NBER WORKING PAPER SERIES

A TEST OF RACIAL BIAS IN CAPITAL SENTENCING

Alberto F. Alesina
Eliana La Ferrara

Working Paper 16981
http://www.nber.org/papers/w16981

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2011

We thank Shamena Anwar, Abhijit Banerjee, Katherine Barnes, Francesco Corielli, Richard Dieter, John Donohue, Je˙rey Fagan, James MacKinnon, Nicola Persico, Stephen Ross, Andrei Shleifer and seminar participants at CIAR, Duke, Warwick, Royal Holloway, IIES Stockholm University, Queen's University, University of Connecticut, NBER Political Economy Conference, CEPR Public Policy Conference and Villa La Pietra Conference on New Directions in Applied Microeconomics for helpful comments. Giulia La Mattina, Lucia Rizzica, Erika Deserranno, Mariagiovanna Di Feo, Damiano Briguglio, Giovanni Rizzo, Matteo Fiorini and Marta Barazzetta provided excellent research assistance. This paper was written while Alesina was visiting IGIER Bocconi. He thanks this institution for hospitality. La Ferrara acknowledges financial support from the European Research Council grant ERC-2007-StG-208661. The usual disclaimer applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Test of Racial Bias in Capital Sentencing
Alberto F. Alesina and Eliana La Ferrara
NBER Working Paper No. 16981
April 2011, Revised March 2013, Revised July 2013
JEL No. K42

ABSTRACT

We propose a test of bias based upon patterns of judicial errors. We model the trial court as minimizing a weighted sum of type I and II errors. We define racial bias a situation where the weight depends on defendant/victim race. If the court is unbiased, the error rate should be independent of the combination defendant/victim race. We test this prediction using an original dataset on all capital appeals in 1973-1995. We find that in the first and last stage of appeal the probability of error is 3 and 9 percentage points higher for minority defendants who killed white (vs. minority) victims.

Alberto F. Alesina
Department of Economics
Harvard University
Littauer Center 210
Cambridge, MA 02138
and IGIER
and also NBER
aalesina@harvard.edu

Eliana La Ferrara
Universita' Bocconi
Dept of Economics
via Roentgen 1
20136 Milano
Italy
eliana.laferrara@unibocconi.it

1 Introduction

One of the arguments against the death penalty in the United States is that it is applied with a racial bias against minorities. Consider for example the following statement, taken from the opening paragraph of a document by one of the most vocal organizations opposing capital punishment:

> "African Americans are disproportionately represented among people condemned to death in the USA. While they make up 12 per cent of the national population, they account for more than 40 per cent of the country's current death row inmates, and one in three of those executed since 1977."[1]

While factually correct, statements like these can hardly be interpreted as evidence of racial bias, because violent crime rates are higher amongst minorities than whites. Accounting for differences in patterns of crime, and more generally in unobservables that may be correlated with race, is crucial if one wants to rigorously test for racial bias. We propose a test of racial bias in capital sentencing that allows for the possibility that members of different racial groups differ along observable and unobservable dimensions, and we collect a large original dataset to implement this test.

We develop a model where courts minimize the probability of making judicial errors and we derive a simple test for racial bias. We build upon the work of Anwar and Fang (2006) and propose a test based upon the same insight. Even if we do not observe in the data all the elements that trial courts consider when imposing a death verdict, if the judicial process is unbiased, ex post we should not observe judicial errors more frequently in cases involving certain combinations of defendant and victim's race. We exploit a feature of the capital sentencing process in the US, namely that all first degree capital sentences are automatically appealed (so we have no selection bias), and we focus upon errors of first degree courts reversed by higher courts. Our test rests on the assumption that superior courts can only improve upon the accuracy of first sentencing and therefore remove part or all racial bias.

Our model allows for the possibility that racial groups differ in their propensity to commit crimes, in the quality of legal assistance they have access to, and in other unobserved dimensions. This implies that a simple test comparing errors in judgements against minority defendants with errors against white defendants is inconclusive, as differences in error rates may reflect

[1] Amnesty International, *USA Death by Discrimination - The Continuiung Role of Race in Capital Cases*, April 2003, p.1.

differences in unobservables that are correlated with defendants' race. Under the assumption that for given defendant's race the distribution of these unobservables does not vary with the race of the victim, we can build a test based on pairs of victim/defendant races.[2] This test relies upon the idea that the ranking of first degree mistakes depending upon these pairs should not violate certain patterns that are consistent with unbiased courts. For example, if courts commit more errors on minority defendants who killed white victims than on those who killed non-white victims, they should also commit more errors on *white* defendants who killed white victims than on those who killed non-white ones. In other words, for each defendant's race the ranking of error rates across victims' race must be the same. Failure to satisfy this condition implies the presence of racial bias in our model. We also discuss an extension of the model that allows for bias in the work of the police, the prosector or the defense, which implies that distribution of the evidence may depend on both victim's and defendant's races. In this case the results of our rank order test cannot be interpreted as bias attributable exclusively to the court, but can be viewed more generally as a bias in the criminal justice system, that is the combination of court, police and trial procedures.

In order to implement our test we embarked on a challenging data collection project. We started from the data on capital appeals assembled by Liebman, Fagan and West (2000) for the period 1973-1995. From their data we take information on reversal of first degree death sentences at two stages of appeal: the Direct Appeal, which is the first stage of appeal that every death sentence undergoes and is ruled by State High Courts, and the Habeas Corpus, which is the final stage of appeal and is ruled by Federal Courts. We supplement this data with information we collected on a case by case basis. An especially difficult variable to reconstruct was the race of the victim for each case (neither defendant's nor victim's race are available in Liebman et al.'s data). As a result, our study is the first to provide even descriptive information on the racial composition of victims in capital cases for the entire US in the period under study. As we report below, 51 percent of defendants in the first stage of appeal are white and 41 percent are African American. On the other hand, 78 percent of these cases involve at least one white victim, and 17 percent at least one African American victim.

When we implement our test we find results consistent with the presence of racial prejudice: ceteris paribus, first degree courts are more severe (i.e., they tend to give more death sentences which are then reversed) against cases involving a minority defendant killing one or more white victims. For Habeas Corpus cases involving a minority defendant, the error rate was

[2] The death penalty applies almost exlusively to homicides so there is always at least one well identified victim.

37.5 percent if the victim was white, and 28.4 percent if it was not white, with a statistically significant difference of 9 percentage points. For cases involving a white defendant the difference indicates higher reversal rates when the victim is non-white, but it is not significant. In the Direct Appeal sample, cases involving a minority defendant had an error rate of 37.7 percent if the victim was white and 34.7 percent if the victim was a minority, with a statistically significant difference of 3 percentage points. In cases involving a white defendant the difference is again in the opposite direction and not significant. This pattern of results is consistent with racial bias according to our rank order test.

When we disaggregate the results by region, we find that the effect is driven by Southern States. The difference in error rates in these States is large: in Habeas Corpus cases, the error rate is 15.5 percentage points higher for minority defendants with white victims as compared to minority defendants with non-white victims (p-value .01). For the Direct Appeal sample the corresponding difference is 3.3 percentage points (p-value .13).

The validity of our test relies upon several assumptions. The first is an assumption about the behavior of the higher courts. If these courts are unbiased and make mistakes uncorrelated with the race of defendant and victim our tests are exactly specified. If higher courts are also racially biased in the same direction of the lower courts but less so, our test *underestimates* the amount of racial bias of first degree courts. Our test would overestimate the level of bias if higher courts actively discriminated in favor of minority defendants who killed white victims.[3] Note that our test would *not* fail if higher courts simply discriminated in favor of minorities, say because of lower quality of their legal counsel: to invalidate our test the "reverse discrimination" should be targeting very specifically the minority defendant / white victim pair. We assess the plausibility of this interpretation empirically, exploiting differences in ideology across appeal courts. We build upon the premise that the judges who would be most likely to reverse discriminate in favor of minority defendants who killed white victims would be those more "left leaning". Using various measures of political orientation of higher courts' judges we do not find any evidence of this effect: both left wing and right wing leaning judges exhibit the same pattern of reversal of first degree sentences.

Another assumption upon which our test rests is that possible unobservable characteristics, such as characteristics of the crime or quality of the evidence, are not systematically different across victims' races, for given defendant's race. To assess the plausibility of this assumption, we test whether the distribution of observable characteristics correlated with the nature of the

[3] See Argys and Mocan (2004) on the issue of reverse discrimination in executions and sentence commutations.

crime (e.g., situation, weapon, etc.) or with its severity (e.g., aggravating and mitigating factors) differs systematically across racial pairs, and we find that for the most part it does not. We also investigate whether the results of our test are robust to conditioning on a number of observables and find that by and large they are. While this does not rule out possible differences in unobservables, the outcome of this battery of tests increases our confidence in the interpretation of our results as consistent with racial bias.

As mentioned above our model builds upon the literature on racial bias in motor vehicle searches and in particular to recent work by Knowles, Persico and Todd (2001), Anwar and Fang (2006) and Antonovics and Knight (2009). However, our model differs from those papers in several ways. In models of car searches the issue is which car to stop and then with certainty either contraband is found or not. In our model the courts have to evaluate guilt or innocence based on a noisy signal and there is a review of the first decision. Guilt or innocence cannot be decided for sure like in a car search. The objective function of our courts is therefore different from that of a trooper stopping cars, in that it trades off the extent of type I and type II errors.[4] We share with Gennaioli and Shleifer (2007) an interest in the effect of bias in judges' decisions. These authors however address a different research question, namely how common law and the accumulation of precedents leads towards an equilibrium without judicial bias. We do not pursue this type of dynamic analysis of bias. Two recent papers exploit random variation to test if there are systematic differences across judges in the racial gap in sentencing for felony crimes (Abrams, Bertrand and Mullainathan, 2012) or if there is evidence of bias in felony trial outcomes depending on the racial composition of the jury (Anwar, Bayer and Hjalmarsson, 2012). This type of random variation is not available for death penalty cases, so our test is built on different grounds, exploiting a prediction on the equilibrium behavior of the court. We should also clarify that, although it would be extremely interesting to disentangle the behavior of different actors (prosecutor, jurors, judge), our data does not allow us to advance in this direction and will only allow us to detect a "combined" bias of the criminal justice system up to the trial stage.

Our paper is also related to the literature on the death sentence and its usefulness. We do not touch upon this issue. We focus only on the question of whether or not the death penalty is applied with a racial bias.[5] There are several early contributions in the law literature on

[4] Both our paper and the literature of motor vehicle searches owe a lot to the path-breaking work by Becker (1957) on rational models of crime.

[5] On the deterrence effect see among others, Erlich (1975), Katz, Levitt and Shustorovich (2003) and Donohue III and Wolfers (2005) for a review.

the role played by race in capital sentencing and execution. The stylized facts described in this literature include: (i) the disproportionate execution of blacks compared to whites; and (ii) the higher likelihood that the death penalty is imposed when the victim is white. Most of these studies rely on small samples and can be criticized on the grounds that important factors affecting the decision of the court may not be observable in the data. This is almost inevitably the case when a direct test of discrimination at the sentencing stage is attempted. Even the most comprehensive data source, in fact, will not possibly include all the information that was available to the court at the moment when the sentence was imposed. One of the most influential early attempts at controlling for observable factors is a study by Gross and Mauro (1984). They constructed an index of aggravating factors and found that, after controlling for them, the race of the victim was still a strong predictor of capital sentencing (the likelihood of a death sentence being higher when the victim was white), but the race of the defendant had no residual explanatory power. Blume, Eisenberg and Wells (2004) combine data on death row cases for eight US States with homicide data for the same States over the period 1976-1998 and find that murders involving black defendants and white victims are significantly more likely to result in death sentences than white defendant-white victim murders. On the other hand, they find that black defendant-black victim cases are significantly under-represented on death row. Compared to this literature, our test is not subject to the omitted variable bias critique (under the assumption of the model). At the same time, our test has a more limited scope, in that it applies to cases that have received the death sentence in the first trial, and cannot estimate bias occurring from exclusion errors, i.e. cases that should have received a death sentence and did not.

The paper is organized as follows. Section 2 offers a brief synthesis of the institutional details useful to understand judicial errors in capital cases in the US. Section 3 describes our model of behavior of the court and derives our test of racial bias. Section 4 describes the data. Section 5 presents our empirical methodology and results. The last section concludes.

2 Institutional background

Today, thirty-four states in the US allow capital punishment.[6] Each state has its own statute but much similarity exists among them. Most statutes are in fact modelled around the Georgia one approved by the Supreme Court in *Gregg v. Georgia* in 1976. That statute prescribed: a)

[6] The Federal Government has two death penalty statutes, one for the military and the other for non military crimes. This section draws on Coyne and Enzeroth (2006).

an independent trial of guilt or innocence; b) a second hearing solely to determine the sentence; c) a finding of at least one aggravating circumstance; d) an automatic review by the Georgia Supreme Court and e) the comparison to similar cases. Even though the statutes are similar, the actual application of the death penalty varies greatly across states.

First trial and sentencing

Trials for capital crimes embed two stages: the guilt phase, where the jury deliberates whether the defendant is guilty or not, and the sentencing stage where, if the defendant is found guilty, the jury (or in some States and until 2002, the judge) weighs the aggravating and mitigating factors presented by the prosecutor and the defense and determines the sentence. The Supreme Court has ruled that no statute can prescribe mandatory capital punishment, that is no one found guilty of a capital crime can be automatically sentenced to death.[7] This implies that the jury always has discretion in choosing between a death sentence or imprisonment, if the defendant is found guilty. A death sentence requires the existence of at least one aggravating circumstance and the consideration of applicable mitigating factors. What constitutes both vary from State to State. Certain aggravating circumstances or mitigating factors are very clear, like killing a police officer (aggravating) or killing under a certain age (mitigating). But other factors are much less clear cut, like a murder being “in cold blood and pitiless” (aggravating) or “acting under duress” (mitigating). The Supreme Court has struggled with unclear and vague definitions of aggravating circumstances and mitigating factors but quite a large latitude remains. About one per cent of the murders committed in a year ends up in a death sentence.[8]

The appeal process

The most important aspect of the capital punishment procedural rules for the purpose of our study is that all capital sentences, with no exceptions, are automatically appealed in state high courts. During our study period, in all but two states the appeal run directly from the trial court to the state court of last resort (typically the state supreme court), while in Alabama and Ohio it went through an intermediate court of criminal appeals before reaching the highest court. Sentences that survive state direct appeals are then inspected by state post-conviction courts and, if they survive this stage too they can be reviewed in federal habeas corpus petitions. The process often lasts several years. At each stage, the appeal court can overturn the sentence

[7] Woodson v. North Carolina (1976) and Roberts v. Louisiana (1976).

[8] See Barnes, Sloss and Thaman (2008) for a recent discussion of criteria according to which prosecutors purse the death penalty in about 4 percent of capital crimes in Missouri.

if “serious error” is found, i.e. “error that substantially undermines the reliability of the guilt finding or death sentence imposed at trial” (Liebman et al., 2000). When all appeals are exhausted, the only hope left for the defendant is an act of clemency from the State Governor.

Liebman, Fagan and West (2000) conducted a study of all 4,578 state capital appeals in the period 1973-1995, plus 248 state post conviction reversals and 599 capital sentences reviewed by federal habeas corpus courts in the same period. Their findings were striking: between 1973 and 1995, the proportion of fully reviewed capital judgments in which “serious error” was found and which were overturned at one of the three stages was 68 percent. This is what happened to overturned cases at retrial: 7 percent were found to be innocent, 75 percent were resentenced to less than death, and 18 percent were resentenced to death. These sentence reversals will play a key role in our empirical test.

3 The model

3.1 Setup

We consider a defendant whom a court can condemn to the death penalty or to a lesser penalty (which includes the case in which the defendant is set free). If the court decides for the death penalty, there is an appeal. In case of a lesser sentence or a no guilt verdict there is no appeal and the decision stands. In appeal, the superior court can either confirm the death penalty or reverse the decision of the lower court because of errors. An error can occur in establishing the guilt of the defendant or in sentencing the death penalty for a crime that did not warrant it. Our assumption (to be discussed below) is that while the lower courts can make mistakes, higher courts make no mistakes. Our empirical test holds identically under the more general assumption that even the highest courts can make mistakes but these are uncorrelated with the race of the defendant and of the victim. Before formally presenting the setup of the model, it is useful to introduce two definitions that we will use to simplify the exposition and that constitute a slight abuse of terminology.

Definition: *We label “guilty” a defendant who has committed a crime deserving of the death penalty according to the law. We define “innocent” a defendant who has committed no crime or has committed a crime which does not deserve the death penalty.*

Thus sentencing an “innocent” may mean sentencing to the death penalty someone for a crime which instead deserved say a life term. Empirically, as we shall see below, most of the “errors” imply sentencing to death somebody guilty of a crime not deserving death.

Definition: *We denote as "court" the combination of judge and jury.*

While in reality judge and jury are obviously distinct, our data does not allow to separately test for bias among the two, hence we consider them jointly in the model because this matches the empirical test we conduct.

We assume that each crime involves one defendant and one victim. Defendants are characterized by their race and by a set of characteristics of the person or the crime or the relationship between the person and crime. Let $r \in \{w, m\}$ be the race of the defendant, where w stands for "white" and m for "minority". Let's define the race of the victim as $R \in \{W, M\}$ where W stands for white and M for minority. The court observes various signals about the characteristics of the defendant and of the crime, provided by the police and the trial proceedings and summarizes them in a single dimension which we denote with x and that we can denote as "the evidence." We normalize the support of x to $x \epsilon \left[0, 1\right]$. The distribution of evidence can depend upon the race of the defendant, for instance because of the quality of his/her legal assistance: if minority defendants (on average poorer) have a lower quality defense, they may carry a less precise signal and face more errors against them in the first trials.[9] While the quality of defense is not explicitly modeled in our framework, it can be incorporated in a different distribution of evidence faced by the courts for minority and white defendants. In the benchmark derivation of our test we allow $F_g^r(x)$ and $F_n^r(x)$ to depend on the race of the defendant but not on the combination of defendant and victim race. In section 3.3 we discuss possible ways of relaxing this hypothesis and the implications for interpreting our test.

We assume that the signal is informative for the court and the densities $f_g^r(x)$ and $f_n^r(x)$ satisfy the strict monotone likelihood ratio property (MLRP), that is:

MLRP: $f_g^r(x)/f_n^r(x)$ is strictly increasing in x, for $r \in \{w, m\}$.

This property implies that higher values of the signal x are associated with a relatively higher probability of guilt. We also assume $f_g^r(x)/f_n^r(x) \longrightarrow +\infty$ as $x \longrightarrow 1$.

The problem of the court

An individual considering whether to commit a capital crime (crime in short) compares the costs and benefits of it. In the Online Appendix we model this problem, but for the purpose of our test we can equivalently assume that there exist an exogenous probability π^r that an

[9] Results by Iyengar (2007) indeed suggest that this may be the case. When comparing the effectiveness of two types of defense lawyers provided for indigent defendants, namely public defenders or court private lawyers compensated by the hours, she finds that the former perform better and minority defendants are disproportionately represented by the latter.

individual of race r commits a crime. We allow this probability to vary across defendant races. In section 3.3 we discuss whether π is likely to also depend on victim's race and how this may affect our results.

The court wants to sentence guilty defendants to the death penalty, but wants to avoid the mistake of sentencing defendants who do not deserve the death penalty (labeled as "innocent").[10] These considerations can be summarized by assuming that the court minimizes a weighted average of the probability of condemning an innocent (type I error) and the probability of letting a guilty person free (type II error). Therefore the court chooses the optimal x_{rR} as the value of x which solves:

$$\min_{x_{rR}} \left\{ \alpha_{rR} \left[1 - \pi^r\right] \left[1 - F_n^r\left(x_{rR}\right)\right] + (1 - \alpha_{rR}) \pi^r F_g^r\left(x_{rR}\right) \right\} \tag{1}$$

with $0 < \alpha_{rR} < 1$, $r \in \{w, m\}$ and $R \in \{W, M\}$. The first and second term in (1) are, respectively, the type I and type II error. The parameter α_{rR} represents the relative weight given by the court to type I error and will be crucial for defining our test of racial bias. The court chooses to sentence the defendant to death if the evidence is above a certain threshold. We define the threshold x_{rR}, which as indicated in the notation could vary with the race of the defendant and of the victim, allowing the court to choose four potentially different thresholds. The probability that an individual of race r killing an individual of race R is sentenced to death is: $p_r\left(x\right) = \Pr\left(x \geqslant x_{rR}\right)$.The optimal decision of the court in a case involving a defendant of race r and a victim of race R is to impose a death sentence if and only if the signal x exceeds the threshold x_{rR}^* given by the court's first order condition:

$$\frac{f_g^r\left(x_{rR}^*\right)}{f_n^r\left(x_{rR}^*\right)} = \frac{\alpha_{rR}}{1 - \alpha_{rR}} \frac{1 - \pi^r}{\pi^r} \tag{2}$$

The cutoff value x_{rR}^* is thus the "standard of proof" applied by the court. Inspection of (2) immediately reveals that x_{rR}^* is increasing in α_{rR}, i.e. the higher the relative concern about condemning an innocent, the higher will be the standard of proof required before imposing a death sentence.

[10] Note that under the assumption that the higher courts never make mistakes the costs of sentencing an innocent for the lower courts are the moral costs of inflicting high costs to innocent defendants and the costs of reputation losses of having made mistakes. In the more general version of the model in which even superior courts can actually make mistakes, although uncorrelated with race, lower courts also have to worry about the possibility that innocent people sentenced to death are never released, which implies a very high moral cost.

Note that even if $\alpha_{rR} = \alpha$ for any r and R, the equilibrium cutoff point chosen by the court can be, and in general will be, different for white and minority defendants. In fact the left hand side of (2) depends on the race of the defendant, r, because f_g^r and f_n^r depend on r; and the right hand side of (2) depends on r through π^r. For example, if minorities have a higher probability of committing crimes, ceteris paribus the court will choose $x_{mR}^* < x_{wR}^*$. On the other hand, under the assumptions of our model the only way the race of the victim R can enter (2) is through the parameter α_{rR}. This is the key insight upon which our test is based. We then discuss how we can relax this assumption.

3.2 A test of racial bias

In our model a court could be biased in different ways.

Definition 1 *Bias only on the race of the defendant:* $\alpha_{rR} = \alpha(r)$

α depends upon the race of the defendant (r), but not the race of the victim (R). A bias against minority defendants is represented by $\alpha_{mM} = \alpha_{mW} < \alpha_{wM} = \alpha_{wW}$, as the court places less weight on the possibility of wrongly condemning a minority defendant.

Definition 2 *Bias only on the race of the victim:* $\alpha_{rR} = \alpha(R)$

α depends on the race of the victim (R), but not on the race of the defendant (r). A bias in favor of white victims is represented by $\alpha_{mW} = \alpha_{wW} < \alpha_{mM} = \alpha_{wM}$, as the court places less weight on the possibility of wrongly condemning someone who has killed a white victim.

Definition 3 *Bias on both the race of the victim (R) and the race of the defendant (r):* $\alpha_{rR} = \alpha(r, R)$

The court applies a differential treatment on the basis of the race of the victim, but it further differentiates depending on who has killed that victim. For example, a situation where the court treats minority defendants who have killed white victims more harshly than whites who have killed whites, but is relatively and equally lenient on both if the victim is non-white, can be represented as: $\alpha_{mW} < \alpha_{wW} < \alpha_{mM} = \alpha_{wM}$.

We now derive a test for racial bias. Define the capital sentencing rate for r, R pairs as:

$$\gamma(r, R) = \pi^r \left[1 - F_g^r\left(x_{rR}^*\right)\right] + \left[1 - \pi^r\right]\left[1 - F_n^r\left(x_{rR}^*\right)\right]. \tag{3}$$

The equilibrium error rate on r, R pairs, which we denote as $E(r, R)$, is:

$$E(r, R) = 1 - \frac{\pi^r \left[1 - F_g^r \left(x_{rR}^*\right)\right]}{\gamma(r, R)} \tag{4}$$

with $\gamma(r, R)$ given by (3). As shown in the Appendix, under MLRP $E(r, R)$ is monotonically decreasing in x_{rR}^*: the higher the standard of proof, the lower the error rate.

The parameter α_{rR} enters the error rate (4) through the optimal threshold x_{rR}^* derived from (2), allowing us to build a test of racial bias. The race of the victim R only affects x_{rR}^* if α_{rR} depends on R. Therefore if the court does not discriminate over the race of the victim, the error rate should be independent of it, even though it could depend upon the race of the defendant. If for given race of the defendant we find higher error rates in cases involving white victims compared to minority victims, this is evidence of racial bias in favor of white victims. In the benchmark version of our model, this bias originates from a lower weight given to type I errors in cases involving a white victim ($\alpha_{rW} < \alpha_{rM}$), which led the court to apply a lower standard of proof, ceteris paribus. The intuition for this result is illustrated in figure 1, which shows the density functions of the signal x for non-guilty and guilty defendants, holding constant the race of the defendant. The type I error is the shaded area to the right of the threshold x^*. When we hold constant the race of the defendant and vary the race of the victim, the assumptions of our model imply that the distributions f_n^r and f_g^r do not shift. Hence the only way one could obtain different error rates is through a shift in the value of x^*, which in turn must reflect different values of α. Figure 1 shows an example where α decreases (e.g., less weight is given to type I errors against defendants who killed white victims), the threshold x^* moves left, and the error rate increases.

On the other hand, a similar inference cannot be made if one holds constant the race of the victim and compares errors against defendants of different races. In other words, we cannot derive from our model the implication that if the court does not discriminate across defendants we should find the same error rate for white and minority defendants when we hold constant the race of the victim. Figure 2 illustrates the point. The top and bottom panels of figure 2 display, respectively, the signal distributions for white and minority defendants, holding constant the race of the victim. Because our model allows f_n^r and f_g^r to differ according to the defendant's race r, error rates (the shaded areas in figure 2) may differ even when the court is unbiased and selects the same threshold x^*.[11]

[11] The fact that the threshold x^* is the same for both defendant races is neither a necessary nor a sufficient condition for unbiasedeness of the court. It is used in figure 2 purely for illustrative purposes. Also, the fact

Thus we cannot test for the presence of bias only on the race of the defendant, while we can test for bias which depends on the race of the victim (bias type 2 and 3 in the definition above). We now derive a test for the relatively more conservative definition that bias is purely a function of victim's race (bias of type 2). This amounts to asking the question: in the presence of bias related to the victim's race, does this bias affect defendants of different races in the same way? In the empirical section we will show that the answer is "no", and that according to our results the behavior of the court is consistent with a bias that depends on the particular combination of defendant and victim race (bias type 3).

Proposition 1 *If $\alpha_{rR} = \alpha_R$ independent of r, then the ranking of average error rates $E(r, M)$ and $E(r, W)$ should not depend on r, for $r \in \{m, w\}$.*

Proof: Suppose without loss of generality that $\alpha_{mW} = \alpha_{wW} < \alpha_{mM} = \alpha_{wM}$. Consider first minority defendants. Because x^*_{rR} is strictly increasing in α_{rR}, $\alpha_{mW} < \alpha_{mM}$ implies $x^*_{mW} < x^*_{mM}$, which in turn implies $E(m, W) > E(m, M)$ due to the fact that $E(r, R)$ is strictly decreasing in x^*_{rR}. The same reasoning applies to white defendants, with $\alpha_{mW} = \alpha_{wW} < \alpha_{wM} = \alpha_{mM}$ implying $E(w, W) > E(w, M)$. $\square$

Thus the following condition must hold if the court discriminates on the basis of victims' race but treats defendants of different races in an unbiased way:[12]

$$E(m, W) > E(m, M) \iff E(w, W) > E(w, M) \tag{5}$$

Expression (5) says that if we find a higher error rate on minority defendants who killed white victims, compared to minority defendants who killed minority victims, then we should also find a higher error rate on white defendants who killed white victims than on white defendants who killed minority victims, and vice versa. This forms the basis for our rank order test.

Our use of rank order tests is in the same spirit of Anwar and Fang (2006), with the difference that in their case one of the two dimensions over which troopers' success rates are computed pertains to the behavior of the agent who may be discriminating (i.e., the police officer), while in our case the two dimensions pertain to offender and victim characteristics, and not features of the court.

that we allow the probability of guilt π^r to differ across races implies that we could not make inference on bias by comparing errors across defendant races even if the signal distributions were the same.

[12] Expression (5) refers to the case of bias in favor of white victims. Obviously for bias in favor of minority victims the inequalities should be reversed.

3.3 Discussion and extensions

In this section we discuss some important assumptions underlying our model, possible extensions, and how they affect the interpretation of our results. We start with the probability of guilt π_r, then move to the shape of the signal distributions $F_n^r(x)$ and $F_g^r(x)$, and finally discuss the behavior of superior courts and the role of plea bargain.

Different probabilities of guilt

It is well known that the frequency of homicides varies depending on the combination of offendant and victim races, namely that intra-racial homicides are more frequent than inter-racial ones. However, the parameter π_r in our model does not represent the *frequency* of crimes but the probability of *guilt*, or more precisely the likelihood that the defendant has actually committed a crime that deserves the death penalty. The question then is: does this probability depend on the four combinations of defendant and victim race in a way that would invalidate our test, leading us to attribute to bias differences in errors that arise from differences in π_{rR}? To answer this question we first present some available empirical evidence and then discuss from a theoretical point of view under what configurations differences in π_{rR} would create a problem for interpreting our results.

We are not aware of a comprehensive dataset including information on guilt rates by race of defendant and victim for the period under study. So we resorted to two alternative sources.[13] The first is a representative sample of murders adjudicated in 1988 in 33 of the largest counties in the US (US Dept. of Justice, 1996). This dataset includes information on race of the defendant and of the victim, as well as the final disposition outcome of the case. We restricted the sample to first degree murders, which are the ones potentially eligible for the death penalty, and to cases that underwent trial, to provide as close a benchmark as possible to our dataset. To investigate whether the differential incidence of intra-racial murders may be indicative of differences in π_{rR}, we regressed the average guilt rate in the county on the fraction of murders in which defendant and victim belonged to the same race group. The estimated coefficient was $-.05$, with a standard error of .31. Figure 3 plots the raw data (with the size of the circles proportional to the number of homicides in the county) and clearly shows that there is no correlation between the two variables. We also exploited the individual level data and

[13] Other than these two sources, to the best of our knowledge existing public use data typically include defendant's race (and in some cases the disposition of the case) but not victims' race. Datasets with information on defendant and victim race typically do not have information on the disposition of the case, which means we could not calculate guilt rates.

regressed the probability that the defendant was found guilty on the race of the defendant, the race of the victim, and an interaction of the two. The coefficient on the latter term was again not statistically different from zero.[14]

A second source of empirical evidence is the work of Blume, Eisenberg and Wells (2004). They calculate death sentence rates for eight US States over the period 1977-2000 as the ratio of number of death sentences and number of murders in the State. While this is not the empirical equivalent of our guilt rate (among other things, because the numerator includes biased decisions of trial courts), a rough correction can be applied by subtracting State and race-pair specific relief rates (which we computed in our data) and assuming that the share of guilt sentences that were affirmed are correct. Again, looking at the pattern of guilt rates computed in this way no systematic correlation emerges between same race pairs and guilt rates.[15]

Even though differences in π_{rR} do not seem warranted on the basis of the above evidence, it is useful to discuss in what direction our results would be affected if such differences were present. This amounts to assessing the sign of the derivative of the error rate (4) with respect to π. In the Online Appendix we show that this derivative can be decomposed in two parts. The first includes a "direct effect" of π on the error E that would obtain if x were treated as exogenous. This effect is negative and captures the fact that, for given standard of proof x, higher values of π imply that fewer of the people condemned will be innocent, hence the error rate is lower. The second part of the derivative is an "indirect effect" that works through endogenous changes in x. To see how, notice that x is decreasing in π (from (2) and MLRP), and E is decreasing in x, hence the indirect effect pushes in the direction of $\frac{\partial E}{\partial \pi} > 0$. Intuitively, this occurs because when π increases the court chooses a lower standard of proof, which $-$for given $\pi-$ increases the size of the Type I error. Which of the two effects dominates depends on the value of π and on the shape of the functions $F_n^r(x)$ and $F_g^r(x)$. In the Online Appendix we

[14] Specifically, our estimated linear probability model is

$$Guilty = \underset{(.089)}{.755} + \underset{(.091)}{.013}\ ND + \underset{(.086)}{.116}\ WV - \underset{(.086)}{.049}\ (ND * WV)$$

where ND is a dummy for non-white defendant, WV is a dummy for white victim, and standard errors (in parenthesis) are clustered at the county level. Results are very similar if we run the regression on the full sample of first degree murders (including cases that did not go to trial and guilty pleas) or if we include county fixed effects.

[15] Results available from the authors. Note that we do not have access to the raw data used by Blume et al. (2004), so we conducted this meta-analysis using the published data in their paper.

provide some simulations where the direct effect dominates when the f_n distribution is relatively more dispersed than f_g.[16]

In summary, what matters four our test is the cross pattern of guilt rates for the four pairs of defendant and victim races. Contrary to perceptions based on the higher *frequency* of intra-racial homicides, the (admittedly scant) evidence that we are aware of does not allow to establish any systematic pattern in this respect. In the remainder of the paper we maintain the assumption that π does not systematically vary with the combination of defendant and victim race.

Bias in the collection of evidence

A second important assumption underlying our test is that the functions $F_n^r(x)$ and $F_g^r(x)$ do not depend both on the race of the defendant and of the victim.[17] Given our test the critical question is whether, in particular, $F_n^{mW}(x)$ differs from $F_n^{mM}(x)$ and $F_g^{mW}(x)$ differs from $F_g^{mM}(x)$. The two pairs of functions may not be the same for two reasons. One is the nature of the crime: the types of crimes involving defendant/victim pairs of different race may be "objectively" different and lead to different type I errors for any given level of evidence x. The second reason is that there may be no intrinsic difference in the nature of the crime but the combination of police work, prosecutor work and defense attornies' work leads the court to face different distributions as a function of the defendant/victim pair.[18]

Let's begin with the second case. More precisely let's assume that if the police, prosecutor and defense attorney were unbiased, the distributions would indeed not depend on the race of the victim. We denote these as the "unbiased" distributions $F_n^r(x)$ and $F_g^r(x)$. Suppose now that a biased "police, prosecutor, or defense" (in short PPD) distort the information available to the court so that the distributions the court uses in its optimization problem, hence in (2), are instead $\widetilde{F}_n^{rR}(x)$ and $\widetilde{F}_g^{rR}(x)$. How might these differ from the true ones and what would be the implications for our test? To answer this question, let us consider the example of a murder

[16] High relative dispersion of f_n implies that decreases of x in response to increases of π translate into relatively small increases in the area corresponding to the type I error as compared to the decreases in the type II error, resulting in a decrease of the overall error rate. Conversely, in cases where f_g has much fatter tails than f_n the opposite effect may prevail.

[17] A similar restriction is common to Anwar and Fang's (2006) test of prejudice.

[18] Empirically, Radelet and Pierce (1985) analyzed a sample of 1017 homicides in Florida in the period 1973-77 and compared the descriptions of the homicides in police reports with the (later) descriptions given by courts. They found that homicides involving African American suspects and white victims were more likely to be described as "felony" by prosecutors.

involving a minority defendant and a white victim. Figure 4 reports the unbiased distributions for this case, $f_n(x)$ and $f_g(x)$ (for simplicity we omit superscripts) and denotes as x^* the solution of an unbiased court in this case. A biased PPD may want to convince the court that for a given threshold of evidence the probability of convicting an innocent is (artificially) low compared to the risk of acquitting a guilty person. In the figure this is represented by the function $\widetilde{f}_n(x)$. In correspondence of x^* the size of the "perceived" Type I error has now decreased relative to type II (which is unchanged in the example). An unbiased court would thus adjust the standard of proof downwards and choose a threshold $\widetilde{x}^* < x^*$. This would result in a higher type I error ex post when the appeal court has available the unbiased distribution $f_n(x)$. In other words, the court would commit more type I error given the "true" distributions while acting on the "wrong" ones. In this case the observed higher error rate for the minority/white pair would be the result of a PPD bias not a court bias. Obviously, if the court were also biased the observed error rate would be the combination if the two. What this discussion suggests is that if we allowed the PPD to manipulate the distribution of the signal differentially depending on the combination of defendant and victim race, we would need to qualify the interpretation of our test. The observed error would not necessarily derive from a bias in the parameter α (i.e., it may not necessarily be attributable to the court). But one would still consider this error as resulting from racial bias in the criminal justice system.

Consider now the other case, namely PPD are unbiased but $F_n^{mW}(x)$ differs from $F_n^{mM}(x)$ and/or $F_g^{mW}(x)$ differs from $F_g^{mM}(x)$ because the type of crime is objectively different. For instance murders with minorities killing whites may involve "worse" crimes or cases where the evidence is less clear cut, something which would go against the validity of our test. In this case more errors might be made by the courts simply because of the nature of the crime. One way of addressing this issue is to examine empirically whether observable indicators of the nature of the crime (e.g., aggravating and mitigating factors) differ systematically across racial pairs, and test if our results are robust to conditioning on observable characteristics. This should give some insights about the potential bias due to differences in unobservables. This is what we do in section 5.2.

To sum up, we have discussed two reasons why the assumption that $F_n^r(x)$ and $F_g^r(x)$ do not depend on R may be violated. One reason relates to the behavior of the police, prosecutor or defense attorney, who may distort the "true" distribution of the signal in a way that produces more type I errors. This possibility requires that we interpret the higher error not as resulting from a bias of the court, but from a bias of another part of the criminal justice system. But it is bias nonetheless. The second potential violation occurs if the severity of the crime is objectively different in inter-racial homicides. This case could generate higher error even if all parts of the criminal justice system are unbiased. We cannot rule out this possibility on theoretical grounds, but we assess empirically how serious a concern it may be and find results that we believe are supportive of our interpretation.

Bias in the decisions of superior courts

Another hypothesis in our model is that superior courts never make mistakes and are racially unbiased. If superior courts made errors which were uncorrelated with the combination of defendant/victim race, this would not invalidate our test of racial bias. We can also relax this assumption in one direction. Suppose for example that superior courts were racially biased in the same direction of lower courts. This would go against finding higher error rates on certain racial pairs because the superior courts would simply reaffirm the first sentence. That is, if we did not find evidence of racial bias based upon our test it could mean that the same bias applies to all levels of courts. Thus not finding a bias could be inconclusive but finding it would not. Note that if the racial bias declines with subsequent stages of revision (from state courts to federal Habeas Corpus courts) then we should find that the difference in errors rates across pairs of defendant's and victim's race should become larger in later stages of appeal. This is what we find below.

What we cannot allow in our model is that superior courts are biased in the opposite direction to lower courts, because in this case higher error rates may be interpreted as "reverse discrimination" rather than evidence of mistakes by lower courts. We are not aware of a literature that documents such bias in opposite directions, and at the same time the pattern of inequalities that we find in our tests (higher error rates on cases in which defendant and victim are from different racial groups) would require a particular pattern of bias by superior courts, not in favor of a particular group but of specific "pairings" of races. Having said this, in the empirical part of the paper we try to address the possibility of bias by superior courts by testing if our results depend on certain characteristics of the appeal court (e.g., political orientation). We do not find evidence that the pattern of reversal we uncover is driven by the ideology of the appeal judges.

Plea bargain

Many potential capital cases are plea bargained. The strength of the evidence against the defendant and the severity of the crime are critical factors in determining the incentive for defense and prosecution to pursue a plea bargain. Comprehensive empirical studies of the nature and characteristics of plea agreements are hard to come by due to data limitations. There is evidence that minority defendants and defendants with previous criminal history receive a harsher plea bargained prison term (Humphrey and Fogarty, 1987). Models of plea bargain typically involve asymmetric information between prosecutor and defendant about the strength of the case, as in for instance Grossman and Katz (1983) and Reinganum (1988). In our model we do not have this asymmetry and only with this extension (which we leave for future research) we could incorporate plea bargain in a meaningful way. As far as our empirical test is concerned, if the likelihood that a case is plea bargained were uncorrelated to the races of the racial pair defendant/victim, our test would be unaffected. If it were not, then this correlation might introduce a bias, but the direction of the bias is unclear, as it would depend among other things on the shape of the signal distribution.

We tried to empirically assess the potential relevance of this source of bias using data on a representative sample of murders adjudicated in 1988 in 33 of the largest counties in the US (US Dept. of Justice, 1996). This dataset includes information on race of the defendant and of the victim, as well as the final disposition outcome of the case (among which guilty plea). When we regress the likelihood of guilty plea on race of the defendant, race of the victim, and the interaction of the two, the coefficient on the interaction is not statistically different from zero.[19]

4 The data

To implement our test of racial bias we could not rely on any readily available dataset. In fact all existing data sets containing information on the race of the defendant and of the victim in capital cases have limited geographical and temporal coverage and – most importantly for our purposes – do not contain information on whether the capital sentence was reaffirmed in appeal. The only comprehensive dataset containing information on judicial errors in capital cases, that is the one used in Liebman et al's (2000) study and compiled by Fagan and Liebman (2002) –from now on FL–, does not contain information on the race of the defendant nor on

[19] Specifically, our estimated linear probability model is

$$Plea = \underset{(.054)}{.307} + \underset{(.063)}{.039}\, ND + \underset{(.049)}{.093}\, WV - \underset{(.072)}{.025}\, (ND * WV)$$

where ND is a dummy for non-white defendant, WV is a dummy for white victim, and standard errors are clustered at the county level. The results are similar if we include county fixed effects.

the race of the victim. We therefore constructed our dataset by examining each individual record in FL's data and searching for information on the race of the defendant and of the victim. For a detailed description of FL's data collection methodology and variable definition we refer the reader to Liebman et al. (2000). In what follows we start by briefly reviewing the characteristics and scope of FL's data, then discuss our search methodology and present some descriptive statistics.

Data coverage

FL's data is the first systematic collection of information on capital appeals in the modern death penalty era in the US. We use two datasets originally compiled by FL:[20]

- DIRECT APPEAL dataset (DA from now on): 4,546 state capital cases whose direct appeal decisions became final between January 1, 1973 and December 31, 1995.[21] This is the universe of all capital sentences that were reviewed on direct appeal by a state high court.
- HABEAS CORPUS dataset (HC from now on): 557 capital cases whose review was finalized by a federal Habeas Corpus court between January 1, 1973 and December 31, 1995. This is the universe of all capital sentences that were finally reviewed over this period.

After eliminating cases for which the name of the defendant could not be identified, we are left with a pool of 4,416 observations in DA and 531 observations in HC.[22]

[20] FL also compiled a "post-conviction" dataset which, however, is incomplete due to the fact that state post-conviction decisions are often not published and includes a selected subset of cases, *all of which resulted in a reversal*. In our analysis we therefore only employ the DA and HC datasets, which comprise the universe of available cases at those stages.

[21] "Became final" should be understood as "the highest state court with jurisdiction to review capital judgments in the relevant state must have taken one or two actions during the study period: (1) affirmed the capital judgment or (2) overturned the capital judgement (either the conviction or the sentence) on one or more grounds" (Liebman et al. (2000), p. 126).

[22] For 26 of the 557 cases in HC, either the sentence indicated in FL's data or the name of the defendant could not be found in Lexis-Nexis, hence we drop those cases. In the DA dataset, the sentence could not be found for 84 of the 4,546 available cases. Also, because some observations in the DA dataset correspond to multiple sentences for the same first degree trial and we want to record error once for each trial, we use one observation per appeal-trial pair and attribute an error if it was found in the first stage appeal (the one automatically granted by all States).

Definition of error

In FL's data, "error" is defined as such only if it led to the reversal of a capital conviction or sentence. If an error was discovered that did not result in a reversal, this is not coded as "error" in the database. For DA cases, a "serious error" that warrants reversal must have three characteristics. First, it must be "prejudicial", in the sense of affecting the outcome of the case (harmless errors do not lead to reversals). Second, it must have been "properly preserved", in the sense that the claim must have been asserted at the time and in the way required by the law. Third, obviously the error must have been discovered. At the federal HC stage, a serious error is reversible if, in addition to satisfying the three conditions required for DA, it violates the federal Constitution.[23]

Collection of the race variables

FL's data does not contain any information on the race of the defendant, nor of the victim. To collect such information, we relied on a number of sources including the Lexis Nexis database, the quarterly publication "Death Row USA" issued by the NAACP Legal Defense Fund, information from the Department of Corrections of several states, FBI UCR Supplementary Homicide Files, the CDC National Death Index, a number of web sites specialized in death penalty issues, plus communications with police officers and defense lawyers.[24] We assembled an almost complete data set for HC and a very extensive one for DA. To the best of our knowledge, this is the only dataset currently spanning two decades of trials for the entire US that contains information on race of defendant and victim.

Table A1 of the Online Appendix reports a tabulation of cases with missing information on the race of defendant and victim for the HC (Panel A) and the DA (panel B) datasets. In the HC data, we achieved almost full coverage of the defendant's race (3 missing cases out of 531), but we are missing information on the race of the victim in 20 cases out of the 528 for which we have the race of the defendant. Thus we have a usable sample of 508 cases out of 531. In the DA data, we have information on defendants' race for 4,146 cases out of 4,416 (94 percent of the sample), and on victims' race for 3,717 cases out of these 4,146 (90 percent). Appendix Tables A2 and A3 contain summary statistics on the share of missing observations by state and by year.

In Appendix Table A4 we try to gauge the extent of possible selection in the pattern of

[23] Some additional technical rules for reversibility at the HC stage are listed in Liebman et al. (2000), p. 130.

[24] A detailed description of the search procedure and of the sources is available from the authors upon request.

missing data for victims' race for all cases in which we have information on the defendant's race. We report the means of several variables related to defendant, victim, and crime characteristics for the sub sample in which we have information on victims' race (column 1) and the cases in which we don't (column 2). We conduct a t-test for the equality of means and report the p-values in column 3. Overall we do not find statistically significant differences across the two samples, with the exception of victim's gender in the Habeas Corpus data, but we show below that our results are robust to excluding female victims. This increases our confidence that there may not be a significant degree of selection on unobservables in the cases for which we have information on victim race.[25]

4.1 Descriptive statistics

Table 1 reports summary statistics on the main variables of interest in the HC (Panel A) and DA (Panel B) datasets. In the HC data the error rate, measured by the variable "Relief" as the fact that relief is granted at some stage of the review process, is .36. Regarding the race of the defendant, 51 percent of the cases involve white defendants, 44 percent African Americans, with the remaining fraction being mostly constituted by Hispanics. In contrast to the relatively even split between white and African American in the defendant's race, 83 percent of the cases involve a white victim, and only 13 percent an African American victim. Cases in which a non-white defendant killed a white victim constitute 36 percent of the total, as opposed to 3 percent for the cases in which a white defendant killed a non-white victim. The remaining cases are split between non-whites who killed non-whites (13 percent) and whites who killed whites (48 percent). The proportions are fairly similar for the DA sample: 37 percent of the sentences are overturned; 51 percent of the defendants are white, 41 percent are African American; 78 percent of the cases involve a white victim, as opposed to 17 percent with an African American victim. In this sample the share of non-white defendants who killed a white victim is .30.

5 Results

The test for racial bias we derived in section 3 required that, in the absence of bias against particular defendant/victim pairs, a difference in error rates for defendants of a given race depending on the victim's race should be maintained in the same direction for defendants of a

[25] One possible reason for the unbalance in the gender variable is greater media coverage of murders involving women, since media coverage makes it easier for us to find information on the race of the victim.

different race. To implement this test we use a rank order test reminiscent of Anwar and Fang's (2006) test for prejudice.

We hold constant the defendant's race r, and compare error rates across victim's race, $R \in \{W, M\}$. Let us denote with $\widehat{E(r,R)}$, the average error rate for cases in which a defendant of race r killed a victim of race R. We test the null $\widehat{E(r,W)} = \widehat{E(r,M)}$ (absence of racial bias) against the alternative $\widehat{E(r,W)} > \widehat{E(r,M)}$ (racial bias in favor of white victims) using the Z-statistic:

$$Z = \frac{\widehat{E(r,W)} - \widehat{E(r,M)}}{\sqrt{\frac{SVar_{rW}}{n_{rW}} + \frac{SVar_{rM}}{n_{rM}}}} \tag{6}$$

where $r \in \{w, m\}$; $SVar_{rR}$ is the sample variance of the error variable in the cases involving a defendant of race r and a victim of race R; and n_{rR} is the number of cases involving a defendant of race r and a victim of race R, with $R \in \{W, M\}$. Under the null hypothesis and given our large sample, Z has a standard normal distribution by the Central Limit Theorem. We will thus reject the null in favor of the alternative if expression (6) exceeds a threshold value z_α, where α is the significance level of the test and $\Phi(z_\alpha) = 1 - \alpha$. Performing this test separately for each defendant race allows us to test the prediction of our model, expression (5).

5.1 Main results

Table 2 contains the outcome of our test for the HC (Panel A) and the DA (Panel B) datasets and the main result of the paper. Each cell reports the average probability of error ("Relief") for a given combination of defendant's and victim's race, $\widehat{E(r,R)}$, and the associated standard error (in parenthesis). The p-values reported at the end of each row are those associated with test statistic (6). They represent the probability that, for a given defendant's race reported in that row, a difference in the error rates between white and minority victims at least as large as the one reported can be found, given that the null (of no racial bias against defendant/victim pairs) is true.

The first row of Table 2, Panel A shows that in cases involving a white defendant the average error rate is 36 percent if the victim is white and 47 percent if it is non-white, with a difference of -11 percentage points.[26] On the other hand, in cases involving a minority defendant, the error rate is 37.5 if the victim is white, and 28.4 percent if it is non-white, with a difference of +9 percentage points (or a 32 percent increase over the the non-white/non-white error rate).

[26] Note that, compared to other combinations, the number of cases involving white defendants and minority victims is quite small.

The differences in error rates across victim's race thus go in opposite directions depending on the defendant's race. For the cases involving minority defendants, we reject the null of no difference against the alternative of a positive difference in error rates with a p-value of .08; for cases involving white defendants we fail to reject the null against the alternative (p-value .80). Based on our rank order test, we therefore reject the hypothesis of no racial bias on defendant/victim racial pairs on behalf of trial courts.

In Panel B we show the same result for the DA sample. In the case of white defendants there is a -2 percentage points difference in error rates between white and non-white victim, though not statistically significant. In the case of minority defendants the difference is $+3$ percentage points (a 9 percent increase over the the non-white/non-white error rate of .35) and is significant at the 10 percent level. Again, we reject the null of no racial bias on defendant/victim racial pairs.

The rank order test implies that the difference in error rates across columns should go in the same direction for both rows in the previous tables (and in all those that follow). We shall see that for the case of minority defendants (second row) the first entry is always larger than the second entry almost always in a statistically significant way, while for white defendants (first row) the pattern of relative sizes of error rates typically goes in the opposite direction. Note also that the fact that the difference in errors is larger for the HC sample is consistent with the possibility that racial bias is eliminated in steps, that is, the DA courts may be less biased than the first degree courts but still biased relative to the final federal panels.

In Table 3 we find that the pattern of racial bias we uncovered is driven by Southern states. In the HC sample when we restrict the sample to sentences imposed by Southern courts we find a very large and statistically significant difference in errors for minority defendants who killed whites compared to minorities who killed non-whites: the difference is striking at 15.4 percentage points (a 66 percent increase over the non-white/non-white error rate of .23), with a p-value of 0.01. The difference goes in the opposite direction and is not significant for white defendants. A similar pattern emerges for DA cases in the South (Panel B), but with a smaller difference (3.3 percentage points for minority defendants, p-value 0.13). Again, the corresponding difference for white defendants has the opposite sign and is not significant.

When we conduct analogous tests for other regions we fail to reject the null for both HC and DA. In HC the error rate is higher with non-white than with white victims both if the defendant is white and if he is not. In DA the sign pattern in the differences is reversed compared to the South, but none of these differences is statistically significant. One caveat about the results for regions other than the South in the HC sample, however, is that they cover a substantially smaller number of cases compared to those for the South.

5.2 Potential confounding factors

We have suggested that our results on the rank order tests show a racial bias on behalf of the criminal justice system up to the level of the trial court. An alternative interpretation would be that the pattern of inequalities in error rates is generated by unobserved characteristics of the crime that are systematically correlated with different combinations of defendant and victim races. In the notation of our model, this would imply that the distribution of the evidence depends on the combination of races, i.e., $F_n^{r,R}(x)$ and $F_g^{r,R}(x)$. Although we cannot test for this possibility explicitly, in this section we aim at providing evidence on the importance of potentially omitted factors in two ways. First, we test whether the distribution of observable crime characteristics is systematically correlated with defendant/victim pairs. Second, we perform our rank order test conditioning on a set of available characteristics that might be correlated with the severity of the crime or the quality of available evidence.

Balance tests on crime characteristics

As we discussed in section 2, the choice between a death sentence and life imprisonment often rests in the relative weight given to aggravating and mitigating circumstances associated with the defendant and/or the crime. We start by examining whether the description of aggravating and mitigating circumstances differs systematically between murders involving different race combinations. Information on a rich set of aggravating and mitigating factors put forward in the trial is available from FL for the HC dataset (not for DA) and includes the following categories:

- Aggravating: heinous and atrocious crime; pecuniary gain motive; attempt to avoid arrest and hinder law enforcement; murder during a violent felony; murder by a person under prison sentence; previous felony convictions; killed a police, fireman, guard, or other public official; multiple victims; young or old victim; great risk of death to many people; cold, calculated, premeditated;

- Mitigating: young age; no prior record; extreme emotional distress; intoxication; mental retardation and limited capacity; deprived or abused background; good prison record; lack of intent; duress.

Table 4 reports the fraction of cases for which a given aggravating or mitigating circumstance is recorded when the defendant is white (columns 1-2) or non-white (columns 4-5) and the victim is white (columns 1, 4) or non-white (columns 2, 5). For each defendant's race, the p-value associated with the difference in average prevalence of a circumstance between the two races of victims is reported in columns 3 and 6. Most important, the "difference in differences" and its associated p-value are reported in columns 7 and 8. The latter is useful to test whether, in case a certain aggravating or mitigating circumstance is more common in murders involving, say, a white victim, this difference is similar across defendant races. For the purpose of our analysis, we would like the differences -if any- to be similar across defendant races because this would imply that the cases we are considering are relatively comparable in terms of aggravating or mitigating circumstances. Columns 9 to 16 in Table 4 report similar statistics, but for the Southern subsample.

As we can see in Panel B of Table 4, all of the mitigating circumstances are balanced across defendant/victim pairs (p-values in the range of $.4 - .9$ in column 8), and this is true also of most aggravating circumstances. In particular, the most common aggravating circumstances -"Heinous or atrocious crime"; "Murder during a violent felony", and "Previous felony conviction"- are perfectly balanced (p-values of .98, .83 and .40, respectively in the full sample and similarly in the South subsample). The only exceptions are the following. "Murder by a person under prison sentence" is more common among whites killing nonwhites in the full sample, but is balanced in the South. "Great risk of death to many people" is more common on same-race pairs, while "Cold, calculated and premeditated" is more common among mix-race pairs. The fact that the unbalance goes in opposite direction in the latter two cases makes it difficult to assess what the net effect may be. Overall, based on the results in Table 4, there does not seem to be a systematic pattern of aggravating or mitigating circumstances being more likely associated with cross racial defendant/victim cases, which increases our confidence in the interpretation of our test as detecting bias.

Aggravating and mitigating circumstances are particularly important when thinking about unobservables potentially correlated with race, because these circumstances are the ones that the court is required to weigh when deciding between a death sentence and a lesser one. We also collected information on other crime characteristics, related to the victim (gender and number) and the weapon used (knife, handgun, shotgun, rifle, or strangulation) for the HC and DA datasets. Furthermore, for HC cases information on the following circumstances is also available: defendant connected to the community where the crime occurred; murder occurred during a burglary, theft, robbery, kidnapping, rape or institutional killing. Appendix Table A5 reports a series of balance tests like the ones in Table 4 for these other variables. We find that in HC cases all the above crime characteristics are balanced. In the DA dataset, the type of weapon used is balanced, while variables related to the victim are not: same race pairs are more likely to invoke multiple victims and women. While ex ante it is not obvious in which direction this may bias our test, in the next section we show that our results are largely robust to excluding these categories of crimes.

Robustness: crime characteristics

We replicate our test for racial bias conditioning on several observable variables that characterize the crime. While this does not constitute direct evidence against the possibility that differences in unobservables are driving our results, it does shed some light on how important a similar concern may be.

In Table 5, Panel A we start from the HC sample, for which relatively detailed information on the crime was collected by FL. The leftmost part of the table uses the full sample, while the righmost part restricts the sample to Southern states. First we test whether the gender of the victim is a significant factor in our results. In the first panel of Table 5A we restrict the attention to cases in which none of the victims was female. We find higher error on minority defendants who killed white men than on those who killed non-white men (the difference is 10.6 percentage points, p-value .12 in the full sample, and 16 percentage points, p-value .05 in the South). The corresponding difference for white defendants is −11.6 and −14.4, not significant.

An aggravating factor that may be responsible for the results we find is the presence of multiple victims. Restricting the analysis to homicides with only one victim shows a difference of 9 percentage points (15 in the South) for nonwhite defendants who killed white versus nonwhite victims (p-values .09 and .02, respectively) and an insignificant difference on the opposite direction for white defendants.

The remaining of Table 5A reports results for other crime characteristics which are available only for the HC sample. A possible aggravating factor is the fact that the defendant killed a policeman, or fireman, or guard, or other public official. One could conjecture that crimes involving minority defendants and white victims are more represented in this category and that this generates the higher error rates we find. When we repeat the analysis considering cases in which none of the victims was one of these public officials (indicated as "no police victim" in the table), we find no significant difference in error rates for white defendants, and a difference of 13 percentage points for minority defendants, with p-value .03, in the full sample. In the South the difference is even larger (18 percentage points) and significant at the 1 percent level.

So our results are not driven by this types of murders.

A commonly held view is that cases in which an outsider who does not know the victim commits a murder are perceived as particularly threatening and sanctioned with more severely. Perhaps cases involving minority defendants and white victims fall disproportionately in this category. In the fourth panel of Table 5A we examine the subset of cases where the defendant was not connected to the community where the crime occurred, according to the information recorded in FL. These cases should be relatively comparable along this dimension. Our results show that in the full sample the likelihood of error is 15 percentage points higher for minority defendants who killed white victims compared to minority defendants whose victims were not white (p-value .03). In the South the difference is 21 percentage points, significant at the 1 percent level. The corresponding difference in error rates for white defendants has the opposite sign and is not statistically significant. In the fifth panel we consider the subset of cases where the victims were not "high status", as classified by FL. We find a difference of 9 percentage points (p-value 0.12) in the full sample and 16 percentage points (p-value .02) in the South for the combination of minority defendants and white victim, and no difference for the opposite combination.

Another way to gauge the role of potentially omitted crime characteristics is to confine our attention to murders that occurred in "similar" environmental conditions. In particular, in the sixth panel of Table 5A we consider murders committed during a robbery. The likelihood of judicial error is 18 percentage points higher for minority defendants who killed at least one white victim during a robbery compared to minority defendants whose victims were all non-white, and is significant at the 5 percent level. Results are even stronger in the South, with a difference of 32 percentage points, significant at the 1 percent level. The difference for white defendants is in the opposite direction and not statistically significant. Finally, when we restrict the sample to cases that are similar in the sense of being classified as "felony murders", we find again a higher error rate for nonwhite defendants who killed white victims (13 percentage points in the full sample, 22 in the South, with p-values .07 and .004, respectively), and no corresponding difference for white defendants.

We have less information on crime characteristics for the DA compared to the HC sample. In Table 5B we begin by testing whether the gender of the victim is a significant factor in our results. In the first panel we restrict the attention to cases in which none of the victims was female. We find 8 percentage points higher error on minority defendants who killed white men than on those who killed nonwhite men (p-values .01 and .02 in the full sample and in the South, respectively). The corresponding difference for white defendants is in the opposite direction and not significant. In the second panel we restrict the analysis to homicides with only one victim we find a difference of 3 percentage points for nonwhite defendants who killed white vs. nonwhite victims, both in the full sample and in the South, but the p-values increase to .16 and .26, respectively.

Robustness: legal assistance

We now analyze whether differences in error rates are due to unequal quality of legal assistance of the defendant. A possible interpretation of our finding is that minority defendants who killed a white victim receive systematically worse legal assistance compared to minority defendants who killed a minority victim. This would actually be another source of racial bias, which we discussed in section 3.3 and which would "distort" the distribution of the signal for given characteristics of the crime. Note that if a minority defendant received a worse defense regardless of the race of the victim, this would not invalidate our test nor change the interpretation of the results.

In Table 6 we repeat our tests restricting the sample to cases that are relatively similar in terms of some trial characteristics. We can only do this for HC cases because no trial characteristic is available in the DA dataset. As a proxy for the quality of legal assistance at the trial stage we use the fact that "ineffective assistance of counsel" in the guilt and sentencing phase was included among the claims for relief. We start by restricting the sample to 220 HC cases in which ineffective assistance of counsel was not raised among the claims in the appeal. In this subset of cases the difference in error rates for minority defendants who killed a white vs. a non-white victim is 19 percentage points (p-value .04) in the full sample, and 29 percentage points (p-value .003) in the South. Comparing these results to those in Tables 2 and 3 suggests that variation in the quality of legal assistance across racial combinations of defendants and victims may actually lead us to *underestimate* the extent of bias.

In the remaining parts of Table 6, we consider the subset of cases in which "prosecutor's suppression or withholding of evidence or other prosecutorial misconduct" was not raised among the claims, nor was "improper interrogation", that is, there was no involuntary confession or guilty plea or request for attorney denied. In both sub-samples the order of magnitude of the differences in error rates and the significance level remain comparable to those of Tables 2 and 3, and the rank order test rejects the null of absence of racial bias according to our model in all cases except for the second panel in the full sample, where the p-value for nonwhite defendants increases to .16.

Note that although the above variables seem reasonably good proxies for the quality of legal assistance, some of them reflect discretionary choices on behalf of the defense in the appeal process (e.g., which claims to present) and in this sense they may not be fully objective. Nonetheless, we take the evidence in Table 6 as suggestive that differences in the quality of legal assistance are not entirely responsible for our results.

Robustness: reverse discrimination of appeal courts

So far we have assumed that the appeal courts are unbiased. As we mentioned above, errors uncorrelated with pairs of defendant/victim races are irrelevant for our empirical test. If the appeal court is biased in the same direction of the trial court, our test will underestimate the extent of racial bias because the (biased) appeal court will reverse the trial court decision less often than an unbiased court would do. The challenge for us would arise from a bias in the opposite direction, namely if the appeal court were inclined to give relief more often than an unbiased court would do. Note that a simple bias of the appeal courts in favor of black defendants (for example on the ground that they are on average poorer and may not be able to afford good legal assistance) would not invalidate our tests of racial bias. What would be problematic for us is a situation where the bias is linked to a particular combination of defendant/victim race, e.g. if the appeal court rules systematically more in favor of non-white defendants who killed white victims. Although we cannot rule this out a priori, we test the plausibility of this scenario by exploiting information on the political orientation of appeal judges. We conjecture that, if a bias in favor of minorities who killed white victims existed, this would more likely be found among liberal judges than among conservative ones, characteristics which we assume to be correlated with party affiliation. Thus we repeat our analysis conditioning on party affiliation of the appeal judges.

Let's begin with the HC sample. For each sentence, we collected the names of the judges who served on the appeal court that decided on that sentence, and recovered information on these judges from the Biographical Directory of Federal Judges available from the Federal Judicial center. This directory contains biographical information on all judges that served on U.S. District Courts, the U.S. Courts of Appeals, the Supreme Court and the U.S. Circuit Courts since 1789. We recorded the year in which each judge was appointed to the relevant court and classified the political orientation of the judge as "Republican" if he or she was appointed under a Republican president and "Democratic" if he or she was appointed under a Democratic president. If our results were driven by "reverse discrimination" on behalf of appeal judges, we should not find discrimination (or find it to a lesser extent) when we look at courts that are predominantly composed of republican judges.

Table 7A reports the results for the subset of HC cases where the majority (first panel) or the totality (second panel) of the judges were appointed under a Republican president. The leftmost part of the table employs the full sample, while the righmost part restricts the analysis to the South. Both sets of results are consistent with our earlier findings, and indicate a higher likelihood of relief for nonwhite defendants who killed white victims. When we consider appeal courts where a majority of the judges are Republican (first panel), the magnitude of the difference in error rates is 7 percentage points in the full sample and 12 percentage points in the South (p-values .20 and .08, respectively). This differences increase to 13 and 16 percentage points when we restrict our test to courts that are entirely composed of republican-appointed judges (second panel, p-values .09 and .06).

In the third panel of Table 7A we consider the possibility that the political climate in a given year may affect relief rates, and restrict the sample to Habeas Corpus appeal sentences that occurred under a Republican administration. We find a difference of 17 percentage points in the full sample and 25 percentage points in the South for non-white defendants who killed white victims, both significant at the 1 percent level. In all three panels the corresponding difference for white defendants is in the opposite direction and not significant.

In Table 7B we conduct a similar exercise for the DA dataset. In this case we have available both the party affiliation of the Direct Appeal judges and the measure of judges' ideology proposed by Brace, Langer and Hall (2000), which they label PAJID.[27] The first panel of Table 7B shows that when we restrict the sample to first stage appeals decided by courts in which at least 50 percent of the judges were Republican, error rates on nonwhite defendants who killed white victims are higher than on those who killed nonwhite victims (7.5 and 24 percentage points in the full sample and in the South, respectively, with p-values .05 and .03). For white defendants, error rates are virtually the same across victim races.

In the remaining panels we rely on the continuous measure of ideology proposed by Brace et al. (2000) and define as "conservative" judges whose ideology score falls in the top 50 percent of the distribution of PAJID. The second panel restricts the sample to courts whose median member (in terms of ideology) is "conservative", while the third does the same but with reference to the Chief Justice. In both cases we find that the direction and the magnitude of the differences in error rates are comparable to our main results in Tables 2 and 3, though we lose statistical significance. Furthermore, this result does not depend on the particular cutoff

[27] Essentially PAIJD measures judges' ideology on a scale from conservative to liberal based upon party affiliation modified by a set of criteria allowing for differences across states. We match this measure to reflect the composition of the state appeal court the year in which the appeal sentence was issued.

for the definition of "conservative". Figure 5 shows that the positive difference in error rates for minority defendants who killed white versus nonwhite victims holds for each and every quartile of the distribution of PAJID, indicating that our main result is not driven by the ideological orientation of the court. The corresponding differences for white defendants are instead sometimes positive, sometimes negative, and vary by quartile.

To sum up, we find no evidence that left liberal leaning judges are those who "correct" more mistakes in pairs involving minority defendants and white victims. In fact we find that our results hold strong when we restrict the sample to relatively conservative appeal courts. Thus, we find no obvious evidence of reverse discrimination by higher courts.

6 Conclusions

This paper proposes a test for racial bias in capital sentencing in the US over the period 1973-1995. We use the share of judicial errors in first degree sentencing as an indicator of racial bias of such courts. Using an originally collected dataset, we uncover a bias against minority defendants killing white victims. The bias is present, according to our test, only in Southern States. More precisely, according to our interpretation first degree courts tend to place less weight on the possibility of condemning an innocent in cases of minority defendants with one or more white victims relative to minority defendants who did not kill whites. The same does not hold for white defendants. This result is not explained by differences in observable characteristics of the crime or of the trial, nor by the ideological orientation of appeal courts.

Appendix - Proof that the error rate is decreasing in x^*

To simplify the notation, here we omit subscripts and superscripts related to race, i.e. we write x instead of x_{rR}, π instead of π^r, and f_g, f_n instead of f_g^r, f_n^r. The error rate (4) is

$$E(r,R) = 1 - \frac{1}{1 + \frac{1-\pi}{\pi}\frac{1-F_n(x^*)}{1-F_g(x^*)}}. \tag{A1}$$

Expression (A1) is decreasing in x^* if and only if $\frac{1-\pi}{\pi}\frac{1-F_n(x^*)}{1-F_g(x^*)}$ is decreasing in x^*. Taking the first derivative of this product with respect to x^*, its sign is the same as the sign of:

$$\frac{1}{\left[1 - F_g\left(x^*\right)\right]^2}\left[\int_{x^*}^{1} f_g(x^*) f_n(x) dx - \int_{x^*}^{1} f_g(x) f_n(x^*) dx\right]. \tag{A2}$$

To see that the term in square brackets is negative, recall that from MLRP we know that $\frac{f_g(x)}{f_n(x)} > \frac{f_g(x^*)}{f_n(x^*)}$ for any $x > x^*$. Because the integrals in (A2) are calculated for $x \in (x^*, 1]$, then in this range $f_g(x^*)f_n(x) < f_g(x)f_n(x^*)$, hence (A2) is negative.

References

[1] Abrams D. M. Bertrand and S. Mullainathan (2012), "Do Judges Vary in their Treatment of Race?", *Journal of Legal Studies,* 41(2), 347-383.

[2] Antonovics, K. and B. Knight (2009), "A New Look at Racial Profiling: Evidence from the Boston Police Department", *Review of Economics and Statistics*, 91(1), 163-175.

[3] Anwar, S., P. Bayer and R. Hjalmarsson (2012), "The Impact of Jury race in Criminal Trials", *Quarterly Journal of Economics*, 127(2), 1017-1055.

[4] Anwar, S. and H. Fang (2006), "An Alternative Test of Racial Profiling in Motor Vehicle Searches: Theory and Evidence", *American Economic Review*, 96(1), 127-151.

[5] Argys, L. and N. Mocan (2004), "Who Shall Live and Who Shall Die? An Analysis of Prisoners on Death Row in the United States", *Journal of Legal Studies*, 33(2), pp. 255-92.

[6] Barnes, K., D. Sloss, and S. Thaman (2008), "Life and Death Decisions: Prosecutorial Discretion and Capital Punishment in Missouri", unpublished.

[7] Becker G. (1957) *The Economics of Discrimination*, Chicago: University of Chicago Press

[8] Blume, J., T. Eisenberg and M.T. Wells (2004), "Explaining Death Row's Population and Racial Composition", *Journal of Empirical Legal Studies*, 1(1), 165-207.

[9] Brace, P., L. Langer and M. Gann Hall (2000), "Measuring the Preferences of Supreme Court Judges", *Journal of Politics,* May, 387-413.

[10] Coyne, R. and L. Entzeroth (2006), *Capital Punishment and the Judicial Process,* Carolina Academic Press.

[11] Donohue III, J. and J. Wolfers (2005), "Uses and Abuses of Empirical Evidence in the Death Penalty Debate", *Stanford Law Review*, 58(3), 791-846.

[12] Erlich, I. (1975), "The Deterrent Effect of Capital Punishment: A Matter of Life and Death", *American Economic Review*, 65(3), 397-417.

[13] Fagan, J. and J. Liebman (2002), *Processing and Outcome of Death Penalty Appeals After Furman v. Georgia, 1973-95.* ICPSR version 3468. New York City, NY: Columbia School of Law and Mailman School of Public Health [producers]. Ann Arbor, MI: ICPSR [distrib.].

[14] Gennaioli N. and A Shleifer (2007), "The Evolution of Common Law", *Journal of Political Economy,* 115(1) 43-67.

[15] Gross, S.R. and R. Mauro (1984), "Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization", *Stanford Law Review*, 37, 27-153

[16] Grossman G. and M. Katz (1983), "Plea Bargain and Social Welfare", *American Economic Review,* 73(4), 749-57

[17] Humphrey, J. and T. J. Fogarty (1987), "Race and Plea Bargained Outcomes: A Research Note," *Social Forces*, 66, 176-85.

[18] Katz, L., S. Levitt and E. Shustorovich (2003), "Prison Conditions, Capital Punishment, and Deterrence", *American Law and Economics Review*, 5(2), 318-343.

[19] Knowles J., N. Persico and P. Todd (2001), "Racial Bias in motor vehicles searches:Theory and Evidence," *Journal of Political Economy*, 109, 203-29.

[20] Iyengar, R. (2007), "An Analysis of Attorney Performance in the Federal Indigent Defense System", NBER Working Paper, n. 13187.

[21] Liebman, J., J. Fagan and V. West (2000), *A Broken System: Error Rates in Capital Cases, 1973-1995*; Columbia University.

[22] Radelet, M. and G. Pierce (1985), "Race and Prosecutorial Discretion in Homicide Cases",19 *Law and Society Review*, 587, 601–15.

[23] Reinganum, J. (1988), "Plea Bargain and Prosecutorial Discretion", *American Economic Review, 78 (4), 713-27.*

[24] U.S. Department of Justice, Bureau of Justice Statistics. *Murder Cases in 33 Large Urban Counties in the United States, 1988.* Distributed by ICPSR 9907, 1996.

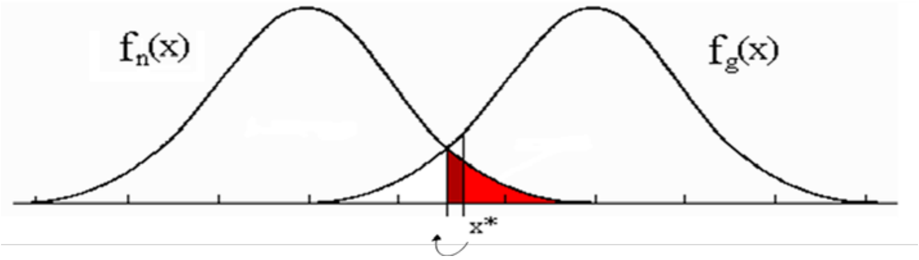


Figure 1: Error rate for given defendant's race

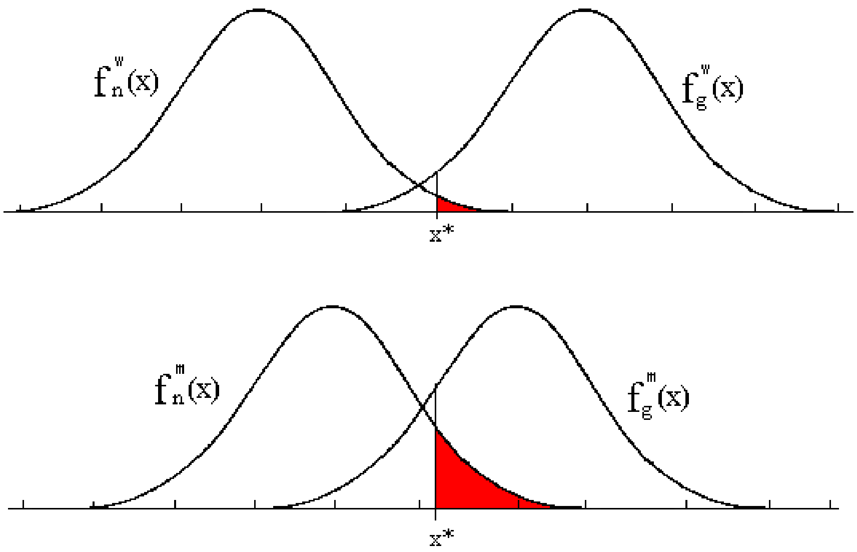


Figure 2: Error rate for given victim's race

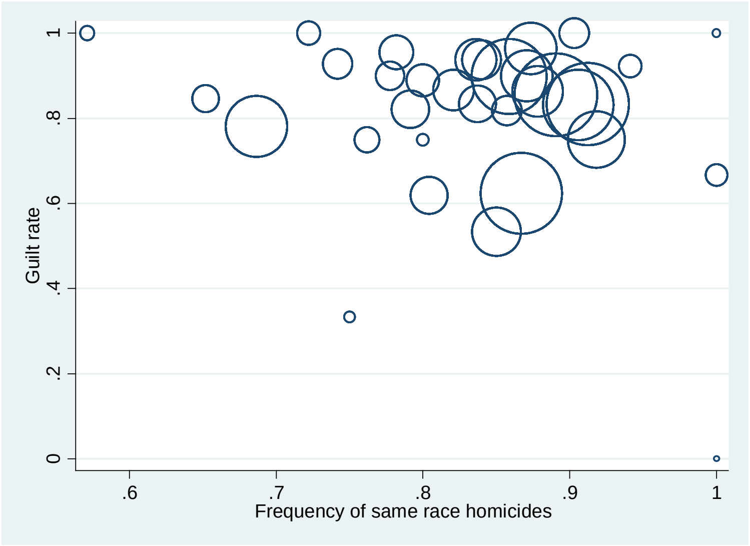


Figure 3: Guilt rate and frequency of homicides

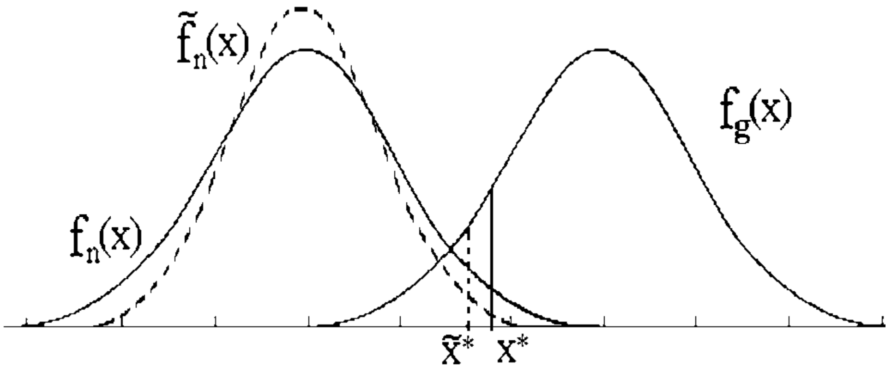
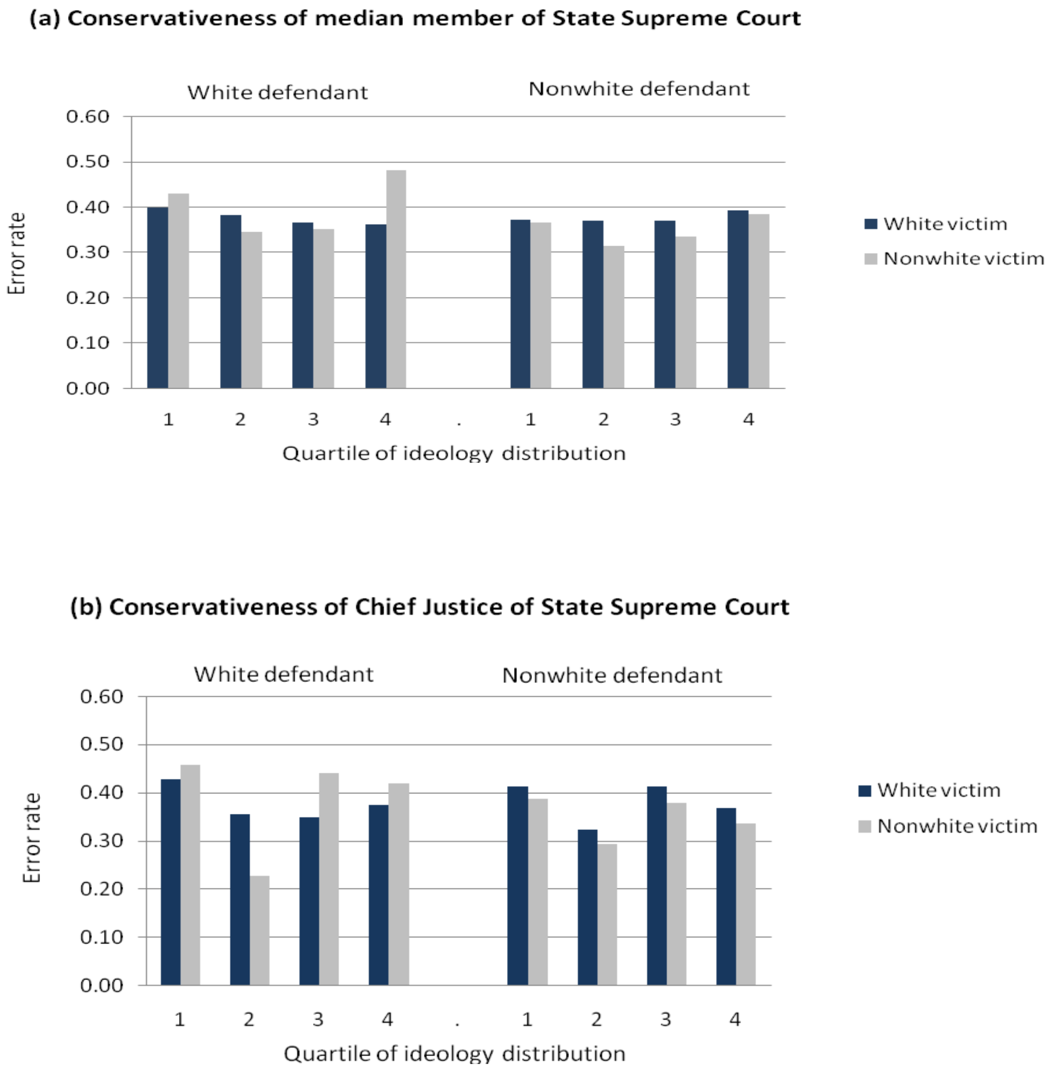


Figure 4: Bias in the collection of evidence


(a) Conservativeness of median member of State Supreme Court

(b) Conservativeness of Chief Justice of State Supreme Court

Figure 5: Error rate and ideology of Appeal Courts

Table 1: Error rate for given defendant’s race

	Panel A: Habeas Corpus			*Panel B: Direct Appeal*		
	No. Obs	Mean	Std. Dev.	No. Obs	Mean	Std. Dev.
Relief	508	0.36	0.48	3717	0.37	0.48
African American defendant	508	0.44	0.50	3717	0.41	0.49
White defendant	508	0.51	0.50	3717	0.51	0.50
African American victim	508	0.13	0.34	3717	0.17	0.38
White victim	508	0.83	0.37	3717	0.78	0.41
White def., Non-white vict.	508	0.03	0.18	3717	0.03	0.17
Non-white def., White vict.	508	0.36	0.48	3717	0.30	0.46
White def., White vict.	508	0.48	0.50	3717	0.48	0.50
Non-white def., Non-white vict.	508	0.13	0.34	3717	0.19	0.39

Table 2: Error rate by defendant and victim race

	Panel A: Habeas Corpus				**Panel B: Direct Appeal**			
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.358	0.471	0.809	260	0.373	0.395	0.673	1908
	(0.031)	(0.125)			(0.011)	(0.046)		
Non-white	0.375	0.284	0.083	251	0.377	0.347	0.097	1809
	(0.036)	(0.055)			(0.015)	(0.018)		
N.obs	427	84			2911	806		

Note: Standard errors of the means in parenthesis

Table 3: Error rate by race and region

Panel A: Habeas Corpus

	South				**Other regions**			
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.350	0.455	0.741	208	0.391	0.500	0.678	52
	(0.034)	(0.157)			(0.073)	(0.224)		
Non-white	0.387	0.232	0.012	219	0.286	0.545	0.917	32
	(0.038)	(0.057)			(0.101)	(0.157)		
N.obs	360	67			67	17		

Panel B: Direct Appeal

	South				**Other regions**			
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.397	0.443	0.785	1305	0.322	0.286	0.324	603
	(0.014)	(0.056)			(0.020)	(0.077)		
Non-white	0.409	0.376	0.134	1234	0.288	0.304	0.657	575
	(0.017)	(0.024)			(0.026)	(0.028)		
N.obs	2048	491			863	315		

Note: Standard errors of the means in parenthesis

Table 4: Balance test on aggravating and mitigating circumstances

	Full sample								South							
	White defendant			Non-white defendant			Diff-in-diff		White defendant			Non-white defendant			Diff-in-diff	
	White victim	*Non-white victim*	*Diff=0 (p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	ΔΔ	*(p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	ΔΔ	*(p-val)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Panel A: Aggravating circumstances																
Heinous, atrocious	0.44	0.41	0.79	0.36	0.33	0.60	0.00	0.98	0.44	0.54	0.48	0.38	0.34	0.58	-0.15	0.38
Pecuniary gain	0.16	0.18	0.86	0.13	0.13	0.88	-0.01	0.93	0.15	0.27	0.26	0.11	0.11	0.98	-0.12	0.28
Avoid arrest, hinder law enforcement	0.15	0.06	0.29	0.12	0.13	0.70	0.11	0.26	0.15	0.09	0.61	0.11	0.14	0.55	0.09	0.46
Murder during violent felony	0.40	0.29	0.39	0.46	0.33	0.06	-0.03	0.83	0.44	0.36	0.61	0.47	0.36	0.15	-0.03	0.85
Murder by person under prison	0.09	0.29	0.01	0.08	0.09	0.87	-0.20	0.02	0.07	0.09	0.80	0.07	0.09	0.62	0.00	0.99
Previous felony conviction	0.23	0.29	0.58	0.22	0.18	0.47	-0.10	0.40	0.20	0.27	0.58	0.19	0.16	0.59	-0.10	0.46
Killed police, fireman or guard	0.01	0.00	0.65	0.06	0.04	0.63	0.00	0.94	0.02	0.00	0.68	0.05	0.05	0.92	0.02	0.76
Killed other public official	0.06	0.06	0.98	0.06	0.04	0.63	-0.02	0.80	0.07	0.09	0.75	0.07	0.04	0.37	-0.06	0.50
Multiple victims	0.02	0.00	0.51	0.02	0.03	0.73	0.03	0.46	0.01	0.00	0.74	0.01	0.04	0.11	0.04	0.29
Young or old victim	0.00	0.00	.	0.01	0.01	0.81	0.00	0.86	0.00	0.00	.	0.01	0.02	0.77	0.01	0.85
Great risk of death to many people	0.13	0.00	0.11	0.05	0.10	0.12	0.19	0.03	0.14	0.00	0.18	0.04	0.11	0.05	0.21	0.04
Cold, calculated, premeditated	0.04	0.12	0.15	0.06	0.01	0.17	-0.12	0.05	0.04	0.18	0.02	0.06	0.02	0.19	-0.19	0.01
Panel B: Mitigating circumstances																
Age	0.04	0.00	0.42	0.06	0.06	0.89	0.04	0.49	0.03	0.00	0.53	0.06	0.07	0.82	0.04	0.55
No prior record	0.05	0.06	0.80	0.05	0.06	0.76	0.00	0.96	0.04	0.09	0.43	0.05	0.07	0.55	-0.03	0.70
Extreme emotional distress	0.03	0.06	0.57	0.01	0.01	0.81	-0.02	0.62	0.02	0.09	0.21	0.01	0.02	0.77	-0.06	0.23
Intoxication	0.02	0.00	0.60	0.01	0.01	0.46	0.03	0.41	0.02	0.00	0.68	0.01	0.02	0.44	0.03	0.47
Mental retardation, limited capacity	0.03	0.06	0.49	0.01	0.00	0.54	-0.04	0.35	0.03	0.09	0.28	0.00	0.00	.	-0.06	0.17
Deprived/abused background	0.02	0.00	0.51	0.01	0.01	0.46	0.03	0.34	0.02	0.00	0.64	0.01	0.02	0.44	0.03	0.44
Good prison/jail record	0.004	0.00	0.79	0.01	0.01	0.80	0.01	0.76	0.01	0.00	0.81	0.01	0.02	0.77	0.01	0.76
Lack of intent	0.004	0.00	0.79	0.00	0.00	.	0.00	0.75	0.00	0.00	.	0.00	0.00	.	0.00	.
Duress	0.004	0.00	0.79	0.00	0.00	.	0.00	0.75	0.01	0.00	0.81	0.00	0.00	.	0.01	0.76

Table 5: Error rate conditional on crime characteristics

Panel A: Habeas Corpus

	Full sample				South			
	No female victim							
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.384	0.500	0.769	124	0.356	0.500	0.264	96
	(0.046)	(0.151)			(0.051)	(0.224)		
Non-white	0.363	0.257	0.117	137	0.370	0.214	0.049	120
	(0.048)	(0.075)			(0.051)	(0.079)		
N.obs	214	47			182	34		
	Single victim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.359	0.471	0.808	254	0.354	0.455		
	(0.031)	(0.125)			(0.034)	(0.157)	0.266	206
Non-white	0.383	0.292	0.089	245	0.389	0.241		
	(0.036)	(0.057)			(0.038)	(0.059)	0.017	216
N.obs	417	82			357	65		
	No police victim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.367	0.438	0.703	242	0.365	0.400	0.416	191
	(0.032)	(0.128)			(0.036)	(0.163)		
Non-white	0.391	0.262	0.030	222	0.399	0.216	0.005	194
	(0.039)	(0.057)			(0.041)	(0.058)		
N.obs	387	77			324	61		
	Defendant not connected to community where crime occurred							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.384	0.500	0.790	191	0.362	0.500	0.211	159
	(0.037)	(0.139)			(0.040)	(0.167)		
Non-white	0.413	0.265	0.027	187	0.432	0.220	0.004	166
	(0.042)	(0.064)			(0.044)	(0.065)		
N.obs	315	63			274	51		
	No high status victim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.398	0.385	0.463	209	0.386	0.375	0.476	166
	(0.035)	(0.140)			(0.039)	(0.183)		
Non-white	0.397	0.309	0.124	186	0.402	0.244	0.024	162
	(0.043)	(0.063)			(0.046)	(0.065)		
N.obs	327	68			275	53		
	Robbery							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.342	0.600	0.848	81	0.292	0.600	0.111	70
	(0.055)	(0.245)			(0.057)	(0.245)		
Non-white	0.430	0.250	0.045	103	0.425	0.105	0.000	92
	(0.056)	(0.090)			(0.058)	(0.072)		
N.obs	155	29			138	24		
	Felony							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.344	0.429	0.658	129	0.305	0.429	0.275	112
	(0.043)	(0.202)			(0.045)	(0.202)		
Non-white	0.385	0.257	0.073	152	0.387	0.167	0.004	136
	(0.045)	(0.075)			(0.048)	(0.069)		
N.obs	239	42			211	37		

Table 5: Error rate conditional on crime characteristics (cont'd)

Panel B: Direct Appeal

	Full sample				South			
	No female victim							
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.358	0.403	0.768	831	0.393	0.463	0.164	576
	(0.017)	(0.058)			(0.021)	(0.068)		
Non-white	0.413	0.333	0.008	926	0.444	0.360	0.022	638
	(0.020)	(0.026)			(0.024)	(0.034)		
N.obs	1343	414			963	251		
	Single victim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.389	0.374	0.389	1419	0.414	0.418	0.474	1012
	(0.013)	(0.051)			(0.016)	(0.061)		
Non-white	0.388	0.361	0.163	1382	0.412	0.391	0.264	1001
	(0.016)	(0.022)			(0.019)	(0.028)		
N.obs	2220	581			1639	374		

Note: Standard errors of the means in parenthesis

Table 6: Error rate conditional on trial characteristics

Habeas Corpus

	Full sample				South			
	Victim's race				Victim's race			
	Ineffective assistance of counsel not in any claim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.458	0.571	0.708	114	0.471	0.500	0.461	91
	(0.048)	(0.202)			(0.054)	(0.289)		
Non-white	0.417	0.227	0.037	106	0.444	0.158	0.003	91
	(0.054)	(0.091)			(0.059)	(0.086)		
N.obs	191	29			159	23		
	Prosecutorial suppression/witholding of evidence not in any claim							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.383	0.600	0.902	177	0.366	0.667	0.081	137
	(0.038)	(0.163)			(0.042)	(0.211)		
Non-white	0.383	0.300	0.162	173	0.387	0.235	0.040	153
	(0.042)	(0.073)			(0.045)	(0.074)		
N.obs	300	50			250	40		
	Improper interrogation not in any claim							
	White	*Non-white*	*p-values*	*N.obs*				
White	0.362	0.500	0.849	234	0.348	0.500	0.187	188
	(0.033)	(0.129)			(0.036)	(0.167)		
Non-white	0.387	0.283	0.070	223	0.396	0.220	0.007	194
	(0.038)	(0.059)			(0.041)	(0.059)		
N.obs	381	76			322	60		

Note: Standard errors of the means in parenthesis

Table 7: Possible bias of Appeal Courts

Panel A: Habeas Corpus

	Full sample				South			
	Majority of final federal panel Republican							
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.355	0.455	0.730	149	0.361	0.500	0.728	114
	(0.041)	(0.157)			(0.046)	(0.224)		
Non-white	0.284	0.216	0.200	146	0.313	0.194	0.084	127
	(0.043)	(0.069)			(0.048)	(0.072)		
N.obs	247	48			204	37		
	All judges appointed under Republican							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.250	0.500	0.799	40	0.212	0.500	0.716	35
	(0.073)	(0.289)			(0.072)	(0.500)		
Non-white	0.216	0.083	0.096	63	0.244	0.083	0.063	57
	(0.058)	(0.083)			(0.065)	(0.083)		
N.obs	87	16			78	14		
	Sentence under Republican administration							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.347	0.438	0.751	183	0.333	0.400	0.654	148
	(0.037)	(0.128)			(0.040)	(0.163)		
Non-white	0.402	0.234	0.013	179	0.412	0.158	0.000	157
	(0.043)	(0.062)			(0.045)	(0.060)		
N.obs	299	63			257	48		

Panel B: Direct Appeal

	Majority of State Supreme Court Republican							
	Victim's race				Victim's race			
Defendant's race	*White*	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.285	0.280	0.477	428	0.365	0.333	0.463	66
	(0.023)	(0.092)			(0.061)	(0.333)		
Non-white	0.292	0.217	0.051	368	0.419	0.176	0.032	48
	(0.031)	(0.034)			(0.090)	(0.095)		
N.obs	619	177			94	20		
	Median ideology of State Supreme Court conservative							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.364	0.406	0.745	1151	0.377	0.421	0.740	974
	(0.015)	(0.062)			(0.016)	(0.066)		
Non-white	0.380	0.354	0.199	1050	0.386	0.367	0.286	957
	(0.018)	(0.025)			(0.019)	(0.027)		
N.obs	1781	420			1555	376		
	Chief justice of State Supreme Court conservative							
	White	*Non-white*	*p-values*	*N.obs*	*White*	*Non-white*	*p-values*	*N.obs*
White	0.363	0.426	0.845	1155	0.380	0.443	0.829	993
	(0.015)	(0.060)			(0.016)	(0.064)		
Non-white	0.385	0.352	0.142	1072	0.391	0.366	0.221	988
	(0.018)	(0.025)			(0.019)	(0.026)		
N.obs	1790	437			1584	397		

Note: Standard errors of the means in parenthesis

Online Appendix - not for publication

A. Endogenizing the behavior of the criminal

This section of the appendix discusses an extension of our model where the probability of guilt, π^r is endogenously derived from an optimization problem of the criminal. We start by presenting the case where the race of the victim is given and known to the criminal, then discuss cases in which the race of the victim is ex ante unknown or in which it can be chosen by the criminal.

A1. Problem of the criminal

An individual who is considering whether to commit a capital crime trades of the expected benefits and costs of it. Let the benefit of committing the crime and getting away with it be $b > 0$; think of it as the money stolen from a bank (with a killing during the robbery) or the pleasure of killing an enemy. The cost of being sentenced to death having committed the crime is $c_g > 0$. The cost of being sentenced to death not having committed the crime is c_n, with $0 < c_n < c_g$.[1] All of the above b, c_g and c_n are public information. For an individual the cost of committing a crime, which may include the moral cost, is v and it is drawn from a distribution $\Im^r(v)$ with support in $\mathbb{R}^+$. The court knows the distribution but not the realization of v which is known only to the individual. We allow the distribution of costs to differ across races, thus allowing a higher propensity of minorities to commit crimes. The individual chooses whether to commit a crime taking into account the likelihood of being convicted and takes x_{rR} as given since it is chosen by the court.

In certain types of crimes the defendant cannot choose the victim and therefore his or her race. One example is a bank robbery with the killing of guards, whose race was unknown to the criminals ex ante. In a second type of crime the defendant wants to kill, say, a relative, in which case he also cannot choose the race of the victim but the race of the victim is known ex ante. In a third type of crime the defendant can choose the race of the victim, say in a rape with murder. We present the second case here. The others are discussed below.

The expected payoff from the crime for an individual with race r and a certain realization v is given by:

[1] Remember that by assumption there are no mistakes in the final ruling of higher courts, therefore no innocent individual is executed. Thus the cost c_n represents the costs of being on death row until the first sentence is reversed.

$$\left[1 - F_g^r\left(x_{rR}\right)\right]\left[-v - c_g\right] + F_g^r\left(x_{rR}\right)\left[b - v\right]$$

The first term represents the cost of being convicted, the second term the benefit of getting away with the crime. The expected payoff from not committing that crime is:

$$-\left[1 - F_n^r\left(x_{rR}\right)\right] c_n.$$

Comparing costs and benefits, an individual commits a crime if and only if:

$$v \leqslant F_g^r\left(x_{rR}\right) b - \left[1 - F_g^r\left(x_{rR}\right)\right] c_g + \left[1 - F_n^r\left(x_{rR}\right)\right] c_n \equiv v^{r*}(x_{rR}) \tag{1}$$

Thus $v^{r*}\left(x_{rR}\right)$ is the threshold of individual cost v below which an individual of race r chooses to commit a crime against a victim of race R. Define:

$$Prob\left(v \leqslant v^{r*}(x_{rR})\right) = \Im^r\left(v^{r*}(x_{rR})\right) \equiv \pi^r\left(x_{rR}\right) \tag{2}$$

as the probability of guilt, i.e. the probability that the realization of v is low enough so that a crime is committed. Note that if the court applied a different standard of proof depending on the race of the victim, e.g. $x_{m,W} < x_{m,M}$, then potential criminals would internalize that and ceteris paribus we would observe fewer crimes involving m, W pairs than m, M pairs. With an endogenous probability of committing a crime the potential criminal would incorporate in his calculations the courts' behavior. With exogenous probabilities, of course he would not.

The equilibrium of the model is given by (**??**) in the text together with:

$$\pi^r\left(x_{rR}^*\right) = \Im^r(v^{r*}(x_{rR}^*)) \tag{3}$$

By Brouwer's fixed point theorem an equilibrium exists. The proof of uniqueness is below.

A.2 Proof of uniqueness of the equilibrium

To simplify the notation, in sections A.2 and A.3 we omit subscripts and superscripts related to race, i.e. we write x instead of x_{rR}, v^* instead of v^{r*}, and f_g, f_n instead of f_g^r, f_n^r.

Claim 1. There exists an $\widehat{x} \in [0,1)$ such that $\frac{\partial v^*(x)}{\partial x} > 0$ for all $x > \widehat{x}$.

Proof. From (*1*) we can calculate the derivative of $v^*(x)$ with respect to x as

$$\begin{aligned} \frac{\partial v^*(x)}{\partial x} &= f_g(x)b + f_g(x)c_g - f_n(x)c_n \\ &= f_n(x)c_n \left[\frac{f_g(x)}{f_n(x)} \frac{(b+c_g)}{c_n} - 1 \right]. \end{aligned} \tag{A1}$$

From MLRP, $\frac{f_g(x)}{f_n(x)}$ is strictly increasing in x. By assumption $\frac{f_g(x)}{f_n(x)} \to +\infty$ as $x \to 1$. Therefore there exists a value $\widehat{x} \in [0,1)$ such that the expression in square brackets in (A1) is positive for all $x > \widehat{x}$.

Claim 2. If x^* is an equilibrium, then $x^* > \widehat{x}$.

Proof. From (1) we have $v^*(0) = c_n - c_g < 0$. From (3) we have $\Im(v^*(0)) = 0$ because $\Im$ has support in $\mathbb{R}^+$. Furthermore, $\Im(v^*(x)) = 0$ for all $x \le \widehat{x}$. Suppose that in equilibrium $x^* < \widehat{x}$. Then we would have $\pi(x^*) = \Im(v^*(x^*)) = 0$. But in this case the optimal response of the court would be to set $x^* = 1 > \widehat{x}$, a contradiction.

Claim 3. The equilibrium x^* is unique.

Suppose there were two equilibria, x_0^* and x_1^*, with $\widehat{x} < x_0^* < x_1^*$. From Claim 1 this would imply $0 < v^*(x_0^*) < v^*(x_1^*)$, and in turn $\pi(x_0^*) < \pi(x_1^*)$. But then the optimal response of the court would involve setting $x_0^* > x_1^*$, a contradiction.

A.3 Proof that the equilibrium error rate is decreasing in x^*

The error rate (4) can be rewritten as

$$E(r,R) = 1 - \frac{1}{1 + \frac{1-\pi(x^*)}{\pi(x^*)} \frac{1-F_n(x^*)}{1-F_g(x^*)}}. \tag{A2}$$

Expression (A2) is decreasing in x^* if and only if $\frac{1-\pi(x^*)}{\pi(x^*)} \frac{1-F_n(x^*)}{1-F_g(x^*)}$ is decreasing in x^*. Taking the first derivative of this product with respect to x^* we obtain:

$$-\frac{1}{[\pi(x^*)]^2} \frac{\partial \pi(x^*)}{\partial x^*} + \frac{1}{[1-F_g(x^*)]^2} \left[\int_{x^*}^1 f_g(x^*) f_n(x) dx - \int_{x^*}^1 f_g(x) f_n(x^*) dx \right]. \tag{A3}$$

The first addendum in (A3) is negative because the equilibrium $\pi(x^*)$ is increasing in x^*, following Claims 1 and 2 above. To see that the second addendum is also negative, recall that from MLRP we know that $\frac{f_g(x)}{f_n(x)} > \frac{f_g(x^*)}{f_n(x^*)}$ for any $x > x^*$. Because the integrals in (A3) are calculated for $x \in (x^*, 1]$, then in this range $f_g(x^*)f_n(x) < f_g(x)f_n(x^*)$, hence the expression in square brackets is negative.

A.4 Case with random race of the victim

Consider the case in which the defendant cannot choose the race of the victim and the latter is unknown ex ante. Define $\beta \in (0, 1)$ as the exogenous probability that the victim of the crime is white. The expected payoff to an individual of race r from committing the crime is:

$$\begin{array}{l} \beta \left\{ \left[1 - F_g^r\left(x_{rW}\right)\right]\left(-v - c_g\right) + F_g^r\left(x_{rW}\right)\left(b - v\right)\right\} \\ +(1-\beta)\left\{ \left[1 - F_g^r\left(x_{rM}\right)\right]\left(-v - c_g\right) + F_g^r\left(x_{rM}\right)\left(b - v\right)\right\}. \end{array}$$

The payoff from not committing a crime is:

$$-\beta\left[1 - F_n^r\left(x_{rW}\right)\right]c_n - (1-\beta)\left[1 - F_n^r\left(x_{rM}\right)\right]c_n.$$

Let us define

$$\begin{aligned} \Gamma_g(x_{rW}, x_{rM}) &\equiv \beta F_g^r\left(x_{rW}\right) + (1-\beta)F_g^r\left(x_{rM}\right) \\ \Gamma_n(x_{rW}, x_{rM}) &\equiv \beta F_n^r\left(x_{rW}\right) + (1-\beta)F_n^r\left(x_{rM}\right) \end{aligned}$$

Following the same procedure as in the text, we obtain the threshold level of v below which a crime is committed.

$$v \leqslant \Gamma_g(x_{rW}, x_{rM})b - \left[1 - \Gamma_g(x_{rW}, x_{rM})\right]c_g + \left[1 - \Gamma_n(x_{rW}, x_{rM})\right]c_n \equiv v^*(\beta, x_{rW}, x_{rM}).$$

Obviously, $v^*(\cdot)$ depends on all the other parameters, namely β, b, c_g and c_n, but the latter do not depend upon the races neither of the defendant nor of the victim and are common knowledge. Relative to the case developed in the text, now the choice of each potential criminal depends on both cutoff points relative to the race of the victim. Repeating the same steps of the proof in the text one reaches the same implications for our test of racial bias.

A.5 Case where the race of the victim can be chosen

Consider now the case in which the criminal can choose the race of the victim. Under the assumptions of our model if the court were biased and this led to setting a lower threshold of evidence x^* for, say, white victims, all potential criminals would choose minority victims. If instead the court were unbiased potential criminals would be indifferent on the race of the victim and would randomize. This implies that under the assumptions of our model in the presence of bias we should not observe in equilibrium a condition (killing white victims) that allows us to test for bias. To be able to derive a test for bias in cases where the race of the victim is a choice variable one should adopt a different theoretical framework, e.g. one in which there are differential benefits to killing victims of different races or the potential criminal was uncertain about the bias of the court or the distribution of the signal. This goes beyond the scope of the current analysis.

B. How the error rate varies with π

In this section we discuss how the error rate changes when the proportion of guilty individuals, π, changes. For the sake of compactness, we omit superscripts for race (and occasionally arguments) in the formulas below.[2]

The error rate E is:

$$E(x) = \frac{(1-\pi)\left[1 - F_n\left(x(\pi)\right)\right]}{\pi\left[1 - F_g\left(x(\pi)\right)\right] + \left[1-\pi\right]\left[1 - F_n\left(x(\pi)\right)\right]} \tag{4}$$

where $x(\pi)$ is implicitly defined by.

$$\frac{f_g\left(x\right)}{f_n\left(x\right)} = \frac{\alpha}{1-\alpha}\frac{1-\pi}{\pi} \tag{5}$$

The numerator of (4) captures the Type I error, as it is the product of (i) the area to the right of the standard of proof x and below the signal distribution for non-guilty defendants, f_n, and (ii) the proportion of non-guilty people in the population, $1-\pi$. The denominator is the sentencing rate, given by the mass of people (guilty and non-guilty) whose realized signal is to the right of x.

We want to assess the sign of the derivative $dE(x)/dx$.

As a preliminary step, it is useful to observe that π has two effects on $E(x)$ that go in opposite directions - we may call them “direct” and “indirect” effect.

[2] In other words, when we write f_g and F_g we mean $f_g(x)$ and $F_g(x)$, etc.

To see the **direct effect** of π, suppose that x were independent of π. In that case:

$$\begin{aligned}
\frac{\partial E(\pi)}{\partial \pi} &= \frac{-(1-F_n)[\pi(1-F_g)+(1-\pi)(1-F_n)]-(1-\pi)(1-F_n)[(1-F_g)-(1-F_n)]}{[\pi(1-F_g)+(1-\pi)(1-F_n)]^2} \\
sign &\approx -\pi(1-F_n)(1-F_g)-(1-\pi)(1-F_n)(1-F_g) \\
&= -(1-F_n)(1-F_g)<0.
\end{aligned}$$

Intuitively, the direct effect captures the fact that, for given standard of proof x, higher values of π imply that fewer of the people condemned will be innocent, hence the error rate is lower.

The **indirect effect** of π on E works through endogenous changes in x. To see how, notice that (i) $\frac{\partial x}{\partial \pi}<0$ (from (5) and MLRP), and (ii) $\frac{\partial E}{\partial x}<0$ (proved in the Appendix of the paper). Together, (i) and (ii) push in the direction of $\frac{\partial E}{\partial \pi}>0$. Intuitively, the indirect effect captures the fact that when π increases the court chooses a lower standard of proof, which $-$for given $\pi-$ increases the size of the Type I error.

The combined effect is obtained by taking the total derivative of (4) with respect to π. The sign of this derivative is determined by:

$$\begin{aligned}
&\frac{dE}{dx} \overset{S}{\approx} -\left[(1-F_g)-\pi f_g\frac{\partial x}{\partial \pi}\right][\pi(1-F_g)+(1-\pi)(1-F_n)]+ \\
&+\pi(1-F_g)\left[(1-F_g)-\pi f_g\frac{\partial x}{\partial \pi}-(1-F_n)-(1-\pi)f_n\frac{\partial x}{\partial \pi}\right] \\
=\ &-(1-F_n)(1-F_g)-\pi(1-\pi)\left[f_n(1-F_g)-f_g(1-F_n)\right]\frac{\partial x}{\partial \pi} \qquad (6)
\end{aligned}$$

The first addendum in (6) is negative and captures the direct effect. The second addendum is positive (notice that the term in square brackets is positive by MLRP) and captures the indirect effect. Which of the two dominates depends on the value of π and the functional form of F_g, F_n.

An example

In what follows we provide an illustration assuming a simple Normal parameterization for F_g and F_n: $F_g(x)=N(\mu_g,\sigma_g)$, $F_n(x)=N(\mu_n,\sigma_n)$ with $\mu_n<\mu_g$ and $\sigma_g=k\sigma_n$, $k>0$.

The equilibrium value of x in this case is given by

$$x=\frac{2(\mu_n\sigma_g^2-\mu_g\sigma_n^2)+\sqrt{(2\mu_n\sigma_g^2-2\mu_g\sigma_n^2)^2-4(\sigma_n^2-\sigma_g^2)\left[\mu_g^2\sigma_n^2-\mu_n^2\sigma_g^2+2\ln\left(\frac{\alpha}{1-\alpha}\frac{1-\pi}{\pi}\frac{\sigma_g}{\sigma_n}\right)\sigma_g^2\sigma_n^2\right]}}{2(\sigma_g^2-\sigma_n^2)}.$$

Denote with $Erfc(z)$ the complementary error function
$Erfc(z) = 1 - \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-\frac{t^2}{2}dt}$ and define

$$\Delta \equiv \sigma_g^2 \sigma_n^2 \left[\mu_g^2 - 2\mu_g \mu_n + \mu_n^2 + 2 \ln \left(\frac{\alpha}{1-\alpha} \frac{1-\pi}{\pi} \frac{\sigma_g}{\sigma_n} \right) (\sigma_g^2 - \sigma_n^2) \right].$$

Then the error rate is

$$E(x) = \frac{(1-\pi)\left[2 - Erfc\left(\frac{\mu_n \sigma_n^2 - \mu_g \sigma_n^2 + \sqrt{\Delta}}{\sqrt{2}(\sigma_n^3 - \sigma_g^2 \sigma_n)}\right)\right]}{2 - \pi Erfc\left(\frac{\mu_g \sigma_g^2 - \mu_n \sigma_g^2 - \sqrt{\Delta}}{\sqrt{2}(\sigma_g^3 - \sigma_g \sigma_n^2)}\right) + (\pi - 1) Erfc\left(\frac{\mu_n \sigma_n^2 - \mu_g \sigma_n^2 + \sqrt{\Delta}}{\sqrt{2}(\sigma_n^3 - \sigma_g^2 \sigma_n)}\right)}.$$

Figure 1 shows how E varies with π for different degrees of relative noise in the signal distributions.[3] The ratio σ_g/σ_n measured on the rightmost axis is allowed to take values in $(0, 4]$. The remaining parameters are set at $\alpha = 0.5$, $\mu_n = 0.3$, $\mu_g = 6$, $\sigma_n = 2$.

In this example when $\sigma_g < \sigma_n$, i.e. $k < 1$, the error rate is monotonically decreasing in π. For intermediate values of k the relationship is non-monotonic, i.e. the error first decreases and then increases with π, and for $\sigma_g >> \sigma_n$ the latter effect dominates. Underlying this pattern is the relative size of the change in the Type I and Type II errors. To understand why, consider figure 2.

Figure 2 depicts the density functions $f_n(x)$ and $f_g(x)$ parameterized as above and the equilibrium value of x, indicated by a vertical line. In panel (a) $\sigma_g = 0.5\sigma_n$. When π increases from 0.1 (top graph) to 0.8 (bottom graph), the equilibrium value of x decreases from 4.7 to 3.4. This leads to a relative small increase in the area corresponding to the type I error ($1 - F_n(x^*)$ goes from .01 to .06) and a relatively larger decrease in the area corresponding to the type II error ($F_g(x^*)$ goes from .09 to .005). The latter declines at a faster rate because the distribution f_g is less dispersed. The overall effect is a decrease in the error rate from .12 to .01.

When we consider panel (b), where $\sigma_g = 2.5\sigma_n$, the pattern is reversed. Here the same increase in π from 0.1 to 0.8 induces a much larger reduction in x^*, from 5.3 to 0.9. This in turn leads to a relatively large increase of the area corresponding to the type I error ($1 - F_n(x^*)$ goes from .01 to .06) relative to the decrease in the area $F_g(x^*)$ (from .45 to .16), with a corresponding increase in the error rate from .09 to .10.

[3] The parameterization used in this figure allows x to take values outside $[0, 1]$ because this improves the readability of the graphs. Qualitatively similar patterns can be obtained using the truncated Normal distribution in $[0, 1]$. Also note that the arguments used in this section are based on MLRP which does not require x to be defined on a compact set.

We can thus summarize the role played by π in our model and the implications for the interpretation of our results. First, the variable π is *not* the proportion of homicides committed across races (in which case one may conjecture that the proportion of intra-racial homicides would be larger). The variable π is the probability that an individual brought to court for a homicide is guilty of the death penalty. Second, based upon the (limited) empirical evidence that we were able to assemble and that we discuss in the text of the paper there is no clear pattern of π_{rR} which would allow to take a stand on their relative size. In particular, there is no evidence that the (imperfectly) predictable patterns of π_{rR} would systematically go in a direction that inficiates the validity of our test. Notice also that even if we had perfect information about the relative size of π_{rR} we would have to take a stand on the shape of the distributions $F_g(x)$ and $F_n(x)$. In the text we then proceed with the assumption that π does not systematically differ across pairs of defendant/victim races.

C. Additional empirical results

This section of the Appendix contains some additional empirical results.

First, we provide descriptive statistics on the missingness of victim's race. We start by tabulating the number of cases with missing race of the victim (Table A1), then we present a breakdown of the number of observations for which we could not find the race of the victim by year and by state, for both the Habeas Corpus (Table A2) and for the Direct Appeal (Table A3) datasets. We also assess the potential selection in missingness of victim's race through a balance test on observables (Table A4). Finally, We report a balance test for crime characteristics across pairs of defendant and victim's races (Table A5).

Appendix Table A1: Missingness of race

Panel A: Habeas Corpus

		Missing victim's race		
		No	*Yes*	*Total*
Missing defendant's race	*No*	508	20	528
	Yes	3	0	3
	Total	511	20	531

Panel B: Direct Appeal

		Missing victim's race		
		No	*Yes*	*Total*
Missing defendant's race	*No*	3,717	447	4,146
	Yes	130	122	252
	Total	3,847	569	4416

Appendix Table A2:
Missingness of victim's race by year and State, Habeas Corpus

Years of 1st sentence	No. Total obs	No. Missing obs	Share missing
1973	4	0	0.00
1974	25	4	0.16
1975	30	5	0.17
1976	25	2	0.08
1977	45	3	0.07
1978	48	1	0.02
1979	51	2	0.04
1980	53	1	0.02
1981	66	0	0.00
1982	68	1	0.01
1983	41	0	0.00
1984	26	1	0.04
1985	25	0	0.00
1986	15	0	0.00
1987	6	0	0.00
1988	2	0	0.00
1989	1	0	0.00
			.

State	No. Total obs	No. Missing obs	Share missing
AL	19	1	0.05
AR	24	0	0.00
AZ	14	1	0.07
CA	4	0	0.00
DE	2	0	0.00
FL	95	4	0.04
GA	84	2	0.02
ID	3	0	0.00
IL	10	0	0.00
IN	4	0	0.00
KY	1	0	0.00
LA	34	0	0.00
MD	1	0	0.00
MO	26	1	0.04
MS	21	0	0.00
MT	4	0	0.00
NC	10	0	0.00
NE	6	0	0.00
NV	4	0	0.00
OK	11	2	0.18
PA	3	0	0.00
SC	7	0	0.00
TN	1	0	0.00
TX	108	9	0.08
UT	3	0	0.00
VA	27	0	0.00
WA	3	0	0.00
WY	2	0	0.00

Appendix Table A3:
Missingness of victim's race by year and State, Direct Appeal

Years od 1^{st} sentence	No. Total obs	No. Missing obs	Share missing
1974	10	1	0.10
1975	27	9	0.33
1976	45	9	0.20
1977	61	22	0.36
1978	77	15	0.19
1979	118	23	0.19
1980	128	25	0.20
1981	161	20	0.12
1982	159	16	0.10
1983	219	17	0.08
1984	238	18	0.08
1985	271	27	0.10
1986	215	17	0.08
1987	246	20	0.08
1988	320	30	0.09
1989	253	30	0.12
1990	238	22	0.09
1991	280	28	0.10
1992	300	27	0.09
1993	258	24	0.09
1994	306	26	0.08
1995	234	21	0.09

State	No. Total obs	No. Missing obs	Share missing
AL	256	50	0.20
AR	75	9	0.12
AZ	189	31	0.16
CA	229	26	0.11
CO	2	0	0.00
CT	2	0	0.00
DE	22	0	0.00
FL	709	80	0.11
GA	269	11	0.04
ID	31	11	0.35
IL	218	0	0.00
IN	67	0	0.00
KY	44	13	0.30
LA	88	4	0.05
MD	42	6	0.14
MO	81	6	0.07
MS	110	3	0.03
MT	13	2	0.15
NC	226	6	0.03
NE	22	5	0.23
NJ	36	3	0.08
NM	8	0	0.00
NV	91	17	0.19
OH	104	0	0.00
OK	186	32	0.17
OR	29	4	0.14
PA	178	42	0.24
SC	115	11	0.10
TN	102	13	0.13
TX	495	51	0.10
UT	13	3	0.23
VA	93	4	0.04
WA	15	4	0.27
WY	4	0	0.00

Appendix Table A4:
Selection in missingness of victim's race

Variable	*Nonmissing victim's race*	*Missing victim's race*	*Diff=0 (p-val)*
Panel A: Habeas Corpus			
Defendant characteristics			
Defendant is White	0.51	0.45	0.59
Defendant is African American	0.44	0.50	0.60
Male defendant	0.99	1.00	0.66
Age of defendant	28.2	41.5	0.11
Prior felony	0.22	0.25	0.79
History of alcohol abuse	0.13	0.00	0.09
History of drug abuse	0.17	0.10	0.44
Deprived/Abused background	0.02	0.00	0.57
Victim characteristics			
Number of victims	1.41	1.22	0.65
Female victim	0.48	0.24	0.05
High status victim	0.23	0.30	0.46
Police victim	0.09	0.00	0.16
Crime characteristics			
Defendant knew victim	0.26	0.25	0.94
Heinous crime	0.40	0.30	0.37
Panel B: Direct Appeal			
Defendant characteristics			
Defendant is White	0.51	0.52	0.68
Defendant is African American	0.41	0.40	0.69
Male defendant	0.98	0.98	0.50
Age of defendant	31.3	32.9	0.00
Victim characteristics			
Number of victims	1.37	1.44	0.22
Female victim	0.52	0.50	0.41

Appendix Table A5: Balance test on crime characteristics

	Full sample								South							
	White defendant			Non-white defendant			Diff-in-diff		White defendant			Non-white defendant			Diff-in-diff	
	White victim	*Non-white victim*	*Diff=0 (p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	ΔΔ	*(p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	*White victim*	*Non-white victim*	*Diff=0 (p-val)*	ΔΔ	*(p-val)*
Panel A: Habeas Corpus																
Victim																
Multiple victims	0.02	0.00	0.51	0.02	0.03	0.73	0.03	0.46	0.01	0	0.74	0.01	0.04	0.11	0.04	0.29
Female victim	0.53	0.29	0.06	0.43	0.44	0.97	0.24	0.10	0.53	0.45	0.61	0.42	0.46	0.64	0.12	0.50
Weapon																
Knife	0.19	0.27	0.48	0.14	0.18	0.50	-0.04	0.77	0.18	0.20	0.86	0.13	0.17	0.57	0.01	0.92
Handgun	0.37	0.20	0.20	0.50	0.41	0.25	0.07	0.65	0.40	0.30	0.53	0.50	0.38	0.17	-0.02	0.91
Shotgun or rifle	0.16	0.27	0.32	0.20	0.18	0.77	-0.12	0.33	0.18	0.30	0.34	0.20	0.21	0.84	-0.11	0.47
Strangulation	0.09	0.07	0.74	0.04	0.08	0.23	0.07	0.39	0.08	0.00	0.37	0.04	0.07	0.48	0.10	0.25
Circumstances																
Defendant connected to	0.27	0.18	0.39	0.24	0.27	0.68	0.12	0.34	0.24	0.09	0.25	0.23	0.27	0.52	0.20	0.19
Burglary or theft	0.12	0.06	0.50	0.11	0.07	0.41	0.02	0.85	0.13	0.00	0.22	0.10	0.08	0.68	0.11	0.32
Robbery	0.36	0.31	0.69	0.47	0.42	0.53	0.00	1.00	0.38	0.50	0.44	0.49	0.39	0.22	-0.22	0.21
Kidnapping	0.07	0.06	0.95	0.07	0.07	0.92	0.01	0.92	0.06	0.10	0.66	0.06	0.08	0.62	-0.02	0.86
Rape or sex related	0.21	0.13	0.40	0.19	0.16	0.56	0.05	0.65	0.23	0.20	0.81	0.18	0.18	1.00	0.03	0.83
Institutional killing	0.04	0.06	0.71	0.10	0.09	0.85	-0.03	0.71	0.05	0.00	0.46	0.09	0.08	0.88	0.04	0.62
Panel B: Direct Appeal																
Victim																
Multiple victims	0.26	0.19	0.12	0.20	0.29	0.00	0.16	0.00	0.23	0.14	0.05	0.15	0.25	0.00	0.19	0.00
Female victim	0.57	0.37	0.00	0.47	0.50	0.28	0.23	0.00	0.57	0.32	0.00	0.46	0.52	0.07	0.31	0.00
Weapon(a)																
Knife	0.28	0.35	0.47	0.27	0.26	0.87	-0.08	0.46	0.26	0.25	0.90	0.28	0.27	0.95	0.01	0.93
Handgun	0.34	0.45	0.32	0.45	0.42	0.60	-0.14	0.27	0.35	0.38	0.83	0.46	0.42	0.51	-0.06	0.64
Shotgun or rifle	0.12	0.15	0.66	0.06	0.08	0.52	-0.02	0.85	0.10	0.19	0.28	0.06	0.09	0.47	-0.06	0.45
Strangulation	0.09	0.00	0.17	0.02	0.00	0.10	0.07	0.23	0.09	0.00	0.22	0.02	0.00	0.16	0.07	0.30

Notes: (a) information available for 935 out of 3717 cases