



DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

**Better that X guilty persons escape than
that one innocent suffer**

Matteo Rizzoli, Margherita Saraceno
No. 168 – July 2009

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
<http://dipeco.economia.unimib.it>

Better that X guilty persons escape than that one innocent suffer*

Matteo Rizzolli[†] and Margherita Saraceno[‡]

July 21, 2009

Abstract

The principle that it is better to let some guilty individuals be set free than to mistakenly convict an innocent person is generally shared by legal scholars, judges and lawmakers of modern societies. The paper shows why this common trait of criminal procedure is also efficient. It extends the standard Polinsky and Shavell (2007) model of deterrence and shows that when the costs of convictions are positive, and guilty individuals are more likely to be convicted than innocent individuals it is always efficient to minimize the number of wrongful convictions, while a more than minimal amount of wrongful acquittals may be optimal.

Keywords: Type I errors, Type II errors, evidence, optimal underdeterrence, Blackstone Pareto distribution, optimal screening

JEL classification: K14, K41, K42.

*The title is an adaptation of a well-known maxim of William Blackstone (1766). The authors thank Libor Dušek, Nuno Garoupa, Luca Stanca, Avi Tabbach, and participants of the trilateral workshop Tel Aviv-Toronto-Siena on law and economics and of the seminar at the University of Milano - Bicocca for useful discussion. All errors remain ours.

[†]University of Milano-Bicocca, Dept of Economics. Piazza dell'Ateneo Nuovo, 1 (U6) 20126 Milano - Italy E-mail: matteo.rizzolli@unimib.it

[‡]University of Milano-Bicocca, Dept of Statistics. Via Bicocca degli Arcimboldi, 8 (U7) 20126 Milan - Italy. E-mail: margherita.saraceno@unimib.it

1 Introduction

A cornerstone of criminal procedures in modern democracies (and also in advanced societies of the past) is the robust protection granted to defendants through several procedural safeguards. Most of these mechanisms protect the innocent from mistaken convictions at the cost of allowing some guilty defendants to be set free. As Blackstone puts it: *it is better that ten guilty persons escape, than that one innocent suffer* (1766). More generally, it can be argued that it is better that “ $X > 1$ ” guilty persons escape punishment than one innocent person suffers it (and this is the reason why the Blackstone citation used for the title of the paper has been emended). From now on the ratio between guilty persons acquitted (type-II errors) and innocent persons convicted (type-I errors) will be referred to as the *Blackstone errors ratio*.

Incontrovertible as it seems, this characteristic of criminal procedure meets with little analysis from law and economics scholars. Standard models of deterrence consider both types of errors and show that they are both detrimental to deterrence (See Png, 1986; Polinsky and Shavell, 2007). They come to this conclusion by observing that acquittals of guilty individuals make crime more favorable as they lower the probability of conviction and thus dilute deterrence, but also convictions of innocents make crime more convenient by lowering the relative benefits of staying honest. According to these models, and under a number of other assumptions¹, the social planner should treat both types of errors as equal evils to deterrence. But if they are both equally bad in terms of deterrence, why should we care about minimizing convictions of innocents even at the cost of allowing many acquittals of guilty individuals? In other words: is the Blackstone errors ratio inefficient or does the standard model overlooks something?

This paper offers an extension of the standard model of optimal deterrence that reconciles the efficiency goals of the judicial system with the Blackstone errors ratio. The paper is organized as follows: section 2 introduces the Blackstone errors ratio and presents a brief overview of the literature. In section 3, the standard model of deterrence is introduced and duly articulated to account for the social welfare implications of both type-I and type-II errors. It is shown that the standard model explains the bias against type-I errors. Section 4 derives some policy implication and concludes.

¹Such as the fact that deterrence is the goal of criminal law, that individuals are rational expected utility maximizers and are neutral to risk.

2 Judicial errors in the literature

The trade-off between the two types of judicial errors has been known and discussed by lawyers and philosophers for a long time. Courts make recurrent mention of it and this seems to point at the case of a conscious and intentional, albeit not systematized, pursuit of a specific ratio of innocents convicted to guilty persons acquitted that is more favorable to innocents. How much more favorable? While every court and scholar would agree that it is desirable to reduce type-I errors, how many more type-II errors are we willing to tolerate in order to achieve this goal?

Every American student of law learns by heart Judge Blackstone's (1766) maxim that gives this paper its title. The United States Supreme Court has recalled the Blackstone's principle although it has never committed to such a precise number². Countless scholars have mentioned a precise number for this trade-off. However, as Volokh (1997) has pointed out, there is a great variety of opinions on what this number should be. Volokh finds mentions of the judicial errors' trade-off that date back to Genesis³ and historically vary at least between $X = 1000$ ⁴ and $X = 1$ ⁵. As seen, Blackstone asserts that the optimal X must be equal to 10. However this is a severe underestimation if compared to, for instance, Benjamin Franklin's figure of $X = 100$ ⁶ and

²The Supreme Court cited Blackstone in "Coffin v. U.S.", 156 U.S. 432 (1895). For direct mention of the trade-off see for instance "Herny v. United States" 61 U.S. 98 (1959): "It is better, so the Fourth Amendment teaches, that the guilty sometimes go free than that citizens be subject to easy arrest", or the concurrent opinion of Judge Harlan in "In re Winship" 397 U.S. 358 (1970) where he states: "I view the requirement of proof beyond a reasonable doubt in a criminal case as bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free".

³In the Genesis (18:23-32), God tells Moses he will save an entire city of guilty individuals if there are at least 10 innocents among them: "*Oh let not the Lord be angry, and I will speak yet but this once: Peradventure ten shall be found there. And he said, I will not destroy it for ten's sake*".

⁴Moses Maimonides, a Jewish Spanish legal theorist, interpreting the commandments of Exodus. Cited in Volokh (1997).

⁵The Justinian's Digest (48.19.5pr. Ulpianus 7 de off. procons.) remarks that a person ought not "to be condemned on suspicion; for it was preferable that the crime of a guilty man should go unpunished than an innocent man be condemned." Also cited in Volokh (1997).

⁶"it is better [one hundred] guilty Persons should escape than that one innocent Person should suffer". Letter from Benjamin Franklin to Benjamin Vaughan (Mar. 14, 1785), in Franklin and Smyth (1970) cited in Volokh (1997).

some other wildly inflated numbers mentioned in the literature⁷; but at the same time it looks pretty generous if compared with – for instance – Hale’s $X = 5$ ⁸ or Ayatollah Hossein Ali Montazeri’s $X = 1$ ⁹. Irony aside, the Blackstone errors ratio, in its extremely variegated declinations, expresses the principle that, for a number of reasons, it is better that the criminal procedure produces more mistakes against society (acquittal of the guilty) than against individuals (conviction of the innocent), under the assumption that the total number of mistakes, given a certain level of forensic technology, is irreducible below a certain threshold. Another way of making this point is to argue, as Posner (1999) does, that the costs of convicting an innocent far exceed the benefits of acquitting one more guilty individual. This may be due to a number of reasons: Hobbes (1660) for instance argues that wrongful convictions overturn the social contract; Craswell and Calfee (1986) argue that they bring about further costs in terms of uncertainty; and Kaplow and Shavell (1994) show how type-I errors lead to suboptimal activity levels.

Whatever the reason may be, there exists an enduring preference, even across different legal systems, for allowing the concern over the possible wrongful conviction of an innocent individual to prevail over the desire for the apprehension of a guilty one. Has Law and Economics explained this characteristic of criminal procedure? There are a number of papers that postulate the Posnerian assertion that type-I errors are far worse than type-II errors, and simply weigh the two errors differently in their functions of social costs (See Miceli, 1991; Lando, 2009). Harris (1970) extends Becker’s (1968) model to include the social costs of type-I errors. Ehrlich (1982) adds an independent term that measures the *costs of miscarriage of justice* to his function of the social costs of crime that society needs to minimize. All these models trigger a Blackstone errors ratio larger than 1 simply by attaching an ethical burden to type-I errors and thus by weighting differently the two errors in the social costs function. As shown in the next section, the model here presented shows how the optimal X is larger than 1 without resorting to different weights.

Some more recent literature tries to endogenize the Blackstone errors ratio. Hylton and Khanna (2007) develop a public-choice account of the

⁷See also Reiman and van den Haag (1990).

⁸Hale and Emlyn (1736) cited in May (1875).

⁹“In Islam, it is better if a guilty person escapes justice than that an innocent man receives punishment.” International news, Reuters, November 23, 1985. Cited in Volokh (1997).

pro-defendant mechanisms in criminal procedure that affects the Blackstone errors ratio as a set of safeguards against the prosecutor's potential misconduct. In their view the Blackstone errors ratio is the result of a second-best equilibrium achieved within the constraint of an irreducible inclination of prosecutors to over-convict defendants (for various public-choice reasons). Persson and Siven (2007) formulate a general equilibrium model of crime deterrence where the Blackstone errors ratio for capital punishment emerges endogenously as a result of a median voter mechanism applied to courts. Both models depart from the standard model of deterrence. The model presented here derives justifications for the Blackstone errors ratio straight from the standard model of public enforcement of law as set out by Polinsky and Shavell (2007). It also extends their framework to model the behavior of the adjudicative authority and shows how the commonly observed reluctance towards the production of type-I errors (even at the cost of producing many type-II errors) is perfectly explained in terms of optimal deterrence.

3 The Model

As in all criminal procedures of modern democracies the defendant is presumed innocent until proven guilty. To put it in statistical terms, the innocence of the defendant is the null hypothesis which the courts are presented with and which prosecutors try to confute¹⁰. Courts, hearing the opposing arguments of the parties, decide on the basis of the evidence presented. An incorrect rejection of the null hypothesis corresponds to the conviction of an innocent person and implies a type-I error whereas an incorrect acceptance of the null hypothesis amounts to the acquittal of a guilty person and is described as a type-II error¹¹.

¹⁰For various reasons it is more common to find in the literature the null hypothesis set on the option that the defendant is guilty. See for instance Polinsky and Shavell (2000); Fon and Schaefer (2007). Hence they obtain definitions of type-I and type-II errors that mirror the ones here described.

¹¹On the use of statistics in evidence see also Yilankaya (2002); Posner (1999); McCauliff (1982); Daughety and Reinganum (2000); Froeb and Kobayashi (2001). For a critical perspective see Allen and Pardo (2007).

3.1 Errors' production

Let $e \in [0, \infty]$ be the prosecutor's differential ability to convince the court of the defendant's guilt. The capacity of the prosecutor to prove guilt varies with each accused as individuals have different abilities to confute the allegations of guilt before the court. These differential abilities depend upon, *inter alia*, i) the capacity to afford good lawyers (which is independent of either actual innocence or guilt) and ii) the ease with which exculpatory evidence can be produced (which in contrast is dependent upon innocence/guilt). It is reasonable to assume that, on average, the prosecutor's ability to prove the guilt of innocent persons is low, whereas it is relatively high in respect of guilty individuals. Thus, let us define $I(e)$ as the positive, continuous and differentiable cumulative function of e for innocent individuals. Furthermore, let us define $G(e)$ as the positive, continuous and differentiable cumulative function of e for guilty individuals. $G(e)$ has first-order stochastic dominance over $I(e)$.

In order to obtain a conviction, a certain threshold of e must be overcome. Suppose the court rejects the null hypothesis when the prosecutor reaches an $e > \tilde{e}$. The threshold \tilde{e} thus becomes a measure of the burden of evidence required by the legal procedure in order to convict the defendant, as the ability of the prosecutor is necessarily linked with the quantity and quality of proofs that must be accumulated in order to reach a conviction¹². The probability of convicting an innocent person is thus $1 - I(\tilde{e})$. The probability of convicting a criminal is $1 - G(\tilde{e})$.

In figure 1, note that when the burden of evidence (and thus the ability of the prosecutor) necessary to obtain a conviction increases ($e_{Low} \rightarrow \tilde{e}_{High}$), the probability of conviction both for guilty and for innocent individuals decrease as $1 - I(\tilde{e}_{High}) < 1 - I(\tilde{e}_{Low})$ and $1 - G(\tilde{e}_{High}) < 1 - G(\tilde{e}_{Low})$. The probability of convicting an innocent individual $1 - I(\tilde{e})$ is the probability of type-I error, and the probability of acquitting a guilty individual $G(\tilde{e})$ represents the probability of type-II error. Thus when the burden of evidence increases, the probability of type-II error increases while the probability of type-I error decreases.

Other changes may happen when better forensic technology is introduced

¹²An alternative way of modelling the production of errors is to simply assume e to be the incriminating proofs for the defendant. The prosecutor can gain more proofs against a guilty defendant than against an innocent one. $G(e)$ and $I(e)$ are the cumulative functions of proofs against the guilty and the innocent respectively.

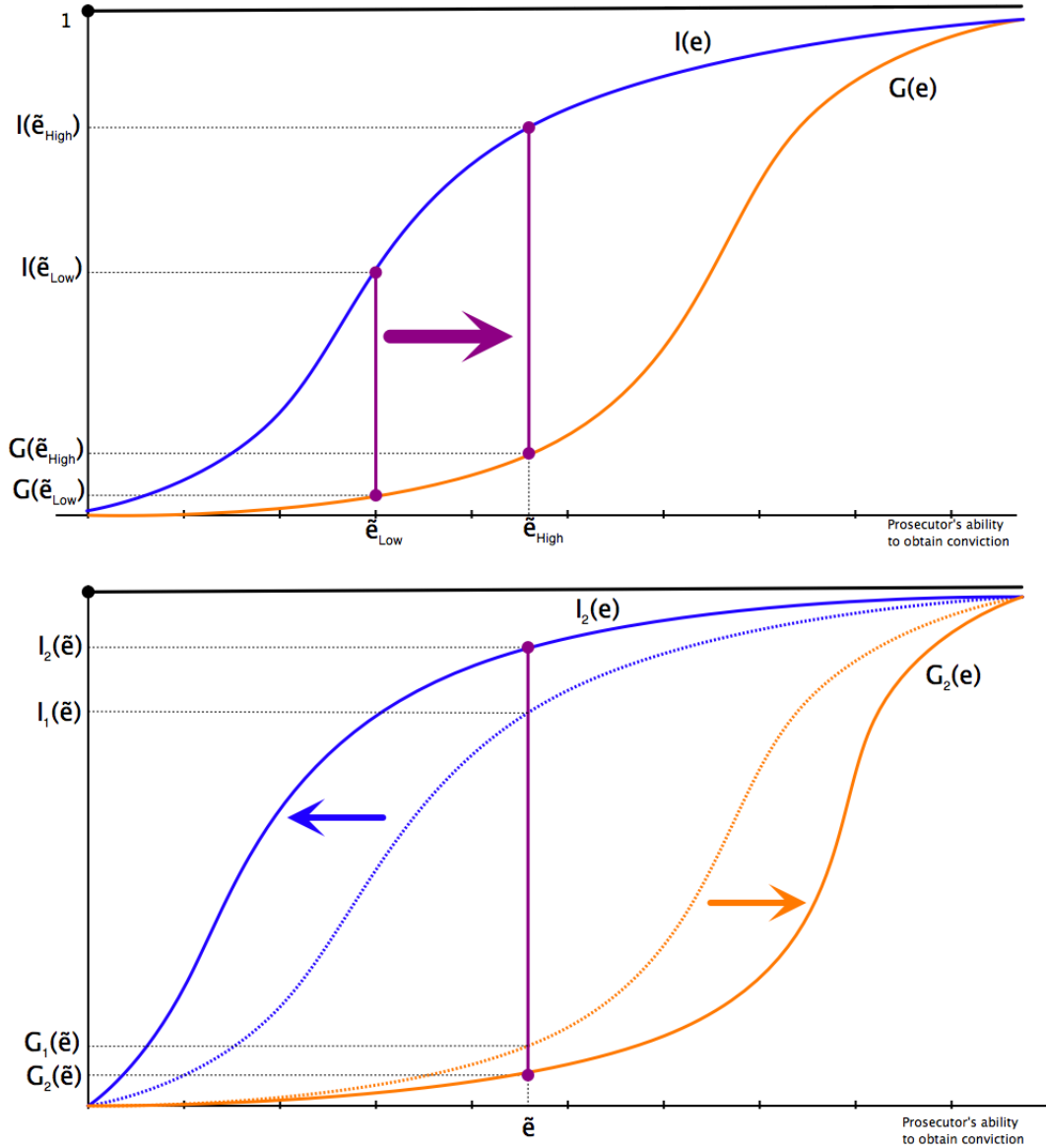


Figure 1: Cumulative distributions of the prosecutor's ability to prove guiltiness, for guilty and innocent persons. The figure above describes what happens to judicial errors when an higher burden of evidence is required. The figure below describes the effects of an improvement in forensic technology.

that allows the prosecutor to collect better evidence of guilt and thus helps the court to better distinguish guilty individuals from innocent ones. For a given burden of evidence (\tilde{e}) required by the procedure, better forensic technology usually improves the ability of the prosecutor to produce incriminating evidence for the guilty and reduces it for the innocent. This implies that the distributions of the differential abilities are less dispersed and more distinguishable (See figure 1, below). Note that improved forensic technology on average reduces both the probability of type-I error as $1 - I_2(\tilde{e}) < 1 - I_1(\tilde{e})$ and the probability of type-II error as $G_2(\tilde{e}) < G_1(\tilde{e})$.

3.2 Standard of evidence and technology

The paper distinguishes between the long-term, where changes in the legal procedures, standard of proofs and forensic technology may happen and affect the prosecutor's ability needed to reach a conviction, and the short-term where procedures and technology are fixed.

Arguably a modern lawmaker should aspire to obtain a legal framework that produces zero errors of any type. In fact, as Harris (1970) shows, in the long term social preferences may favor different regimes and the political process may build on one extreme some "law and order" or "zero tolerance" type of equilibrium compatible with a high number of type-I errors, or at the other extreme a very radical approach to civil liberties that imposes a high number of type-II errors. Depending on these long-term trends, a society sets (by statute or case law) the standard of evidence necessary for a court to convict. The court thus in the short term faces an exogenously determined \tilde{e} .

Of course in the long term forensic technology also improves. Knowledge and technical discoveries can be thought of as irreversible so the distance between the two distribution necessarily widens in time. In the short term courts may voluntarily forgo the use of more accurate technology in order to contain costs. Further, statutes and procedural rules can set limits and constraints on forensic technologies which can be used by the courts and prosecutors¹³.

As said above, in the short term, which is the main horizon of this analysis, the set of rules, the standard of proof and the forensic technologies are given. In fact, changes in the standard of evidence needed to reach a conviction, in

¹³For instance telephone tapping may be limited in order to protect individuals' privacy

the training of prosecutors and in forensic technologies are usually quite slow. These technological and legal constraints imply a lower bound on errors: in the short term the court system generates at best a given probability of type-I error equal to $\varepsilon_{1\min} \in [0, 1)$, and a minimum probability of type-II error $\varepsilon_{2\min} \in [0, 1)$. In other words, the best judge with the best technology can produce no fewer than $\varepsilon_{1\min}$ convictions of innocents and no less than $\varepsilon_{2\min}$ wrongful acquittals. Courts may produce higher error levels because they still retain a certain discretion. For instance, prosecutors and judges may be particularly unskilled, ignorant or ideologically biased. Or they might wilfully fail to make use of all the legal tools or accurate forensic technologies at their disposal. Therefore a very good court (that can potentially achieve the lower bound of errors) can deliberately produce more type-II errors than $\varepsilon_{2\min}$ ¹⁴. and an inadequate court (in terms of training or technology) may accidentally produce more type-I errors than $\varepsilon_{1\min}$.

3.3 The Blackstone errors ratio

Recall that $(1 - \varepsilon_{1\min}) > \varepsilon_{2\min}$, because of the assumption of first-order stochastic dominance. As defined above, the ratio of the two errors is $X = \frac{\varepsilon_2}{\varepsilon_1} = \frac{G(\cdot)}{(1-I(\cdot))}$.

In the short term, the Blackstone errors ratio X falls within the interval $\left[\varepsilon_{2\min}, \frac{1}{\varepsilon_{1\min}} \right]$, and the ratio that maximizes the accuracy of adjudication (Kaplow, 1994) is the one that minimizes the sum of the two errors: $\tilde{X} = \frac{\varepsilon_{2\min}}{\varepsilon_{1\min}}$. Note that \tilde{X} falls within the interval. In the next paragraph \tilde{X} (the Blackstone errors ratio that maximizes accuracy) will be compared with the optimal Blackstone errors ratio X^* : the one that instead maximizes the short-term social welfare.

3.4 Individual choice to commit crime

Define w as the individual gain from crime; w is known only by the person who decides whether to commit the crime.

¹⁴Levitt (1997) and Levitt (2009) show that prison population may change following short term policy needs (electoral cycles in the first case and state budget crises in the second case). Particularly, prison overcrowding or government budget balance requirements affect the courts' inclination to enhancing prison population. This evidence illustrates that courts can opportunely tune their aptitude for conviction.

Let q be the probability of detection, $0 < q \leq 1$, that is to say the probability that the police checks an individual (either innocent or guilty) in connection with a crime. In this model, q is exogenous as it simply depends on police efficiency or on nature (see further discussion in section 3.7). Once the police detects an individual, it brings him to court where he goes through the procedure that establishes whether he is to be convicted and sentenced.

Note that the probability of detection q and the probabilities of type-I and type-II errors are assumed to be independent. In fact the probabilities of error depend only on innocence/guilt and on institutional and technological settings.

Let c_p be the private cost of the sanction imposed after a verdict of guilt. This could be either a monetary loss, in case of a fine, or the disutility suffered because of imprisonment.

Individuals, who are assumed to be rational and risk-neutral, choose to stay innocent or to commit a crime by comparing the expected utility of staying innocent $[-q(1 - I(\tilde{e}))c_p]$ and the expected utility of committing the crime $[w - q(1 - G(\tilde{e}))c_p]$. Thus, the individual commits the crime if:

$$w - q(1 - G(\tilde{e}))c_p > -q(1 - I(\tilde{e}))c_p \quad (1)$$

that is to say if $w > \tilde{w}$ where

$$\tilde{w} = qc_p[I(\tilde{e}) - G(\tilde{e})] \quad (2)$$

A “deterrence effect” à la Becker (1968) can be identified. Note that \tilde{w} increases with both the probability of detection and the private costs of punishment ($\uparrow q, c_p \Rightarrow \uparrow \tilde{w}$). An “underdeterrence effect” of judicial acquittals of guilty individuals can also be observed: \tilde{w} decreases when the probability of mistaken acquittal increases ($\uparrow G(\tilde{e}) \Rightarrow \downarrow \tilde{w}$). Similarly a “compliance effect” of type-I errors ($\uparrow I(\tilde{e}) \Rightarrow \uparrow \tilde{w}$) is identifiable, because \tilde{w} increases when correct convictions increase (and thus type-I errors decrease). Finally, a “screening effect” can be identified. Let us define $(I(\tilde{e}) - G(\tilde{e})) = \Delta(\cdot)$ as the difference between the probabilities of being acquitted respectively when innocent and guilty. $\Delta(\cdot)$ represents the ability of the court to distinguish innocent from guilty individuals: the better the court can discriminate, the higher the advantages of staying honest ($\uparrow \Delta(\cdot) \Rightarrow \uparrow \tilde{w}$). Note that equation 2 can also be rewritten as: $I(\tilde{e}) - G(\tilde{e}) \geq \frac{w}{qc_p}$. From an individual perspective, the left term of the inequality represents the minimum *court’s screening ability* that induces an individual to stay honest. Note that the

court must guarantee some positive amount of screening ability in order to convince at least some people to stay innocent. This distance increases with w and decreases with the expected private cost of conviction (qc_p).

3.5 Social perspective

Criminal activities and the enforcement system imply three costs. First, the conviction of a defendant implies a private cost (c_p) for the individual: the disutility suffered because of the conviction. Second the conviction also implies a cost (c_s) to society: the cost of imprisonment and, more generically, the enforcement cost of both monetary and non-monetary penalties. Third, each crime carried out implies an harm (h) to society. For simplicity, w is assumed as a transfer from the victim to the criminal that cancels out in the social welfare function. However, this forced transfer produces h which is the socially relevant externality caused by the crime (see further discussion in section 3.7). Notice that the social planner cannot know the individual differential gain from crime for every person. Thus, the social planner knows that w is distributed according to a probability distribution z and a cumulative distribution Z . The population is normalized to 1. The social planner is assumed to be risk-neutral and to have the goal of maximizing social welfare. Thus, in the short term the problem of the social planner lies in defining the optimal ratio of judicial errors that minimizes the expected total costs from crime (TC).

$$TC = \underbrace{[1 - Z(\tilde{w})]h}_A + \underbrace{[1 - Z(\tilde{w})]q(1 - G(.)) [c_s + c_p]}_B + \underbrace{Z(\tilde{w})q(1 - I(.)) [c_s + c_p]}_C \quad (3)$$

Term A of equation (3) represents the expected social harm from crime; Term B represents the expected total (private and social) costs of convicting criminals; and Term C represents expected total costs of convicting innocent people. By looking at equation (3) some preliminary considerations can be made. Consider Term A : as already know from Png (1986) a drop of both errors lessens crime because it reinforces deterrence¹⁵. Thus in term A , both

¹⁵See the appendix (section A.1 and A.2) for the derivation of the result when applying the Pareto distribution.

errors equally decrease the social costs of crime $[h]$. However, terms B and C in the equation (3) show that there are differences in the way the two errors affect deterrence. In term B , type-II errors $[G(\cdot)]$ ambiguously affect the expected social costs of convicting criminals: fewer type-II errors decrease the number of criminals $[1 - Z(\tilde{w})]$ but at the same time they increase the number of criminals that are actually punished $[1 - G(\cdot)]$. Finally, in term C , fewer type-II errors increase the number of innocent people and thus the number of innocents who are eventually convicted. Let us now focus on type-I errors $[1 - I(\cdot)]$. Term B shows that fewer type-I errors decrease the number of criminals and thus the social costs of convicting criminals. The effect of fewer type-I errors in term C is more ambiguous: on the one hand they decrease the probability of convicting innocents but on the other hand there are now more innocents among which to commit the judicial error $Z(\tilde{w})$. Thus, type-I and type-II errors determine multiple and complex effects on the three addenda of the social function. Table 1 summarizes these effects.

Error effects on the addenda:	A	B	C
Type-I ($1 - I(\tilde{\epsilon})$)	↑	↑	↑↓
Type-II ($G(\tilde{\epsilon})$)	↑	↑↓	↓

Table 1: The effects of the two errors on the addenda of equation 3.

However also notice that when the population of innocents shrinks and the population of criminals increases there is a substitution effect. Subjects with a $1 - I(\cdot)$ probability of being convicted (innocents) are substituted with subjects with a $1 - G(\cdot)$ probability of being convicted (guilty). However, $1 - I(\cdot) < 1 - G(\cdot)$ because of stochastic dominance. Thus the expected social costs associated with convicting innocents are lower than the expected social costs of convicting guilty individuals (given the same pool size). Thus it is always efficient to decrease the number of type-I errors as far as possible (down to $\varepsilon_{1\min}$) because deterrence is enhanced and, consequently the population of criminals shrinks and the population of innocents increases.

Lemma 1. $I^*(\cdot) = 1 - \varepsilon_{1\min}$. *In order to minimize the social costs of crime, no innocent person should be convicted because the optimal probability of type-I error is the smallest possible one.*

We restate the reasoning again: when type-I errors decrease i) deterrence increases and thus harm is reduced; ii) fewer of those innocents that were

anyhow deterred are punished and therefore social costs are lower; iii) fewer of those who change behavior from crime to honesty are convicted (because and therefore social costs are lower and finally iv) the behavior of those who are not deterred anyway is not affected. Therefore fewer type-I errors always increase social welfare. This finding is formally derived in the appendix (section A.3)¹⁶.

Implications. The normative implication of this lemma is that courts must convict as few innocent people as possible given the procedure and forensic technology that are in place. This is not only fair but also socially efficient.

In order to further study the interactions of both judicial errors and disentangle the different effects seen in table 1 the model is developed by specifying the distribution of w and by deriving equation (3).

3.6 The optimal Blackstone errors ratio with a Pareto distribution

For a given set of rules (embodied by the prosecutor's ability threshold \tilde{e}) and a given forensic technology, the short-term goal of the social planner is to identify the optimal level of type-I and type-II errors in order to minimize the social costs of crime. Thus

$$\begin{aligned} \min_{I(\cdot); G(\cdot)} TC & & (4) \\ \text{s.t. } \tilde{w} = qc_p \Delta(\cdot) & \\ \text{s.t. } \Delta(\cdot) > 0 & \end{aligned}$$

In order to derive this result in the following part of the paper w is assumed to be distributed with a Pareto distribution $z(w; k, w_{\min})$, where, as usual, $k > 0$ ¹⁷, $w_{\min} > 0$ and the distribution is supported in the interval

¹⁶

Lemma 2. By looking at equation 6, $\frac{\partial TC}{\partial (1-I(\cdot))} > 0 \implies (1 - I(\cdot))^* = \varepsilon_{1 \min}$. The optimal level of type-I error is a corner solution in the interval $[\varepsilon_{1 \min}, 1)$.

¹⁷ k is the Pareto index $(0, 1)$. The larger the Pareto index, the smaller the proportion of very high-gain people.

$[w_{\min}, \infty]$. The Pareto distribution is well known to fit the usually observed distribution of wealth and income (Pareto, 1896; Levy and Solomon, 1997). This is because the long right tail describes inequality, that is to say the possibility of a few extremely large outcomes. The Pareto distribution also fits well with the gains from crime we try to model in this paper. In fact the Pareto distribution well describes the possibility of having quite a large number of individuals who can extract minor gains from petty crime and a small number of criminals that can get extremely large outcomes from serious crime¹⁸ (see further discussion in section 3.7). Moreover, the Pareto distribution is also mathematically convenient. However the adoption of the Pareto distribution does not affect the generality of our results: by assuming a generic, well-behaved distribution function the paper's results do not change, but the analysis becomes more cumbersome.

Note that z is independent of $I(\cdot)$ and $G(\cdot)$. Given a Pareto distribution of the gains from crime w , the probability that an individual commits the crime is equal to:

$$\Pr(w > \tilde{w}) = 1 - Z(\tilde{w}) = \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} \quad (5)$$

Lemma 3. $G^*(\cdot) < 1$. *In order to minimize the social costs of crime, at least some criminals must be convicted because the optimal probability of type-II errors is always smaller than 1.*

As shown in the Appendix (A.4) the first and second-order conditions imply $\frac{\partial TC}{\partial G(\cdot)} = 0$ and $\frac{\partial^2 TC}{\partial G^*(\cdot)^2} > 0$. According to the first and second-order conditions $G^*(\cdot) = I^*(\cdot) - \Delta^*(\cdot)$. From Lemma 1, I^* can be substituted by $(1 - \varepsilon_{1 \min})$, thus $G^*(\cdot) = (1 - \varepsilon_{1 \min}) - \Delta^*(\cdot)$ which is always lower than 1.

¹⁸Consider for instance a homicide: some criminals could extract very high gains because the homicide allows them to control a gang or extract a high rent while if the same homicide is carried out by a drug addict suffering withdrawal symptoms it could lead to few and very short-term gains. Consider also the peddling of a given amount of cocaine: a well organized criminal may extract high gains from it while it could easily land a clumsy crook in trouble.

Implications. This lemma simply states that optimal deterrence needs at least some convictions. This implies that there must be an upper bound in our Blackstone errors ratio X and thus some of the severely inflated numbers shown in section 2 are likely not to lead to optimal deterrence.

Proposition 1. $G^*(.) \geq \varepsilon_{2\min}$. *If the costs of conviction are sufficiently high (relative to the social harm of crime), then some under-deterrence in the form of type-II errors is efficient.*

Even when the legal system is able to produce the ideal value of $\varepsilon_{1\min} = 0$, $G^*(.)$ is equal to zero if and only if $\Delta^*(.) = 1$. Since $\Delta^*(.) = \frac{k}{1-k} \frac{h}{q[c_s+c_p]}$ (see appendix A.4), $\Delta^*(.) = 1$ if $c_s+c_p = \frac{h}{q} \frac{k}{1-k}$. Therefore if the costs of conviction are high enough ($c_s + c_p > \frac{h}{q} \frac{k}{1-k}$) the optimal $G^*(.)$ is an interior solution.

Implications. This proposition suggests that whenever punishment is costly, in order to minimize the social costs of crime at least some criminals must be acquitted because the optimal probability of type-II errors may be greater than the smallest possible type-II error.

Corollary 1. *Optimal accuracy should be different for serious vis-à-vis less severe crimes.*

Only when the social harm of crime is sufficiently high with respect to the total costs of conviction, it is then efficient to have type-II errors at their smallest possible number $G^*(.) = \varepsilon_{2\min}$.

Implications. This corollary shows that no avoidable wrongful acquittal should be allowed for serious crimes (those that produce high h) while some wrongful acquittals may be efficient for less severe crimes.

Now, focusing on the error trade-off, recall that, for a given set of rules, the optimal Blackstone errors ratio $X^* = \frac{G^*(.)}{(1-I^*(.))}$.

Proposition 2. $X^* \geq \tilde{X}$. *The optimal Blackstone errors ratio that minimizes the social costs of crime may be larger than $\tilde{X} = \frac{\varepsilon_{2\min}}{\varepsilon_{1\min}}$.*

Recall that while $I^*(.) = \varepsilon_{1 \min}$ (Lemma 1), $G^*(.) \geq \varepsilon_{2 \min}$ (Proposition 1), at least when the costs of convictions are sufficiently high *vis-à-vis* the severity of the crime. Therefore the optimal Blackstone errors ratio is $X^* = \frac{G^*(.) \in [\varepsilon_{2 \min}, 1]}{\varepsilon_{1 \min}}$. Note that $X_{\min} < \tilde{X} \leq X^* < X_{\max}$, where $X_{\min} = \varepsilon_{2 \min}$ and $X_{\max} = \frac{1}{\varepsilon_{1 \min}}$.

Finally, the short-term equilibrium, when $\varepsilon_{1 \min}$ and $\varepsilon_{2 \min}$ are given, is characterized by:

$$\begin{cases} I^*(.) = 1 - \varepsilon_{1 \min} \\ \Delta^*(.) = \frac{k}{1-k} \frac{h}{q[c_s + c_p]} \end{cases}$$

Proposition 3: $\Delta^*(.)$ negatively depends on the social costs of conviction and positively depends on the harm caused by crime and on the Pareto index.

Note that court's optimal screening ability $\Delta^*(.)$ is non-negative and that this is consistent with stochastic dominance. Furthermore, its first derivatives with respect to its determinants respectively are:

$$\begin{cases} \frac{\partial \Delta^*(.)}{\partial h}, \frac{\partial \Delta^*(.)}{\partial k} > 0 \\ \frac{\partial \Delta^*(.)}{\partial c_s}, \frac{\partial \Delta^*(.)}{\partial c_p}, \frac{\partial \Delta^*(.)}{\partial q} < 0 \end{cases}$$

The optimal distance $\Delta^*(.)$ increases with the severity of crime (h).

Implications. This proposition supports the idea that in case of serious crimes the court should strive to better distinguish between innocent and guilty persons (for instance by using better technology) . Not only serious crimes must be penalised with higher sentences, but their cases must also be investigated more thoroughly.

The optimal distance $\Delta^*(.)$ increases also with the Pareto index (k). A high level of k means that there are few criminals able to extract a high w and many petty crooks with a low w (see section 3.7). In this case the ability to identify and convict major lawbreakers becomes crucial. In fact,

the probability of a high level of w increases with k . Thus the probability of an individual becoming a criminal increases with k . The optimal distance $\Delta^*(.)$ indicates that, from a social perspective, when k is large, the optimal ability to correctly identify criminals must be high and the consequent optimal probability of type-II error must be low. Obviously, when private and social costs of conviction are relatively high (with respect to h and k) even a low ability to identify criminals and consequently a high level of type-II error can be socially efficient.

3.7 Discussion and extensions

The model departs as little as possible from the standard framework of analysis of optimal deterrence in order to facilitate the comparison and highlight the novelty of some results. This subsection discusses, and whenever possible relaxes, the main assumptions of the model.

The probability of detection q . In the model the probability with which a defendant is brought in front of the judge is the same both for the guilty and for the innocent. This implies that defendants are detected by *monitoring*. With monitoring Mookherjee and Png (1992) intend the enforcement activity where resources must be committed before information concerning the offence can be received. As such, detection by monitoring is common to the whole population of potential criminals. An example is a speed check to drivers, or a random tax check. This technology of detection may sound odd in the context of crime. A more likely alternative is *investigation*: the enforcement activity which the authority can deploy after the crime has happened. In this case the authority checks only a subset of the population (the suspects) and conditional upon the severity of the crime (Mookherjee and Png, 1992). If the authority detects crime *by monitoring* then the probability of detection for the guilty and for the innocent are equal. If instead it uses *investigation* then the probability of detection conditioned on guilt is greater than the probability of detection conditioned on innocence. In the paper the monitoring technology is assumed because it simplifies calculation without loss of generality. In fact, with investigation the pool of guilty defendants brought to court would get larger and the one of innocents would get smaller. This should qualitatively affect the results only as long as a Bayesian judge may infer its subjective probability of guilt from the fact that the guilty were more likely to be detected in the first place. However judges are required

to presume the defendant's innocence until proven guilty exactly in order to avoid this mechanism.

The costs of conviction. In the model convictions are assumed to be socially costly. In fact the result of the model crucially depends on this assumption. It is trivial to show that when the costs of conviction are zero, judicial errors affect only the probability of crime and the optimal type-I and type-II errors are equal to the lowest possible values ($X_{(c_s=c_p=0)}^* = \tilde{X}$). Although monetary sanctions are often modelled as costless transfers, this is generally considered an overly simplifying assumption. There are at least two reasons why this is the case: first, criminal convictions are often punished with non-monetary sanctions that are by definition, socially costly (Polinsky and Shavell, 1984; Shavell, 1987). Second, even in the case of fines, sanctions are socially costly in as far as their imposition implies the organization of the court system and the authorities that comminate and administer this form of punishment (Polinsky and Shavell, 1992).

Efficiency of crime and underdeterrence. The model assumes that w is simply a transfer from the victim to the criminal and that every crime produces a negative externality h . This implies that there is no crime for which the private benefits to the criminal are higher than the social costs for society (including the private costs of the victim). Were this the case, allowing a certain level of crime would be efficient. This was the notion of efficient crime in the original Becker's (1968) model. However in the model here presented, when h is not excessively high and c_s is not negligible, a positive level of crime is also efficient because convicting all guilty individuals would be too socially expensive. Therefore some underdeterrence is efficient. This resembles the conclusions of Polinsky and Shavell (1992) where efficient underdeterrence is the result of positive enforcement costs¹⁹. However, while Polinsky and Shavell (1992) focus on the detection probability at the police level, we focus on the trade-off of errors at the court level. Although the standard notion of efficient crime can be easily implemented, the model, as it is now, presents a cleaner result as all efficient crime is due to the underdeterrent effect and there is no need to disentangle it from the efficient level of crime à-la Becker.

¹⁹Elsewhere Polinsky and Shavell (1984) find that underdeterrence (or overdeterrence) may be the result of the use of imprisonment as a sanction but this is a different matter.

Risk neutrality. As in most models of optimal deterrence, individuals are here assumed to be risk neutral. It is well known that, when risk aversion is considered, the fine that achieves optimal deterrence is less than maximal (Polinsky and Shavell, 1979; Block and Sidak, 1980; Kaplow, 1992). Moreover, risk aversion entails that -*ceteris paribus*- a conviction following from a type-I error brings more disutility than a conviction resulting from a crime (Nicita and Rizzolli, 2009; Rizzolli and Stanca, 2009). It thus makes sense to trade a type-I error (that imply higher disutility) for a type-II error (with less disutility). The introduction of risk aversion thus would reinforce the result of the paper as type-I errors would be both more privately costly and more socially costly than type-II errors.

The costs of accuracy. The model is built on the premises that, in the short run, the best accuracy that can be achieved is given by $\varepsilon_{1\min}$ and $\varepsilon_{2\min}$. As said these lower bounds depend on the ability of the prosecutors and on the technology available. In the model there are no direct costs attached to achieving the best accuracy. One may question this point: after all better training and technology is likely to come at a cost. However the introduction of a cost term linked with lower levels of both errors does not change qualitatively the results. In fact an interior solution is found for the derivation of equation (3) with respect to $I(\cdot)$ (whereas now it is a corner solution) implying that $I^{**}(\cdot) < I^*(\cdot) = 1 - \varepsilon_{1\min}$. When looking at type-II errors, the introduction of accuracy costs lead to $G^{**}(\cdot) > G^*(\cdot) \geq \varepsilon_{2\min}$. Optimal accuracy would thus be less than maximal $[(1 - I^{**}(\cdot)) + G^{**}(\cdot)] > [(1 - I^*(\cdot)) + G^*(\cdot)] > [\varepsilon_{1\min} + \varepsilon_{2\min}]$. Whether the optimal Blackstone errors ratio would be higher or lower instead depends on the relative costs of the two errors. However it would still be higher than \tilde{X} .

The different “ethical” weight of the two errors. As briefly explained in the literature review, other authors put different weights on the two errors in the social welfare function. This reflects the intuition that convictions of innocents are inherently bad and unethical (and in any case more so than wrongful acquittals) The (un)ethical weight of type-I error can be easily implemented in the model and simply reinforces the main argument that type-I errors should be exchanged for type-II errors as far as the marginal costs of deteriorated deterrence equal the marginal benefits of reduced costs of conviction. By adding the ethical weight to type-I errors the Blackstone

errors ratio simply becomes larger. The merit of the approach here considered is that X^* is shown to be larger than 1 net of the ethical considerations that the authors nevertheless share.

First order stochastic dominance. In section 3.1 the production of the two errors has been modelled using the concept of first order stochastic dominance. This is a novel approach to modelling judicial error production in the Law and Economics literature on optimal deterrence. Assuming that, on average, convincing the court of the guilt of a real criminal requires less ability than convincing the court of an innocent culpability seems reasonable and sound. It is also not in contradiction with assuming q to be equal across innocents and guilty as it concerns the conviction by the court and not the detection by the enforcement authority. It is clear that when this assumption does not hold (for instance for extremely low or extremely high levels of prosecutor's ability) the results collapse and type-I and type-II errors become equally detrimental.

The Pareto distribution. The other aspect worth noticing in the model is the implementation of the Pareto distribution for w . The reasons in support of the application of the Pareto distribution have been already discussed in section 3.6. Despite it is seldom used in the law and economics literature, the Pareto distribution is generally considered to fit well the need to model wealth varying among individuals²⁰ as well as gains from crime. It is however necessary to emphasize that the results of the paper are general and do not depend on the choice of this specific distribution. The Pareto distribution has been used both because it is positively descriptive and for the sake of easy illustration of the results. Further, this distribution allows to fully catch the effects and implications of the efficiency/ability of criminals to produce illegal gains, and offers a new perspective in terms of policies.

²⁰The Pareto distribution also describes a range of situations including insurance (where it is used to model claims where the minimum claim is also the modal value, but where there is no set maximum). Other social, scientific, geophysical and actuarial phenomena are sometimes seen as Pareto-distributed. For instance the sizes of human settlements - few cities, many villages; the standardized price returns on individual stocks; the severity of large casualty losses for certain lines of business such as general liability and so on.

4 Conclusions and Policy Implications

The criminal procedure is inherently exposed to the risk of producing type-I and type-II errors. The pro-defendant safeguards are set against the occurrence of type-I errors (wrongful convictions) although this inevitably implies that more type-II errors (wrongful acquittals) are produced. In other words, modern procedures are constructed in order to produce a Blackstone errors ratio $X > 1$. This paper offers an efficiency-based argument in support of the full array of pro-defendant features (such as mandatory disclosure, double jeopardy, the right to silence, the high burden of proof and so on) that produce a large Blackstone errors ratio. .

This paper shows that the standard model of public enforcement of law (as mastered by Polinsky and Shavell, 2007) can be extended as to explain this bias and that even a large Blackstone errors ratio is compatible with the standard model of deterrence.

The intuition of the model is quite simple. Suppose that one wrongful conviction can be traded off against one wrongful acquittal (for instance by slightly increasing the burden of evidence). On balance deterrence remains constant while social costs of conviction decrease as more guilty defendants are acquitted and more innocents avoid wrongful punishment. However, deterrence remains constant only for negligible changes of errors' levels. In fact, since the trade-off between errors is not linear, at a certain point one fewer type-I error will be traded off against too many type-II errors, causing a drop of deterrence that cannot be compensated any further by the saved costs of conviction in our social welfare function. The paper thus explains how the associated costs in terms of decreased deterrence should be balanced against lower conviction costs.

Our main finding is that while courts should always strive to minimize type-I errors, they may let some type-II errors happen even where these could be avoided, if the costs of convictions are relatively high. In other words the paper shows that some under-deterrence may be optimal if the costs of conviction are significant relative to the social harm. This could happen when, for instance, the prison system is inefficient or whenever petty crimes do not cause great social harm. This result confirms the Polinsky and Shavell (1992) findings that an expensive system of deterrence implies an optimal level of underdeterrence and shows that their results are robust against the introduction of type-I errors. However the model is novel inasmuch as it focuses on courts "fine-tuning" judicial errors in order to minimize in the

short term the social costs of crime.

The propositions derived in the paper and the following corollary have important policy implications. Some under-deterrence is efficient if the costs of conviction offset the social harm of crime (Proposition 1). Furthermore, the Corollary 1 explains why courts should focus more on serious crime (for which fewer avoidable type-II errors should be allowed) while they could acquit some criminals accused of petty crimes (those with low h). When the costs of conviction (c_p) are high, the court may allow more type-II errors to happen. This could represent an alternative tool to the clemency bills used in some countries to contain public deficits (as it was the case of Italy in 2007 -see Drago et al. 2009- and of California in 2009 -see Levitt 2009-), which have adverse consequences for overall deterrence and the perception of the rule of law. Court's screening ability between innocents and guilty people is crucial in connection with the severity and the efficiency of crime (Proposition 3). From a policy perspective, identifying and convicting criminals who can extract large gains from their activities and/or can seriously damage the society becomes a priority over the correct prosecution of petty offenders. Further, the model highlights the role of forensic technology in improving courts' ability to optimally screen between the innocent and the guilty.

Appendix

First and second-order conditions

The first and second-order conditions of equation 4 are derived. Recall that the gain's threshold is $\tilde{w} = qc_p(I(\cdot) - G(\cdot))$ and that the probability of committing a crime, given the Pareto distribution is $\left(\frac{\tilde{w}}{w_{\min}}\right)^{-k} = \left(\frac{qc_p(I(\cdot) - G(\cdot))}{w_{\min}}\right)^{-k}$. In order to see the effects of the two errors on the criminal population the Pareto distribution is applied as in equation 5 to equation 3. Partial derivatives of the equations in respect of the two errors show that both errors dilute deterrence as predicted by Png (1986).

A.1. The first derivative of the probability of committing a crime with respect to the complement to one of type-I error (probability of acquitting an innocent person) is:

$$\frac{d}{dI(\cdot)} = -k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}}\right)^{-k-1} < 0 \quad (6)$$

A.2. The first derivative of the probability of committing a crime with respect to the probability of type-II error is:

$$\frac{d}{dG(\cdot)} = +k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}}\right)^{-k-1} > 0 \quad (7)$$

A.3. The social function is:

$$TC = \left(\frac{\tilde{w}}{w_{\min}}\right)^{-k} h + \left(\frac{\tilde{w}}{w_{\min}}\right)^{-k} q (1 - G(\cdot)) [c_s + c_p] \\ + \left[1 - \left(\frac{\tilde{w}}{w_{\min}}\right)^{-k}\right] q (1 - I(\cdot)) [c_s + c_p]$$

A.3.1. The first-order conditions are calculated as the first derivative of TC with respect to correct acquittals (the complement to one of the probability of type-I error).

$$\begin{aligned}
\frac{\partial TC}{\partial I(\cdot)} &= -k \frac{qc_p}{w_{\min}} [h + q(1 - G(\cdot)) [c_s + c_p]] \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} \\
&+ k \frac{qc_p}{w_{\min}} q(1 - I(\cdot)) [c_s + c_p] \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} - \left[1 - \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} \right] q [c_s + c_p] = \\
&= -k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} [h + q [c_s + c_p] (I(\cdot) - G(\cdot))] \\
&\quad - \left[1 - \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} \right] q [c_s + c_p] < 0 \\
\implies \frac{\partial}{\partial(1-I(\cdot))} > 0 &\implies (1 - I(\cdot))^* = \varepsilon_{1 \min}
\end{aligned}$$

A.3.2. The first-order conditions are calculated as the first derivative of TC with respect to the probability of type-II error:

$$\begin{aligned}
\frac{\partial TC}{\partial G(\cdot)} &= -k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} h - k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} q(1 - G(\cdot)) [c_s + c_p] \\
&\quad - \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} q [c_s + c_p] + k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} q(1 - I(\cdot)) [c_s + c_p] = \\
&= +k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} [h + q [c_s + c_p] (I(\cdot) - G(\cdot))] \\
&\quad - \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k} q [c_s + c_p] \stackrel{!}{=} 0
\end{aligned}$$

$$(I(\cdot) - G(\cdot)) = \frac{h}{q[c_s + c_p]} \left(\frac{k}{1-k} \right)$$

$$\Delta^*(\cdot) = \frac{k}{1-k} \frac{h}{q[c_s + c_p]} \geq 0$$

$$G^*(\cdot) = I^*(\cdot) - \Delta^*(\cdot)$$

$$G^*(.) = (1 - \varepsilon_{1 \min}) - \Delta^*(.) < 1$$

A.3.3. Also the second-order condition must be calculated with respect to the probability of type-II error:

$$\begin{aligned} \frac{\partial^2 TC}{\partial G(.)^2} &= (k+1)k \frac{qc_p}{w_{\min}} \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-2} [h + q[c_s + c_p] (I(.) - G(.))] \\ &\quad - 2k \frac{qc_p}{w_{\min}} \left(\frac{\tilde{w}}{w_{\min}} \right)^{-k-1} q[c_s + c_p] > 0 \end{aligned}$$

$$(k+1)k \frac{qc_p}{w_{\min}} \frac{qc_p}{\tilde{w}} [h + q[c_s + c_p] (I(.) - G(.))] > 2k \frac{qc_p}{w_{\min}} q[c_s + c_p]$$

$$h > q[c_s + c_p] (I(.) - G(.)) \left[\frac{2}{(k+1)} - 1 \right]$$

$$h > q[c_s + c_p] (I(.) - G(.)) \frac{1-k}{(k+1)}$$

$$(I(.) - G(.))^* = \Delta^*(.) = \frac{k}{1-k} \frac{h}{q[c_s + c_p]}$$

$$h > q[c_s + c_p] \frac{k}{1-k} \frac{h}{q[c_s + c_p]} \frac{1-k}{(k+1)}$$

$$h > \frac{k}{(k+1)} h \quad h(k+1) > kh \text{ always true.}$$

The second-order condition always holds in (G^*, I^*) .

References

- Allen, R. J. and M. S. Pardo**, “The Problematic Value of Mathematical Models of Evidence,” *Journal of Legal Studies*, 2007, 36.
- Becker, Gary**, “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 1968, 76, 169–217.
- Blackstone, Willian**, *Commentaries on the laws of England.*, Vol. 4, Oxford: Clarendon Pr, 1766.
- Block, Michael K. and Joseph Gregory Sidak**, “The Cost of Antitrust Deterrence: Why not Hang a Price Fixer Now and Then?,” *Georgetown Law Journal*, 1980, 68 (5), 1131–1139.
- Craswell, R. and J.E. Calfee**, “Deterrence and uncertain legal standards,” *Journal of Law, Economics, and Organization*, 1986, 2 (2), 279–303.
- Daughety, A. F. and J. F. Reinganum**, “On the economics of trials: adversarial process, evidence, and equilibrium bias,” *Journal of Law, Economics, and Organization*, 2000, 16 (I2).
- Drago, F., R. Galbiati, and P. Vertova**, “The deterrent effects of prison: evidence from a natural experiment,” *Journal of Political Economy*, 2009, 117 (2), 257–280.
- Ehrlich, Isaac**, “The Optimum Enforcement of Laws and the Concept of Justice: A Positive Analysis,” *International Review of Law and Economics*, 1982, 2, 3–27.
- Fon, V. and H.B. Schaefer**, “State Liability for Wrongful Conviction: Incentive Effects on Crime Levels,” *Journal of Institutional and Theoretical Economics*, 2007, 163 (2), 269–284.
- Franklin, Benjamin and Albert Henry Smyth**, *The writings of Benjamin Franklin*, New York,: Haskell House, 1970.
- Froeb, L. M. and B. H. Kobayashi**, “Evidence Production in Adversarial vs. Inquisitorial Regimes,” *Economics Letters*, 2001, 70 (2), 267–272.

- Hale, Matthew and Sollom Emlyn**, *Historia placitorum coronæ. The history of the pleas of the crown*, [London]: In the Savoy, Printed by E. and R. Nutt, 1736.
- Harris, J.R.**, “On the Economics of Law and Order,” *Journal of Political Economy*, 1970, 78 (1), 165.
- Hobbes, Thomas**, *Leviathan* 1660.
- Hylton, Keith N. and Vikramaditya S. Khanna**, “A Public Choice Theory of Criminal Procedure,” *Supreme Court Economic Review*, 2007, 15, 61–118.
- Kaplow, Louis**, “The optimal probability and magnitude of fines for acts that definitely are undesirable,” *International Review of Law and Economics*, March 1992, 12 (1), 3–11.
- , “The Value of Accuracy in Adjudication: An Economic Analysis,” *Journal of Legal Studies*, 1994, 23 (1), 307–401.
- Kaplow, Luis and Steven Shavell**, “Accuracy in the Determination of Liability,” *Journal of Law and Economics*, 1994, 37 (1), 1–15.
- Lando, Henrik**, “Prevention of Crime and the Optimal Standard of Proof in Criminal Law,” *Review of Law and Economics*, 2009, 5 (1), 33–52.
- Levitt, S.**, “The Great California Prison Experiment,” February 10 2009.
- Levitt, Steven D.**, “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime,” *The American Economic Review*, 1997, 87 (3), 270–290.
- Levy, M. and S. Solomon**, “New evidence for the power-law distribution of wealth,” *Physica A*, 1997, 242 (1-2), 90–94.
- May, Wilder**, “Some Rules of Evidence: Reasonable Doubt in Civil and Criminal Cases,” *American Law Review*, 1875, 10, 642–672.
- McCauliff, Catherine M.A.**, “Burdens of proof, degrees of belief, quanta of evidence, or Constitutional guarantees?,” *Vanderbilt Law Review*, 1982, 35, 1293.

- Miceli, Thomas J.**, “Optimal criminal procedure: Fairness and deterrence,” *International Review of Law and Economics*, 1991, 11 (1), 3–10.
- Mookherjee, D and Ivan P. L. Png**, “Monitoring vis-a-vis Investigation in Enforcement of Law,” *The American Economic Review*, 1992, 82 (3), 556–565.
- Nicita, Antonio and Matteo Rizzolli**, “Risk Aversion and the standrd of proof,” *mimeo*, 2009.
- Pareto, Vilfredo**, *Cours d’economie politique*, Vol. 2 v, F. Rouge, Lusanne, 1896.
- Persson, M. and C.H. Siven**, “The Becker Paradox And Type I Versus Type Ii Errors In The Economics Of Crime,” *International Economic Review*, 2007, 48 (1), 211–233.
- Png, Ivan P. L.**, “Optimal Subsidies and Damages in the Presence of Judicial Error,” *International Review of Law and Economics*, 1986, 6 (1), 101–05.
- Polinsky, A. Mitchell and Steven Shavell**, “The optimal use of fines and imprisonment,” *Journal of Public Economics*, 1984, 24 (1), 89–99.
- **and** – , “The Economic Theory of Public Enforcement of Law,” *Journal of Economic Literature*, 2000, 38 (1), 45–76.
- **and** – , “The Theory of Public Enforcement of Law,” in A. Mitchell Polinsky and Steven Shavell, eds., *Handbook of Law and economics*, Amsterdam: Elsevier, 2007.
- Polinsky, A.M. and S. Shavell**, “The optimal tradeoff between the probability and magnitude of fines,” *The American Economic Review*, 1979, pp. 880–891.
- **and** – , “Enforcement costs and the optimal magnitude and probability of fines,” *Journal of Law and Economics*, 1992, pp. 133–148.
- Posner, Richard A.**, “An Economic Approach to the Law of Evidence,” *Stanford Law Review*, 1999, 51 (6), 1477–1546.

- Reiman, Jeffrey and Ernest van den Haag**, “On the Common Saying That It Is Better That Ten Guilty Persons Escape Than That One Innocent Suffer: Pro and Con,” *Social Philosophy and Policy. Issue Crime, Culpability, and Remedy*, 1990, 7 (2), 226.
- Rizzolli, Matteo and Luca Stanca**, “Judicial errors and deterrence: theory and experimental evidence,” *mimeo*, 2009.
- Shavell, Steven**, “The Optimal Use of Nonmonetary Sanctions as a Deterrent,” *American Economic Review*, 1987, 77 (4), 584–592.
- Volokh, Alexander**, “n Guilty Men,” *University of Pennsylvania Law Review*, 1997, 146 (1), 173–216.
- Yilankaya, Okan**, “A Model of Evidence Production and Optimal Standard of Proof and Penalty in Criminal Trials,” *Canadian Journal of Economics*, 2002, 35 (2), 385–409.