Do Wrongfully Convicted Defendants Receive Higher Punishments?

Roee Sarel

Institute of Law and Economics, University of Hamburg

Draft version: April 9, 2020

Abstract

This paper explores theoretically and empirically whether wrongfully convicted defendants receive different punishments than correctly convicted defendants. I develop a theoretical model in which judges observe a signal – the level of incriminating evidence – and then engage in 'compromise verdicts' by adjusting the punishment size to the probability that the defendant is guilty. Due to this adjustment, wrongfully convicted defendants systematically receive lower punishments, as long as judges' intrinsic cost of punishing the innocent is sufficiently high. Type I errors (wrongful convictions) and type II errors (wrongful acquittals) will then generally affect crime deterrence asymmetrically. I test the model's prediction in an empirical analysis of criminal sentences in the U.S and find that exonerated defendants receive lower punishments compared to non-exonerated defendants. In line with the theoretical prediction, I also find that the punishment of exonerees increases in the level of the incriminating evidence.

JEL Classification: K14, K42, D02

Keywords: Wrongful convictions, compromise verdicts, sentencing, crime deterrence, judicial errors.

Introduction

A corner stone of criminal justice is the idea that "it is better that ten guilty persons escape than that one innocent suffers" (see e.g. Epps, 2015). Preventing the suffering of innocents, however, can be achieved in two alternative ways. First, one can reduce the *probability* that innocents will suffer, by minimizing the frequency of wrongful convictions. Second, one can reduce the *degree* of suffering, so that if wrongful convictions do occur - the punishment imposed on innocents will be less severe.

The criminal justice system usually opts for the first alternative, e.g. by setting a high standard of proof in criminal cases. At the same time, unified punishments are typically declared for each crime. However, judges may use their discretion to indirectly adopt the second alternative by engaging in a strategy known as "compromise verdicts" (Lundberg, 2016). In this strategy, judges balance the punishment size and the probability of guilt, so that defendants convicted of the same crime will nonetheless receive different punishments, depending on how guilty they seem to be. Subsequently, judges who are unsure of the defendant's guilt may prefer to still convict, but compensate for their uncertainty by reducing the penalty. Such a strategy will usually lead judges to impose lower sentences on innocents (who usually seem less guilty), so that wrongfully convicted defendants will receive lower punishments compared to correctly convicted defendants.

However, a compromise verdicts strategy may also backfire: in misleading cases, where an innocent defendant seems very guilty, the (seemingly) high probability of guilt may induce the judge to both convict and impose a higher penalty. Innocents may look very guilty due to bad luck, but also due to more systematic effects, such as cases where defendants are framed, coerced into a (false) confession, or improperly represented at trial. If such phenomena were frequent, then – conditional on conviction – one would expect that the wrongfully convicted would actually receive higher punishments. Innocents may also receive higher punishments due to other reasons, e.g. because they are less likely to accept plea bargains (Tor, Gazal-Ayal, and Garcia, 2010; Lundberg, 2018) and to express remorse during sentencing (Heath, 2009), both of which typically yield more lenient sentences.

If wrongful convictions were an extremely rare event, disparities in punishment could perhaps be neglected. Unfortunately, that is not the case: Existing studies estimate the frequency of wrongful convictions to be anywhere between a minor rate of 0.016% and a staggering rate of 15% (see Gross and O'Brien, 2008; Gould and Leo, 2010; Clow and Leach, 2015; Cassell, 2018; Thomas III, 2018), where for the harshest sentence – the death penalty – the wrongful conviction rate is estimated at 4% (Gross, 2016).

Knowing whether the innocent and the guilty face similar punishments is important not only for criminal justice, but also for crime deterrence. Following the canonical Becker (1968) model, in which rational individuals compare expected payoffs from crime vs. innocence, the literature assumes that individuals respond to the probability of apprehension and conviction (p) and the punishment size (f). Most papers then treat p as two conditional probabilities: one p applying to those who commit a crime and another p to those who abstain. This respectively captures the probabilities of wrongful acquittals ("type II errors") and wrongful convictions ("type I errors").[1] Both error types are then argued to lower deterrence, where wrongful acquittals make crime more worthwhile whereas wrongful convictions make innocence less worthwhile. Although several criticisms of this approach have arisen (see e.g. Lando, 2006; Rizzolli and Stanca, 2012; Nicita and Rizzolli, 2014; Lando and Mungan, 2017), the discussion is still based on an implicit assumption that, for a given crime, all convicted defendants receive the same punishment.

There are two possible reasons for why the same f is assumed for all convicted defendants. First, one might assume that as judges do not know which defendants are guilty, they behave dichotomously: convict and impose a pre-determined penalty when they are sufficiently convinced of the defendant's guilt, and acquit otherwise. Second, one might assume that judges either have no discretion at all (e.g. because the penalty is exogenously set by law) or no incentive to impose heterogeneous punishments (e.g. because the punishment should simply "fit the crime" irrespective of the probability of guilt).

Such assumptions are sufficient for some analyses, but are somewhat unrealistic. Although judges do not know which defendants are guilty, the evidence presented at trial do provide an indication for the probability of guilt. Furthermore, penalties are seldom exogenous, as adjudicators usually have wiggle room in their decision of which sanction to impose. For instance, judges are sometimes bound by a maximal penalty set by law but can usually deviate downwards;[2] juries typically do not determine the sentence (King and Noble, 2004), but can decide (not) to convict for crimes that bear a lower (higher) sanction (e.g. convict for manslaughter but not murder). Finally, as the compromise verdicts theory suggests, judges may have an incentive to impose different punishments even for identical crimes.

I develop a theoretical model, in which judges receive an imperfect signal on the defendant's probability of guilt and only then decide whether to convict and which penalty to impose. Following the law and economics literature (see e.g. Posner, 1993; Knight and Epstein, 2017; Feess and Sarel, 2018), I assume that judges minimize their private costs, which take the form of aversions to wrongful convictions, wrongful acquittals, and extreme punishments. The model yields several insights. First, if judges are strongly averse to punishing innocents, then wrongfully convicted defendants will face lower expected punishments. Ceteris paribus, this increases deterrence. Second, however, the (de-facto) standard of proof ("SOP") is also determined by the judge's aversions, which may either reinforce or countervail the overall impact on deterrence. Namely, if judges are strongly averse to wrongful convictions, then the SOP and the penalty size are strategic *complements*, so that a lower penalty increases the willingness to convict. Conversely, if judges mainly care about punishing the guilty, the SOP and penalty are strategic *substitutes*. Then, those who seem guilty are more likely to be convicted and receive higher punishments. As the signal is imperfect, some innocents still generate a signal of guilt and suffer twice: they are (wrongfully) convicted because they seem guilty and receive higher punishments for the same reason.[3]

The model is followed by a two-part empirical analysis. Part I tests whether wrongful convictions are correlated with lower punishments. As wrongful convictions are not directly observable, I use exonerations as a proxy, which seems appropriate for estimating differences in punishment (see Gross and Shaffer, 2012; Gross et al., 2014; Olney and Bonn, 2014; Bjerk and Helland, 2017). I find that exonerated defendants receive lower punishments at the time of their conviction compared to other convicted defendants. This result is robust also when controlling for possible non-random treatment assignment, using a matching estimator and an endogenous treatment model.

Part II then tests whether punishments increase in the level of evidence. I construct an index from indicators of the defendant's apparent guilt, such as false confessions or misleading forensic evidence. I find that, conditional on being wrongfully convicted, the sentence is positively correlated with the index, i.e. with a higher level of evidence.

The paper's contribution is three-fold. First, to my knowledge, this paper is the first to empirically test whether wrongfully convicted defendants receive different punishments. My finding that wrongfully convicted defendants receive lower punishments therefore seems novel. Second, my finding that punishments increase in the level of evidence provides empirical support for the compromise verdicts theory. Third, I provide a new argument for why type I and II errors may affect deterrence asymmetrically.

The rest of the paper is organized as follows: Section I reviews related literature. Section II presents the model. Section III describes the data. Section IV presents the empirical analysis. Section V concludes.

I. Related literature

The paper is closely related to the literature on the relationship between the SOP and severity of the sanction. The most closely related paper is Lundberg (2016), which demonstrates how compromise verdicts lead self-interested judges to increase the sanction when the probability of guilt is higher (see also Teichman, 2017; Garoupa, 2017) and analyzes possible effects on deterrence and welfare. Lando (2005) shows that imposing higher sanctions on defendants whose probability of guilt is higher can increase deterrence. Schauer and Zeckhauser (1996) argue that the SOP should increase with the sanction size. My model entails these insights as special cases, but my extensions (in Appendix B) also show how compromise verdicts may lead to higher punishments for the innocents and thus to less deterrence.

A series of papers by Kaplow (2011a,b, 2012) argues that the SOP should depend on the costs of wrongful convictions and wrongful acquittals, in order to avoid a chilling effect of benign behavior. In my model, these costs also affect the de-facto SOP adopted by judges, but only insofar that they overlap with judges' private aversions. In the model by Andreoni (1991), the SOP is similarly determined by aversions but the adjudicator has no discretion over the sanction size, which is exogenously set by a regulator. The adjudicator then responds to (exogenous) changes in the penalty size by reducing the probability of conviction (i.e. the penalty and the SOP are complements) which can then reduce deterrence. Such an effect is also supported by experimental (e.g. Feess et al., 2018; Baumann and Friehe, 2017) and empirical (Bindler and Hjalmarsson, 2018) evidence. In my model, the SOP and the penalty may be either strategic complements or substitutes, depending on the judge's aversions.

The paper is also related to the literature that considers how judicial errors affect crime deterrence. While there is some debate on whether wrongful convictions decrease deterrence at all,[4] experimental evidence (e.g. Rizzolli and Stanca, 2012) as well as recent empirical evidence (Sarel, 2018) support the claim that wrongful convictions have an adverse effect on deterrence. However, scholars also diverge on whether wrongful convictions and wrongful acquittals affect deterrence *symmetrically*. Most papers assume that the two types of error do affect deterrence symmetrically, as both errors reflect the same kind of marginal changes in probabilities, which then yield symmetric punishments. Conversely, Nicita and Rizzolli (2014) and Rizzolli and Stanca (2012) argue that type I errors have a stronger effect than type II errors, due to risk or loss aversion, e.g. because defendants perceive wrongful convictions (acquittals) as a loss (gain). In contrast, Lando and Mungan (2017) assert that type II errors will generally have a stronger effect than type I errors, e.g. because choosing innocence reduces the probability of adjudication. Hynes (2018) shows that asymmetry can arise if the innocent and the guilty have different risk preferences. My model contributes to this debate by providing a new argument for why type I and II errors may have an asymmetric effect on deterrence: Penalties themselves may differ, even without different weighing of probabilities.

On the empirical side, my findings are closely related to Lundberg (2017), which analyzes whether convictions and sentences in the U.S. federal courts are driven by compromise verdicts. He finds that when judges enjoy larger discretion (depending on time-varying limitations set by case law), they tend to convict more - with some weaker evidence that judges also impose lower sentences. This finding seems overall consistent with (and complementary to) my finding that defendants who are wrongfully convicted receive lower punishments. My empirical analysis is further related to the many papers that analyze the rate of wrongful convictions (e.g. Spencer, 2007; Kim et al., 2013; Gould et al., 2014; Loeffler, Hyatt, and Ridgeway, 2019) and in particular papers which focus on exonerations (Gross and O'Brien, 2008; Gross and Shaffer, 2012; Gross et al., 2014; Olney and Bonn, 2014; Gould and Leo, 2015; Bjerk and Helland, 2017). Finally, the paper also relates to other empirical studies on sentencing in the U.S. federal courts (see e.g. Starr, 2014; Kaiser and Spohn, 2018; Ulmer, 2019) that do not directly consider wrongful convictions.[5]

II. The model

II.1. Preamble: deterrence models with judicial errors

Before presenting the specifics of my model, it is useful to generalize the conceptual framework that is used in deterrence models with judicial errors. Consider a Becker-style model, where potential offenders must make a dichotomous choice: commit the crime or remain innocent. Potential offenders are heterogeneous in their benefit from crime, denoted by b, which is randomly drawn from a uniform distribution with the support $(0, 1)$. The probability of apprehension and conviction p_t is conditional on the individual's choice, which is determined by their "type" (guilty or innocent) $t \in \{G, I\}$.[6] A conviction leads to a sanction f_t, which is again conditional on the individual's type.

The expected payoff from crime is $b - p_G * f_G$ while the expected payoff from innocence is $-p_I * f_I$. The expected sanction for each type is thus $E(f_t) = p_t * f_t$. A potential offender commits the crime if:

$$b \geq \widetilde{b} \equiv \underbrace{p_G * f_G}_{E(f_G)} - \underbrace{p_I * f_I}_{E(f_I)} \tag{1}$$

Note that $\widetilde{b}$ reflects the level of deterrence. Given the uniform distribution of b, the probability that a defendant j is innocent is $Pr(b_j \leq \widetilde{b}) = \widetilde{b}$. Respectively, the probability that the defendant is guilty is $1 - \widetilde{b}$. Differentiating by the probabilities of correct and wrongful convictions shows how deterrence is affected by each component:

$$\frac{\partial \widetilde{b}}{\partial p_t} = \begin{cases} f_G > 0 & \text{if } t = G \\ -f_I < 0 & \text{if } t = I \end{cases} \tag{2}$$

Thus, deterrence increases (decreases) in the probability of a correct (wrongful) conviction.[7] The effect size, however, depends on the penalty size (i.e. whether $f_G > f_I$). Only the special case of a fixed sanction, where $f_G = f_I = f$, breeds the symmetrical argument, where $\widetilde{b} \equiv (p_G - p_I)f$ and thus $|\frac{\partial \widetilde{b}}{\partial p_G}| = |\frac{\partial \widetilde{b}}{\partial p_I}| = f$.

In other words, deterrence decreases in the probability of a type I and II error equally only in the special case of $f_I = f_G$. However, this captures only to the *partial effect* of the probability of each error, which may differ from the *general effect* of the expected sanction. In the remaining parts of the paper, I show that the general effect is not always straightforward, as it depends on the relationship between the penalty size and the probability of conviction.

II.2. The basic model

My model builds on the framework above but slightly changes the setup: instead of the dichotomous penalties (f_G and f_I), potential offenders must calculate their expected penalty using distribution functions. Namely, when committing the crime, each offender must consider the probability that a certain amount of incriminating evidence would accumulate in his case, where each level of evidence may lead to both a different probability of being punished and - importantly - to a different penalty. To show how these are determined, I use backward induction and solve for the judges' decisions first.

Suppose that after individuals decide whether to commit the crime, a judge is randomly assigned a defendant who may have committed the crime. The judge does not know the defendant's type, but does observe the level of (incriminating) evidence e which serves as an (imperfect) signal. Specifically, I assume that e is drawn from a distribution with density $g_t(e)$ that is conditional on the type t. The two distributions $g_G(e)$ and $g_I(e)$ have a monotone likelihood ratio property (MLRP) in e, so that $\frac{\partial}{\partial e}(\frac{g_G(e)}{g_I(e)}) > 0$. Additionally, I assume that there exists a critical level of evidence $\widetilde{e} \geq 0$, such that

$$\frac{g_G(e)}{g_I(e)} \in \begin{cases} < 1 & \text{if } e < \widetilde{e} \\ = 1 & \text{if } e = \widetilde{e} \\ > 1 & \text{if } e > \widetilde{e} \end{cases} \tag{3}$$

Note that e is continuous, but includes three ranges: low signals ($e < \widetilde{e}$) that indicate innocence, a medium signal ($e = \widetilde{e}$) that does not provide useful information,[8] and a high signal ($e > \widetilde{e}$) that indicates guilt.

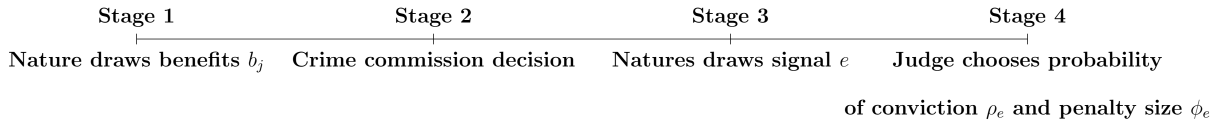
The posterior probability of guilt, after observing e, is denoted by[9]

$$\omega_e \equiv Pr(t = G|e) = \frac{g_G(e)(1-\widetilde{b})}{g_G(e)(1-\widetilde{b}) + g_I(e)\widetilde{b}}, \tag{4}$$

where the MLRP ensures that $\frac{\partial \omega_e}{\partial e} > 0$.[10] In other words, as the level of evidence increases, the (posterior) probability that the defendant is guilty increases as well.

After observing the signal, the judge can influence what happens to the defendant through two strategic variables: the probability of convicting $\rho_e \in [0,1]$ and the penalty size $\phi_e \in [0, \infty)$. As the judge first observes the signal, these choices can be conditioned on its realization.

Figure 1: Timing of the model



The timing of the model, depicted in Figure 1, is thus as follows: in stage 1, nature draws the benefits b. In stage 2, individuals decide whether to commit a crime (as determined by whether the benefit exceeds the threshold $\widetilde{b}$). In stage 3, the judge is randomly assigned a defendant and nature draws the signal e, which is observed by the judge. In stage 4, the judge chooses the conviction probability ρ_e and the penalty ϕ_e.[11]

Note that given the timing, the judge must actually treat deterrence as exogenous - as the defendants have already made up their mind prior to being judged and there is no repeated action. This means that, in the basic model, I assume that the judge does not have explicit deterrence concern (this assumption is later relaxed in Appendix B). Of course, even under this assumption, defendants will still care about the expected sanction and can anticipate the judge's decision when deciding whether to commit a crime, so that the expected sanction does matter for deterrence.

II.3. Adjudication

The judge's goal is to minimize the costs of her decisions, where both a conviction and an acquittal may be costly. The expected cost of conviction is

$$EC_e^{Convict} \equiv \omega_e c^G + (1 - \omega_e) c^I,$$

where c^t is the cost of convicting a defendant of type t. Note that these costs are both a function of the penalty ϕ_e, but for brevity, I avoid using the explicit notation of $c^t(\phi_e)$.

Convictions may be costly for several reasons, including pure-morality concerns, sensitivity to society's monetary cost of enforcing the sanction, sensitivity to the defendant's disutility from the imposed sanction, and the fear of reversal by an appellate court. These costs, however, depend on the penalty size ϕ_e. I adopt two key assumptions: First, I assume that there exists a $\hat{\phi}^G$ that minimizes c^G. From the judge's perspective, this implies that minimizing the cost of a correct conviction c^G has an interior solution, e.g. because increasing the sanction is "just" up to a certain point but becomes unjust (or too costly) when the optimal threshold $\hat{\phi}^G$ is crossed.

This is expressed by the following assumption:

Assumption 1 *The cost of a correct conviction has a minimum at the point $\hat{\phi}^G$, such that:*

$$\frac{\partial c^G}{\partial \phi_e} \in \begin{cases} < 0 & \textit{if } \phi_e < \hat{\phi}^G \\ = 0 & \textit{if } \phi_e = \hat{\phi}^G \\ > 0 & \textit{if } \phi_e > \hat{\phi}^G \end{cases} \tag{5}$$

A similar assumption is adopted in Lundberg (2016) and is necessary in order to exclude a corner solution in which the judge never punishes. Second, I assume that:

Assumption 2 *The cost of a wrongful conviction increases in the punishment, i.e.:*

$$\frac{\partial c^I}{\partial \phi_e} > 0 \tag{6}$$

so that punishing the innocent is always costly.[12] In addition, I assume that these two costs are twice-differentiable in ϕ_e and convex (i.e. $\frac{\partial^2 c^t}{\partial \phi_e^2} > 0, \forall t \in \{G, I\}$), e.g. because the judge's marginal disutility increases more and more as she imposes more severe punishments.[13]

The expected cost of an acquittal is

$$EC_e^{Acquit} \equiv \omega_e a^G,$$

where a^G is the cost of acquitting a guilty defendant (for simplicity, acquitting an innocent defendant is assumed to be costless). Note that a^G does not depend on the punishment size. The judge minimizes her cost function $J(\rho_e, \phi_e)$ by solving:

$$\text{(7)} \qquad \min_{\rho_e,\phi_e} J(\rho_e, \phi_e) \equiv \qquad \rho_e EC_e^{Convict}(\omega_e, \phi_e) + (1-\rho_e) EC_e^{Acquit}(\omega_e)$$

$$\text{(8)} \qquad = \qquad \rho_e(\omega_e c^G + (1-\omega_e)c^I) + (1-\rho_e)\omega_e a^G$$

so that with probability ρ_e the judge convicts and incurs the expected cost of conviction, and with probability $1-\rho_e$ the judge acquits and incurs the expected cost of acquittal.[14]

II.3.1. Sentencing.

Taking the partial derivative w.r.t. the penalty yields the following first order condition:

$$\text{(9)} \qquad \frac{\partial J(\rho_e, \phi_e)}{\partial \phi_e} = \rho_e \left(\omega_e \frac{\partial c^G}{\partial \phi_e} + (1-\omega_e) * \frac{\partial c^I}{\partial \phi_e} \right) = 0$$

The expression in parentheses in equation (9) reflects the changes in the expected costs of a correct and a wrongful conviction due to a marginal increase in punishment. Note that (an interior solution for) the cost-minimizing penalty in this condition is independent of ρ_e, but is dependent on the probability of guilt ω_e. Proposition 1 summarizes how the cost-minimizing punishment behaves in the level of evidence (all proofs are in Appendix A).

Proposition 1 *The cost-minimizing punishment (ϕ_e^*) increases in the level of incriminating evidence (e) if and only if marginally increasing the punishment increases the cost of a wrongful conviction more than the cost of a correct conviction. i.e. if* $\frac{\partial \phi_e^*}{\partial e} > 0 \Leftrightarrow \frac{\partial c^I}{\partial \phi_e} > \frac{\partial c^G}{\partial \phi_e}$.

The intuition behind proposition 1 is as follows: as the judge does not know whether the defendant is innocent, she must consider both potential costs when increasing the penalty. Then, if the marginal cost of a wrongful conviction is higher, the judge reduces the sentence whenever the defendant seems less guilty. Yet, if Assumptions 1 and 2 hold, the condition in

Proposition 1 is always fulfilled. Stronger evidence then always lead to enhanced penalties, yielding the following corollary:[15]

Corollary 1 *Given assumptions 1 and 2, the cost-minimizing punishment always increases in the level of evidence and therefore increases in the probability of the defendant's guilt.*

The Corollary reflects compromise verdicts: if the judge believes that the defendant is more likely to be guilty, she feels more comfortable with imposing a higher punishment.

II.3.2. Conviction decision.

Taking the partial derivative of the cost function w.r.t. the probability of conviction yields

$$\frac{\partial J(\rho_e, \phi_e)}{\partial \rho_e} = EC_e^{Convict} - EC_e^{Acquit} \tag{10}$$

$$= \omega_e c^G + (1 - \omega_e) c^I - \omega_e a^G \tag{11}$$

Recalling that the cost-minimizing penalty ϕ_e^* is independent of ρ_e, it follows that the expression in equation (11) has only a corner solution in which the judge chooses either the highest or the lowest possible ρ_e, i.e. either 0 or 1. In other words, for a given level of evidence (and a given penalty size), the judge simply chooses between (always) convicting and (always) acquitting. Note that this is the case only in the basic model (in the extensions in Appendix B, I later show that the judge may also use a mixed strategy).

The judge will convict if and only if

$$\omega_e > \widetilde{\omega}_e \equiv \frac{c^I}{c^I - c^G + a^G}, \tag{12}$$

where the conviction threshold $\widetilde{\omega}_e$ reflects the de-facto SOP used by the judge. However, the SOP indirectly depends on the level of evidence through c^I and c^G. Thus, different signals also yield different standards of proof. This occurs because the judge balances the cost of higher punishments through a lower willingness to convict.

II.3.3. Possible outcomes.

As the level of evidence is a stochastic signal, there are several possible cases to consider:

(1) **Case A**: All defendants are acquitted (i.e. $\rho_e = 0, \forall e$). This corner solution arises if $\widetilde{\omega}_e > 1, \forall e$, which can occur if the costs of (correct or wrongful) convictions are extremely high.

(2) **Case B:** Exactly one signal realization yields a conviction, whereas the other realizations yield an acquittal. For instance, this occurs if the judge is only willing to convict if $e = 1$ so that she is 100% sure that the defendant is guilty.

(3) **Case C:** More than one signal realization yields a conviction. This encompasses both a corner solution where all defendants are convicted (i.e. $\rho_e = 1, \forall e$) and an interior solution where some defendants are acquitted and others are convicted.

In Case A there is no punishment, such that a discussion of punishing the innocent is moot. In Case B there can be no asymmetric punishment, as all convicted defendants share the same signal. Case C is therefore more interesting and may imply different punishments for different signals. However, as the full expected sanction includes also the probability of being punished (which depends on the SOP), it is important to first check whether punishments and convictions are substitutes or complements. In other words, one must check whether or not a judge who gives higher punishments will decrease the SOP, as captured by the following proposition:

Proposition 2 *If* $\frac{\partial c^I}{\partial \phi_e}(a^G - c^G) + \frac{\partial c^G}{\partial \phi_e} * c^I < 0$, *then the cost-minimizing SOP* $\widetilde{\omega}_e$ *decreases in both (i) the punishment size* ϕ_e *and (ii) the level of evidence* e.

Proposition 2 provides a condition that determines whether the SOP and the punishment are strategic complements or substitutes. This condition shows that the penalty size has two countervailing effects on the SOP. The first effect ("WC effect"), expressed by $\frac{\partial c^I}{\partial \phi_e}(a^G - c^G) > 0$, is the judge's fear of wrongful convictions, which implies that the judge will be less willing to convict.[16] The second effect ("CC effect"), expressed by $\frac{\partial c^G}{\partial \phi_e} c^I < 0$, is the judge's marginal reduction in the cost of correct convictions, which increases the willingness to convict.[17] It is important to note that the name I assign to the two effects relates to the judge's *concerns* regarding the type of conviction, i.e. whether the judge is more averse to wrongful conviction (WC effect) or correct convictions (CC effect). Part (i) of the proposition shows that if the condition holds, then ϕ_e and $\widetilde{\omega}_e$ are substitutes. This occurs when the CC effect dominates, so that the judge cares more about convicting the guilty. As a result, the judge actually benefits more from convicting when she also imposes a harsher sentence. Part (ii) then suggests that if the condition holds, the willingness to convict also increases in the level of evidence, for a similar reason.[18] Corollary 2 summarizes the impact of the WC and CC effects:

Corollary 2

(i) If the CC effect dominates, defendants who seem more guilty are more likely to be convicted and subsequently also receive higher punishments (i.e. $\frac{\partial \rho_e^*}{\partial e} > 0$ *and* $\frac{\partial \phi_e^*}{\partial e} > 0$*).*

(ii) If the WC effect dominates, defendants who seem more guilty may be either more or less likely to be convicted, but receive higher punishments if they are convicted (i.e. $\frac{\partial \rho_e^*}{\partial e} \gtreqless 0$ *and* $\frac{\partial \phi_e^*}{\partial e} > 0$*).*

The corollary demonstrates the importance of considering the full expected sanction on deterrence rather than looking only at the partial effect. Namely, part (i) of the corollary shows that when the CC effect dominates, both the probability of conviction and the accompanying penalties are lower for innocents. This suggests that the two components of the expected sanction reinforce one another. On the contrary, part (ii) shows that if the WC effect dominates, looking only at the partial effect would be misleading. The reason is that while innocents still receive lower punishments, their probability of conviction might be even higher, as their lighter sentence implies that convicting them is less detrimental.

Summing up, analyzing the judge's decision reveals that convicted innocents generally receive lower sentences, but that the full expected sanction depends on whether the judge cares more about wrongful or correct convictions.

II.3.4. Deterrence.

For the main result of the basic model, as well as for the empirical analysis that follows (which restrices attention to penalties), it is not crucial to explictly write down the deterrence equation. However, for the sake of providing a complete picture, I briefly summarize the effects on deterrence below.

Recall from the preamble, that individuals generally commit the crime if and only if $b > \widetilde{b} = E(f_G) - E(f_I)$. In terms of my model, the expected sanction for type t can be rewritten as:

$$E(f_t) = \int_0^1 \rho_e^* \phi_e^* g_t(e) de \tag{13}$$

where ρ_e^* and ϕ_e^* are the probability of conviction and penalty size chosen by the judge in equilibrium for each level of evidence e. Deterrence is then

$$\widetilde{b} = E(f_G) - E(f_I) = \int_0^1 \rho_e^* \phi_e^* (g_G(e) - g_I(e)) de \tag{14}$$

Note that isolating the partial effect (as done in traditional deterrence models) is not useful here - instead what matters is the overall expected sanction. Depending on which case (A,B,C) applies and which effect (WC effect or CC effect) dominates, different results can emerge. If no one is convicted (case A), the penalty is never imposed and does not matter. If both types face the exact same expected penalty (case B, where only one signal yields a conviction), my model would converge to the traditional models that assume symmetrical punishments. However, if several levels of evidence yield a conviction (case C), the bottom line is more complex. If the CC effect dominates, guilty defendants clearly face a higher expected sanction - as they are more likely to be assigned a higher level of incriminating evidence and subsequently both their probability of conviction and sanction size are higher (recall part (ii) of Corollary 2). Conversely, if the WC effect dominates, the penalty size for guilty defendants is still higher but this might be (partially or fully) countervailed by a lower probability of conviction. At the extreme, it might even occur that $\widetilde{b} < 0$ so that everyone commits the crime. In more moderate cases, the lower penalty size for innocents, which positively contributes to deterrence, might be enough to ensure that not all individuals commit the crime.

In any case, the *general* effect of judicial errors - which has an impact on both the penalty size and the probability of conviction - either amplifies the deterrence problem or mitigates it, depending on whether the judge is more concerned with wrongful convictions or with correct convictions.

II.4. Higher punishments for the wrongfully convicted?

Under the assumptions above, wrongfully convicted defendants receive lower punishments on expectation. The intuition for this outcome is simple: guilty defendants are more likely to generate a signal of higher guilt and therefore when judges observe such a signal, they assume that the defendant can be more severely punished. Nonetheless, wrongfully convicted defendants may end up receiving higher punishments in several different ways, which are either systematic or non-systematic.

Starting from non-systematic ways, in the (relatively rarer) cases where an innocent defendant generates a high signal, the punishment would be higher. For example, defendants who just happen to be at the "wrong place at the wrong time" and therefore by chance look extremely guilty would receive a higher punishment.[19] Alternatively, inculpatory evidence might accumulate for some innocent defendants in a non-random way. For example, defendants who are framed by others or receive ineffective legal counsel at trial may appear to be very guilty due to actions of third parties. Such defendants would be more likely to appear to be guilty and will thus receive higher punishments. Note that non-systematic ways are still in line with the basic model, as the *expected* punishment for the wrongfully convicted remains lower.

In Appendix B (Sections B.1 and B.2) I relax some assumptions in the model and show how wrongfully convicted can also, at least in theory, receive higher punishments as a systematic (general) result. This occurs either if the judge gains some utility from punishing the innocent or if the timeline is reversed, so that the judge can influence her likelihood of facing a guilty defendant. The judge is then affected by a "pool effect", which gives an incentive to strategically manipulate deterrence by purposefully punishing the innocent. This then leads to a lower probability of facing innocent defendants, which implies less wrongful convictions. In an additional extension (Section B.3), I consider a stylized variant that allows for plea bargains and explains why the main mechanism identified in the model remains robust even with plea bargains. Namely, I show that plea bargains cause a portion of defendants to drop out of the pool, but the judge's aversions remain the same. Note, however, that an empirical analysis must take into account the fact that innocents are possibly less likely to accept a plea bargain (as argued in the literature). If innocents rarely take pleas, then the average penalty among all convicted defendants (including those with a plea) may be higher among the innocent. However, after excluding (or controlling for) pleas, the model still predicts lower sentences for innocents who do not take a plea bargain.

III. Data

Data on exonerated defendants are derived from the "National Registry of Exonerations" (NRE) database, which provides an overview of U.S. cases in both the state and the federal courts, in which the defendant was factually proven to be innocent.[20] Part I of my empirical analysis uses 107 *federal* cases from the years 1983-2017, which serve as the treatment group.[21] As a control group, I use case-level data on sentenced defendants in the federal courts from the series "Monitoring of Federal Criminal Sentences", as disseminated by the Inter-university Consortium for Political and Social Research (ICPSR).[22] Note that the control group and the treatment group are both taken from the *same* courts - the federal courts. Combining the 107 (federal) exonerations with the sentencing database yields a total of 1,695,539 observations (for further details see Table C1 in Appendix C).[23] Data regarding judicial attributes is derived additionally from the Federal Justice Center's "Biographical Directory of Article III Federal Judges".

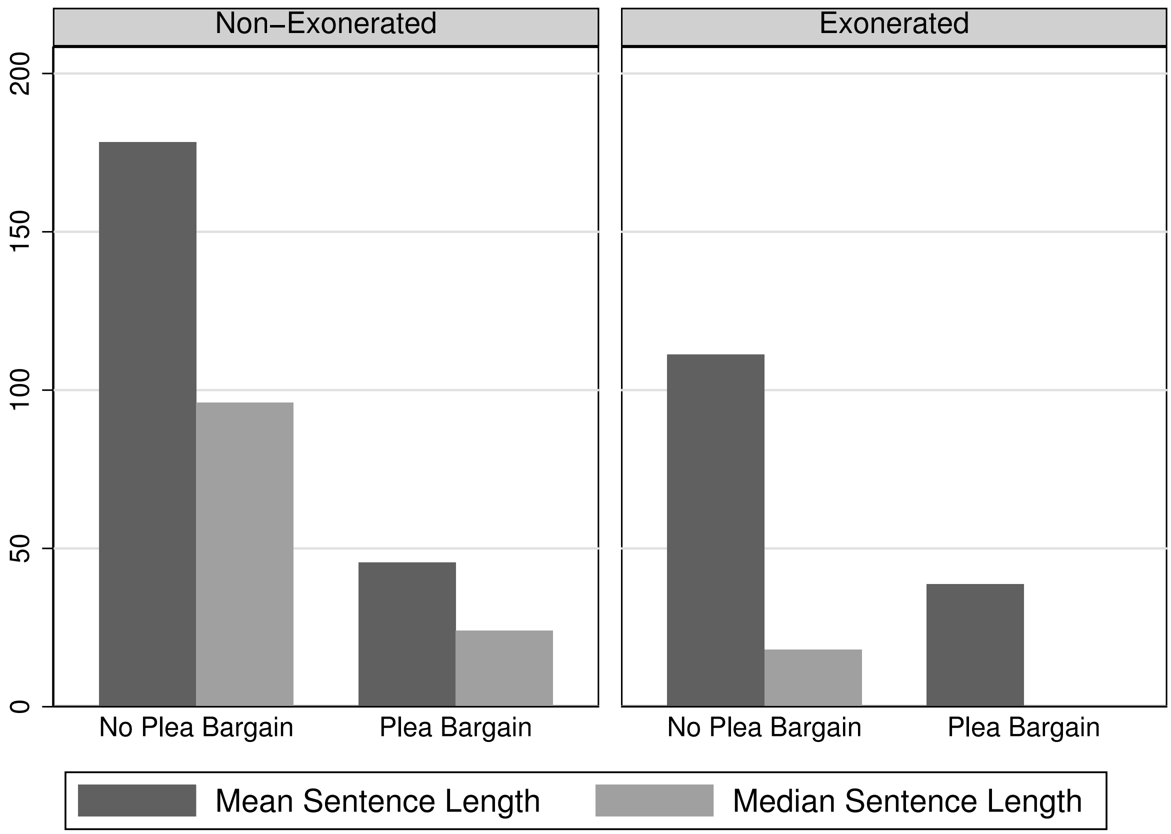
Part II of the analysis uses instead only exonerations, but includes the cases from both the federal and the state courts from the NRE database (1983-2017), summing up to 2,468 observations.

III.1. Variables and descriptive statistics

III.1.1. Main variables

The dependent variable is *totsent* - the length (in months) of the defendant's prison sentence.[24] The independent variable is *Exonerated* - a dummy variable assigning 1 if the defendant was (wrongfully convicted and) exonerated and 0 otherwise. Figure 2 visually contrasts *totsent* (mean and median) between exonerated (on the right) and non-exonerated (on the left) defendants. Within each group, data is separated by whether the defendant took a plea bargain (a pooled version is also available in Figure C2 in Appendix C). The figure yields three insights. First, comparing only those who *do not* take a plea between groups reveals that exonerees receive lower sentences. Second, comparing those who *do* take a plea between groups shows that the means are similar, but the median sentence of exonerated is zero (i.e. no prison time) and hence much lower.[25] Third, sentences of exonerees with no plea bargain are higher than the sentences of non-exonerated who do take a plea. This, taken in conjunction with the fact that exonerees are less likely to take a plea,[26] yields a first interesting result: the population of exonerees as a whole actually receives *higher* punishments on average (see also Figure C2), but this is entirely driven by the fact that these defendants are less likely to take a plea. In contrast, once attention is restricted to

Figure 2: Comparison of sentence



defendants who are sentenced in a full trial (=no plea), the sentence of exonerees turns out to be lower, as predicted by the theoretical model.

Figure C1 in Appendix C also presents a histogram of the sentence lengths, separated again by whether the defendant was exonerated and whether a plea bargain was taken. The figure similarly shows that exonerees generally receive lower sentences (conditional on plea bargains), where for those who do not take a plea - one can identify a few outliers with higher sentences. This is again consistent with the basic model and the possibility of (non-systematic) cases in which a few innocents receive higher punishments.

III.1.2. Control Variables

Several groups of control variables are considered in the analysis that follows below. First, a dummy variable capturing whether the defendant took a plea bargain, for reasons discussed above.[27] Second, defendants' demographics (age, age-squared, gender, and race) are included, as these may be correlated with the judge's private costs. Controlling for the defendant's gender and race seems particularly important, as the literature finds that minorities have a larger probability of being wrongfully convicted (e.g. Olney and Bonn, 2014; Bjerk and Helland, 2017) and receive harsher sentences (see e.g. Mustard, 2001). Third, offense-type dummies are included.[28] This ensures that observed differences are *not* due to variation in the type of crime committed. Fourth, court and year dummies are added, to capture unobservable time (court) invariant effects. Fifth, I include court-by-year averages of judicial demographics (age, age-squared, gender, and race), which serve as a proxy for the identity of the judge (for a similar approach, see Lundberg, 2017) and capture additional unobservables that do vary by time (court).

Table 1 compares descriptive statistics between exonerated and non-exonerated defendants. Exonerees differ significantly in their attributes, offense-type, tendency to take a plea bargain, and judicial demographics (except for judge's gender). This seems consistent with the model, as judicial attributes presumably capture differences in the judges' private cost functions. Tables C2 and C3 in Appendix C provide further comparisons of descriptive statistics, depending on the offense type or category.

Table 1:
DESCRIPTIVE STATISTICS

Factor	Level	Non-Exonerated	Exonerated	p-value	Test
Observations (n)		1695432	107		
Total Sentence (totsent), mean (SD)		52.6 (110.1)	100.4 (237.1)	0.0391	Two-sided unequal variance t-test
Plea Bargain		1603199 (95.0%)	16 (15.0%)	<0.001	Pearson's χ^2
DEFENDANT'S ATTRIBUTES					
Gender	Male	823522 (48.6%)	94 (87.9%)	<0.001	Pearson's χ^2
	Female	136854 (8.1%)	13 (12.1%)		
	Unknown	735056 (43.4%)	0 (0.0%)		
Race	White	498389 (29.4%)	61 (57.0%)	<0.001	Pearson's χ^2
	Black	383675 (22.6%)	21 (19.6%)		
	Hispanic	640270 (37.8%)	15 (14.0%)		
	Native	25893 (1.5%)	1 (0.9%)		
	Asian	34054 (2.0%)	5 (4.7%)		
	Other/Unknown	113151 (6.7%)	4 (3.7%)		
Age, mean (SD)		35.0 (10.9)	42.3 (11.1)	<0.001	Two-sided unequal variance t-test
Offense Type (categories)	Violent	168080 (9.9%)	18 (16.8%)	<0.001	Pearson's χ^2
	Property	59384 (3.5%)	6 (5.6%)		
	Sex	8156 (0.5%)	2 (1.9%)		
	Administration of justice	25584 (1.5%)	3 (2.8%)		
	Immigration	380044 (22.4%)	4 (3.7%)		
	White Collar	242754 (14.3%)	42 (39.3%)		
	Drugs	553625 (32.7%)	28 (26.2%)		
	Other	257805 (15.2%)	4 (3.7%)		
JUDGES' ATTRIBUTES (COURT-BY-YEAR MEAN)					
Gender: Female, mean (SD)		16.7 (11.2)	16.2 (11.6)	0.65	Two-sided unequal variance t-test
Race: White, mean (SD)		.812 (.164)	.866 (.111)	<0.001	Two-sided unequal variance t-test
Race: Black, mean (SD)		7.3 (7.2)	8.6 (7.9)	0.09	Two-sided unequal variance t-test
Race: Hispanic, mean (SD)		10.6 (16.7)	3.5 (6.6)	<0.001	Two-sided unequal variance t-test
Race: Native, mean (SD)		0.1 (1.3)	0.0 (0.0)	< 0.001	Two-sided unequal variance t-test
Race: Asian, mean (SD)		0.8 (2.7)	1.3 (2.8)	0.07	Two-sided unequal variance t-test
Race: Pacific Islander, mean (SD)		0.1 (1.2)	0.0 (0.0)	< 0.001	Two-sided unequal variance t-test
Age, mean (SD)		6440.7 (375.7)	6501.9 (372.9)	0.09	Two-sided unequal variance t-test

NOTE.– This table compares descriptive statistics for the two groups: Exonerated (registry of exonerations database) and Non-Exonerated (ICPSR database). For each variable, the test is specified alongside a p-value. For continuous variables, a two-sided unequal variance t-test is used. Offense categories are coded as follows: (1) Violent - Arson, Assault, Firearms, Kidnapping/Hostage Taking, Manslaughter, Murder; (2) Property: Auto theft, Burglary/Breaking and Entering, Larceny, Robbery; (3) Sex - Sexual Abuse, Obscenity, Pornography/Prostitution; (4) Administration of Justice - Administration of justice and obstruction of justice; (5) Immigration - Immigration crimes; (6) White Collar - Fraud, Embezzlement, Forgery/Counterfeiting, Bribery, Tax Offenses, Money Laundering, Racketeering/Extortion, Gambling/Lottery; (7) Drugs - Trafficking, Manufacturing, and Importing; Communication Facilities; Simple Possession.

IV. EMPIRICAL RESULTS

IV.1. Part I: Exonerees vs. others

I run a set of OLS regressions testing the following model:

$$totsent = \beta_0 + \beta_1 Exonerated + \beta_x' X' + \epsilon,$$

where X' is a vector of control variables and ϵ is the error term. Table 2 presents the results. Column (1) excludes controls. Columns (2) adds a plea-bargain dummy. Column (3) adds attributes (gender, race, age, and age-squared). Column (4) adds offense-type dummies.[29]

Table 2:
OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exonerated	47.89** (22.81)	-58.01** (22.78)	-52.71** (22.84)	-53.49*** (20.11)	-51.37** (20.05)	-51.38** (20.08)	-60.35*** (23.38)
Plea Bargain		-132.68*** (0.88)	-126.07*** (0.88)	-116.78*** (0.84)	-115.29*** (0.84)	-115.28*** (0.84)	-115.29*** (0.84)
Exonerated × Plea Bargain							59.89** (26.62)
DEFENDANT'S ATTRIBUTES							
Race: Black			28.13*** (0.25)	23.34*** (0.24)	21.40*** (0.25)	21.40*** (0.25)	21.40*** (0.25)
Race: Hispanic			-2.30*** (0.20)	1.75*** (0.22)	2.74*** (0.23)	2.92*** (0.23)	2.92*** (0.23)
Race: Native			3.29*** (0.68)	-5.67*** (0.71)	-1.09 (0.76)	-1.03 (0.76)	-1.03 (0.76)
Race: Asian			-8.75*** (0.58)	-1.34** (0.56)	-2.19*** (0.57)	-2.23*** (0.57)	-2.23*** (0.57)
Race: Other/Unknown			-19.80*** (0.37)	0.54 (0.38)	1.64*** (0.39)	2.06*** (0.39)	2.06*** (0.39)
Gender: female			-29.85*** (0.28)	-18.35*** (0.28)	-18.18*** (0.28)	-18.19*** (0.28)	-18.19*** (0.28)
Gender: unknown			-3.04*** (0.17)	-2.25*** (0.16)	-2.15*** (0.16)	-3.27*** (0.27)	-3.27*** (0.27)
Defendant's age			1.96*** (0.04)	2.20*** (0.04)	2.06*** (0.04)	2.08*** (0.04)	2.08*** (0.04)
Defendant's age × Defendant's age			-0.03*** (0.00)	-0.03*** (0.00)	-0.02*** (0.00)	-0.02*** (0.00)	-0.02*** (0.00)
JUDGE'S ATTRIBUTES (COURT-BY-YEAR MEAN)							
Judge's Gender: Female			-0.35*** (0.01)	-0.29*** (0.01)	-0.12*** (0.01)	0.03 (0.02)	0.03 (0.02)
Judge's Race: African American			0.09*** (0.01)	0.11*** (0.01)	0.01 (0.03)	0.04 (0.03)	0.04 (0.03)
Judge's Race: Hispanic			-0.12*** (0.01)	-0.12*** (0.01)	-0.20*** (0.02)	-0.03 (0.02)	-0.03 (0.02)
Judge's Race: Asian American			-0.32*** (0.03)	-0.27*** (0.03)	-0.54*** (0.05)	-0.20*** (0.06)	-0.20*** (0.06)
Judge's Race: Native American			-0.36*** (0.06)	0.03 (0.06)	-0.21* (0.12)	0.07 (0.12)	0.07 (0.12)
Judge's Race: Pacific Islander			1.07*** (0.07)	0.48*** (0.07)	-1.11*** (0.20)	-0.86*** (0.20)	-0.87*** (0.20)
Judge's Age			0.27 (0.50)	4.49*** (0.48)	6.01*** (0.54)	5.01*** (0.54)	5.01*** (0.54)
Judge's Age × Judge's Age			-0.00 (0.00)	-0.04*** (0.00)	-0.05*** (0.00)	-0.04*** (0.00)	-0.04*** (0.00)
R-squared	< 0.01	0.07	0.10	0.18	0.19	0.19	0.19
Control: Offense type	No	No	Yes	Yes	Yes	Yes	Yes
Court dummies	No	No	No	No	Yes	Yes	Yes
Year dummies	No	No	No	No	No	Yes	Yes
Observations	1691696	1685378	1658887	1655965	1655965	1655965	1655965

NOTE– This table presents OLS regression result of sentencing length on being wrongfully convicted and exonerated, and additional controls. Robust standard errors are in parentheses. Coefficients of offense-type, court, and year dummies, and the constant are not reported. Reference categories for the defendant's attributes are a white race and male gender. The (continuous) rate of white judges is excluded and similarly reflects a reference category. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Column (5) adds court dummies. Column (6) adds year dummies.[30] Column (6) adds an interaction term of *Exonerated* with the plea-bargain dummy, which seems appropriate given the apparent differences in Figure 2 above. Robust (Huber-White Sandwich) standard errors are used in all regressions. Note that clustering the standard errors by district court yields qualitatively similar results (available upon request), but seems less appropriate, as the treatment (*Exonerated*) is administered at the individual level and the sample includes the entire population of convicted defendants (apart from those who appealed) and not a random sub-sample (for discussion of when clustering is appropriate, see Abadie et al., 2017).

The resulting coefficient in Column (1) demonstrates the danger of neglecting plea bargains: When only the *Exonerated* dummy is included, its coefficient is positively significant. However, once plea bargains are controlled for, the coefficient changes sign and becomes negatively significant at the 5% level or below in all specifications. Recall that this is precisely what the descriptives demonstrate: innocents are less likely to take pleas and thus receive higher (non-plea) penalties on average, but once one conducts a ceteris paribus comparison of defendants at trial, innocents receive lower sentences.

Next, the coefficient of the interaction term in Column (7) is also significant at the 5% level, with a positive sign and a similar size as the coefficient of the main term (of *Exonerated*).[31] The implication is as follows: defendants who do not take a plea (i.e. the interaction term equals zero) receive significantly shorter sentences (approximately 60.35 months less). However, defendants who do take a plea receive only mildly lower punishments, as the main effect and interaction term almost cancel each other out. Hence, the effect of *Exonerated* is (mostly) driven by defendants who do not take a plea. Again, this is fully consistent with the theoretical model.

Further exploring this point, Table 3 presents regression results of a split-sample, separated into those who did not (columns (1)-(2)) and those who did (column (3)-(4)) take a plea bargain. The coefficient of *Exonerated* is again negatively significant for those with no plea, but is insignificant for those with a plea. Note that the model's extension (section B.3 in Appendix B) predicts precisely this result.

Summing up, the results indicate that the wrongfully convicted receive shorter sentences when they do not take a plea bargain, but receive similar sentences when a plea is taken.

IV.1.1. Robustness tests

The basic results are subject to some limitations: First, defendants with a longer sentence may be more likely to be exonerated for various reasons, e.g. because only those with a long sentence pursue a costly exoneration (Gross, 2016, pp. 767); because false convictions are more likely to be detected when the sentence is severe (Gross, 2016, pp. 785); or because judges are more likely to exonerate those with a higher sentence. Second, defendants who appealed their sentence are excluded from the ICPSR database, such that the control group

Table 3:
OLS: Split-sample

	No Plea		Plea	
	(1)	(2)	(3)	(4)
Exonerated	-67.01** (26.43)	-49.05** (23.67)	-6.87 (20.62)	0.52 (10.72)
R-squared	< 0.01	0.20	< 0.01	0.14
Control: Defendant attributes	No	Yes	No	Yes
Control: Judge attributes (court-by-year mean)	No	Yes	No	Yes
Control: Offense type	No	Yes	No	Yes
Court dummies	No	Yes	No	Yes
Year dummies	No	Yes	No	Yes
Observations	84957	82759	1600421	1573206

Note.– This table presents OLS regression results of sentence length (*totsent*) on *Exonerated* and other variables. Columns (1) and (2) include only defendants who do not take a plea bargain, whereas columns (3) and (4) includes defendants who do take a plea bargain. The controls "Defendant attributes" and "Judge attributes" are the same as in Table 2, i.e. age, age-squared, and gender. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

is imperfect. It is important to note that any of these reasons would only lead to an *under-estimation* of the true effect, because they all imply that the exonerated group might be observed with *longer* sentences. As I nonetheless find that exonerees receiving lower punishments, such effects cannot drive my results (i.e. at most, the true effect is stronger and exonerees actually receive even lower sentences).

Furthermore, as the purpose of the analysis is to contrast the punishments of the treated and the untreated, it is sufficient to restrict attention to correlations (see, however, Part II of the analysis below, which does shed more light on the causal mechanism).[32] However, what may still matter is the potential concern of non-random treatment assignment - as the the treated and the untreated units must be comparable.

In order to account for this issue, I use two robustness tests. The first robustness test, described below, uses a *matching* estimator. The second robustness test, described in Appendix D, employs an *endogenous treatment model*. Both tests reinforce the basic results.

Matching estimator. The first robustness test matches exonerees with defendants in the control group. Matching techniques are a common solution when there is a concern that treatment assignment is non-random and, in particular, when the treatment group is small (see Rosenbaum and Rubin, 1983). The approach I take combines several matching methods, following the suggestion of King and Nielsen (2019) who propose to first match individuals exactly on important features and complement the remaining process with additional matching techniques.

The matching algorithm I use thus combines exact matching ("EM"), multivariate distance ("MD") kernel matching, regression adjustment ("RA") and post-matching entropy balancing ("EB"). In a nutshell, this process works as follows: First, a group of variables is selected for exact matching, i.e. each defendant in the treatment group (exonerees) is matched with (one or more) defendants from the control group who have the exact same value. As my control group is very large, it is possible to locate multiple defendants which are (simultaneously): the same race, same gender, convicted by the same court, and for the same offense. Such exact matches can be found for 104 out of the 107 exonerees. Second, the algorithm minimizes the distance between the ages, sentencing years, and judicial attributes relating to the matched defendants.[33] Third, the algorithm allows to add a post-matching bias-adjustment (a.k.a. "regression adjustment"; see Abadie and Imbens (2011)) using an auxiliary regression, with *totsent* as the dependent variable, the relevant covariates (without the treatment) as independent variables, and probability weights derived from the matching.[34] Finally, post-matching entropy balancing ("EB") (Hainmueller, 2012; Hainmueller and Xu, 2013) can be used to reduce remaining imbalances unaccounted for by the MD matching.[35] Standard errors are bootstrapped, as recommended by Jann (2017).

The goal of the matching process is the estimate the average treatment effect (of the treatment *Exonerated*) on the treated ("ATT"), i.e. to predict the sentence decrease for exonerees in comparison to a counter-factual.[36] The algorithm assumes 'conditional independence', i.e. that, conditional on the covariates, the treatment assignment is as good as random. Including the (mean) court-by-year attributes of judges for matching makes this assumption plausible: as judges are generally randomly assigned to cases (e.g. via lottery, see Samaha, 2009), whether two (otherwise identical) defendants are treated should depend only on the cost function of a randomly drawn judge.

Table 4:
MATCHING

	(1)	(2)	(3)
ATT	-48.49***	-50.97***	-51.84***
	(17.90)	(18.00)	(19.55)
Observations	1655965	1655965	1655965
Exact matching	Yes	Yes	Yes
MD matching	Yes	Yes	Yes
Regression Adjustment	No	Yes	Yes
Entropy Matching	No	No	Yes

NOTE.– This table presents the results of the matching algorithm, using Stata's kmatch command. Bootstrapped standard errors are in parentheses. MD matching (and Entropy matching in column 3) is made on the following variables: Sentencing year, Defendant's age, Judges' (court-by-year mean) gender, age, and race. Exact matching is made on the following variables: Plea Bargain dummy, Defendant's Gender, Defendant's Race, Offense type, Court dummies. In column 2, the regression adjustment equation includes "totsent" as the dependent variable, and the following independent variables: Plea-bargain, offense-type, court, and attributes (gender, race, age, age-squared) of the defendant and of judges. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 4 presents results. Column (1) includes only EM and MD matching. Column (2) adds RA. Column (3) adds EB. The first row presents the ATT and indicates that exonerees receive a sentence that is 48.5-51.8 months lower on average. The effect is significant at the 1% level and similar in size to the OLS. Other variants (not reported for brevity) yield similar results.[37] Figure C3 in Appendix C presents balancing and common support diagnostics for the covariates that are included in the distance (entropy) matching. Parts (a) and (c) plot post-matching standardized mean differences (SMDs) and variance ratios (VRs) for the specification with (EB) and without (No EB) entropy balancing.[38] The figures demonstrates that the matching works, so that covariates are well-balanced (as SMDs are small and VRs are close to 1).[39] Parts (b) and (d) compare the 104 matched exonerees with the full group of the 107 exonerees. SMDs are again small and VRs are close to 1, i.e. the exclusion of the 3 unmatched observations does not bias the results.

IV.2. Part II: Evidence Level Index

The finding that wrongfully convicted defendants receive lower punishments provides only implicit evidence on the mechanism behind it. In this part, I focus on this mechanism, i.e. on the relationship between the level of evidence and the sentence.

Although the NRE database includes some information about the reasons leading to the wrongful conviction, the variation of these indicators within the group of 107 exonerees in the federal courts is quite low. Thus, simply including these indicators as additional regressors seems moot. Instead, one can restrict attention to the entire population of exonerees, i.e. also those in *state* courts, and omit the control group (the ICPSR database). The goal is then to examine whether, conditional on being wrongfully convicted, defendants who appear to be less guilty also receive lower sentences.

I construct an index, which counts several indicators in the NRE database, each corresponding to a different aspect that plausibly increases the level of inculpatory evidence. These indicators can be divided into four categories: a (false) accusation or testimony by a third party (co-defendant, witness, or jailhouse informant), a false confession by the defendant, misleading evidence (including misidentification by a witness), and situations in which no crime has actually happened.[40] A higher index value suggests a higher level of inculpatory evidence. Definitions of the indicators are provided in Table C4 in Appendix C. Descriptive statistics can be found in Table C5 in Appendix C.

The regression model is then:

$$totsent = \beta_0 + \beta_1 Index + \beta_x X' + \varepsilon,$$

where $totsent$ is the sentence (as before), $Index$ is the constructed index, X' is a vector of (varying) control variables and ε is the error term. OLS regression results are presented in Table 5. Column (1) excludes controls. Column (2) adds a plea-bargain dummy and a court-type (federal/state) dummy. Column (3) adds defendants' attributes.[41] Column (4)

Table 5: OLS: Evidence Level Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Evidence Level Index	144.06*** (5.33)	108.67*** (5.99)	106.39*** (6.14)	40.56*** (6.34)	31.54*** (6.71)	36.53*** (6.17)	37.51*** (6.53)
Plea Bargain		-319.61*** (17.71)	-316.85*** (18.16)	-173.52*** (22.43)	-151.33*** (22.14)	-224.69*** (24.55)	-201.76*** (41.87)
Plea Bargain × Evidence Level Index							-8.64 (14.48)
Federal		-351.83*** (27.06)	-315.02*** (30.29)	-86.96** (35.52)	-66.09* (35.56)	204.75* (112.63)	
Race: Black			14.45 (19.18)	27.62 (17.52)	30.01 (18.25)	41.96** (17.22)	41.89** (17.22)
Race: Hispanic			-68.62** (26.87)	-59.72** (23.88)	-40.60* (24.47)	12.09 (23.97)	11.88 (23.90)
Race: Native			-95.79 (110.72)	-53.31 (91.76)	11.91 (100.55)	90.78 (83.45)	89.07 (83.51)
Race: Asian			44.43 (82.98)	60.35 (63.73)	79.32 (71.21)	118.07* (60.55)	116.53* (60.91)
Race: Other/Unknown			-147.47** (57.87)	-60.87 (65.89)	-20.28 (64.76)	37.92 (64.25)	37.50 (64.11)
Female			-57.25** (28.16)	-28.63 (26.10)	-21.22 (26.16)	-40.93 (26.60)	-40.72 (26.60)
Age			11.21** (4.48)	15.59*** (4.02)	15.67*** (4.05)	19.08*** (3.67)	19.05*** (3.66)
Age × Age			-0.18*** (0.06)	-0.22*** (0.06)	-0.22*** (0.06)	-0.26*** (0.05)	-0.26*** (0.05)
R-squared	0.20	0.28	0.29	0.47	0.52	0.65	0.65
Control: Offense type	No	No	No	Yes	Yes	Yes	Yes
Court dummies	No	No	No	No	No	Yes	Yes
Year dummies	No	No	No	No	Yes	Yes	Yes
Observations	2466	2466	2460	2460	2460	2460	2460

Note.– This table presents OLS regression result of the sentence length on the evidence level index and other covariates. Note that 6 observations drop out in column (3)-(7), due to missing information about the defendant's age. Also note that the federal dummy drops out in column (7) as it is replaced by court dummies. Robust standard errors are in parentheses. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

adds offense-type dummies. Column (5) adds dummies for year of conviction and year of crime commission. Column (6) adds court dummies. Column (7) includes an interaction of plea bargains and the index.

The coefficient of the index is positively significant at the 1% level in all specifications, i.e. those with higher index values receive higher sentences. The interaction in column (6) is insignificant, implying that judicial discretion (rather than plea deals) drive the impact of the index. These results suggest that the sentence increases (decreases) when the level of incriminating evidence is high (low), which is consistent with the theoretical model.

V. Conclusion

The first key insight of my analysis is that wrongfully convicted defendants may receive different punishments than correctly convicted defendants. The basic model predicts that punishments of the wrongfully convicted will be lower, as judges impose lighter sentences on defendants who seem less guilty. This result is a straightforward corollary to the general theory on compromise verdicts (e.g. Lundberg, 2016). However, I also show that the general effect on expected sanctions depends also on whether the penalties and the willingness to convict are substitutes or complements, which in turn is determined by the judge's private costs from correct and wrongful convictions.[42]

Part I of the empirical analysis suggests that exonerated defendants receive lower punishment, which supports the theoretical prediction. Part II then suggests that compromise verdicts indeed drive the sentence of the (wrongfully) convicted, where a higher level of evidence corresponds to a higher punishment.

A second key insight of the model (which I do not-test empirically here), is that the asymmetry in punishments feeds back into deterrence. A lower punishment for the wrongfully convicted, ceteris paribus, increases the payoff from innocence and thus increases deterrence. Hence, the mechanism which allows judges to reduce the sentence may be welfare-enhancing - as it directly contributes to deterrence. At the same time, the judicial discretion to reduce sentences also results in more wrongful convictions – as lower sentences may imply a higher willingness to convict. In that sense, my model is a generalization of the well-known result that higher mandatory sentences may induce judges to acquit as the only available alternative (Andreoni, 1991), but in my model, the ability to reduce sentences typically prevents the extreme result of acquitting. I also show how higher sentences may instead increase the willingness to convict, if the judge's marginal utility from correct convictions is large.

Irrespective of whether judicial discretion contributes to deterrence, the diagnosis is far less clear from a criminal justice perspective. If lower sentences go along with a higher willingness to convict, then wrongful convictions are likely to be more frequent. It then depends how one weighs the enhanced frequency of wrongful convictions versus the decreased disutility for the defendants due to lower punishments. In other words, if justice is more heavily damaged by wrongful convictions per se, then compromise verdicts are undesirable. If, however, justice is sensitive to the defendant's utility, then compromise verdicts are beneficial.

The analysis is subject to several limitations. As the model focuses on the judge's decision, I neglect some aspects which are discussed in the literature, such as the roles of detection, arrests and prosecutions (see e.g. Bjerk, 2005) (the latter is partially addressed in Appendix B). I further neglect explicit deterrence concerns on the part of judges, which may produce a countervailing incentive to the desire to face more guilty defendants. A different countervailing effect may be monetary compensation for exonerees (see e.g. Fon and Schäfer, 2007; Mungan and Klick, 2016), which has been argued to be insufficiently effective in the U.S. (Kahn, 2010). Risk and loss aversion (see e.g. Rizzolli and Stanca, 2012; Nicita and Rizzolli, 2014) are also neglected. While I formally neglect juries, introducing a second adjudicator (a jury) who decides on convictions prior to sentencing will not change the main effect of interest, as the sentence in the (basic) model is anyway independent of the probability of conviction. Furthermore, if a jury holds similar preferences as the judge and correctly anticipates the judge's choice of sentence, the results will remain largely unaffected for that reason as well (see also Lundberg, 2016).

While the model is not explicitly framed as a signaling game, it is helpful to note that the defendant's choice of innocence is, by itself, a form of a costly signal: by abstaining from crime (and giving up the benefit) the innocent implicitly sends a noisy signal (the level of evidence) on his type, leading to either a pooling equilibrium or a separating equilibrium (depending on which of the cases discussed in the paper applies). In this sense, my model does remains well within the framework of standard dynamic models with asymmetric information.

For the empirical part, the use of exonerations limits the ability to disentangle the effect of error occurrence and error correction (for a discussion, see Sarel, 2018). In particular, relying on the punishment of exonerees neglects the punishment of those who were wrongfully convicted but either did not challenge their conviction or failed to convince the court of their innocence post-conviction. Yet, as such defendants are also estimated by the judges to be guilty, this has no bearing on the mechanism that considers the level of evidence. Other limitations of the empirical part are addressed in the robustness tests, where I find that selection effects do not drive the results.

Acknowledgments

I am grateful to Sophie Bienenstock, Marco Fabbri, Eberhard Feess, Oren Gazal-Eyal, Jonah Gelbach, Samuel Gross, Jerg Gutmann, Roland Kirstein, Ronald Klingebiel, Hashem Nabas, Hans-Bernd Schäfer, Christoph Rößler, Stefan Voigt, and Franziska Weber for useful comments. I thank the German Law & Economics Association for their support in the form of an award for best paper presented by a young scholar. I also thank the participants of the following seminars and conferences: European Association of Law & Economics (Milan, 2018), LIEN seminar (Paris-Nanterre, 2018), Empirical Legal Studies Workshop (Amsterdam Law School, 2018), EMLE Midterm meeting (Hamburg, 2019), European Public Choice Society (Jerusalem, 2019), Jour Fixe of the ILE (Hamburg, 2019), and the German Law & Economics Association (Hannover, 2019).

Notes

[1]Separating p is usually attributed to Png (1986) (for earlier models, see Harris, 1970; Posner, 1973).
[2]For example, federal judges are guided by ranges set in sentencing guidelines, but can deviate (see *U.S. v. Booker*, 543 U.S. 220 (2005), *Gall v. U.S.*, 552 U.S. 38 (2007), and discussion in Kaiser and Spohn (2018))
[3]In Appendix B, I show that wrongfully convicted defendants may also systematically receive higher punishments if judges gain utility from punishing the innocents or if judges are able to manipulate the pool of defendants by altering deterrence strategically. In the same appendix, I demonstrate why compromise verdicts can occur at trial even if plea bargains are allowed.
[4]Many papers (e.g. Png, 1986; Kaplow, 1994; Polinsky and Shavell, 2000, 2001, 2007; Rizzolli and Saraceno, 2013) argue that wrongful convictions decrease deterrence, as mentioned above. Other scholars argue that wrongful convictions increase deterrence, e.g. if one has a continuous action space and can over-comply (Craswell and Calfee, 1986). Some argue that type I errors do not affect deterrence at all, e.g. when the defendant is convicted for someone else's crime (Lando, 2006; Lando and Mungan, 2017), which can happen irrespective of the decision to commit another crime (for counter arguments, see Garoupa and Rizzolli, 2012; Obidzinski and Oytana, 2019).
[5]The lion's share of this literature analyzes the impact of the federal sentencing guidelines. As my paper focuses on the comparison between the wrongfully and correctly convicted defendants, the sentencing guidelines - which apply to both defendant types - is not of main interest.
[6]With a slight abuse of language, I refer to the individual's self-selection into crime as the individual's type, as the self-selection decision directly follows from the exogenous draw of benefits.
[7]Denoting the probability of a type II error as $\epsilon_2 \equiv 1 - p_G$ shows that $\frac{\partial b^*}{\partial \varepsilon_2} = -f_G < 0$.
[8]The critical level of evidence $\widetilde{e}$ captures one specific level of evidence, which both types are equally likely to generate. I do not impose strong assumptions on what $\widetilde{e}$ is, so that even $\widetilde{e} = 0$ is possible, i.e. that both types are equally likely to generate no evidence at all.
[9]The equation follows from Bayes' rule given the MLRP (see e.g. Klemens, 2007).
[10]Note that rewriting yields $\omega_e = \frac{\widetilde{b}}{\widetilde{b} + \frac{g_I(e)}{g_G(e)}(1-\widetilde{b})}$, where $\frac{\partial}{\partial e}(\frac{g_I(e)}{g_G(e)}) < 0$ and $\frac{\partial \widetilde{b}}{\partial e} = 0$.
[11]I model the judge's decisions as simultaneous (rather than sequential), as they are made in the same information space.
[12] As $\frac{\partial c^I}{\partial \phi_e} > 0$, the first order condition below can only be satisfied when $\frac{\partial c^G}{\partial \phi_e} < 0$, i.e. when $\phi_e^* < \hat{\phi}^G$.
[13]The assumption that both functions are convex is made for simplicity and is sufficient, but not necessary, for the second order condition to hold (see the proof of Proposition 2 in Appendix A). It is, however, necessary that at least one of the functions is convex.
[14]In the basic model, the judge will never actually mix between convicting and acquitting. However, in the extensions in Appendix B, there is possible mixing, so that it is sensible to consider the probability of convicting.

[15]Recall that $\frac{\partial c^I}{\partial \phi_e} > 0$ and that in order to satisfy the first order condition, it must hold that $\frac{\partial c^G}{\partial \phi_e} < 0$. Thus, the condition $\frac{\partial c^I}{\partial \phi_e} > \frac{\partial c^I}{\partial \phi_e}$ always holds.

[16]$a^G > c^G$ holds by assumption. Note that if this were not the case, the judge would be better off by acquitting everyone, as convicting the guilty would be too costly.

[17]Recall that, in equilibrium, $\frac{\partial c^G}{\partial \phi_e} < 0$ (see footnote 12 above).

[18]On a technical level, this occurs because inequality (12) becomes less restrictive as e increases, as the probability that the defendant is guilty (on the LHS) increases whereas the SOP (on the RHS) decreases. Conversely, if the condition in the proposition is violated, the effect becomes ambiguous as then the LHS and RHS both increase in e, so that it depends which effect dominates. Note that the said condition is sufficient and necessary for part (i) and sufficient for part (ii).

[19]One good example of such a scenario can be found in the real-life case of Michael Peterson, which received media attention due to Netflix's documentary "The Staircase". Peterson was found next to his wife's bloody body, but maintained his innocence. After many legal battles, the case ended in a plea bargain, but some believe that a wild owl actually killed the defendant's wife's, shortly before her body was found by the defendant. If this is true, Petersen was indeed simply extremely unlucky.

[20]"Exoneration" is defined in the database as follows: "A person has been exonerated if he or she was convicted of a crime and, following a post-conviction re-examination of the evidence in the case, was either: (1) declared to be factually innocent by a government official or agency with the authority to make that declaration; or (2) relieved of all the consequences of the criminal conviction by a government official or body with the authority to take that action. The official action may be: (i) a complete pardon by a governor or other competent authority, whether or not the pardon is designated as based on innocence; (ii) an acquittal of all charges factually related to the crime for which the person was originally convicted; or (iii) a dismissal of all charges related to the crime for which the person was originally convicted, by a court or by a prosecutor with the authority to enter that dismissal. The pardon, acquittal, or dismissal must have been the result, at least in part, of evidence of innocence that either (i) was not presented at the trial at which the person was convicted; or (ii) if the person pled guilty, was not known to the defendant and the defense attorney, and to the court, at the time the plea was entered. The evidence of innocence need not be an explicit basis for the official action that exonerated the person. A person who otherwise qualifies has not been exonerated if there is unexplained physical evidence of that person's guilt." Exoneree is defined as "A person who was convicted of a crime and later officially declared innocent of that crime, or relieved of all legal consequences of the conviction because evidence of innocence that was not presented at trial required reconsideration of the case".

[21]Numbers are correct for the database's version as of 29.6.18. The database is continuously updated, so that a few new cases are discovered and added from time to time. The database lists 113 cases, but 6 cases were litigated in federal military courts and are therefore excluded.

[22]United States Sentencing Commission. Monitoring of Federal Criminal Sentences, Ann Arbor, MI: Inter-university Consortium for Political and Social Research. All files can be found at http://doi.org/10.3886/ICPSR*[XXXXX.vX]*, where "*[XXXX.vX]*" should be replaced with the following: 09317.v5, 03106.v1, 03496.v1, 03497.v1, 04110.v1, 04290.v3, 04633.v2, 04630.v2, 20120.v2, 22623.v2, 25424.v2, 28602.v1, 35336.v1, 35339.v1, 35342.v1, 35345.v1.

[23]Note that the 107 exonerations seem to be excluded in the ICPSR database although they have been litigated in the same (federal) courts. Most likely, this is because all innocents have appealed their cases (despite attempts, I have not been able to verify the official reason).

[24]Extremely long sentences are normalized to 99 years, to stay consistent with the NRE database's coding.

[25]The between-group difference in medians is significant both when comparing only those who did not take a plea ($p = 0.0098$, Wilcoxon ranksum) and those who did ($p < 0.0001$, Wilcoxon ranksum).

[26]The between-group difference in proportions of plea bargains is significant at the 1% level, see Table 1

[27]The exoneration database includes a plea-bargain dummy. For the ICPSR database, I classify all cases where the defendant plead guilty or did not contest the charges as a plea bargain.

[28]For defendants who are convicted of multiple offenses, the most severe offense is used. As the offense type and race definitions differ between the NRE and ICPSR databases, I adopted the nearest definition available. For example, "Tax offenses" and "Tax evasion" were merged into one category.

[29]Note that dummies for all offense types that appear in Table C3 are included, i.e. I do not use the pooled crime categories presented in Table 1.

[30]To avoid singleton dummies, one observation from the year 1983 (2017) is pooled with 1986 (2016).

[31]A specification that omits the control variables and includes only *Exonerated*, the plea-bargain dummy and their interaction yields similar coefficients.

[32]Note that I also do not make a causal claim here: It is not the exoneration per se which drives the sentence, but rather the mechanism of balancing the probability of guilt and sentence length, as a function of judicial concerns (which are unobservable).

[33]The matching is implemented Stata's user-developed 'Kmatch' command (Jann, 2017). The algorithm uses Mahalanobis Distance (Mahalanobis, 1936), which estimates a distance relative to a central point where the means of all variables intersect. This accounts also for the covariance between variables (see also Wooldridge, 2010, at §21.3.5).

[34]Regression adjustments with matching weights allow to reduce bias, in particular given rare outcomes (Franklin et al., 2017). As wrongful convictions are rare events, this seems appropriate. I include all covariates in the regression adjustment, except for year dummies (as years are otherwise included as continuous variables for the purpose of the matching).

[35]EB uses a reweighing scheme that imposes balancing constraints into the target function. Units are reweighed to achieve balance, but weights are kept as close as possible to the base weights, to avoid information loss. EB has been previously used to analyze sentencing (MacDonald and Donnelly, 2017).

[36]The average treatment effect ("ATE") is less relevant, as it calculates how much sentences would differ on average if *all* defendants were wrongfully convicted, whereas I focus on the marginal decrease in punishments for innocents. In particular, the ATE assumes that all individuals homogeneously respond to the treatment, which is unlikely to hold if innocents differ from the guilty on the level of evidence. Additionally, as the treatment group is small, the ATE may also be biased due to imprecise extrapolation.

[37]Replacing MD with propensity score matching with kernel density yields qualitatively similar results. Restricting attention to defendants who did not take a plea also does not qualitatively change the results.

[38]These correspond to columns (3) and (2) of Table 4 above.

[39]VRs all fall within the range of 0.5 - 2 and therefore deemed acceptable (see Rubin, 2001; Franklin et al., 2017). Conventions regarding the acceptable SMDs vary (see Linden and Samuels, 2013) with cutoff thresholds between 0.1-0.25. Post-matching SMDs are all below 0.25, except for the specification without EB, where the sentencing year has an SMD of 0.3. However, even in that specification, the average SMD of covariates is very low (0.09), suggesting an acceptable overall balance.

[40]The level of evidence is presumably weaker when no crime has occurred as the incident itself is in dispute, whereas in the case of mistaken identity, the dispute is only with respect to who the perpetrator is.

[41]Judges' demographics are unavailable for the state courts, and therefore excluded.

[42] The extensions further show that relaxing some assumptions may lead to a reversed outcome, where judges impose higher penalties (systematically or non-systematically) when the defendant seems less guilty.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" NBER Working Paper 24003. URL: `https://www.nber.org/papers/w24003.pdf`.

Abadie, Alberto and Guido W Imbens. 2011. "Bias-corrected matching estimators for average treatment effects." *Journal of Business & Economic Statistics* 29 (1):1–11.

Andreoni, James. 1991. "Reasonable doubt and the optimal magnitude of fines: Should the penalty fit the crime?" *RAND Journal of Economics* 22 (3):385–395.

Baumann, Florian and Tim Friehe. 2017. "Crime and Punishment under Evidentiary Uncertainty: Laboratory Evidence." *Supreme Court Economic Review* 25 (1):65–104.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2):169–217.

Bindler, Anna and Randi Hjalmarsson. 2018. "How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments." *American Economic Journal: Economic Policy* 10 (4):36–78.

Bjerk, David. 2005. "Making the crime fit the penalty: The role of prosecutorial discretion under mandatory minimum sentencing." *The Journal of Law and Economics* 48 (2):591–625.

Bjerk, David and Eric Helland. 2017. "Using a Ratio Test to Estimate Racial Differences in Wrongful Conviction Rates." *IZA Discussion Paper 10631. URL: https://ssrn.com/abstract=2940621* .

Cassell, Paul G. 2018. "Overstating America's Wrongful Conviction Rate: Reassessing the Conventional Wisdom about the Prevalence of Wrongful Convictions." *Arizona Law Review* 60:815–864.

Chen, Daniel L, Tobias J Moskowitz, and Kelly Shue. 2016. "Decision making under the gamblers fallacy: Evidence from asylum judges, loan officers, and baseball umpires." *The Quarterly Journal of Economics* 131 (3):1181–1242.

Clow, Kimberley A and Amy-May Leach. 2015. "After innocence: Perceptions of individuals who have been wrongfully convicted." *Legal and Criminological Psychology* 20 (1):147–164.

Craswell, Richard and John E Calfee. 1986. "Deterrence and uncertain legal standards." *Journal of Law, Economics, and Organization* 2:279–303.

Epps, Daniel. 2015. "Consequences of Error in Criminal Justice, The." *Harvard Law Review* 128:1065–1151.

Feess, Eberhard and Roee Sarel. 2018. "Judicial Effort and the Appeal System: Theory and Experiment." *The Journal of Legal Studies* 47 (2):269–294.

Feess, Eberhard, Hannah Schildberg-Hörisch, Markus Schramm, and Ansgar Wohlschlegel. 2018. "The Impact of Fine Size and Uncertainty on Punishment and Deterrence: Theory and Evidence from the Laboratory." *Journal of Economic Behavior and Organization* 149:58–73.

Fon, Vincy and Hans-Bernd Schäfer. 2007. "State liability for wrongful conviction: Incentive effects on crime levels." *Journal of Institutional and Theoretical Economics (JITE)* 163 (2):269–284.

Franklin, Jessica M, Wesley Eddings, Peter C Austin, Elizabeth A Stuart, and Sebastian Schneeweiss. 2017. "Comparing the performance of propensity score methods in healthcare database studies with rare outcomes." *Statistics in Medicine* 36 (12):1946–1963.

Garoupa, Nuno. 2017. "Explaining the standard of proof in criminal law: A new insight." *Supreme Court Economic Review* 25 (1):111–122.

Garoupa, Nuno and Matteo Rizzolli. 2012. "Wrongful convictions do lower deterrence." *Journal of Institutional and Theoretical Economics (JITE)* 168 (2):224–231.

Gould, Jon B, Julia Carrano, Richard A Leo, and Katie Hail-Jares. 2014. "Predicting Erroneous Convictions." *Iowa Law Review* 99:471–2299.

Gould, Jon B and Richard A Leo. 2010. "One hundred years later: Wrongful convictions after a century of research." *Journal of Criminal Law and Criminology* 100 (3):825–868.

———. 2015. "The path to exoneration." *Albany Law Review* 79:325–372.

Gross, Samuel R. 2016. "What we think, what we know and what we think we know about false convictions." *Ohio State Journal of Criminal Law* 14:753.

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

Gross, Samuel R, Barbara OBrien, Chen Hu, and Edward H Kennedy. 2014. "Rate of false conviction of criminal defendants who are sentenced to death." *Proceedings of the National Academy of Sciences* 111 (20):7230–7235.

Gross, Samuel R and Michael Shaffer. 2012. "Exonerations in the United States, 1989–2012." University of Michigan Public Law Working Paper No. 277. URL: https://ssrn.com/abstract=2092195.

Gutmann, Jerg, Katharina Pfaff, and Stefan Voigt. 2017. "Banking crises and human rights." *Applied Economics Letters* 24 (19):1374–1377.

Hainmueller, Jens. 2012. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis* 20 (1):25–46.

Hainmueller, Jens and Yiqing Xu. 2013. "Ebalance: A Stata package for entropy balancing." *Journal of Statistical Software* 54 (7):1–18.

Harris, John R. 1970. "On the economics of law and order." *Journal of Political Economy* 78:165–174.

Hartje, Rebecca and Michael Hübler. 2017. "Smartphones support smart labour." *Applied Economics Letters* 24 (7):467–471.

Heath, Wendy P. 2009. "Arresting and convicting the innocent: The potential role of an inappropriate emotional display in the accused." *Behavioral Sciences & the Law* 27 (3):313–332.

Hynes, Richard. 2018. "The Optimal Evidence Threshold: Balancing the Frequency and Severity of Wrongful Convictions." *Supreme Court Economic Review* 26:113–134.

Jann, Ben. 2017. "Kmatch: Stata module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment." Working Paper. Institute of Sociology, University of Bern, Switzerland. URL: https://ideas.repec.org/c/boc/bocode/s458346.html.

Kahn, Daniel S. 2010. "Presumed guilty until proven innocent: The burden of proof in wrongful conviction claims under state compensation statutes." *University of Michigan JL Reform* 44:123–168.

Kaiser, Kimberly and Cassia Spohn. 2018. "Why Do Judges Depart? A Review of Reasons for Judicial Departures in Federal Sentencing." *Criminology, Criminal Justice, Law & Society* 19:43–62.

Kaplow, Louis. 1994. "The value of accuracy in adjudication: An economic analysis." *The Journal of Legal Studies* 23 (1):307–401.

———. 2011a. "On the Optimal Burden of Proof." *Journal of Political Economy* 119 (6):1104–1140.

———. 2011b. "Optimal proof burdens, deterrence, and the chilling of desirable behavior." *American Economic Review* 101 (3):277–80.

———. 2012. "Burden of Proof." *Yale Law Journal* 121:738–859.

Kim, Sangjoon, Jaihyun Park, Kwangbai Park, and Jin-Sup Eom. 2013. "Judge-Jury Agreement in Criminal Cases: The First Three Years of the Korean Jury System." *Journal of Empirical Legal Studies* 10 (1):35–53.

King, Gary and Richard Nielsen. 2019. "Why propensity scores should not be used for matching." *Political Analysis* 27 (4):435–454.

King, Nancy J. and Rosevelt L. Noble. 2004. "Felony jury sentencing in practice: A three-state study." *Vanderbilt Law Review* 57:885–962.

Klemens, Ben. 2007. "When Do Ordered Prior Distributions Induce Ordered Posterior Distributions?" Working Paper, U.S. Department of the Treasury. URL: https://ssrn.com/abstract=964720.

Knight, Jack and Lee Epstein. 2017. "Strategic Accounts of Judging." In *Routledge Handbook of Judicial Behavior*. Routledge, 48–61.

Lando, Henrik. 2005. "The size of the sanction should depend on the weight of the evidence." *Review of Law & Economics* 1 (2):277–292.

———. 2006. "Does wrongful conviction lower deterrence?" *The Journal of Legal Studies* 35 (2):327–337.

Lando, Henrik and Murat C Mungan. 2017. "The Effect of Type-1 Error on Deterrence." *International Review of Law and Economics* 53:1–8.

Linden, Ariel and Steven J Samuels. 2013. "Using balance statistics to determine the optimal number of controls in matching studies." *Journal of Evaluation in Clinical Practice* 19 (5):968–975.

Loeffler, Charles E, Jordan Hyatt, and Greg Ridgeway. 2019. "Measuring Self-Reported Wrongful Convictions Among Prisoners." *Journal of Quantitative Criminology* 35:259–286.

Lundberg, Alexander. 2016. "Sentencing Discretion and Burdens of Proof." *International Review of Law and Economics* 46:34–42.

———. 2017. "Resolving the Conviction Paradox: Federal Judges and Shifting Burdens of Proof Decisions." Working paper. West Virginia University, West Virginia, VA. URL: http://alex-lundberg-economics.com/wp-content/uploads/2017/05/lundberg_judges.pdf.

———. 2018. "When Do The Innocent Plead Guilty?" Working Paper. West Virginia University, WV. URL: https://ssrn.com/abstract=3100845.

MacDonald, John M and Ellen A Donnelly. 2017. "Evaluating the Role of Race in Sentencing: An Entropy Weighting Analysis." *Justice Quarterly* :1–26.

Mahalanobis, Prasanta Chandra. 1936. "On the generalized distance in statistics." *Proceedings of the National Institute of Science of India* 12:49–55.

Malo, Miguel Á, Ángel Martín-Román, and Alfonso Moral. 2018. ""Peer effects" or "quasi-peer effects" in Spanish labour court rulings." *European Journal of Law and Economics* 45 (3):497–525.

Mungan, Murat C and Jonathan Klick. 2016. "Reducing False Guilty Pleas and Wrongful Convictions through Exoneree Compensation." *The Journal of Law and Economics* 59 (1):173–189.

Mustard, David B. 2001. "Racial, ethnic, and gender disparities in sentencing: Evidence from the US federal courts." *The Journal of Law and Economics* 44 (1):285–314.

Nicita, Antonio and Matteo Rizzolli. 2014. "In Dubio Pro Reo. Behavioral explanations of pro-defendant bias in procedures." *CESifo Economic Studies* .

Obidzinski, Marie and Yves Oytana. 2019. "Identity errors and the standard of proof." *International Review of Law and Economics* 57:73–80.

Olney, Maeve and Scott Bonn. 2014. "An Exploratory Study of the Legal and Non-Legal Factors Associated With Exoneration for Wrongful Conviction: The Power of DNA Evidence." *Criminal Justice Policy Review* 26 (4):1–21.

Png, Ivan PL. 1986. "Optimal subsidies and damages in the presence of judicial error." *International Review of Law and Economics* 6 (1):101–105.

Polinsky, A Mitchell and Steven Shavell. 2000. "The Economic Theory of Public Enforcement of Law." *Journal of Economic Literature* 38 (1):45–76.

———. 2001. "Corruption and optimal law enforcement." *Journal of Public Economics* 81 (1):1–24.

———. 2007. "The theory of public enforcement of law." *Handbook of Law and Economics* 1:403–454.

Posner, Richard A. 1973. "An economic approach to legal procedure and judicial administration." *The Journal of Legal Studies* 2 (2):399–458.

———. 1993. "What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does)." *Supreme Court Economic Review* 3:1–41.

Rizzolli, Matteo and Margherita Saraceno. 2013. "Better that ten guilty persons escape: Punishment costs explain the standard of evidence." *Public Choice* 155 (3-4):395–411.

Rizzolli, Matteo and Luca Stanca. 2012. "Judicial errors and crime deterrence: Theory and experimental evidence." *The Journal of Law and Economics* 55 (2):311–338.

Rosenbaum, Paul R and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1):41–55.

Rubin, Donald B. 2001. "Using propensity scores to help design observational studies: Application to the tobacco litigation." *Health Services and Outcomes Research Methodology* 2 (3-4):169–188.

Samaha, Adam M. 2009. "Randomization in adjudication." *William & Mary Law Review* 51:1–86.

Sarel, Roee. 2018. "Judicial Errors, Crime Deterrence and Appeals: Evidence from US Federal Courts." Working Paper. Institute of Law & Economics, University of Hamburg, URL: `https://ssrn.com/abstract=2739674`.

Schauer, Frederick and Richard Zeckhauser. 1996. "On the degree of confidence for adverse decisions." *The Journal of Legal Studies* 25 (1):27–52.

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

Starr, Sonja B. 2014. "Estimating gender disparities in federal criminal cases." *American Law and Economics Review* 17 (1):127–159.

Teichman, Doron. 2017. "Convicting with Reasonable Doubt: An Evidentiary Theory of Criminal Law." *Notre Dame Law Review* 93:757–810.

Thomas III, George C. 2018. "Where Have All the Innocents Gone." *Arizona Law Review* 60:865–890.

Tor, Avishalom, Oren Gazal-Ayal, and Stephen M Garcia. 2010. "Fairness and the willingness to accept plea bargain offers." *Journal of Empirical Legal Studies* 7 (1):97–116.

Tsiboe, Francis, Yacob A Zereyesus, and Evelyn Osei. 2016. "Non-farm work, food poverty, and nutrient availability in northern Ghana." *Journal of Rural Studies* 47:97–107.

Ulmer, Jeffery T. 2019. "Criminal Courts as Inhabited Institutions: Making Sense of Difference and Similarity in Sentencing." *Crime and Justice* 48 (1):483–522.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT press, second ed.

Appendix A: Proofs

Proof of Proposition 1. Applying the implicit function theorem to the first order condition $Q(\phi_e, \omega_e) \equiv \rho_e \left(\omega_e \frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e}\right) = 0$ yields:

(A1) $$\frac{\partial \phi_e^*}{\partial e} = -\frac{\frac{\partial Q(\phi_e,\omega_e)}{\partial e}}{\frac{\partial Q(\phi_e,\omega_e)}{\partial \phi_e}}$$

The denominator is identical to the SOC w.r.t. the penalty ϕ_e and is required to be positive by assumption for an interior solution, i.e.

(A2) $$\frac{\partial^2 J(\rho_e, \phi_e)}{\partial \phi_e^2} = \rho_e \frac{\partial Q(\phi_e, \omega_e)}{\partial \phi_e} = \rho_e \left(\omega_e \frac{\partial^2 c^G}{\partial \phi_e^2} + (1-\omega_e)\frac{\partial^2 c^I}{\partial \phi_e^2}\right) > 0$$

as $\frac{\partial^2 c^G}{\partial^2 \phi_e} > 0$ and $\frac{\partial^2 c^I}{\partial^2 \phi_e} > 0$ by assumption. Next, the numerator is

(A3) $$\frac{\partial Q(\phi_e, \omega_e)}{\partial e} = \frac{\partial \omega_e}{\partial e}\rho_e(\frac{\partial c^G}{\partial \phi_e} - \frac{\partial c^I}{\partial \phi_e}) + \frac{\partial \rho_e}{\partial e}\left(\omega_e \frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e}\right)$$

The FOC $Q(\phi_e, \omega_e) = 0$ implies that, in equilibrium, the second term equals zero. Thus, the numerator is negative if and only if the first term is negative, i.e. if $\frac{\partial c^I}{\partial \phi_e} > \frac{\partial c^G}{\partial \phi_e}$.■

Proof of Corollary 1. The corollary directly follows from Proposition 1 and $\frac{\partial \omega_e}{\partial e} > 0$.■

Proof of Proposition 2. *Part (i).* Recall that ρ_e only has a corner solution, depending on whether inequality (12) holds. As the LHS ($=\omega_e$) does not depend on ϕ_e, it is sufficient to consider the SOP on the RHS ($= \widetilde{\omega}_e = \frac{c^I}{c^I - c^G + a^G}$). Deriving the SOP yields $\frac{\partial \widetilde{\omega}_e}{\partial \phi_e} = \frac{\frac{\partial c^I}{\partial \phi_e}(a^G - c^G) + \frac{\partial c^G}{\partial \phi_e} * c^I}{(c^I - c^G + a^G)^2}$. The denominator is positive, so attention can be restricted to the numerator $N \equiv \frac{\partial c^I}{\partial \phi_e}(a^G - c^G) + \frac{\partial c^G}{\partial \phi_e} * c^I$. Thus, $\frac{\partial \widetilde{\omega}_e}{\partial \phi_e} < 0 \Leftrightarrow N < 0$. *Part (ii).* $\frac{\partial \widetilde{\omega}_e}{\partial e} = \frac{\frac{\partial \phi_e}{\partial e}\left(\frac{\partial c^I}{\partial e}(a^G - c^G) + \frac{\partial c^G}{\partial e} c^I\right)}{(c^I - c^G + a^G)^2} < 0 \Leftrightarrow \frac{\partial c^I}{\partial \phi_e}(a^G - c^G) + \frac{\partial c^G}{\partial \phi_e} * c^I < 0$, given that $\frac{\partial \phi_e}{\partial e} > 0$ as shown above in the proof of Proposition 1.■

Proof of Corollary 2. The corollary directly follows from Proposition 2 and the explanation provided in footnote 18.■

Proof of Proposition 3. *Part (i).* the first order conditions are

$$\text{(A4)}\quad \frac{\partial J(\rho_e,\phi_e)}{\partial \phi_e} = \rho_e\left(\omega_e\frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e}\right) + \frac{\partial \omega_e}{\partial \phi_e}\left(a^G + \rho_e\left(c^G - c^I - a^G\right)\right) = 0$$

$$\text{(A5)}\quad \frac{\partial J(\rho_e,\phi_e)}{\partial \rho_e} = \omega_e\left(c^G - a^G\right) + (1-\omega_e)\,c^I + \frac{\partial \omega_e}{\partial \rho_e}\left(a^G + \rho_e\left(c^G - c^I - a^G\right)\right) = 0$$

In the case of $e = \widetilde{e}$, we get $\frac{\partial \omega_e}{\partial \phi_e} = \frac{\partial \omega_e}{\partial \rho_e} = 0$, which converges to the basic model.

Part (ii). If $e \neq \widetilde{e}$, we get instead $\frac{\partial \omega_e}{\partial \phi_e} = -\frac{g_G(e)g_I(e)\frac{\partial \widetilde{b}}{\partial \phi_e^*}}{\widetilde{b}(g_G(e)-g_I(e))+g_G(e))^2}$ and $\frac{\partial \omega_e}{\partial \rho_e} = -\frac{g_G(e)g_I(e)\frac{\partial \widetilde{b}}{\partial \rho_e^*}}{\widetilde{b}(g_G(e)-g_I(e))+g_G(e))^2}$. The deterrence threshold is then $\widetilde{b} = E(f_G) - E(f_I) = \int_0^1 \rho_e^*\phi_e^*(g_G(e) - g_I(e))de$ and its partial derivatives w.r.t. to ϕ_e and ρ_e are $\frac{\partial \widetilde{b}}{\partial \phi_e} = \int_0^1 \rho_e^* g_I(e)(\frac{g_G(e)}{g_I(e)} - 1)$ and $\frac{\partial \widetilde{b}}{\partial \rho_e} = \int_0^1 \phi_e^* g_I(e)(\frac{g_G(e)}{g_I(e)} - 1)$. Next, equation (3) in the main text implies that $\frac{\partial \omega_e}{\partial \phi_e} \in \begin{cases} > 0 & \text{if } e < \widetilde{e} \\ 0 & \text{if } e = \widetilde{e} \\ < 0 & \text{if } e > \widetilde{e} \end{cases}$. The same holds for $\frac{\partial \omega_e}{\partial \rho_e}$. Rewriting in terms of $\widetilde{\omega}_e$ and combining equations (A4) and (A5) yields: $\rho_e(\omega_e\frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e} = \frac{\frac{\partial \omega_e}{\partial \phi_e}}{\frac{\partial \omega_e}{\partial \rho_e}} c^I(1 - \frac{\omega_e}{\widetilde{\omega}_e})$, as specified in the proposition. Note that the RHS is positive, as $\frac{\partial \omega_e}{\partial \phi_e} > 0 \Leftrightarrow \frac{\partial \omega_e}{\partial \rho_e} > 0$.

Proof of Corollary 3. Both parts of Corollary 3 follow from the ambiguity described in the text w.r.t to whether ϕ_e^* and $\widetilde{\omega}_e$ increase or decrease in e.■

Appendix B: Extensions

B.1. Extension 1: Interior solution for punishing the innocent

Suppose that instead of Assumption 2 one would assume as follows:

Assumption 2B *The cost of a wrongful conviction has a minimum at the point $\hat{\phi}^I$, such that:*

(B1)
$$\frac{\partial c^I}{\partial \phi_e} \in \begin{cases} < 0 & \text{if } \phi_e < \hat{\phi}^I \\ = 0 & \text{if } \phi_e = \hat{\phi}^I \\ > 0 & \text{if } \phi_e > \hat{\phi}^I \end{cases}$$

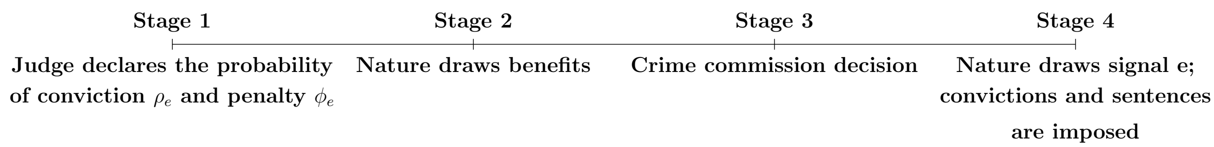
where $\hat{\phi}^I$ functions as an "optimal" penalty for the innocent similar to $\hat{\phi}^G$ for the guilty. There are several reasons why judges may want to increase the punishment (up to a certain point) even if the defendant is innocent, e.g. because the judge fears of appearing to be "too lenient" (see e.g. Chen, Moskowitz, and Shue, 2016, pp. 1188) or because setting an extremely low sentence may increase the probability of appellate review.[43] Then, if $\hat{\phi}^G < \phi_e^* < \hat{\phi}^I$ it can occur that $\frac{\partial c^I}{\partial \phi_e} \leq 0 < \frac{\partial c^G}{\partial \phi_e}$. Proposition 1 would then imply that $\frac{\partial \phi_e^*}{\partial e} < 0$.

This outcome is counter-intuitive, as it requires that the judge would weakly gain utility from increasing the punishment if the defendant is innocent, but pay a preference cost if the defendant is guilty. However, the fear of reversal by a higher court may bring about such a scenario. For example, suppose that innocent defendants are more likely to appeal (or even always appeal) the judge's decision. Then, once the judge has decided to convict, increasing the penalty will not affect the probability that an appeal by an innocent is filed. However, if the defendant is guilty, he may be more likely to appeal if the punishment is higher. Hence, if the fear of reversal (which I do not explicitly model) dominates, it may occur that wrongfully convicted defendants receive higher punishments.

B.2. Extension 2: Reversed-timing

In the basic model, the judge acts after crimes have already been committed and therefore takes deterrence as given. Suppose instead that the timing of the model is reversed (see Figure B1 below): In stage 1, the judge declares whether she convicts and what is the penalty for each signal realization. In stage 2, nature draws the benefits. In stage 3, individuals decide whether to commit a crime. In stage 4, nature draws the signal and then convictions and punishments are realized.

Figure B1: Reversed timing



To ensure that the probability ω_e remains bounded between 0 and 1, assume that for all e, $\widetilde{b} \leq 0 \Rightarrow \omega_e = 1$ and $\widetilde{b} \geq 1 \Rightarrow \omega_e = 0$. This ensures that if everyone (no one) commits the crime then the signal is uninformative, as the probability of facing guilty (innocent) defendants is simply the base probability of 1.

B.2.1. Sentencing: A pool effect

As the judge can now affect deterrence, her minimization of the cost function w.r.t. the penalty yields the following first order condition:

$$\text{(B2)} \qquad \frac{\partial J(\rho_e, \phi_e)}{\partial \phi_e} = \rho_e \left(\omega_e \frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e} \right) + \frac{\partial \omega_e}{\partial \phi_e} \left(\rho_e (c^G - c^I) + (1-\rho_e) a^G \right) = 0$$

The first term in equation (B2) is identical to the basic model. The second term captures the judge's change in costs due a marginal change in the likelihood of facing guilty defendants. This change can be interpreted as a "pool effect": by altering deterrence, the judge is able to affect the composition of the pool of defendants, which implies that the probability of making a type I (II) error changes. As part of the proof of Proposition 3 below, I show that

$$\text{(B3)} \qquad \frac{\partial \omega_e}{\partial \phi_e} \in \begin{cases} > 0 & \text{if } e < \widetilde{e} \\ 0 & \text{if } e = \widetilde{e} \\ < 0 & \text{if } e > \widetilde{e} \end{cases}$$

The cases in (B3) show that the direction of the pool effect depends on the level of evidence. For an interpretation, suppose first that the judge punishes those who are likely to be innocent (when $e < \widetilde{e}$). This decreases deterrence, so that the base probability of facing a guilty defendant increases (i.e. $\frac{\partial \omega_L}{\partial \phi_e} > 0$). Then, if the judge convicts (this occurs with probability ρ_e), she is more likely to incur the cost of a correct conviction c^G but less likely to incur the cost of a type I error c^I, so that $(c^G - c^I)$ is the net change in costs. As the cost of a wrongful conviction is typically higher, $c^G - c^I < 0$ can be assumed, so that if $\frac{\partial \omega_e}{\partial \phi_e} > 0$ then the overall effect is a reduction in the judge's costs. This means that the judge may have a distorted incentive: by inducing crime, the judge can avoid facing innocents and thus avoid wrongful convictions. Moreover, if the judge acquits (this occurs with probability $1 - \rho_e$), she is also marginally more likely to incur the cost of a wrongful acquittal a^G, as the defendant is now (marginally) more likely to be guilty due to lower deterrence.

Respectively, the judge may have an incentive to impose very *lenient* sentences on those who seem guilty (when $e > \widetilde{e}$) as this again reduces the base probability of facing an innocent defendant (as then $\frac{\partial \omega_e}{\partial \phi_e} < 0$).

Finally, increasing the punishment for a 'medium' signal $e = \widetilde{e}$ has no effect on deterrence ($\frac{\partial \omega_e}{\partial \phi_e} = 0$), as this particular signal is equally likely for the innocent and the guilty (for a model with a similar intuition, see Lando (2006)).

B.2.2. Cost-minimizing penalty and probability of conviction

In contrast to the basic model, the cost-minimizing penalty depends on the probability of guilt, as long as $e \neq \widetilde{e}$. To account for this dependency, combining the first order conditions w.r.t. ϕ_e and ρ_e yields the following proposition (see proof in Appendix A above):[44]

Proposition 3

(i) If $e = \widetilde{e}$, the cost-minimizing penalty and probability of conviction are the same as in the basic model.

(ii) if $e \neq \widetilde{e}$ then the cost-minimizing penalty and probability of conviction satisfy the condition $\rho_e(\omega_e \frac{\partial c^G}{\partial \phi_e} + (1-\omega_e)\frac{\partial c^I}{\partial \phi_e} = \frac{\frac{\partial \omega_e}{\partial \phi_e}}{\frac{\partial \omega_e}{\partial \rho_e}} c^I(1 - \frac{\omega_e}{\widetilde{\omega}_e})$

Part (i) shows that the problem for a medium signal is the same as in the basic model, as the expected sanction then has no effect on deterrence (which holds by assumption in the basic model). Part (ii) specifies a condition for the minimization of the judge's costs when the signal is not medium, i.e. when the penalty and probability of convictions affect deterrence. For an interpretation, note that the LHS is the expected marginal costs of conviction and identical to the LHS of the first order condition in the basic model. The RHS is then the expected marginal benefit from the pool effect. Namely, this benefit arises when the judge reduces the expected costs of a wrongful conviction as a result of reducing the probability of facing a guilty defendants (expressed by the cross derivative of ω_e w.r.t. ϕ_e and ρ_e. The term in parentheses relates to whether the probability of facing a guilty defendant exceeds the SOP (whether $\frac{\omega_e}{\widetilde{\omega}_e} > 1$). Then, the condition has two alternative solutions, depending on whether the judge focuses on ex-ante or ex-post costs: As a first option, the judge may decide to set the parameters so that those who seem guilty receive *low* punishments (and vice versa), due to the pool effect, as explained above.[45]

As a second option, the judge can neglect the deterrence effect and reduce the expected sanction for those who seem to be innocent, as in the basic model. Without additional assumptions, it remains ambiguous which option minimizes the judge's costs, in particular because the chosen parameters also affect the SOP. As shown in the basic model, the effect on the SOP depends on whether the CC effect or WC effect dominates.[46]

Summing up, if the judge can affect deterrence, wrongfully convicted defendants may systematically receive either higher or lower penalties, but the result is ambiguous in the absence of additional assumptions. This is summarized in the following Corollary:

Corollary 3 *If the judge can affect deterrence, then:*

(i) Defendants who seem more guilty may be either more, less, or equally likely to be convicted and may either receive lower, higher, or equal punishments if they are convicted.

(ii) Type I errors may have either a lower, higher, or the same marginal effect on deterrence compared to type II errors.

B.3. Extension 3: Plea bargains

In the final extension, I briefly consider a (stylized) variant that includes plea bargains. As plea bargains are not of direct interest and are mostly relevant as a possible confounder in the empirical analysis, I do not develop a full-fledged model, as this would drastically complicate the analysis (for such a model, see Lundberg, 2018). In particular, I assume that prosecutors cannot condition the plea offer on the level of evidence. Instead, prosecutors only offer a binary "take it or leave it" offer. The timing is as in the basic model, with one addition: Each defendant first observes his own level of evidence and can decide whether or not to take a plea bargain prior to trial. If the defendant takes the plea, he is convicted and incurs a fixed sanction $\bar{\Phi}$. Otherwise, the defendant proceeds to trial. The effect of plea bargains is then straightforward, yielding several possible scenarios, depending on $\bar{\Phi}$:

- Option 1: All defendants take the plea ($\bar{\Phi} < \phi_e^*, \forall e$). Then, trivially, all defendants receive the same penalty $\bar{\Phi}$.

- Option 2: No defendant takes the plea ($\bar{\Phi} > \phi^*_e, \forall e$). Then, everything remains the same as in the basic model.

- Option 3: Some defendants take the plea, i.e. there exists a cutoff level of evidence $\hat{e}^{plea}$ such that only those with a higher level take a plea.

Suppose that $\bar{\Phi}$ is endogenously decided by a prosecutor. The prosecutor then faces a similar dilemma as the judge: if the prosecutor is somewhat averse to wrongful convictions, she will offer a relatively high penalty, so that most innocents will refuse.[47] For instance, the prosecutor may set $\bar{\Phi} = \phi^*_{\tilde{e}}$, so that only defendants whose e higher than the threshold that indicates innocence will take a plea. Conversely, if the prosecutor cares more about the conviction rate and does not care in particular about the punishment size, she will offer a low penalty.

As options 1 and 2 are trivial for the judge's decision, suppose that the prosecutor chooses option 3. The implication for the judge is then only that the range of observed signals shrinks, but the decision's mechanism remains otherwise exactly the same. Namely, as guilty defendants are still more likely to generate more evidence, those who take the plea are, on average, still more likely to be guilty. The only difference is then that the probability of facing innocent defendants increases compared to the basic model. Hence, more evidence still yield a higher sentence, so that the results remain qualitatively unaffected.

Appendix C: Additional tables and figures

Figure C1: Histogram

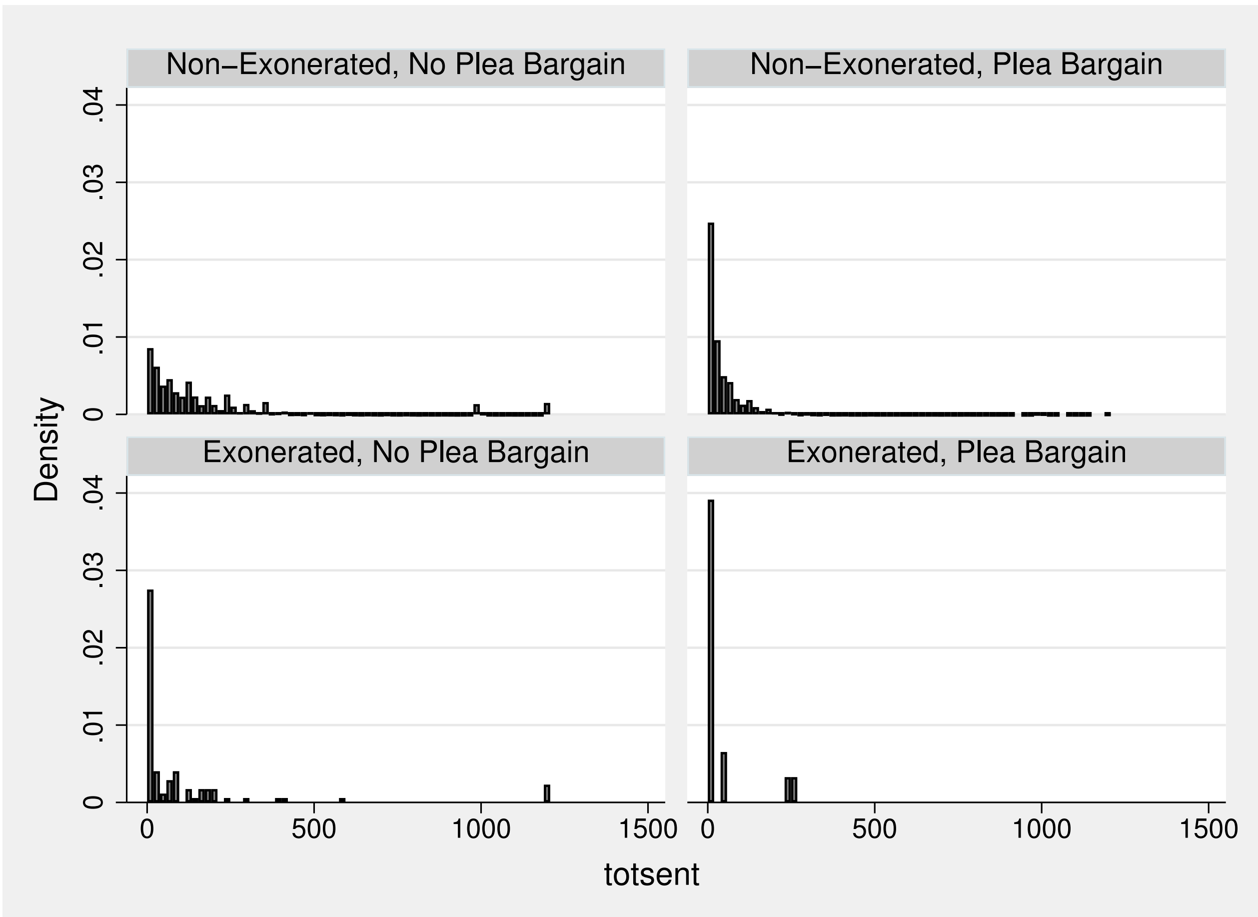


Figure C2: Comparison of sentences: pooled (no separation by plea bargains)

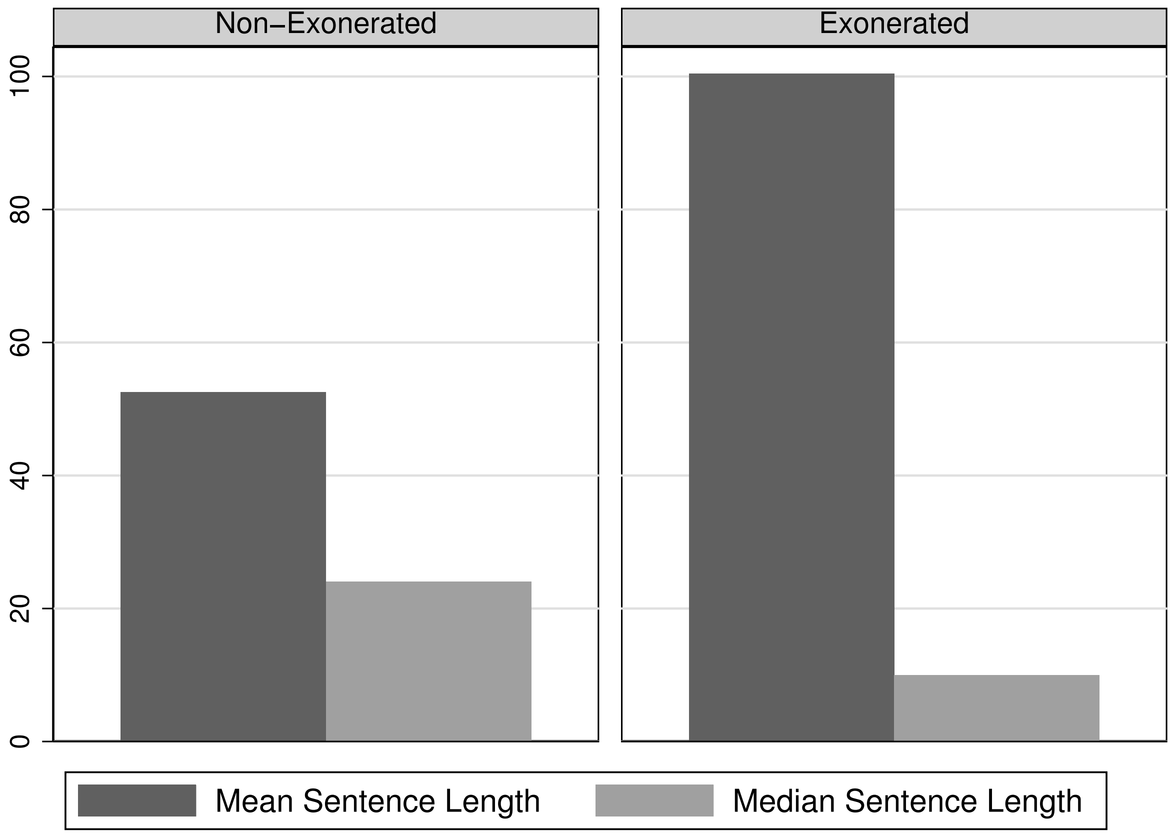


Table C1:
SAMPLE SELECTION PROCESS

Database/Source	N	Cumulative observations	Years range
Registry of Exonerations	113	113	1983-2017
Military cases	-6	107	
Federal Sentencing (ICPSR)	1,695,432	1,695,539	1987-2017
Missing values (totsent)	-3,843	1,691,696	
Additional Missing values (plea bargain)	-6,318	1,685,378	
Additional Missing values (defendant's age)	-21,246	1,664,132	
Additional Missing values (judge's attributes)	-5,245	1,658,887	
Additional Missing values (offense type)	-2,922	1,655,695	
Additional Missing values (unknown gender treated as missing)	-714,274	941,691	
TOTAL		941,691	1983-2017

NOTE.– This table presents the origins of the observations used in the regressions, where some observations drop due to missing values.

Table C2:
Descriptives by Offense Category

Factor	Level	Other	Violent	Property	Sex	Administration of justice	Immigration	White Collar	Drugs	p-value	Test
Observations (n)		244367	168098	59390	21600	25587	380048	242796	553653		
Total Sentence (totsent), mean (SD)		37.1 (117.3)	80.7 (134.0)	68.1 (120.4)	122.1 (144.4)	20.8 (81.8)	22.6 (69.0)	20.2 (78.4)	82.6 (120.0)	<0.001	ANOVA
Exonerated		4 (<1%)	18 (<1%)	6 (<1%)	2 (<1%)	3 (<1%)	4 (<1%)	42 (<1%)	28 (<1%)	<0.001	Pearson's χ^2
Defendant's Gender	Gender: male	122065 (50.0%)	91034 (54.2%)	38333 (64.5%)	10031 (46.4%)	10292 (40.2%)	171864 (45.2%)	102507 (42.2%)	277490 (50.1%)	<0.001	Pearson's χ^2
	Gender: female	26317 (10.8%)	4338 (2.6%)	8501 (14.3%)	283 (1.3%)	3916 (15.3%)	13192 (3.5%)	40230 (16.6%)	40090 (7.2%)		
	Gender: unknown	95985 (39.3%)	72726 (43.3%)	12556 (21.1%)	11286 (52.3%)	11379 (44.5%)	194992 (51.3%)	100059 (41.2%)	236073 (42.6%)		
Defendant's race	Race: White	100842 (41.3%)	54822 (32.6%)	28154 (47.4%)	14076 (65.2%)	11028 (43.1%)	22358 (5.9%)	119934 (49.4%)	147236 (26.6%)	<0.001	Pearson's χ^2
	Race: Black	48044 (19.7%)	72564 (43.2%)	23075 (38.9%)	1507 (7.0%)	5386 (21.0%)	8652 (2.3%)	66762 (27.5%)	157706 (28.5%)		
	Race: Hispanic	46426 (19.0%)	25948 (15.4%)	4135 (7.0%)	1982 (9.2%)	6253 (24.4%)	303754 (79.9%)	30931 (12.7%)	220856 (39.9%)		
	Race: Native	3964 (1.6%)	9013 (5.4%)	910 (1.5%)	3540 (16.4%)	811 (3.2%)	770 (0.2%)	1820 (0.7%)	5066 (0.9%)		
	Race: Asian	7358 (3.0%)	2123 (1.3%)	1294 (2.2%)	308 (1.4%)	779 (3.0%)	3210 (0.8%)	10095 (4.2%)	8892 (1.6%)		
	Race: Other/Unknown	37733 (15.4%)	3628 (2.2%)	1822 (3.1%)	187 (0.9%)	1330 (5.2%)	41304 (10.9%)	13254 (5.5%)	13897 (2.5%)		
Defendant's age, mean (SD)		36.3 (12.2)	32.8 (10.0)	34.0 (11.1)	39.5 (13.1)	37.9 (12.4)	33.8 (9.1)	39.9 (12.3)	33.5 (9.9)	<0.001	ANOVA
Plea Bargain		223971 (94.1%)	154207 (91.8%)	55521 (93.6%)	20123 (93.2%)	24194 (94.7%)	376124 (99.0%)	229944 (94.8%)	519131 (93.8%)	<0.001	Pearson's χ^2
Judge's Gender: Female, mean (SD)		14.0 (10.9)	16.0 (11.2)	12.6 (10.4)	18.8 (11.7)	15.3 (11.1)	20.9 (10.6)	16.1 (11.0)	15.9 (11.1)	<0.001	ANOVA
Judge's Race: African American, mean (SD)		7.6 (7.4)	8.2 (8.3)	7.7 (7.4)	7.8 (8.2)	6.9 (7.4)	6.2 (5.9)	8.6 (7.5)	7.2 (7.3)	<0.001	ANOVA
Judge's Race: Hispanic, mean (SD)		6.6 (13.5)	6.5 (15.4)	3.9 (9.6)	6.7 (13.1)	9.5 (17.3)	19.8 (16.7)	6.0 (12.7)	10.3 (18.0)	<0.001	ANOVA
Judge's Race: Native American, mean (SD)		0.2 (1.4)	0.1 (1.8)	0.1 (1.5)	0.1 (1.4)	0.2 (2.0)	0.1 (0.7)	0.1 (1.6)	0.1 (1.2)	<0.001	ANOVA
Judge's Race: Pacific Islander, mean (SD)		0.1 (1.5)	0.1 (1.0)	0.2 (1.9)	0.1 (0.9)	0.0 (0.9)	0.0 (0.6)	0.1 (1.1)	0.1 (1.3)	<0.001	ANOVA
Judge's Race: Asian American, mean (SD)		()	()	()	()	()	()	()	()	<0.001	ANOVA
Judge's Age, mean (SD)		64.3 (3.9)	65.2 (3.7)	63.7 (3.6)	66.7 (3.3)	64.7 (4.1)	64.0 (3.5)	64.8 (3.6)	64.3 (3.9)	<0.001	ANOVA

Note– This table compares descriptive statistics for the variables used in the analysis across offense categories. For each variable, the test is specified alongside a p-value.

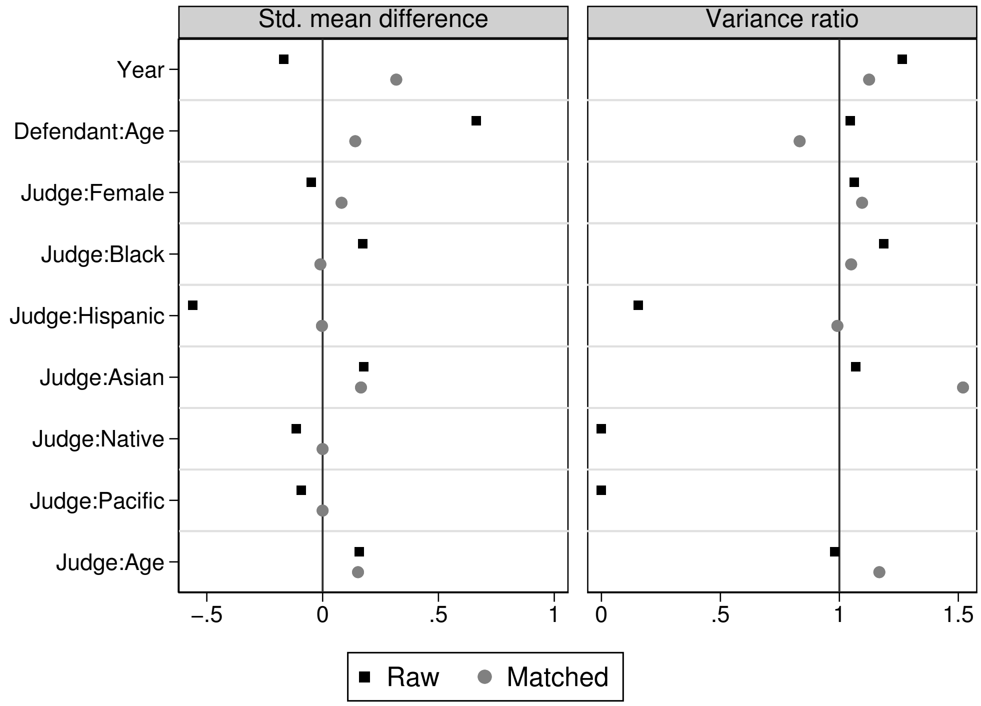
Table C3:
DESCRIPTIVES: OFFENSE TYPE

Factor	Level	Non-Exonerated	Exonerated	p-value	Test
Observations (n)		1695432	107		
Offence type	Murder	2149 (0.1%)	4 (3.7%)	<0.001	Pearson's χ^2
	Manslaughter	1433 (0.1%)	1 (0.9%)		
	Kidnapping/Hostage Taking	1480 (0.1%)	0 (0.0%)		
	Sexual Abuse	8156 (0.5%)	2 (1.9%)		
	Assault	14552 (0.9%)	4 (3.7%)		
	Robbery	35163 (2.1%)	6 (5.6%)		
	Arson	1916 (0.1%)	4 (3.7%)		
	Drugs - Trafficking, Manufacturing, and Importing	551709 (32.6%)	24 (22.4%)		
	Drugs - Communication Facilities	9985 (0.6%)	0 (0.0%)		
	Drugs: - Simple Possession	27052 (1.6%)	0 (0.0%)		
	Firearms (Includes Firearms Use, Possession, and Trafficking)	148466 (8.8%)	5 (4.7%)		
	Burglary/Breaking and Entering	1370 (0.1%)	0 (0.0%)		
	Auto Theft	3294 (0.2%)	0 (0.0%)		
	Larceny	52947 (3.1%)	0 (0.0%)		
	Fraud	177599 (10.5%)	32 (29.9%)		
	Embezzlement	17216 (1.0%)	0 (0.0%)		
	Forgery/Counterfeiting	24831 (1.5%)	0 (0.0%)		
	Bribery	5909 (0.3%)	3 (2.8%)		
	Tax Offenses	17199 (1.0%)	7 (6.5%)		
	Money Laundering	21995 (1.3%)	0 (0.0%)		
	Racketeering /Extortion	17866 (1.1%)	0 (0.0%)		
	Gambling/Lottery	3229 (0.2%)	0 (0.0%)		
	Civil Rights Offenses	2169 (0.1%)	0 (0.0%)		
	Immigration	380044 (22.5%)	4 (3.7%)		
	Pornography/Prostitution	12803 (0.8%)	0 (0.0%)		
	Prison Offenses	9333 (0.6%)	0 (0.0%)		
	Administration of Justice	25584 (1.5%)	3 (2.8%)		
	Environmental, Game, Fish, and Wildlife Offenses	4369 (0.3%)	0 (0.0%)		
	National Defense Offenses	1305 (0.1%)	2 (1.9%)		
	Antitrust Violations	500 (<1%)	0 (0.0%)		
	Food and Drug Offense	1947 (0.1%)	0 (0.0%)		
	Traffic Violations and Other Offenses	37889 (2.2%)	0 (0.0%)		
	Child Pornography	13442 (0.8%)	0 (0.0%)		
	Obscenity	185 (<1%)	0 (0.0%)		
	Prostitution	1079 (0.1%)	0 (0.0%)		
	Other/Unknown	54357 (3.2%)	6 (5.6%)		

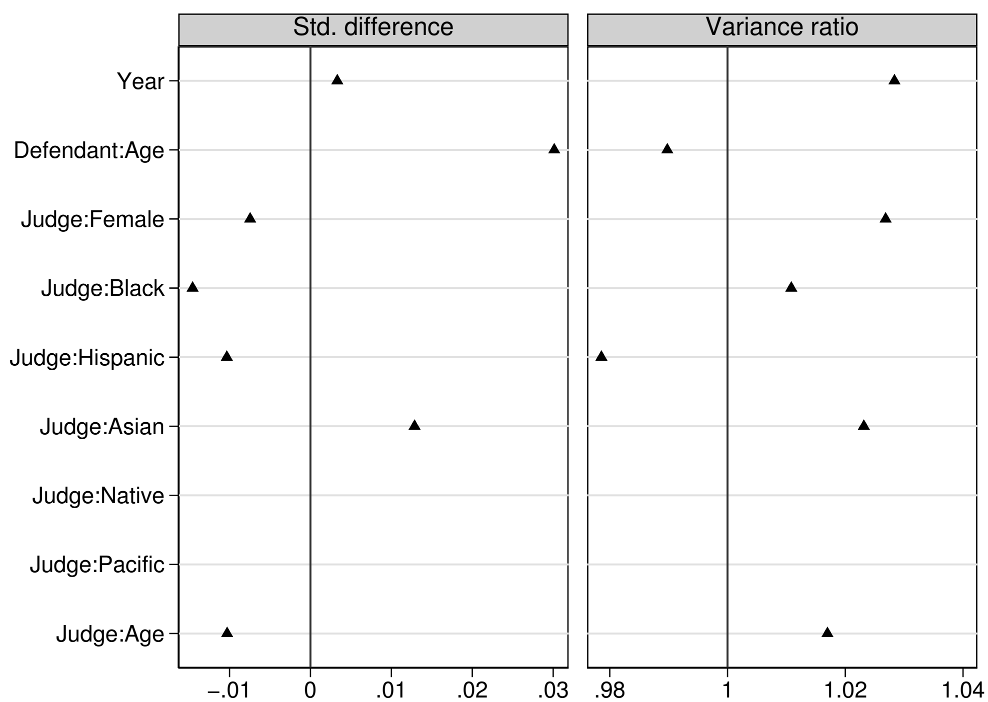
NOTE.– This table compares the detailed offense types between the two groups: Exonerated and Non-Exonerated. For each variable, a test is specified alongside a p-value.

Figure C3: Balancing and Common Support

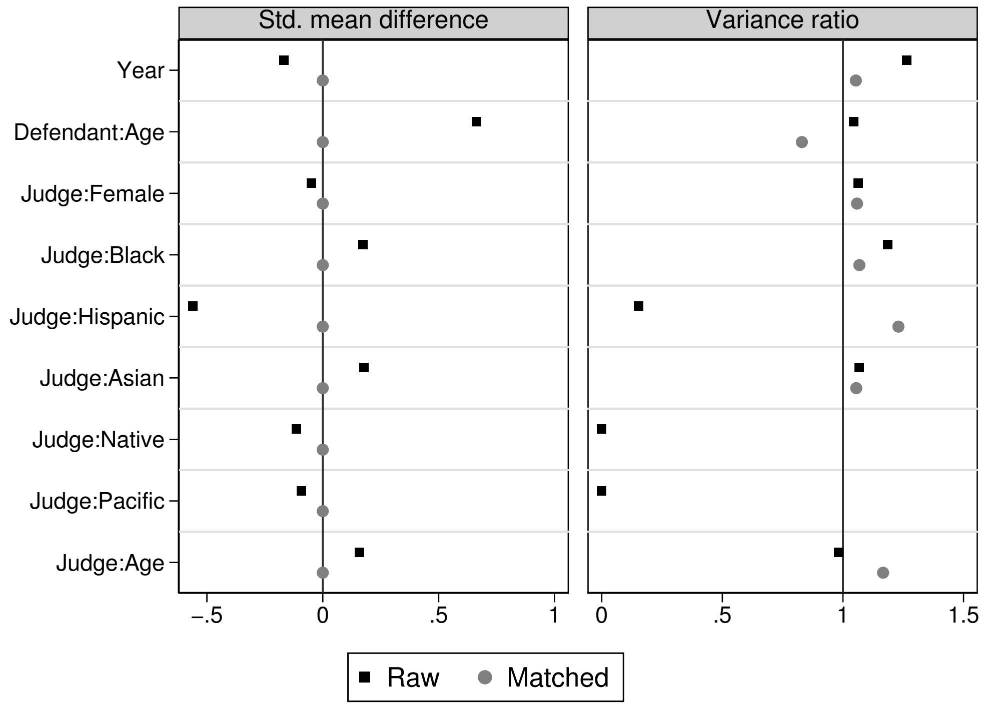
(a) Balancing (No EB)



(b) Common Support (No EB)



(c) Balancing (EB)



(d) Common Support (EB)

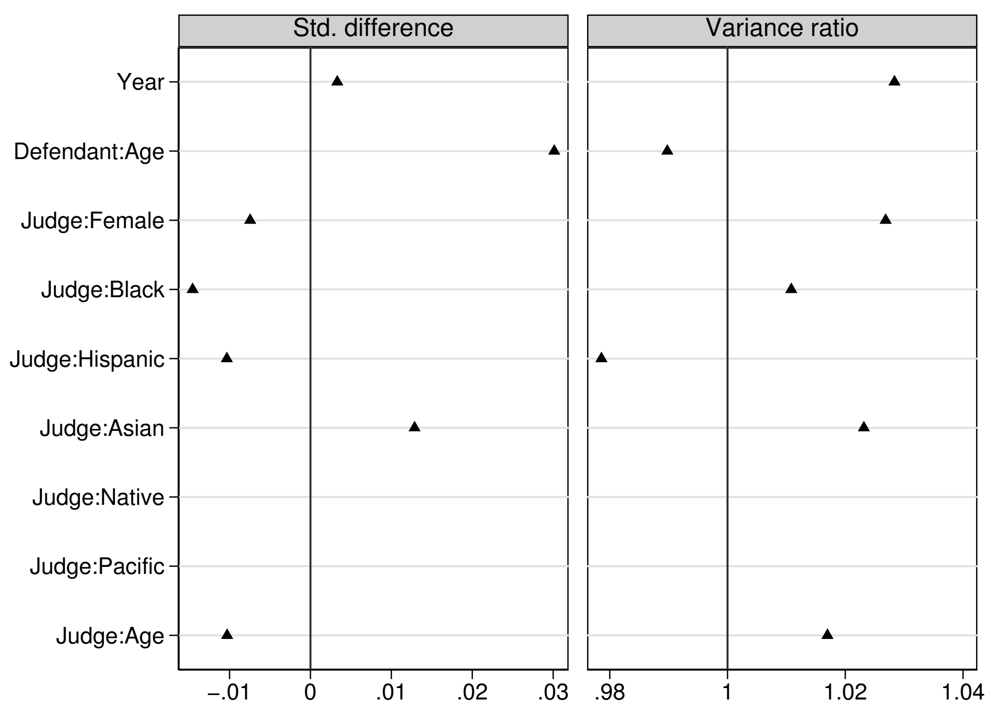


Table C4:
INDICATOR DEFINITIONS

Indicator	Definition (NRE database)	Impact on index
Co-defendant confessed	A co-defendant of the exoneree, or a person who might have been charged as a codefendant, gave a pre-conviction confession that also implicated the exoneree. This does not apply if the only confession is a guilty plea from the co-defendant.	+1 if applies, 0 otherwise.
Child sex abuse hysteria or satanic ritual hysteria	The exoneree was convicted of child sex abuse as part of a wave of child sex abuse prosecutions in the 1980s and 1990s based on aggressive and suggestive interviews of children who were thought to be victims. These cases generally included bizarre and implausible claims by the supposed victims, frequently featuring satanic rituals.'	+1 if applies, 0 otherwise.
False confession	The exoneree made a statement to law enforcement at any point during the proceedings which was interpreted or presented by law enforcement as an admission of participation in or presence at the crime, even if the statement was not presented at trial. A statement is not a confession if it was made to someone other than law enforcement. A statement that is not at odds with the defense is not a confession. A guilty plea is not a confession.'	+1 if applies, 0 otherwise.
False or Misleading Forensic Evidence	The exoneree's conviction was based at least in part on forensic information that was (1) caused by errors in forensic testing, (2) based on unreliable or unproven methods, (3) expressed with exaggerated and misleading confidence, or (4) fraudulent.	+1 it applies, 0 otherwise.
Jailhouse informant	A witness who was incarcerated with the exoneree testified or reported that the exoneree confessed to him or her.	+1 if applies, 0 otherwise.
Mistaken Witness Identification	At least one eyewitness affirmatively and mistakenly said that he or she saw the exoneree commit the crime, or saw the exoneree under circumstances that suggest that the exoneree participated in the crime (e.g. witness claims he saw the exoneree flee from the scene).	+1 if applies, 0 otherwise.
Official Misconduct	Police, prosecutors, or other government officials significantly abused their authority or the judicial process in a manner that contributed to the exoneree's conviction.	+1 if applies, 0 otherwise.
Perjury or False Accusation	A person other than the exoneree committed perjury by making a false statement under oath that incriminated the exoneree in the crime for which the exoneree was later exonerated, or made a similar unsworn statement that would have been perjury if made under oath.	+1 if applies, 0 otherwise.
No Crime	The exoneree was convicted of a crime that did not occur, either because an accident, act of self-defense, or a suicide was mistaken for a crime, or the exoneree was accused of a fabricated crime that never happened.	0 if applies, +1 otherwise.
Inadequate Legal Defense	The exonoree's lawyer at trial provided obviously and grossly inadequate representation.	+1 if applies, 0 otherwise.
Shaken baby syndrome	The exoneree was convicted of injuring or killing an infant by violent shaking, based on a medical diagnosis that is now highly controversial.	+1 if applies, 0 otherwise.

NOTE.– This table provides the definitions for the indicators used for the construction of the index, as taken from the NRE database's codebook. The last column specifies the impact of the indicator on the index.

Table C5: Descriptives: Federal vs. State exonerees

Factor	Level	State	Federal	p-value	Test
Observations (n)		2355	113		
Evidence Level Index, median (IQR)		3.0 (2.0, 4.0)	2.0 (2.0, 3.0)	<0.001	Wilcoxon rank-sum
Co-defendant confessed		304 (12.9%)	19 (16.8%)	0.23	Pearson's χ^2
Jailhouse informant		155 (6.6%)	9 (8.0%)	0.56	Pearson's χ^2
No Crime		860 (36.5%)	50 (44.2%)	0.096	Pearson's χ^2
Mistaken Witness Identification		693 (29.4%)	7 (6.2%)	<0.001	Pearson's χ^2
False Confession		297 (12.6%)	2 (1.8%)	<0.001	Pearson's χ^2
Perjury or False Accusation		1367 (58.0%)	70 (61.9%)	0.41	Pearson's χ^2
False or Misleading Forensic Evidence		559 (23.7%)	6 (5.3%)	<0.001	Pearson's χ^2
Official Misconduct		1253 (53.2%)	69 (61.1%)	0.10	Pearson's χ^2
Inadequate Legal Defense		623 (26.5%)	12 (10.6%)	<0.001	Pearson's χ^2
Shaken Baby Syndrome		17 (0.7%)	0 (0.0%)	0.36	Pearson's χ^2
Child Sex abuse Hysteria		58 (2.5%)	0 (0.0%)	0.091	Pearson's χ^2
Total Sentence (totsent), mean (SD)		497.7 (484.4)	97.9 (231.4)	<0.001	Two sample t-test
Female	Male	2148 (91.2%)	100 (88.5%)	0.32	Pearson's χ^2
	Female	207 (8.8%)	13 (11.5%)		
Race	Race: White	857 (36.4%)	66 (58.4%)	<0.001	Pearson's χ^2
	Race: Black	1180 (50.1%)	22 (19.5%)		
	Race: Hispanic	272 (11.5%)	15 (13.3%)		
	Race: Native	17 (0.7%)	1 (0.9%)		
	Race: Asian	18 (0.8%)	5 (4.4%)		
	Race: Other/Unknown	11 (0.5%)	4 (3.5%)		
Age, mean (SD)		29.9 (10.2)	41.9 (11.0)	<0.001	Two sample t-test
Plea Bargain		478 (20.3%)	16 (14.2%)	0.11	Pearson's χ^2
Offense type	Accessory to Murder	2 (0.1%)	0 (0.0%)	<0.001	Pearson's χ^2
	Arson	19 (0.8%)	4 (3.5%)		
	Assault	92 (3.9%)	4 (3.5%)		
	Attempt, Violent	3 (0.1%)	1 (0.9%)		
	Attempted Murder	54 (2.3%)	0 (0.0%)		
	Bribery	2 (0.1%)	3 (2.7%)		
	Burglary/Unlawful Entry	15 (0.6%)	0 (0.0%)		
	Child Abuse	9 (0.4%)	0 (0.0%)		
	Child Sex Abuse	272 (11.5%)	0 (0.0%)		
	Dependent Adult Abuse	1 (<1%)	0 (0.0%)		
	Destruction of Property	2 (0.1%)	0 (0.0%)		
	Drug Possession or Sale	298 (12.7%)	26 (23.0%)		
	Failure to Pay Child Support	2 (0.1%)	0 (0.0%)		
	Filing a False Report	1 (<1%)	0 (0.0%)		
	Forgery	3 (0.1%)	0 (0.0%)		
	Fraud	8 (0.3%)	32 (28.3%)		
	Harassment	1 (<1%)	0 (0.0%)		
	Immigration	0 (0.0%)	4 (3.5%)		
	Kidnapping	15 (0.6%)	0 (0.0%)		
	Manslaughter	46 (2.0%)	1 (0.9%)		
	Menacing	2 (0.1%)	0 (0.0%)		
	Military Justice Offense	0 (0.0%)	1 (0.9%)		
	Murder	937 (39.8%)	4 (3.5%)		
	Obstruction of Justice	0 (0.0%)	1 (0.9%)		
	Official Misconduct	1 (<1%)	1 (0.9%)		
	Other	9 (0.4%)	0 (0.0%)		
	Other Nonviolent Felony	13 (0.6%)	0 (0.0%)		
	Other Violent Felony	6 (0.3%)	3 (2.7%)		
	Perjury	3 (0.1%)	2 (1.8%)		
	Possession of Stolen Property	3 (0.1%)	0 (0.0%)		
	Robbery	122 (5.2%)	6 (5.3%)		
	Sex Offender Registration	26 (1.1%)	0 (0.0%)		
	Sexual Assault	322 (13.7%)	5 (4.4%)		
	Solicitation	2 (0.1%)	0 (0.0%)		
	Stalking	2 (0.1%)	0 (0.0%)		
	Supporting Terrorism	0 (0.0%)	2 (1.8%)		
	Tax Evasion/Fraud	0 (0.0%)	7 (6.2%)		
	Theft	19 (0.8%)	0 (0.0%)		
	Threats	3 (0.1%)	1 (0.9%)		
	Traffic Offense	6 (0.3%)	0 (0.0%)		
	Weapon Possession or Sale	34 (1.4%)	5 (4.4%)		
Year of Conviction, mean (SD)		1996.9 (10.8)	2003.4 (8.3)	<0.001	Two sample t-test
Year of Crime, mean (SD)		1995.3 (11.1)	2000.3 (8.9)	<0.001	Two sample t-test

Note.– This table compares the variables used in the analysis of Part II between exonerated defendants in state and federal courts (including 6 cases in military federal courts). For each variable, the test is specified alongside a p-value.

Appendix D: Endogenous Treatment Model

Endogenous treatment models ("ETM") apply a correction that accounts for the impact of unobservables that simultaneously affect the treatment assignment and the outcome (see e.g. Tsiboe, Zereyesus, and Osei, 2016; Hartje and Hübler, 2017; Gutmann, Pfaff, and Voigt, 2017; Malo, Martín-Román, and Moral, 2018).[48] This includes an estimation of a primary and a secondary equation. The primary equation is the same as in the OLS, i.e.:

$$totsent = \beta_0 + \beta_1 Exonerated + \beta_x`X' + \epsilon_1 \tag{B4}$$

The secondary equation addresses the assignment of the treatment "*Exonerated*", which is assumed to be determined by a latent variable $Exonerated^*$, as follows:

$$Exonerated = \begin{cases} 1 & \text{if } Exonerated^* \equiv \gamma_0 + \gamma_Z`Z' + \epsilon_2 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B5}$$

Where Z' is a vector of predictors which may (or may not) overlap with X' (from the primary equation); and the error terms ϵ_1 and ϵ_2 are allowed to be correlated.[49] Table D1 consists of three main parts: an upper part with coefficients in the primary equation; a middle part with coefficients in the secondary equation; and a lower part showing the correlation between the error terms.[50] Column (1) excludes controls and uses only the plea-bargain dummy to predict the treatment assignment. Column (2) adds the plea-bargain dummy to the primary equation. Column (3) adds the attributes of defendants and of the (mean) judge.[51] Column (4) adds offense-type dummies. Column (5) includes all controls.[52]

As in the OLS, the coefficient of *Exonerated* is positively significant without controls (column (1)), but changes signs once the plea-bargain dummy is added to the primary equation. The coefficient is then negatively significant at the 5% level in columns (2)-(4), but insignificant in column (5). When controls are included (columns (2)-(5)), the correlation

Table D1:
ENDOGENOUS TREATMENT MODEL

	(1)	(2)	(3)	(4)	(5)
Primary equation. Dependent Variable = totsent					
Exonerated	254.48*** (9.47)	-58.65** (23.04)	-52.03** (23.27)	-51.27** (22.02)	-35.52 (27.60)
Plea Bargain		-132.68*** (0.88)	-121.09*** (1.01)	-110.88*** (0.97)	-109.49*** (0.98)
DEFENDANT'S ATTRIBUTES					
Gender: female			-29.94*** (0.28)	-18.91*** (0.28)	-18.82*** (0.28)
Race: Black			28.48*** (0.34)	23.07*** (0.32)	21.40*** (0.33)
Race: Hispanic			-2.38*** (0.28)	0.19 (0.31)	1.88*** (0.32)
Race: Native			4.09*** (0.92)	-5.55*** (0.96)	0.04 (1.02)
Race: Asian			-8.03*** (0.78)	-1.20 (0.75)	-2.05*** (0.76)
Race: Other/Unknown			-5.10*** (0.68)	7.14*** (0.67)	8.25*** (0.69)
Defendant's age			1.96*** (0.05)	2.13*** (0.05)	2.00*** (0.05)
Defendant's age × Defendant's age			-0.03*** (0.00)	-0.02*** (0.00)	-0.02*** (0.00)
JUDGE'S ATTRIBUTES (COURT-BY-YEAR MEAN)					
Judge's Gender: Female			-0.32*** (0.01)	-0.29*** (0.01)	0.06*** (0.02)
Judge's Race: African American			0.07*** (0.02)	0.12*** (0.02)	0.11*** (0.03)
Judge's Race: Hispanic			-0.07*** (0.01)	-0.06*** (0.01)	-0.25*** (0.03)
Judge's Race: Asian American			-0.35*** (0.04)	-0.27*** (0.04)	-0.45*** (0.08)
Judge's Race: Native American			-0.49*** (0.08)	0.01 (0.07)	0.07 (0.14)
Judge's Race: Pacific Islander			0.83*** (0.09)	0.29*** (0.08)	-1.07*** (0.28)
Judge's Age			3.44*** (0.66)	7.01*** (0.63)	5.78*** (0.72)
Judge's Age × Judge's Age			-0.02*** (0.01)	-0.05*** (0.00)	-0.04*** (0.01)
Secondary equation. Dependent Variable = Exonerated					
Plea Bargain	-1.10*** (0.14)	-1.20*** (0.06)	-1.22*** (0.07)	-1.23*** (0.07)	-1.23*** (0.07)
DEFENDANT'S ATTRIBUTES					
Gender: female			0.00 (0.09)	-0.04 (0.09)	-0.04 (0.09)
Race: Black			-0.29*** (0.08)	-0.23*** (0.08)	-0.23*** (0.08)
Race: Hispanic			-0.17** (0.09)	-0.08 (0.08)	-0.07 (0.08)
Race: Native			-0.31 (0.29)	-0.62*** (0.24)	-0.62*** (0.24)
Race: Asian			0.12 (0.15)	0.12 (0.16)	0.12 (0.16)
Race: Other/Unknown			-0.03 (0.15)	0.00 (0.15)	0.00 (0.15)
Defendant's age			0.04** (0.02)	0.04** (0.02)	0.04** (0.02)
Defendant's age × Defendant's age			-0.00* (0.00)	-0.00* (0.00)	-0.00* (0.00)
JUDGE'S ATTRIBUTES					
Judge's Gender: Female			0.01*** (0.00)	0.01** (0.00)	0.01** (0.00)
Judge's Race: African American			0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Judge's Race: Hispanic			-0.01** (0.00)	-0.01* (0.00)	-0.01* (0.00)
Judge's Race: Asian American			0.02*** (0.01)	0.02** (0.01)	0.02** (0.01)
Judge's Race: Native American			-0.61*** (0.02)	-0.62*** (0.02)	-0.68*** (0.03)
Judge's Race: Pacific Islander			-0.28*** (0.02)	-0.28*** (0.02)	-0.30*** (0.02)
Judge's Age			0.04 (0.14)	0.04 (0.15)	0.04 (0.15)
Judge's Age × Judge's Age			-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Correlation ($\varepsilon_1, \varepsilon_2$)	-0.65*** (0.11)	0.00* (0.00)	0.01*** (0.00)	0.01 (0.01)	0.01 (0.01)
Offense type dummies	No	No	No	Yes	Yes
Sentence year dummies	No	No	No	No	Yes
Court dummies	No	No	No	No	Yes
Observations	1,685,378	1,685,378	944,579	941,691	941,691

NOTE.– This table presents the results of the Extended Regression Model using Stata's "eregress, entreat ( )" command. The upper part of the table presents the coefficients of the main regression, with sentence length as the dependent variable. The second part of the table presents the coefficient in the auxiliary regression, with *Exonerated* as the dependent variable. The third part presents the correlation of the error terms in the two equations. The fourth part reports on whether dummies are included in the main regression. Year and court dummies are excluded from the secondary regression in all columns, given that there are only 107 treated units. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

between the error terms is weak (1% or less) and in some specifications (columns (4)-(5)) insignificant.[53] Note that an insignificant correlation implies that OLS is superior to the ETM (i.e. equations are better off estimated separately). As the ETM does not change the results much, the OLS results seem to be robust and are not driven by omitted variables.