of crime, which a strong assumption. Nonetheless, since previous literature finds different elasticities for violent and property crime (see Lee and McCrary, 2009; Chalfin and McCrary, 2014), it is worthwhile to explore this alternative estimation as well.

Differentiating violent and property crimes is particularly relevant for comparing UCR and NCVS, which have been previously found to diverge on violent crime. To attain state-level estimates of NCVS, I interpolate the 3-year average of the 'Small Area Estimates' (Diallo et al., 2015), to achieve a representative year-level average.[54] Table 5 reports results.

The direction of the effects is simialr to the base result, but the spread coefficient is insignificant for UCR violent crime and NCVS total and property crime. Conversely, affirmance rates (squared or cubic) coefficients are significant across NCVS and UCR. Given the differences between how UCR and NCVS are measured (see Diallo et al., 2015, pp. 9-10) and the aforementioned aggregation issues, it is clear that these findings should be taken with caution. Nonetheless, for the purpose of evaluating the strength of the base results, two insights arise. First, the effect of error *occurrence* on deterrence generally holds irrespective of measurement method (UCR/NCVS). Second, error correction matters for both crime types, but significance

[54]The SAE contain 3-year averages. Thus, each year is repeated in the sample *twice*. For example, the year 2006 appears in the range 2005-2007 and 2006-2008. Thus, I add up each adjacent 3-year average and divide by 6, to extract an estimate for each year.

Table 5
ROBUSTNESS CHECK: COMPARISON OF VIOLENT AND PROPERTY CRIME

	(1) UCR total crime	(2) UCR violent crime only	(3) UCR property crime only	(4) NCVS total crime	(5) NCVS violent crime only	(6) NCVS property crime only
Criminal Reversal-Remand spread	-0.216***	-0.002	-0.002***	-0.147	-0.287*	-0.085
	(0.004)	(0.172)	(0.004)	(0.127)	(0.060)	(0.325)
Criminal Affirmance rate (centered)	0.072	0.002	0.001	0.141	0.442	0.027
	(0.592)	(0.552)	(0.608)	(0.470)	(0.162)	(0.880)
Criminal affirmance rate (centered) squared	-0.015*	-0.000*	-0.000*	-0.006	-0.020*	-0.001
	(0.065)	(0.094)	(0.097)	(0.327)	(0.058)	(0.826)
Criminal affirmance rate (centered) cubic	-0.001	-0.000	-0.000	-0.001**	-0.001*	-0.000*
	(0.103)	(0.420)	(0.113)	(0.043)	(0.068)	(0.066)
Conviction rate						
Conviction rate (centered) squared	-0.004	-0.000	-0.000	-0.005	0.009	-0.010
	(0.641)	(0.668)	(0.752)	(0.598)	(0.502)	(0.359)
Conviction rate (centered) cubic	-0.000	-0.000	-0.000	-0.000	0.000	-0.000
	(0.243)	(0.551)	(0.390)	(0.302)	(0.848)	(0.240)
Criminal affirmance * Conviction (both rates centered)	0.011**	0.000	0.000*	0.007	0.003	0.009*
	(0.043)	(0.347)	(0.063)	(0.167)	(0.770)	(0.082)
Criminal Dismissal rate	0.126					
	(0.357)					
Criminal Remainder rate	-0.463					
	(0.372)					
SC full affirmance rate (centered)	-0.106***	-0.001**	-0.001***	-0.115	-0.129	-0.116
	(0.001)	(0.048)	(0.001)	(0.101)	(0.180)	(0.108)
SC reversal rate (centered)	-0.086***	-0.000	-0.001***	-0.124*	-0.134	-0.126*
	(0.005)	(0.224)	(0.003)	(0.075)	(0.163)	(0.075)
SC partial affirmance rate (centered)	-0.136***	-0.000	-0.002***	-0.078	-0.080	-0.070
	(0.005)	(0.861)	(0.003)	(0.320)	(0.468)	(0.367)
SC dismissal rate (centered)	-0.092**	0.000	-0.001**	-0.072	-0.047	-0.095
	(0.022)	(0.721)	(0.027)	(0.334)	(0.668)	(0.210)
Dependent variable	ln_cp100	ln_vp100	ln_pp100	ln_vp100	ln_vp100	ln_vp100
Court and year fixed effects	Yes	Yes	Yes	Yes		Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.945	0.932	0.942	0.938	0.922	0.934
Log likelihood	-4494.604	1165.850	2032.466	-1843.710	-2050.456	-1868.787
Number of observations	1424.000	1424.000	1424.000	623.000	623.000	623.000

NOTE.– This table compares regression results for violent, property and total crime as measured by the Uniform crime reports (columns 1-3) and the National Victimization Survey (columns 4-6). Standard errors are two-way clustered by court and year. Control variables are identical to those used in table 5. All regressions include court and year fixed effects. P.values are in parentheses under each coefficient. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

levels vary. Further research, using disaggregated rates, is however necessary for verification.

8.3. Disentangling appeal results - using rates of sentencing appeals

A disaggregation of affirmance rates is required in order to rule out the possibility that the effect of affirmance rates is driven by the presence of (1) partially-affirmed-partially-reversed decisions and (2) government-filed appeals (of sentence).I therefore use an alternative estimation of appeal results, using databases of the U.S. sentencing commission, which differ from the sample analyzed thus far in several aspects. First, rates are measured at the district court level, obviating the need to rely on representativeness assumptions of the appellate court mean. Second, appeals are separated into defendant-filed, government-filed and cross-appeals.[55] Third, affirmances are separated into "full" and "partial".[56]

This alternative measurement suffers from some limitations. Namely, only *sentencing appeals* are included, i.e. where the defendant appealed the sentence (with or without appealing the conviction). Rates of "conviction-only" appeals are thus omitted. However, the *share* of conviction-only appeals can still be used as a control variable. Restricting the scope to sentencing appeals entails however an advantage: the exclusion of interlocutory appeals (of interim decisions), which are included in the original sample but are less relevant for estimating final

[55] Cross-appeals refers to cases where both the defendant and the government appealed the decision.
[56] The "remainder" category is non-existent in the sentencing database.

expected sanctions. A second limitation arises since the alternative database contains only a specific fraction of sentencing appeals, namely those identified as "constitutional".

A different limitation lies in the difficulty to establish the timeline with which information was made available to the public. Specifically, the sentencing database which allows for extraction of appeal rates for the full sample-range (1997-2013) seems to have been made available (by the ICPSR) only during 2014, casting doubt on whether this information was in fact available in real time. While the USSC does additionally publish statistics using an online data-tool, it covers only the range 2006-2013 and was possible also unavailable in earlier years as well (when the internet was at its infancy). As we do not know how potential offenders acquire their information and given the effort cost of extracting rates from case-level data, it is unclear whether the effect of early rates is comparable to that of later years. Moreover, the 2006-2013 range happens to overlap with the post-Booker-case range, i.e. when sentences could be reopened and the appeal system plausibly became more salient.

8.3.1. Descriptive statistics of sentencing appeal results

Table 6 presents descriptive statistics of the alternative sample. Sentencing appeal rates seem quite different than the appellate court mean, especially when comparing defendant and government appeals. For example, recall that the affirmance rate mean was 80.56% in the large sample. Defendant-appeals affirmance rate is a bit lower (76.5%), but the government-appeals affirmance rate is much lower (29.3%). Furthermore, sentencing reversal rates are higher than

Table 6
Descriptive statistics - sentencing commission database

	mean	sd	min	max
Full affirmance rate (senetence)	75.15	14.00	0.00	100.00
Partial affirmance rate (sentence)	4.61	4.79	0.00	31.03
Reversal rate (sentence)	9.70	7.75	0.00	71.43
Remand rate (sentence)	2.13	5.07	0.00	45.95
Dismissal rate (sentence)	8.42	8.91	0.00	44.44
Reversal-remand spread (sentence)	7.57	8.31	-39.13	57.14
Affirmance rate (defendant appeals)	76.23	14.12	0.00	100.00
Reversal rate (defendant appeals)	8.74	7.51	0.00	71.43
Partial affirmance rate (defendant appeals)	4.38	4.67	0.00	31.03
Remand rate (defendant appeals)	2.09	5.00	0.00	45.95
Dismissal rate (defendant appeals)	8.56	9.04	0.00	46.67
Reversal-Remand spread (defendant appeals)	6.65	7.98	-39.13	57.14
Affirmance rate (government appeals)	29.08	41.23	0.00	100.00
Reversal rate (government appeals)	64.56	43.07	0.00	100.00
Partial affirmance rate (government appeals)	4.33	7.01	0.00	103.13
Remand rate (government appeals)	2.07	13.52	0.00	100.00
Dismissal rate (government appeals)	0.40	4.79	0.00	100.00
Reversal-Remand spread (government appeals)	66.63	42.48	0.00	100.00
Gov. appeals share	1.67	2.71	0.00	23.53
Observations	1420			

NOTE.– This table presents descriptive statistics for the appeal rates of sentencing appeals taken from the U.S sentencing commission online database over the range 1997-2013. Appeal results are measured at the trial court level (i.e. each district court is represented by its respective rates). Note that only appeals results regarding sentence are included (no conviction only appeals).

in the large sample: 8.7% for defendant-filed and 64.3%(!) for government-filed appeals. A higher reversal rate for government filed appeals is consistent with findings of Hettinger and Lindquist (2012, pp. 140), who suggest that a selection effect - where the government hand-picks cases expected to end in reversal - may be driving this difference. Government appeals constitute less than 2% of appeals, but are an outlier. The combined sentencing affirmance rate (full an partial) is not significantly different from the h the appellate court average used in the larger sample shows no significant difference, supporting the assumption that appeal level data is in fact representative of each district court. However, partial affirmances constitute around 4.6% of overall affirmances, which seems substantial.[57] I thus run additional regressions, to verify that these outliers are not driving the effects. I interact conviction rates separately with full and partial affirmances. Independent variables are centered to reduce collinearity with the interaction terms.[58]

8.3.2. Analyzing the effect of judicial errors using sentencing appeal results

Table 7 compares regression results for the full sample and the sub-range 2005-2013. [59], where appeal rates are still combined for both appellant types (defendant/government). The basic results hold qualitatively, but the interaction effect is only significant using the method that accounts for cross sectional correlation and in the sub-range (column (4)). This may suggest that spillover effects became relevant only after the Booker-Case, when the appeal system perhaps gained salience in the public eyes. Partial affirmances are also negatively correlated with crime rates, supporting the argument that reversals and affirmance affect deterrence in the same direction.

Table 8 compares additional regression results where appeal rates are separated by appellant type: defendant, government or cross-appeals. Since government and cross appeals are rare, missing values are replaced by 1-year lagged values, under the assumption that criminals could rely on earlier rates as an estimate for contemporaneous rates even when no government appeals were filed in a specific year. The effect of reversal-remand spreads is once more reinforced for defendant-filed appeals. Affirmance rates have a negative coefficient (as in the base results) but insignificant at the 10% level. However, coefficients of SC affirmance rates are negative and significant at, or slightly above, the 5% level. Government appeal rates are all insignificant when defendant and cross-appeal rates are controlled for (column 4). When the other rates are excluded (column 3), however, the reversal-remand spread has a positive effect (significant at the 10% level). Since government appeals are necessarily about sentences - i.e. about the Beckerian f rather than p - this does not derogate from the base findings. Note that a reversal in a government-filed appeal implies a higher f, which should typically increase deterrence. While some argue that higher punishments may in fact decrease deterrence (see above), a straightforward interpretation can also be derived from the EBM. Namely, if the government

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

[58] For example, interactions are correlated with the main terms: 0.92 for full affirmances and 0.99 for partial affirmances (each with their interaction with conviction rates).

[59] The year 2005 is included, since the Booker verdict was already decided during that year.

Table 7

ROBUSTNESS CHECK: SENTENCING DATABASE REGRESSIONS (FULL SAMPLE V. POST-2005)

	(1) Two-way cluster (full sample)	(2) Two-way cluster (2005-2013)	(3) PCSE-AR1 (full sample))	(4) PCSE-AR1 (2005-2013))	(5) Two-way cluster (no polynomials, full sample)
Reversal-remand spread (sentence) (centered)	-0.061**	-0.078**	-0.022*	-0.063***	-0.050**
	(0.023)	(0.022)	(0.091)	(0.002)	(0.047)
Full affirmance rate (senetence) (centered)	-0.064*	-0.090*	-0.041***	-0.078***	-0.051*
	(0.081)	(0.057)	(0.000)	(0.000)	(0.061)
Full affirmance (centered) squared	-0.001	0.002	-0.000	0.001	
	(0.590)	(0.305)	(0.319)	(0.389)	
Full affirmance (centered) cubic	-0.000	0.000	-0.000	0.000	
	(0.914)	(0.296)	(0.986)	(0.160)	
Partial affirmance rate (sentence) (centered)	-0.019	-0.069	-0.051*	-0.082	-0.083
	(0.822)	(0.407)	(0.059)	(0.131)	(0.216)
Partial affirmance (centered) squared	-0.016	-0.019	-0.005	-0.006	
	(0.355)	(0.261)	(0.309)	(0.585)	
Partial affirmance (centered) cubic	0.000	0.001	0.000	0.000	
	(0.731)	(0.221)	(0.238)	(0.520)	
Full affirmance * conviction (both rates centered)	0.004	0.006	0.002	0.006***	0.004
	(0.177)	(0.141)	(0.360)	(0.001)	(0.304)
Partial affirmance * Conviction (both rates centered)	0.001	0.004	0.002	0.007	0.003
	(0.899)	(0.596)	(0.677)	(0.167)	(0.621)
Conviction rate (centered) squared	-0.004	0.004	0.000	0.004	
	(0.704)	(0.710)	(0.940)	(0.486)	
Conviction rate (centered) cubic	-0.000	0.000	-0.000	0.000	
	(0.380)	(0.918)	(0.952)	(0.814)	
Criminal appeals decided on the merits (centered)	0.004*	0.000	0.001***	0.000	0.004
	(0.091)	(0.955)	(0.001)	(0.939)	(0.101)
SC full affirmance rate (centered)	-0.111**	-0.060	-0.051***	-0.060	-0.103**
	(0.024)	(0.567)	(0.000)	(0.596)	(0.032)
SC reversal rate (centered)	-0.093*	-0.062	-0.051***	-0.067	-0.085*
	(0.053)	(0.552)	(0.000)	(0.562)	(0.077)
SC partial affirmance rate (centered)	-0.142**	-0.156	-0.069***	-0.134	-0.133**
	(0.023)	(0.263)	(0.000)	(0.333)	(0.030)
SC dismissal rate (centered)	-0.103**	-0.071	-0.053**	-0.065	-0.097**
	(0.036)	(0.519)	(0.021)	(0.576)	(0.042)
Court and year fixed effects	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.943	0.958			0.943
Number of observations	1424.000	801.000	1424.000	801.000	1424.000

NOTE.– This table compares regression results of two-way clustered and panel-corrected standard errors (AR1) for the full range (1997-2013) and after the Booker case (2005-2013). Column (6) excludes polynomial terms. Appeals results are at the district court level. Control variables are identical to those used in tables 5 and 6. All regressions include court and year fixed effects. Constant and fixed effect coefficients are not reported. P.values are in parentheses under each coefficient. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

is prone to appeal sentences that accompany wrongful convictions, then an increase in this conditional f would decrease deterrence. A tendency to appeal the sentences of wrongfully convicted defendants may reflect misconduct (e.g. overly eager prosecutors seeking to ensure that innocent defendant get fully punished), but may also emerge if such sentences are simply usually lower (for example, due to "compromise verdicts" of judges and juries, as mentioned above). Cross-appeal rates are also insignificant when other rates are controlled for (column 4), but its reversal-remand spread is significant at the 10% level when these rates are excluded (column 3). Two coefficients of cross-appeals (with p.values marginally above 10%) deserve attention. First, full affirmance (squared) is negatively associated with crime (as in the base results). Since cross-appeals presumably correspond to those complex cases, where both parties believe that a mistake has been made, affirmances are a strong signal that district courts are accurate in both conviction and punishment imposition. Second, the interaction of *partial affirmance* and conviction rates has a positive coefficient, which can have two interpretations.

First, the same reasoning behind the basic results may apply, where high rates of convictions and affirmance cause a loss of faith in the system. This effect should be particularly strong for cross-appeals, if the public perceives such appeals as likely to be on solid grounds (since both parties claim an error occurred). However, this provides little guidance as to why *partial* affirmances interact with conviction rates. A second interpretation thus focuses on the fact that (1) both the defendant and the government appealed and (2) the decision was partially reversed. Specifically, the presence of a government appellant implies that the sentence is most likely very low. If defendants manage to achieve a reversal of such a sentence (while the conviction is affirmed), a lower expected sanction is implied. An interaction effect would then be related to the punishment rather than convictions.

8.3.3. An Instrumental variable approach: judical ideology as an instrument

In order to address any remaining endogeneity, I also employ an IV approach. For those readers unfamiliar with IV, the idea is using an additional variable ("an instrument") which is: (1) "exogenous," i.e. uncorrelated with the regressions' error term, and (2) "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 use IV for other stages of the legal chain (e.g. arrests (Cornwell and Trumbull, 1994), police hiring (Levitt, 1997, 2002) and prison population (Levitt, 1996)) and indeed, appeal results may also be suspected to be endogenous. However, the suspicion is stronger for the reversal-remand spread than for affirmance rates. Namely, changes in crime rates could, in theory, cause appellate judges to alternate between reversals and remand, making the choice of tool correlated with the regression's error term. Conversely, affirmance rates are less likely to be affected by crime rates, since appellate judges are constrained in their ability to manipulate affirmance rates by declaring that an error took place where it did not (or vice versa), due to the potential for Supreme court review.

To instrument for the *spread*, I use the ideology of judges in the appellate and Supreme courts, as captured by the Judicial Common Space scores ("JCS") (Epstein et al., 2007). In a nutshell, these scores build on the identity of appointing political actors to extract a measurement of ideological preferences on a continuous scale between "conservative" and "liberal". Supreme court ideology is measured through transformed Martin-Quinn scores (Martin and Quinn, 2002), which are based on judicial votes (rather than appointments). Judicial ideologies are known predictors of the tendencies to reverse (Hettinger and Lindquist, 2012; Iaryczower et al., 2013) and to remand (Boyd, 2015; Borochoff, 2008) in the federal courts. Some suggest that ideology is especially relevant in criminal cases, given their salience and tendency to invoke strong emotions (e.g. Borochoff, 2008, pp. 47). Since the assignment of appellate judges is generally random, it is exogenous to crime rates. Thus, insofar as ideology determines how cases are decided (i.e. reversed/remand), it should be both relevant and exogenous.

Table 8
ROBUSTNESS CHECK: DEFENDANT V. GOVERNMENT APPEALS

	(1) Defendant appeals	(2) Government appeals	(3) Cross appeals	(4) All appeals
Defendant Reversal-Remand spread	-0.060**			-0.061*
	(0.034)			(0.050)
Defendant Partial Affirmance rate	-0.032			-0.184
	(0.748)			(0.455)
Defendant Full Affirmance rate	0.091			0.671
	(0.854)			(0.215)
Defendant Full Affirmance × Conviction rate	0.004			0.003
	(0.245)			(0.389)
Defendant Partial Affirmance × Conviction rate	-0.001			-0.002
	(0.844)			(0.856)
Defendant Full Affirmance squared	0.001			0.008
	(0.835)			(0.157)
Defendant Partial Affirmance squared	-0.022			-0.025
	(0.450)			(0.500)
Defendant Full Affirmance cubic	-0.000			-0.000
	(0.749)			(0.185)
Defendant Partial Affirmance cubic	0.000			0.000
	(0.709)			(0.650)
Government Reversal-Remand spread		0.022*		0.006
		(0.082)		(0.748)
Government Partial Affirmance rate		0.011		0.131
		(0.878)		(0.596)
Government Full Affirmance rate		0.093		0.041
		(0.229)		(0.657)
Government Full Affirmance × Conviction rate		-0.001		-0.001
		(0.543)		(0.581)
Government Partial Affirmance × Conviction rate		-0.008		-0.004
		(0.141)		(0.698)
Government Full Affirmance squared		0.003		0.002
		(0.337)		(0.570)
Government Partial Affirmance squared		-0.007		-0.001
		(0.423)		(0.943)
Government Full Affirmance cubic		-0.000		-0.000
		(0.334)		(0.616)
Government Partial Affirmance cubic		0.000		-0.000
		(0.404)		(0.968)
Cross-Appeals Reversal-Remand spread			-0.021*	-0.021
			(0.091)	(0.136)
Cross-Appeals Full Affirmance rate			-0.084	-0.066
			(0.194)	(0.312)
Cross-Appeals Partial Affirmance rate			-0.011	-0.015
			(0.852)	(0.784)
Cross-Appeals Full Affirmance × Conviction rate			0.001	0.001
			(0.684)	(0.641)
Cross-Appeals Partial Affirmance × Conviction rate			0.002	0.002
			(0.114)	(0.102)
Cross-Appeals Full Affirmance squared			-0.002	-0.002
			(0.124)	(0.143)
Cross-Appeals Partial Affirmance squared			0.001	0.001
			(0.394)	(0.437)
Cross-Appeals Full Affirmance cubic			0.000	0.000
			(0.116)	(0.181)
Cross-Appeals Partial Affirmance cubic			-0.000	-0.000
			(0.776)	(0.870)
Criminal appeals decided on the merits (centered)	0.004*	0.003	0.003	0.002
	(0.063)	(0.114)	(0.182)	(0.250)
SC full affirmance rate (centered)	-0.109**	-0.110**	-0.098	-0.116**
	(0.027)	(0.034)	(0.133)	(0.042)
SC reversal rate (centered)	-0.092*	-0.090*	-0.077	-0.095*
	(0.055)	(0.078)	(0.228)	(0.089)
SC partial affirmance rate (centered)	-0.141**	-0.138**	-0.132*	-0.154**
	(0.023)	(0.029)	(0.068)	(0.026)
SC dismissal rate (centered)	-0.101*	-0.114**	-0.111*	-0.121**
	(0.054)	(0.020)	(0.098)	(0.042)
R-squared	0.943	0.942	0.946	0.946
Court and Year fixed effects	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
Number of observations	1424.000	1390.000	1319.000	1292.000

NOTE.– This table shows results for two-way clustered (by court and year) regression, using separate variables for defendant-filed, overnment-filed and cross appeals. Column (1) includes only defendant appeals. Column (2) includes only government appeals. Column (3) includes only cross-appeals. Column (4) includes all rates together. Appeals results are at the district court level. Control variables are the same as previous tables (5,6,8). 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. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

I utilize the following combinations of JCS as instruments:

1. *JCS of the median appellate judge:* captures time-variant ideology of the appellate court.
2. *JCS of the median Supreme court judge:* captures time-variant ideology of the SC. This affects the reversal-remand spread by constraining appellate judges. For example, if a conservative SC tends to affirm decisions more than a liberal SC, appellate judges are less likely to fear a reversal by the SC. This weaker monitoring constraint then affects the choice between reversing and remanding at the appellate court.
3. *The interaction of JCS at the appellate and Supreme courts:* allows the effects of ideology on the spread to depend on ideological differences. For example, a conservative SC may tend to affirm conservative decisions (issued by a conservative appellate court) more then liberal decisions. The appellate court may then respond by more reversals/remands.
4. *Standard deviation of median appellate judge JCS:* captures the degree of ideological heterogeneity. Since appellate cases are decided in judicial panels, the need for compromises and fear from reversal in an en-banc review,[60] can alter the rate of reversals/remands (see also Kastellec (2011) for a similar argument regarding the decision to dissent). Such effects depend, inter alia, on intra-court ideological heterogeneity.

To verify the validity and strength of these instruments, I use standard econometric tests.[61] Table 9 presents the IV results. Findings 1 and 2 of the base results are qualitatively reaffirmed: the spread negatively affects crime rates and affirmance rates have a negative marginal effect on crime rates (note the negative coefficients on the quadratic and cubic terms). The interaction effect has a positive effect on crime, as in Finding 3, but is insignificant. However, the endogeneity test indicates that endogeneity may not pose a problem to start with.[62] Other tests indicate that the instruments are relevant and exogenous.[63]

8.3.4. Using alternative instruments

While judicial ideology is arguably exogenous precisely since it predicts appellate judges' behavior irrespective of the case merits, the aspect of error occurrence at the trial court perhaps should not be excluded. Therefore, as a final robustness test, I run additional IV regressions

[60]"En-banc" reviews are procedures in which the appellate court reconsiders a case with a full quorum of all judges in the court.

[61]The following tests are used:

- 'Overidentifying restrictions' test' (Hansen, 1982), to ensure that instruments are jointly valid. Note that the null hypothesis is that of validity, thus a failure to reject the null indicates that instruments are indeed valid.
- 'Endogeneity test', to check whether there is a need to treat the spread as endogenous. The endogeneity test is implemented using the Stata command xtivreg2, endog() (Schaffer, 2012). And an "orthogonality test" (which is simply the flip-side of the endogeneity test, see Baum et al. (2007a, pp. 15)), to verify that all other control variables that are suspicious of endogeneity - including affirmance - can be treated as exogenous.
- Underidentification LM test (Kleibergen and Paap, 2006), where a rejection of the null hypothesis means that the model is not underidentified.
- Kleibergen-Paap F-statistic,(Kleibergen and Paap, 2006) which is the heteroskedasticity robust version of the Cragg-Donald statistic (Cragg and Donald, 1993). Instruments are deemed relevant if this statistic exceeds a threshold (Stock and Yogo, 2005) of a relative bias compared to OLS. For example, a threshold of 10% means that the IV estimator bias is smaller by at least 90 percentage points than the OLS bias.
- Anderson-Rubin (AR) test for significance of endogenous regressors (Anderson and Rubin, 1949). A rejection of the null hypothesis means that inference can be made even if instruments are only weakly relevant.

[62]The endogeneity Hausman-Durbin-Wu statistic is significant at the 10% but not the 5% level

[63]The Kleibergen-Paap F. value exceeds the Stock-Yogo 10% relative bias threshold of 9.08. The Hansen J p.value exceeds 10%.

Table 9

Instrumental variables: Ideology as an instrument

	(1)
Criminal Reversal-Remand spread	-0.949***
	(0.001)
Criminal Affirmance rate (centered)	0.693**
	(0.018)
Criminal affirmance * Conviction (both rates centered)	0.000
	(0.957)
Criminal affirmance rate (centered) squared	-0.035***
	(0.001)
Criminal affirmance rate (centered) cubic	-0.002***
	(0.003)
Criminal Dismissal rate (centered)	0.510**
	(0.041)
Criminal Remainder rate (centered)	-1.443**
	(0.013)
Hansen J (p.value)	0.280
Kleibergen-Paap LM (underidentification)	0.033
Kleibergen-Paap Wald (weak-IV)	11.359
Anderson-Rubin (weak-IV test) p.value	0.000
Endogeneity test	0.0931
Court & Year fixed effect	Yes
Control variables	Yes
Quadratic and Cubic terms	Yes
Number of observations	1424.000

NOTE.– This table shows regression results for the IV approach. Control variables are identical to those used in table 9, except for appeal results decided on the merits, which was excluded. Appeal rates are taken from the appellate-court level, as in the basic results. Instruments are: (1) Judicial common score of the median judge in the appellate court. (3) Judicial common score of the median Supreme Court judge (3) Interaction between the Judicial Common Scores of the median judges in the apellate and Supreme court. (4) Standard deviation of the Judicial Common Score of the median judge in the appeallte court. Standard errors are two-way clustered by court and year. In order to allow for the two-way clustering, year fixed effects and control variables are partialled-out. A two-step GMM estimator is implemented in the regression. P.values are in parentheses under each coefficient. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

using the sentencing appeal rates (which avoids granularity problems), while simultaneously instrumenting for the spread and affirmance rates - but using different instruments.

The different instrument set consists of reversal-remand spreads *in non-criminal cases*, relying on those general factors that plausibly carry over from criminal to non-criminal cases. For example, remands may be used as a restraining tool, to ensure that trial judges comply or otherwise suffer the effort cost of reconsidering the case (Boyd, 2015), irrespective of the nature of proceeding. Namely, I include three spreads, of: (1) administrative (2) private-civil and (3) other-civil cases. In light of the affirmance-conviction interaction, I interact also these instruments with the conviction rate, as suggested by Bun et al. (2014).

Administrative appeals are presumably exogenous, since they are unlikely to be affected by crime, especially not by the types of offenses considered here.[64] However, administrative appeals typically originate from administrative agencies rather than district courts.[65] and thus may miss factors relating to trial courts. Civil appeals, conversely, originate from district courts, providing higher predictive power while mainting a distance from crime related issues.

[64] Recall that 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.

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

However, it cannot be ruled out that some civil litigation that involves damages caused by crimes may render civil appeals imperfectly exogenous. However, the main concern for all non-criminal spreads is perhaps the possibilty that these are correlated with some unobserved factor, such as type II errors. For example, administrative appeals are a manifestation of a conflict between a private entity and a public authority, where an incorrect ruling in favor of the administrative petitioner can bear some resemblance to a wrongful acquittal.

Therefore, to ensure the validity of the additional IV regressions, I enlarge the instrument set by employing the Lewbel (2012) approach, which uses higher order moments to generate valid instruments using heteroskedasticy.[66] This methods is able to produce valid estimators where external instruments are unavailable (Mishra and Smyth, 2015, pp. 167).[67] The additional IV estimations are presented in table 10.[68] The tests indicate that, once more, no endogeneity issues arise and that instruments are relevant and exogenous.[69] Otherwise, results are qualitatively similar to the base findings.

Summarizing, the base results are generally robust to different specifications, especially for defendant-filed appeals - which are the theoretically relevant kind for testing type I errors. An interaction effect between affirmance and conviction rates seems to be in play, but it is unclear what is driving the effect - disbelief in the system when convictions and affirmance rates are simultaneously high or a side-effects of either reduced sentencing or government appellants.

9. Implications

9.1. General

My results offer several implications. First, the EBM's argument that judicial errors decrease deterrence seems to hold empirically. However, a theoretical adjustment is needed, such that error correction is incorporated. Theory should further reflect how errors are handled by appellate courts, as reversals and remand have opposite effects. Since appeal results are correlated with other stages of the legal process, the results also imply that some previous evidence on deterrence may be somewhat inaccurate, insofar that the omitted variable bias caused by ignoring appeal results is unaccounted for. My results further imply a need for caution when choosing how to correct errors, such that appellate judges who care about deterrence should prefer reversals to remands.[70] Trial judges with deterrence concerns should

[66] The regression is run using the user-developed command ivreg2h (Baum and Schaffer, 2014).

[67] One limitation of this approach, however, is its reliance on the assumption that heteroskedastic covariates are exogenous (see also Baum et al., 2012).While it is unclear to which extent control variables are exogenous, some seem more likely to satisfy the exogeneity assumption. For example, previous studies find that higher crime rates do not necessarily lead to higher conviction rates (Jensen and Heller, 2003, pp. 83), indicating that convictions (at least) do not suffer from reversed causality. Subsequently, interaction terms with conviction rates are plausibly exogenous as well, since interactions of endogenous variables with exogenous variables are exogenous, as long as the endogenous variable is controlled for. (Bun et al., 2014; Nizalova and Murtazashvili, 2016)

[68] I use a two-stage GMM estimator, which is asymptotically more efficient than the traditional 2SLS estimator (Baum et al., 2007a; Lee, 2003). Variables are centered. I exclude the control variable of appeals decided on the merits, to avoid confounding effects that emerge since this variable is the denominator of the appeal rates. The exclusion does not change the results qualitatively.

[69] Since the regressions include multiple variables, I report on F-tests as developed by Angrist and Pischke (2008) and modified by Sanderson et al. (2016) for individual regressors. These F-tests can be compared to Stock-Yogo critical values. All variables fulfill the condition $F > 10.968$, which is the relevant cutoff value (10% relative bias). Note that since I only have three endogenous regressors, regular Stock-Yogo critical values are also still available for the overall model.

[70] Theoretically, judges may already be aware of this effect and respond, 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).

Table 10
INSTRUMENTAL VARIABLES: HETEROSKEDASTICITY-BASED AND EXTERNAL INSTRUMENTS

	(1) Generated instruments only	(2) All instruments
Defendant Reversal-Remand spread	-0.066***	-0.047***
	(0.000)	(0.000)
Defendant Full Affirmance	-0.064***	-0.064***
	(0.000)	(0.000)
Defendant Partial Affirmance	-0.252***	-0.269***
	(0.000)	(0.000)
Defendant Full Affirmance * Conviction	0.006***	0.005***
	(0.000)	(0.000)
Defendant Partial Affirmance * Conviction	0.023***	0.024***
	(0.000)	(0.000)
Hansen J (p.value)	0.411	0.253
Anderson-Rubin (weak-IV test) p.value	0.000	0.000
Endogeneity test	0.918	0.926
Control variables	Yes	Yes
Court fixed effects	Yes	Yes
Year fixed effects	No	No
Number of observations	1424.000	1424.000

	Weak-IV F-test (Sanderson-Windmeijer)	
Defendant Reversal-Remand spread	Defendant Full Affirmance rate	Defendant Partial Affirmance rate
152.1579	118.6081	38.05601

NOTE.– This table shows regression results for the IV approach. Three variables are instrumented: Defendant reversal-remand spread, Defendant Full affirmance and Defendant partial affirmance. Column 1 presents IV results using only heteroskedasticity-based generated instruments. Column 2 presents results using all instruments: (I) heteroskedasticity-based instruments (II) administrative reversal-remand spread (III) private civil appeals reversal-remand spread (IV) other civil appeals reversal-remand spread and (V) the interactions of conviction rates with (II), (III) and (IV). Instruments are taken from the appeal court level. Control variables are the same as previous tables. Variables are all centered by court to include court fixed effects. Standard errors are clustered at the district court level. A two-step GMM estimator is implemented in all regressions. P.values are in parentheses under each coefficient. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

be similarly error averse, especially since my findings indicate that preventing errors a-priori may have a much larger positive effect on deterrence than correcting errors ex-post. Trial judges who are considering to free-ride on the appellate court's accuracy (Feess and Sarel, 2016), should thus be aware that a retrospective remedy is not as efficient. For policy makers, the findings imply that judicial errors and their correction should be integrated into a cost-benefit analysis, when deciding on resource allocation. From a social welfare perspective, errors should only be avoided or corrected when the benefit from correction exceeds the social cost, but calculating the cost requires a recognition of the effect of appeals. My results are then especially important as a first indication for the numerical magnitude of the effect.

9.2. Final remarks on the effect of remands on deterrence

Remands are usually justified by either (1) "institutional superiority" of the first instance to correct errors cheaper; (2) A narrow jurisdiction of appellate 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 of trial court judges (Boyd, 2015). The various costs of remands identified in the literature do not include deterrence effects. The deterrence effects of remands as reviewed in section 3 above, can be roughly aggregated into three competing arguments: (1) A delay in the sanction (2) loss of faith in the judicial system and (3) cost asymmetry for the guilty and the innocent.[71]

If my results are driven by delayed sanctions, there is perhaps little to be done other than limiting the scope of remands or speeding up the legal process, which can be 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 stays of sentence 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 appellate courts do not only decide whether to remand but also *how* to remand, the problem can be attacked on several fronts. First, by strengthening appellate courts' role in giving guidance to lower courts. For example, providing more specific instructions can reduce ambiguity (Hessick, 2012, pp. 5-6) and encourage error correction (Boyd, 2015). If the source of public disbelief are remands back to the same panel, appellate courts should perhaps be restricted from doing so. Hannibal and Worth (2012) note, that the authority to enforce reassignment to another judge (28 U.S.C 2106) is reluctantly used, as to avoid implicit votes of no-confidence in the trial judge. Nonetheless, some courts have developed rules or criteria that dictate when and if cases are reassigned (Scheinfeld and Bagley, 2013). Unfortunately, my results would suggest that these measures are insufficient to prevent declines in deterrence.

Finally, if a cost-asymmetry explains the effect, additional compensation for wrongful convictions following a remand process may be considered. A cost reimbursement for acquitted defendants can balance the effect of judicial errors (Fon and Schäfer, 2007), but has been argued to be ineffective in the U.S. (Kahn, 2010). Improving reimbursement mechanisms for post-remand acquittals may thus provide relief.

There exists of course also one extreme measure that would circumvent these problems at once: abolishing 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 judges to alternate between affirmance, reversals and dismissal. It is unlikely that judges will switch remand-worthy cases to affirmance 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 to lower prosecution rates. Conversely, incentivizing prosecutors to focus on strong cases may prevent wrongful prosecutions, thereby reducing wrongful convictions. In an alternative scenario, judges may dismiss appeals that would otherwise be remanded. This might lead to a higher loss of faith in the system, as convictions will be less

[71] 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.

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

Deterrence models aiming at effective predictions of how rational criminals will behave as a function of expected sanctions need to consider how a multi-tier court system influences crime commission incentives. The analysis in this paper suggests that the effects of appellate courts on deterrence are complex, given the countervailing influences of reversals and remand on the components of the expected sanction (p and f). While the choice between the two paths to error correction also involves considerations other than deterrence (e.g. effort cost, procedural constraints and division of labor between lower and upper courts), the effects on deterrence should clearly be incorporated in the cost-benefit analysis of appellate process design.

The results of this paper are subjected to Several limitations: Measurement errors may occur due to aggregations of crime rates and appeal results; categorization of appeal results may not be precise enough to capture all effects; the sample may be imperfectly random, since only years with available data have been included; crime rates may not fully reflect deterrence[72]; Ideological scores may be imperfectly measured[73] and dynamic effects are not considered. I address most concerns throughout the paper, but further research should be conducted in order to address remaining issues. Nonetheless, my results are important as a first step in the empirical analysis of the effects of appellate courts on deterrence.

Acknowledgements

I thank Eberhard Feess, Ronald Wintrobe, Darwyn Deyo, Ansgar Wohlschlegel, Eric Talley, Robin Christmann, Yun-Chien Chang, Hartmut Kliemt and participants of the following conferences: European Law & Economics association (2016), German Law & Economic association (2016), First Conference on Empirical Legal Studies in Europe (2016), Public choice society (2016) and European Public Choice Society (2016), for their useful comments. This work was supported by the Frankfurt School of Finance & Management.

References

Afshartous, D. and Preston, R. A. (2011). Key results of interaction models with centering. *J. Stat. Educ.*, 19(3):1–24.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals Math. Stat.*, pages 46–63.

Andreoni, J. (1991). Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime? *RAND J. Econ.*, pages 385–395.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harm-*

[72] Among else, crime rates may decrease artificially, although deterrence did not change. For example, if the definition of crimes is changed, the same activity might take place but crime rates will be different.

[73] For a historical review and recent criticism of these and other ideological scores, see Bonica et al. (2016)

less econometrics: An empiricist's companion. Princeton university press.

Baker, T., Harel, A., and Kugler, T. (2003). Virtues of uncertainty in law: An experimental approach, the. *Iowa L. Rev.*, 89:443.

Balli, H. O. and Sørensen, B. E. (2013). Interaction effects in econometrics. *Empir. Econ.*, 45(1):583–603.

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

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*, pages 13–14.

Baum, C. F. and Schaffer, M. E. (2014). Ivreg2h: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments. *Stat. Softw. Components*.

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

Baum, C. F., Schaffer, M. E., Stillman, S., et al. (2007b). ivreg2: Stata module for extended instrumental variables/2sls, gmm and ac/hac, liml and k-class regression. *mimeo*.

Baumann, F. and Friehe, T. (2015). Proof beyond a reasonable doubt: Laboratory evidence. *mimeo*.

Beck, N. and Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *Am. Polit. Sci. Rev.*, 89(03):634–647.

Becker, G. S. (1968). Crime and punishment: An economic approach. *J. Polit. Econ.*, 76(2):169–217.

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

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

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

Bonica, A., Chilton, A. S., Goldin, J., Rozema, K., Sen, M., and Kennedy, J. F. (2016). Measuring judicial ideology using law clerk hiring. *Am. L. Econ. Rev.*, page ahw013.

Borochoff, E. (2008). Lower court compliance with supreme court remands. *Touro L. Rev.*, 24:849.

Boyd, C. L. (2015). The hierarchical influence of courts of appeals on district courts. *J. Leg. Stud.*, 44(1):113–141.

Bun, M. J., Harrison, T. D., et al. (2014). Ols and iv estimation of regression models including endogenous interaction terms. *Univ. Amsterdam Discuss. Pap.*, 2.

Cameron, A. C. and Miller, D. L. (2015). A practitioners guide to cluster-robust inference. *J. Hum. Resour.*, 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? *Manag. Sci.*, 39(4):458–476.

Chalfin, A. and McCrary, J. (2014). Criminal deterrence: A review of the literature. *J Econ Lit*.

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

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

Cornwell, C. and Trumbull, W. N. (1994). Estimating the economic model of crime with panel data. *Rev. Econ. Stat.*, pages 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. *Stat. Softw. Components*.

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

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

Craswell, R. and Calfee, J. E. (1986). Deterrence and uncertain legal standards. *J. Law, Econ. Organ.*, 2:279.

Cummins, T., Aft, A., and Cumby, J. C. (2015). Appellate review iii. *J. Leg. Metrics*.

Dalla Pellegrina, L. (2008). Court delays and crime deterrence. *Eur. J. L. Econ.*, 26(3):267–290.

Dari-Mattiacci, G. and Deffains, B. (2007). Uncertainty of law and the legal process. *J. Institutional Theor. Econ. (JITE)*, 163(4):627–656.

Davis, M. L. (1988). Time and punishment: an intertemporal model of crime. *J. Polit. Econ.*, pages 383–390.

DeAngelo, G. and Charness, G. (2012). Deterrence, expected cost, uncertainty and voting: Experimental evidence. *J. Risk Uncertain.*, 44(1):73–100.

Dhami, S. and al Nowaihi, A. (2013). An extension of the becker proposition to non-expected utility theory. *Math. Soc. Sci.*, 65(1):10–20.

Di Tella, R. and Schargrodsky, E. (2013). Criminal recidivism after prison and electronic monitoring. *J. Polit. Econ.*, 121(1):28–73.

Diallo, M., Fay, R., et al. (2015). Developmental estimates of subnational crime rates based on the national crime victimization survey. *mimeo.*

Dickens, W. T. (1986). Crime and punishment again: the economic approach with a psychological twist. *J. Public Econ.*, 30(1):97–107.

Domenech, G. and Puchades, M. (2014). Better that ten innocent persons suffer than that one guilty scape. Technical report, Working paper, University of Valencia, Valencia, Spain.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Rev. Econ. Stat.*, 80(4):549–560.

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

Dušek, L. (2015). Time to punishment: The effects of a shorter criminal procedure on crime rates. *Int. Rev. L. Econ.*, 43:134–147.

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

Ehrlich, I. (1982). The optimum enforcement of laws and the concept of justice: a positive analysis. *Int. Rev. L. Econ.*, 2(1):3–27.

Entorf, H. and Spengler, H. (2000). Criminality, social cohesion and economic performance. *WEP Wuerzburg Econ. Pap. No. 00-22.*

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

Epstein, L., Landes, W. M., and Posner, R. A. (2013). *The behavior of federal judges.* Harvard University Press.

Epstein, L., Martin, A. D., Segal, J. A., and Westerland, C. (2007). The judicial common space. *J. Law, Econ. Organ.*, 23(2):303–325.

Evans, W. N. and Owens, E. G. (2007). Cops and crime. *J. Public Econ.*, 91(1):181–201.

Feess, E. and Sarel, R. (2016). Judicial effort and the appeal system: theory and experiment. *mimeo.*

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

Fon, V. and Schäfer, H.-B. (2007). State liability for wrongful conviction: Incentive effects on crime levels. *J. Institutional Theor. Econ. (JITE)*, pages 269–284.

Friesen, L. (2012). Certainty of punishment versus severity of punishment: An experimental investigation. *South. Econ. J.*, 79(2):399–421.

Garoupa, N. and Rizzolli, M. (2012). Wrongful convictions do lower deterrence. *J. Institutional Theor. Econ. (JITE)*, 168(2):224–231.

Gathmann, C. (2008). Effects of enforcement on illegal markets: Evidence from migrant smuggling along the southwestern border. *J. Public Econ.*, 92(10):1926–1941.

Gay, G. D., Grace, M. F., Kale, J. R., and Noe, T. H. (1989). Noisy juries and the choice of trial mode in a sequential signalling game: theory and evidence. *Rand J. Econ.*, pages 196–213.

Glaeser, E. L. and Sacerdote, B. (1999). Why is there more crime in cities? *J. Polit. Econ.*, 107(6 pt 2).

Gneezy, U. and Rustichini, A. (2004). Incentives, punishment and behavior. *Adv. Behav. Econ.*, pages 572–89.

Gould, E. D. et al. (2002). Crime rates and local labor market opportunities in the united states: 1979–1997. *Rev. Econ. Stat.*, 84(1):45–61.

Gould, J. B., Carrano, J., Leo, R. A., and Hail-Jares, K. (2014). Predicting erroneous convictions. *Iowa L. Rev.*, 99:471–2299.

Gross, S. R. and O'Brien, B. (2008). Frequency and predictors of false conviction: Why we know so little, and new data on capital cases. *J. Empir. Leg. Stud.*, 5(4):927–962.

Gross, S. R., OBrien, B., Hu, C., and Kennedy, E. H. (2014). Rate of false conviction of criminal defendants who are sentenced to death. *Proc. Natl. Acad. Sci.*, 111(20):7230–7235.

Gross, S. R. and Shaffer, M. (2012). *Exonerations in the United States, 1989 Through 2012: Report by the National Registry of Exonerations*. University of Michigan Law School.

Hannibal, K. and Worth, J. A. (2012). Judicial reassignment: a proposal. *The Natl. L. J.*.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica.*, pages 1029–1054.

Harris, J. R. (1970). On the economics of law and order. *J. Polit. Econ.*, pages 165–174.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica.*, pages 1251–1271.

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

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

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

Iaryczower, M., Lewis, G., and Shum, M. (2013). To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *J. Public Econ.*, 97:230–244.

Ihlanfeldt, K. R. (2007). Neighborhood drug crime and young males' job accessibility. *Rev. Econ. Stat.*, 89(1):151–164.

Iyengar, R. (2009). Does the certainty of arrest reduce domestic violence? evidence from mandatory and recommended arrest laws. *J. Public Econ.*, 93(1):85–98.

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

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

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

Kahan, D. M. (1996). Between economics and sociology: The new path of deterrence. *Mich. L. Rev.*, 95:2477.

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. *J. Leg. Stud.*, pages 307–401.

Kaplow, L. (2011). On the optimal burden of proof. *J. Polit. Econ.*, 119(6):1104–1140.

Kastellec, J. P. (2011). Hierarchical and collegial politics on the us courts of appeals. *The J. Polit.*, 73(02):345–361.

Khadjavi, M. (2014). Deterrence works for criminals. *Eur. J. L. Econ.*, pages 1–14.

Khadjavi, M. (2015). On the interaction of deterrence and emotions. *J. Law, Econ. Organ.*, 31(2):287–319.

Kim, S., Park, J., Park, K., and Eom, J.-S. (2013). Judge-jury agreement in criminal cases: The first three years of the korean jury system. *J. Empir. Leg. Stud.*, 10(1):35–53.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *J. Econom.*, 133(1):97–126.

Korobkin, R. B. and Ulen, T. S. (2000). Law and behavioral science: Removing the rationality assumption from law and economics. *California L. Rev.*, pages 1051–1144.

Kovandzic, T. V. and Vieraitis, L. M. (2006). The effect of county-level prison population growth on crime rates. *Criminol. Public Policy*, 5(2):213–244.

Lando, H. (2006). Does wrongful conviction lower deterrence? *J. Leg. Stud.*, 35(2):327–337.

Lee, D. S. and McCrary, J. (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. *Econom. Rev.*, 22(4):307–335.

Levitt, S. D. (1996). The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. *Q. J. Econ.*, 111(2):319–351.

Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The Am. Econ. Rev.*, pages 270–290.

Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *Am. Econ. Rev.*, pages 1244–1250.

Levy, G. (2005). Careerist judges and the appeals process. *RAND J. Econ.*, pages 275–297.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. & Econ. Stat.*, 30(1):67–80.

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

Lin, M.-J. (2008). Does unemployment increase crime? evidence from us data 1974–2000. *J. Hum. Resour.*, 43(2):413–436.

Livernois, J. and McKenna, C. J. (1999). Truth or consequences: enforcing pollution standards with self-reporting. *J. Public Econ.*, 71(3):415–440.

Lundberg, A. (2016). Sentencing discretion and burdens of proof. *Int. Rev. L. Econ.*.

Machin, S. and Meghir, C. (2004). Crime and economic incentives. *J. Hum. Resour.*, 39(4):958–979.

Martin, A. D. and Quinn, K. M. (2002). Dynamic ideal point estimation via markov chain monte carlo for the us supreme court, 1953–1999. *Polit. Analysis*, 10(2):134–153.

Mastrobuoni, G. and Rivers, D. A. (2016). Criminal discount factors and deterrence. *Available at SSRN 2730969*.

Masur, J. and Bronsteen, J. (2015). The overlooked benefits of the blackstone principle. *Harv. L. Rev. Forum*, 128:289.

Mishra, V. and Smyth, R. (2015). Estimating returns to schooling in urban china using conventional and heteroskedasticity-based instruments. *Econ. Model.*, 47:166–173.

Montag, J. and Tremewan, J. (2016). Let the punishment fit the criminal: an experimental study. Technical report.

Mungan, M. C. and Lando, H. (2015). The effect of type-1 error on deterrence. *FSU Coll. Law, Public L. Res. Pap. 687*.

Mustard, D. B. (2003). Reexamining criminal behavior: the importance of omitted variable bias. *Rev. Econ. Stat.*, 85(1):205–211.

Nicita, A. and Rizzolli, M. (2014). In dubio pro reo. behavioral explanations of pro-defendant bias in procedures. *CESifo Econ. Stud.*.

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

Olney, M. and Bonn, S. (2014). An exploratory study of the legal and non-legal factors associated with exoneration for wrongful conviction: The power of dna evidence. *Crim. Justice Policy Rev.*, pages 1–21.

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *J. Quant. Criminol.*, 16(1):21–43.

Ouss, A. and 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? *L. Contemp. Probl.*, pages 129–148.

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

Pfander, J. E. (2007). Removing federal judges. *The Univ. Chic. L. Rev.*, pages 1227–1250.

Plerhoples, C. and Summit, C. F. P. (2012). The effect of vacant building demolitions on crime under depopulation. *Econ. job market paper. Lansing: Mich. State Univ. Dep. Agric. Food, Resour.*.

Png, I. P. (1986). Optimal subsidies and damages in the presence of judicial error. *Int. Rev. L. Econ.*, 6(1):101–105.

Polinsky, A. M. and Shavell, S. (2000). The economic theory of public enforcement of law. *J. Econ. Lit.*, 38(1):45–76.

Polinsky, A. M. and Shavell, S. (2001). Corruption and optimal law enforcement. *J. Public Econ.*, 81(1):1–24.

Polinsky, A. M. and Shavell, S. (2007). The theory of public enforcement of law. *Handb. L. Econ.*, 1:403–454.

Posner, R. A. (1973). An economic approach to legal procedure and judicial administration. *J. Leg. Stud.*, pages 399–458.

Rizzolli, M. and 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. and Stanca, L. (2012). Judicial errors and crime deterrence: theory and experimental evidence. *J. L. Econ.*, 55(2):311–338.

Rizzolli, M. and Tremewan, J. (2016). Hard labour in the lab: Are monetary and non-monetary sanctions really substitutable? *mimeo*.

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

Sah, R. K. (1991). Social osmosis and patterns of crime: A dynamic economic analysis. *J. Polit. Econ.*, 99(6).

Sanderson, E., Windmeijer, F., et al. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *J. Econom.*, 190(2):212–221.

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. *Stat. Softw. Components*.

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

Schildberg-Hörisch, H. and Strassmair, C. (2012). An experimental test of the deterrence hypothesis. *J. Law, Econ. Organ.*, 28(3):447–459.

Scott, K. M. (2006). Understanding judicial hierarchy: Reversals and the behavior of intermediate appellate judges. *L. & Soc. Rev.*, 40(1):163–192.

Shavell, S. (1995). The appeals process as a means of error correction. *J. Leg. Stud.*, 24(2):379–426.

Shavell, S. (2006). The appeals process and adjudicator incentives. *J. Leg. Stud.*, 35(1):1–29.

Simcoe, T. (2008). Xtpqml: Stata module to estimate fixed-effects poisson (quasi-ml) regression with robust standard errors. *Stat. Softw. Components*.

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

Soares, Y. and Sviatschi, M. M. (2010). Does court efficiency have a deterrent effect on crime? evidence for costa rica. *mimeo*.

Spaeth, H. J., Epstein, L., et al. (2016). 2016 supreme court database, version 2016 release 1, url: http://supremecourtdatabase.org.

Sparrow, M. K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Soc. networks*, 13(3):251–274.

Spencer, B. D. (2007). Estimating the accuracy of jury verdicts. *J. Empir. Leg. Stud.*, 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 Univ. Chic. L. Rev.*, pages 1–61.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, chapter 5, pages 80–109.

Thaler, R. (1977). An econometric analysis of property crime: interaction between police and criminals. *J. Public Econ.*, 8(1):37–51.

Torre, A. (2003). The impact of court delays on the prosecutor and the defendant: An economic analysis. *Eur. J. L. Econ.*, 16(1):91–111.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cogn. Psychol.*, 5(2):207–232.

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

Tversky, A. and Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica.*, pages 263–291.

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

Weatherburn, D. (2012). The effect of arrest and imprisonment on crime. *Crime Justice Bull.*, 158:1–20.

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

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

Worrall, J. L. and Pratt, T. C. (2004). Estimation issues associated with time-series-cross-section analysis in criminology. *W. Criminol. Rev.*, 5:35.

Zamir, E. and Ritov, I. (2012). Loss aversion, omission bias, and the burden of proof in civil litigation. *J. Leg. Stud.*, 41(1):165–207.

A. Data annex

The following datasets have been combined in the analysis:

Supreme court database Harold J. Spaeth, Lee Epstein et al. 2016 Supreme Court Database, Version 2016 Release 1, URL: http://Supremecourtdatabase.org

Buro of Justice statistics

- Federal Criminal Case Processing Statistics. URL: http://www.bjs.gov/fjsrc/
- Corrections Statistical Analysis Tool (CSAT) - Prisoners. URL: http://www.bjs.gov/index.cfm?ty=nps

Federal courts statistics Judicial business of the U.S. courts (1997-2013). URL: http://www.uscourts.gov/

ICPSR database: appellate courts

- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 1993-1998 . ICPSR06559-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004-11-05. http://doi.org/10.3886/ICPSR06559.v4
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 1999 . ICPSR03105-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. http://doi.org/10.3886/ICPSR03105.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2000 . ICPSR03493-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2002. http://doi.org/10.3886/ICPSR03493.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2001 . ICPSR03494-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2002. http://doi.org/10.3886/ICPSR03494.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2001 . ICPSR03494-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2002. http://doi.org/10.3886/ICPSR03494.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2002 . ICPSR04108-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. http://doi.org/10.3886/ICPSR04108.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2003. ICPSR04632-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-03-20. http://doi.org/10.3886/ICPSR04632.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2004. ICPSR04629-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-03-20. http://doi.org/10.3886/ICPSR04629.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2005. ICPSR04627-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-03-21. http://doi.org/10.3886/ICPSR04627.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2006. ICPSR20101-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-09-26. http://doi.org/10.3886/ICPSR20101.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2007. ICPSR22624-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-03-08. http://doi.org/10.3886/ICPSR22624.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2008. ICPSR25425-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-03-04. http://doi.org/10.3886/ICPSR25425.v2

- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2008. ICPSR25425-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-03-04. http://doi.org/10.3886/ICPSR25425.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2009. ICPSR28601-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-09-24. http://doi.org/10.3886/ICPSR28601.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2010. ICPSR35337-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-09-24. http://doi.org/10.3886/ICPSR35337.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2011. ICPSR35340-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-11-25. http://doi.org/10.3886/ICPSR35340.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2012. ICPSR35343-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-12-03. http://doi.org/10.3886/ICPSR35343.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Convictions and Sentences: Appeals Data, 2013. ICPSR35346-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-12-03. http://doi.org/10.3886/ICPSR35346.v1

ICPSR database: district courts

- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 1987-1998. ICPSR09317-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012. http://doi.org/10.3886/ICPSR09317.v5
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 1999. ICPSR03106-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. http://doi.org/10.3886/ICPSR03106.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2000. ICPSR03496-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2002. http://doi.org/10.3886/ICPSR03496.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2001. ICPSR03497-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. http://doi.org/10.3886/ICPSR03497.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2002. ICPSR04110-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004. http://doi.org/10.3886/ICPSR04110.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2003. ICPSR04290-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007. http://doi.org/10.3886/ICPSR04290.v3
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2004. ICPSR04633-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007. http://doi.org/10.3886/ICPSR04633.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2005. ICPSR04630-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007. http://doi.org/10.3886/ICPSR04630.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2006. ICPSR20120-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007. http://doi.org/10.3886/ICPSR20120.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2007. ICPSR22623-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009. http://doi.org/10.3886/ICPSR22623.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2008. ICPSR25424-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009. http://doi.org/10.3886/ICPSR25424.v2
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2009. ICPSR28602-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014. http://doi.org/10.3886/ICPSR28602.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2010. ICPSR35336-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014. http://doi.org/10.3886/ICPSR35336.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2011. ICPSR35339-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014. http://doi.org/10.3886/ICPSR35339.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2012. ICPSR35342-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014. http://doi.org/10.3886/ICPSR35342.v1
- United States Sentencing Commission. Monitoring of Federal Criminal Sentences, 2013. ICPSR35345-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014. http://doi.org/10.3886/ICPSR35345.v1