Appendix B. Online Appendix—Not for publication

B.1. **Additional reduced form analysis.** An empirical implication of the theoretical model discussed in Section 4 is that the probability of settlement is decreasing in the length of the trial sentence. In this Appendix I present evidence supporting such a prediction. I also show further evidence that the observable case characteristics do not vary systematically with the harshness of the assigned judge. The latter result complements the analysis in Section 3.3 of the paper.

A simple way of testing whether the settlement probability decreases in the length of the trial sentence is by comparing the settlement rates across cases of different severity. Table 12 shows settlement rates and average incarceration sentences for several categories of crimes in the data. The table separately reports the average sentence lengths for cases resolved at trial and by plea bargain. Here, I consider as settled all cases decided by plea bargain—independent of whether the sentence includes incarceration time. That is because, otherwise, the settlement rates would largely capture differences in the the probabilities of incarceration across crime categories. The table shows that, as expected, sentences assigned to defendants convicted of homicide are very long.[51] Among the categories displayed in the table, non-homicide violent crimes have the second longest average sentences, followed by drug-related crimes and property crimes. More interestingly, the table suggests a negative relationship between average sentences and the likelihood of settlement. The settlement ratios for homicides, non-homicide violent crimes, drug-related crimes and property crimes are, respectively, 81.83 percent, 89.94 percent, 96.95 percent and 98.26 percent. It is also worth noticing that cases in which the average trial sentences are long tend to settle for relatively long sentences. This observation provides further evidence that settlement negotiations take place in the shadow of the trial.

Another method of testing whether cases with long potential trial sentences are settled less often is to explore the correlation between settlement ratios and the sentencing patterns of different judges. As explained in the main text, an important feature of the North Carolina justice system is the rotation of judges across districts within the same Superior Court divisions. This rotation makes it plausible to assume that cases are randomly assigned to judges—which allows me to treat the caseloads of different judges as identical and to attribute any variation in the sentencing patterns to judges' characteristics. I can then verify whether the settlement ratios of cases

[51]The numbers in table 12 underestimate the average homicide sentence since death sentences are not accounted for in the computation.

TABLE 12. Settlement rates and sentences by type of offense

Type of offense	Obs.	% Settled	Average sentence (settled)†	Average sentence (trial)†
Violent (non-homicide)	118948	89.94%	38.98	100.73
Homicide	6611	81.83%	134.71	236.76
Property	215505	98.26%	13.36	49.68
Drugs	199692	96.95%	18.63	51.44
Other	313622	96.09%	18.47	37.03
Total	854378	95.86%	25.05	81.30

Notes: Non-homicide violent crimes include assault, robbery and sexual assault. Property crimes include burglary, larceny and arson. Drug-related crimes include both trafficking and possession.
In this table I classify as settled all cases resolved by plea bargain—that is, I consider cases that resulted in incarceration or alternative sentences.
† Measured in months.

decided by harsher judges are smaller than those of cases decided by more lenient ones.[52]

Consider the specification

$$settled_i = \vartheta_{\mathbf{3}}\mathbf{X_{3i}} + \zeta_{\mathbf{3}}\ \mathbf{Judge_i} + \epsilon_{3i}, \tag{B.1}$$

where $settled_i$ is a dummy indicating whether case i is resolved by plea bargain, $\mathbf{X_{3i}}$ is a vector of controls, $\mathbf{Judge_i}$ is a vector of judge-specific dummies defined as in specification (3.1) in Section 3, and ϵ_{3i} is an error term. The controls $\mathbf{X_{3i}}$ are identical to $\mathbf{X_{1i}}$ from (3.1), except that the former exclude the dummy indicating settlement. I estimate this specification by OLS, using only cases in which the main offense is a non-homicide violent crime. Table 13 presents the results. I am interested in the relation between $\hat{\zeta}_{\mathbf{1}}$ and $\hat{\zeta}_{\mathbf{3}}$, the vectors of estimated coefficients for the judge-specific dummies in (3.1) and (B.1), respectively. A negative correlation between these vectors suggests that cases decided by harsh judges are less likely to be resolved by a plea bargain, as predicted by the model in Section 4. I find the Pearson's correlation coefficient to be -0.23 and significant at the ten percent level.[53]

[52]For an analogous exercise using data on federal criminal cases, see Boylan (2012). Waldfogel (1998) undertakes a similar analysis using data on civil cases. The results of both papers are similar to the ones presented in this section.
[53]The standard deviation of the correlation coefficient, calculated from 1000 bootstrap samples, is 0.1242. The associated p-value is 0.067. Replicating the exercise with the entire data set (i.e., not only non-homicide violent crimes), I find a correlation coefficient of -0.18 that is significant at all conventional levels.

TABLE 13. Determinants of settlement

	Dependent variable: *settled*
age	-0.0097
	(0.0007)
age^2	-0.0001
	(0.0000)
female	- 0.0601
	(0.0049)
black	0.0025
	(0.0035)
hispanic	0.1861
	(0.0115)
private attorney	-0.0875
	(0.0056)
public defender	-0.0195
	(0.0068)
Judge dummies	Yes
County dummies	Yes
Superior Court division dummies	No
Observations	97942
R^2	0.2663

Notes: OLS estimates. The dummy *settled* indicates whether the case results in incarceration by plea bargain.
Other controls: Year of disposition, offense severity and defendant's criminal record.
Standard errors (robust to clustering at the judge level) in parenthesis.

In Section 3.3, I show evidence that observable case characteristics do not vary systematically with the harshness of the assigned judge, which supports the identifying assumption that cross-judge variation in sentencing patterns is independent from other aspects of the cases. In specification (3.2), in particular, I regress a dummy variable indicating whether the judge is harsh on a large set of case-level covariates, and find that such covariates have little explanatory power. Here I conduct a similar exercise, in which I replace the harsh indicator by the estimated judge fixed effects from (3.1) as the dependent variable in (3.2).

Table 14 reports the results. None of the variables shown in the table is significant at the five percent level.[54] The only variables that are significant at ten percent are

[54]As in the regression shown in the main text, the only regressor that is significant at this level is a dummy indicating whether the case is a class three misdemeanor. An F-test excluding this dummy fails to reject the null hypothesis that all other variables are jointly insignificant.

Table 14. Judge assignment—Determinants of judge fixed effects

	$\hat{\zeta}_1$ (Judge FE)
age	0.0119
	(0.0090)
age^2	0.0002
	(0.0001)
female	-0.0890
	(0.0608)
black	-0.0946
	(0.0870)
hispanic	-0.2033
	(0.1218)
private attorney	0.0429
	(0.0726)
public defender	-0.3782
	(0.2102)
Superior Court division dummies	Yes
Observations	97942
R^2	0.3266

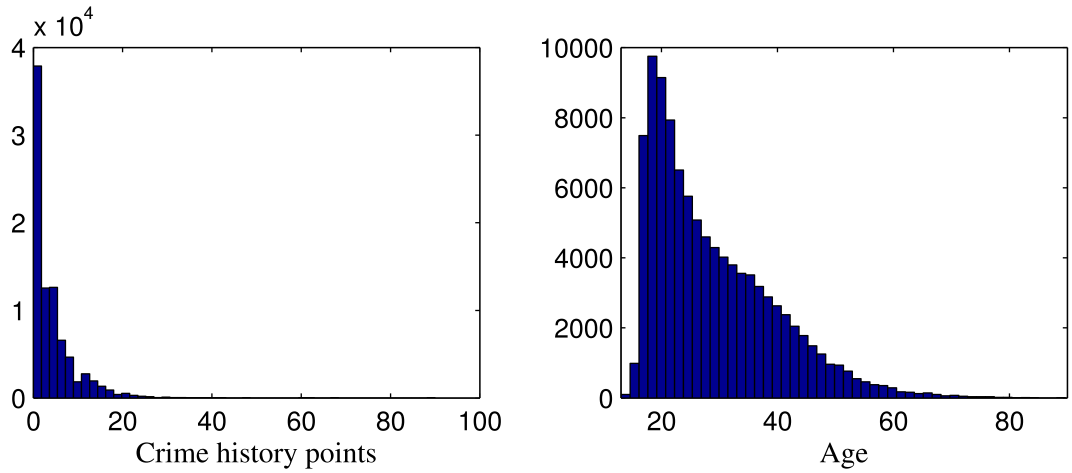
Notes: OLS estimates. The variable $\hat{\zeta}_1$ is obtained from the OLS estimation of specification (3.1).
Other controls: Disposition year, offense severity, defendant's criminal record and Superior Court division.
Standard errors (robust to clustering at the judge level) in parenthesis.

the dummies indicating whether the defendant is Hispanic and whether the defense attorney is a public defender. The point estimates of the coefficients associated to these variables are small in magnitude: -0.20 for the Hispanic dummy and -0.37 for the public defender one. As a reference, it is useful to compare such values with moments from the judge fixed effect distribution, as reported in table 4. Its mean is 31.58, its standard deviation is 6.26 and its interquartile range is 8.06. Thus, the estimated coefficients associated with the hispanic and public defender dummies correspond to only 3.19 and 5.91 percent of the dependent variable's standard error, respectively.

B.2. **Data appendix.**

B.2.1. *Additional descriptive statistics.* Figure 5 shows the histograms of the defendants' criminal history points and age. The mode of the criminal history points

Figure 5. Defendants' criminal history points and age



Notes: Criminal history points are assigned by the North Carolina justice system for the purpose of setting sentencing guidelines. Age measured in years.

distribution is zero, indicating no previous criminal history. Relatively few defendants have more than ten points. Regarding age, very few defendants are less than 18 years old, and the vast majority of them are between 18 and 40.

Table 15 presents descriptive statistics on trial outcomes by defendant's race. Conditional on a trial, African-American defendants are 9.16 p.p. more likely to be convicted than their non-African-American counterparts. These differences in conviction probability are consistent with the estimation results reported in Section 6, which suggest that the distribution of defendant's types for covariate group two (African-American defendants) places more mass at high types than that for group one (non-African-American defendants).

B.2.2. *Reducing multiple-counts cases to a single count.* My unit of analysis is a case. Some cases in the data are associated with multiple counts—that is, multiple charges against the same defendant. To reduce such cases to a single count, I employ the following procedure: If sentences are assigned to more than one count of the same case, I consider only the count with the longest sentence.[55] If no sentence is assigned to any count in a case, I classify the arrest offenses for each count according to their

[55]I observe in the data whether a sentence consists of incarceration time, intermediate punishment (such as probation) or community service. In order to classify the sentences, I use a lexicographic order, so that any incarceration time is considered a longer sentence than any intermediate punishment. Similarly, any intermediate sentence is considered longer than any community service.

Table 15. Descriptive statistics—Trial outcomes by race

Frequencies of method of resolution by race, conditional on trial		
Method of resolution	Non-African-Americans	African-Americans
Trial conviction		
Incarceration	32.61%	45.04%
Alternative sentence	8.40%	5.12%
Total	41.01%	50.17%
Trial acquittal / dismissed		
Absolved by jury	46.03%	40.12%
Dismissed by judge	12.96%	9.71%
Total	58.99%	49.83%

Length of incarceration sentences by race, conditional on trial†		
	Non-African-Americans	African-Americans
Mean	106.59	97.25
Standard deviation	116.14	100.03

Notes: The table is based on all 11,801 cases that have a non-homicide violent crime as the main arrest charge and that result in a trial. In 6,478 of these cases, the defendant is African-American, while in 5,323 the defendant is not.
† Measured in months.

severity (using the same classification adopted by the structured sentencing guidelines in North Carolina) and consider only the count with the most severe arrest offense.

B.2.3. *Classification of offenses.* Each count in the data is associated with an offense code assigned by the North Carolina Justice System. The offense codes employed in North Carolina are based on the Uniform Offense Classifications, organized by the National Crime Information Center (NCIC). Like in the NCIC code system, the first two digits of the North Carolina codes classify the offenses into relatively broad categories (e.g., robbery, fraud, vehicle theft, etc). Based on these two digits, I determine whether each count in my data qualifies as a robbery, an assault or a sexual assault—the offense categories that I use in my structural analysis.

B.2.4. *Identification of judges.* In the main, case-level data, judges are identified only by their initials. In most cases, three initials are used. I match the initials to the full names of the judges, as reported annually in the North Carolina Manual. In the period comprised by the sentencing data, only two pairs of judges have the same three initials. Cases decided by these judges were excluded from the data. I also excluded all the cases in which the judge was either identified by fewer than three initials or not identified at all.

B.2.5. *Life sentences.* To convert life sentences into a length of incarceration time, I consider the life expectancy in North Carolina for individuals of age 29.04, which is the average defendant's age in my sample. This life expectancy is 77.14 years (Buescher and Gizlice, 2002). To ensure that any life sentence is at least as long as the longest non-life sentence—which, in North Carolina, is forty years—I define the length of a life sentence as Max {77.14 - defendant's age ; 40}. Cases resulting in a death sentence are excluded from the analysis.[56] Cases whose sentence length is missing in the data are treated as cases in which only an alternative sentence is assigned.

[56]Cases of non-homicide violent crimes, which constitute the subsample considered in the structural analysis, never result in a capital sentence.

B.3. **Settlement offer estimator – Consistency.** In this section I show that the estimator for the settlement offer function presented in the main text is uniformly consistent. With this intent, I first state and prove a lemma, offering sufficient conditions for the consistency of two-steps estimators in which the second step employs sieve methods. The lemma is adapted from the well known results on consistency of two-steps estimators by Newey and McFadden (1994), and from the results on the consistency of sieve estimators by Chen (2007). After that, I show that its conditions are satisfied by the estimator of $\tilde{s}(\cdot)$ from Section 5.

LEMMA 2. *Let Ξ and Γ be two (possibly infinite-dimensional) parameter spaces endowed with metrics d_ξ and d_γ, respectively. Consider a data-generating process that can be described by the true parameters $\xi_0 \in \Xi$ and $\gamma_0 \in \Gamma$. An estimate $\hat{\gamma}_n$ of γ_0 is available from a previous estimation procedure. Let $\hat{Q}_n : \Xi \times \Gamma \to \Re$ be an empirical criterion, and Ξ_k be a sequence of approximating spaces to Ξ. Also, let*

$$\hat{\xi}_n = \underset{\xi \in \Xi_k}{argmax}\ Q_n(\xi, \hat{\gamma}_n).$$

Assume that the following conditions are true:

(a) *(i) Under the metric d_ξ: Ξ is compact; and $Q(\xi, \gamma_0)$ is continuous on ξ_0 and upper semi-continuous on Ξ*

(ii) $\xi_0 = \underset{\xi \in \Xi}{argmax}\ Q(\xi, \gamma_0)$ *and* $Q(\xi_0, \gamma_0) > -\infty$

(b) *(i) For any $\xi \in \Xi$ there exists $\pi_k \xi \in \Xi_k$ such that $d_\xi(\xi, \pi_k \xi) \to 0$ as $k \to \infty$*

(ii) Under d_ξ, and for all $k \geq 1$: Ξ_k is compact and $Q_n(\xi, \gamma_0)$ is upper semi-continuous on Ξ_k

(c) For all $k \geq 1$, $\underset{n\to\infty}{plim} \sup_{(\xi,\gamma) \in \Xi_k \times \Gamma} |Q_n(\xi, \gamma) - Q(\xi, \gamma)| = 0$

(d) *(i) $\hat{\gamma}_n \to^P \gamma_0$ under d_γ*

(ii) $\sup_{\xi \in \Xi} |Q(\xi, \gamma) - Q(\xi, \gamma')| \to 0$ *as $\gamma \to \gamma'$ under d_γ*

Then $\hat{\xi} \to^P \xi_0$ under d_ξ.

Proof. Consider any $\epsilon > 0$, and notice that, by assumption (a), $d_\xi(\hat{\xi}_n, \xi_0) > \epsilon$ implies $Q(\hat{\xi}_n, \gamma_0) - Q(\xi_0, \gamma_0) < -2\eta$ for some $\eta > 0$. By assumptions (a.i) and (b), for k high enough, there is $\pi_k \xi_0 \in \Xi_k$ such that $Q(\xi_0, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < \eta$.

$$\begin{aligned} d_\xi(\hat{\xi}_n, \xi_0) > \epsilon &\Rightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) + Q(\pi_k \xi_0, \gamma_0) - Q(\xi_0, \gamma_0) < -2\eta \\ &\Leftrightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -2\eta + Q(\xi_0, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) \\ &\Rightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -\eta. \end{aligned}$$

Define $A_n \equiv \left\{ Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -\eta \right\}$. Clearly, $P\left[d_\xi(\hat{\xi}_n, \xi_0) \right] \leq P\left[A_n \right]$. To complete the proof, I just need to show that $A_n = o_p(1)$. Define

$$\begin{aligned} B_n &\equiv \{ |Q(\pi_k \xi_0, \hat{\gamma}_n) - Q_n(\pi_k \xi_0, \hat{\gamma}_n)| > \eta/5 \} \\ C_n &\equiv \{ |Q(\pi_k \xi_0, \gamma_0) - Q(\pi_k \xi_0, \hat{\gamma}_n)| > \eta/5 \} \\ D_n &\equiv \left\{ |Q(\hat{\xi}_n, \hat{\gamma}_n) - Q_n(\hat{\xi}_n, \hat{\gamma}_n)| > \eta/5 \right\} \\ E_n &\equiv \left\{ |Q(\hat{\xi}_n, \gamma_0) - Q(\hat{\xi}_n, \hat{\gamma}_n)| > \eta/5 \right\}. \end{aligned}$$

Notice that $B_n = o_p(1)$ and $D_n = o_p(1)$ (by assumption (c)). Similarly, $C_n = o_p(1)$ and $E_n = o_p(1)$ (by assumption (d.ii)).

Now I argue that $A_n \cap B_n^C \cap C_n^C \cap D_n^C \cap E_n^C = \emptyset$. Indeed,

$$\begin{aligned} D_n^C &\Rightarrow Q_n(\hat{\xi}_n, \hat{\gamma}_n) \leq Q(\hat{\xi}_n, \hat{\gamma}_n) + \eta/5 \\ E_n^C &\Rightarrow Q(\hat{\xi}_n, \hat{\gamma}_n) \leq Q(\hat{\xi}_n, \gamma_0) + \eta/5 \\ A_n &\Rightarrow Q(\hat{\xi}_n, \gamma_0) < Q(\pi_k \xi_0, \gamma_0) - \eta \\ C_n^C &\Rightarrow Q(\pi_k \xi_0, \gamma_0) \leq Q(\pi_k \xi_0, \hat{\gamma}_n) + \eta/5 \\ B_n^C &\Rightarrow Q(\pi_k \xi_0, \hat{\gamma}_n) \leq Q_n(\pi_k \xi_0, \hat{\gamma}_n) + \eta/5. \end{aligned}$$

Hence, $A_n \cap B_n^C \cap C_n^C \cap D_n^C \cap E_n^C$ implies

$$\begin{aligned} Q_n(\hat{\xi}_n, \hat{\gamma}_n) &\leq Q(\hat{\xi}_n, \gamma_0) + 2(\eta/5) < Q(\pi_k \xi_0, \gamma_0) + 2(\eta/5) - \eta \\ &\leq Q(\pi_k \xi_0, \hat{\gamma}_n) + 3(\eta/5) - \eta \leq Q_n(\pi_k \xi_0, \hat{\gamma}_n) + 4(\eta/5) - \eta \\ &= Q_n(\pi_k \xi_0, \hat{\gamma}_n) - \eta/5. \end{aligned}$$

That contradicts $\hat{\xi}_n = \underset{\xi \in \Xi_k}{\operatorname{argmax}} \; Q_n(\xi, \hat{\gamma}_n)$.

Finally, notice that

$$
\begin{aligned}
P[A_n] &\leq P[A_n \cup (B_n \cup C_n \cup D_n \cup E_n)] \\
&= P\left[A_n \cap (B_n \cup C_n \cup D_n \cup E_n)^C\right] + P[B_n \cup C_n \cup D_n \cup E_n] \\
&= P[\emptyset] + P[B_n] + P[C_n] + P[D_n] + P[E_n].
\end{aligned}
$$

Therefore, $A_n = o_p(1)$. □

I can now show that the offer function estimator is consistent. Consider the following functions:

$$
\begin{aligned}
&\bar{b}_r(s) \equiv \frac{b[s|\Psi=1, Z=h]}{b[s|\Psi=1, Z=l]} \frac{P[\Psi=1|Z=h]}{P[\Psi=1|Z=l]} \\
\text{and} \qquad &\bar{g}_r(t) \equiv \frac{g(t|\Psi=2, Z=h)}{g(t|\Psi=2, Z=l)} \frac{P[\Psi=2|Z=h]}{P[\Psi=2|Z=l]}.
\end{aligned}
$$

PROPOSITION 3. *Assume that (i) $\bar{b}_r(\cdot)$ is differentiable and the absolute value of its derivative is bounded by $\mathring{b}_r$; and (ii) $\bar{b}_r(\cdot)$ and $\bar{g}_r(\cdot)$ are positive and bounded by $\bar{B}_r$ and $\bar{G}_r$, respectively. Then the estimator of $\tilde{s}(\cdot)$ in Section 5 is uniformly consistent.*

Proof. My goal is to show that the offer function estimator presented in Section 5 satisfies the conditions of Lemma 2. A feature of that estimator is that, given the first-stage estimates for the conditional distribution of Ψ and the censored densities of trial sentences and settlement offers, the objective function of the second stage is not random. In the notation of Lemma 2, $Q_n(\cdot,\cdot) = Q(\cdot,\cdot)$. As a consequence, in order to apply the lemma, I do not need to verify condition (c).

From 4.1 and the boundedness of $[\underline{t}, \bar{t}]$, I can assume, without loss of generality, that the function $\tilde{s}(\cdot)$ is bounded from above by $\bar{s}$, and its derivative is bounded from above by a constant $\mathring{s}$. The kernel density estimators of $b[s|\Psi=1, Z=z]$ and $g(t|\Psi=2, Z=z)$ are uniformly consistent, for $z \in \{l, h\}$. The estimators of $P[\Psi=1]$ and $P[\Psi=2]$ are consistent, as well. From Slutsky's theorem, therefore, the first stage of the estimation procedure returns uniformly consistent estimates of $\bar{b}_r(s)$ and $\bar{g}_r(t)$.

The objective function in the second stage of estimation is given by

$$
Q(s(\cdot), b_r(\cdot), g_r(\cdot)) = E\left\{[b_r(s(t)) - g_r(t)]^2\right\}
$$

where $s(\cdot)$ is an element from the space of increasing and convex functions on $[\underline{t}, \bar{t}]$, and $g_r(\cdot)$ and $b_r(\cdot)$ are positive valued functions. In order to show that the estimator in the second step is consistent, I need to prove that $Q(s(\cdot), b_r(\cdot), g_r(\cdot))$ satisfies conditions (a) and (d) from Lemma 2.

I start with condition (a). Using (5.2), it is easy to verify that

$$Q\left(\tilde{s}(\cdot), \bar{b}_r(s), \bar{g}_r(t)\right) = 0$$

and that the objective function is strictly positive for any other continuous function $s(\cdot)$. Condition (a.ii) is then trivially verified. Let Ξ be the space of functions defined on $[\underline{t}, \bar{t}]$ that are increasing and convex, uniformly bounded by zero and $\bar{s}$, and whose derivative is uniformly bounded by zero and $\mathring{b}_r$. To verify condition (a.i), notice first that, by the Arzela-Ascoli theorem, the space of differentiable functions on $[\underline{t}, \bar{t}]$ that are uniformly bounded and have a uniformly bounded derivative is compact under the sup norm. The space Ξ is the intersection of that space and the space of increasing and convex functions on $[\underline{t}, \bar{t}]$, which is closed. Hence, Ξ is compact. It remains to verify the upper semi-continuity of the objective function on Ξ and the continuity at $\tilde{s}(\cdot)$. Here, I show that the objective function is continuous on Ξ under the sup norm. Let $s(\cdot)$ and $\breve{s}(\cdot)$ be two functions in Ξ such that

$$\sup_{t \in [\underline{t}, \bar{t}]} \quad |s(t) - \breve{s}(t)| \leq \eta.$$

I can write

$$\begin{aligned}
&\left| E\left\{\left[\tilde{b}_r(s(t)) - \tilde{g}_r(t)\right]^2\right\} - E\left\{\left[\tilde{b}_r(\breve{s}(t)) - \tilde{g}_r(t)\right]^2\right\}\right| \\
&\quad = \left|E\left[b_r(s(t))^2 - b_r(\breve{s}(t))^2\right] + 2E\left\{g_r(t)\left[b_r(s(t)) - b_r(\breve{s}(t))\right]\right\}\right| \\
&\quad \leq \left|E\left\{\left[b_r(s(t)) - b_r(\breve{s}(t))\right]\left[2\, g_r(t) + b_r(s(t)) + b_r(\breve{s}(t))\right]\right\}\right|.
\end{aligned}$$

Some algebra shows that the right-hand side of the last inequality is lower than $2\,\mathring{b}_r\left[\bar{B}_r + \bar{G}_r\right]\eta$, so that the objective function is continuous on the whole space Ξ.

Now I show that condition (d) holds. Since I have uniformly consistent estimates from the first stage, condition (d.i) holds under the sup norm. To verify condition (d.ii), assume that $b_r(\cdot)$, $\breve{b}_r(\cdot)$, $g_r(\cdot)$ and $\breve{g}_r(\cdot)$ are such that

$$\begin{aligned}
&\sup_s \quad \left|b_r(s) - \breve{b}_r(s)\right| \leq \eta \\
\text{and} \qquad &\sup_t \quad |g_r(t) - \breve{g}_r(t)| \leq \eta
\end{aligned} \tag{B.2}$$

for some $\eta > 0$. I can write

$$\begin{aligned}
&\left| E\left\{[b_r(s(t)) - g_r(t)]^2\right\} - E\left\{\left[\breve{b}_r(s(t)) - \breve{g}_r(t)\right]^2\right\}\right| \\
&\quad = \left| E\left[b_r(s(t))^2 - \breve{b}_r(s(t))^2\right] + E\left[g_r(t)^2 - \breve{g}_r(t)^2\right] + 2E\left[\breve{b}(s(t))\breve{g}(t) - \underline{(}s(t))g(t)\right]\right| \\
&\quad \leq E_1(t) + E_2(t) + E_3(t),
\end{aligned}$$

where

$$\begin{aligned}
E_1(t) &\equiv \left| E\left[b_r(s(t))^2 - \breve{b}_r(s(t))^2\right]\right| \\
E_2(t) &\equiv \left| E\left[g_r(t)^2 - \breve{g}_r(t)^2\right]\right| \\
\text{and} \qquad E_3(t) &\equiv \left| 2E\left[\breve{b}(s(t))\breve{g}(t) - b(s(t))g(t)\right]\right|.
\end{aligned}$$

Some algebra shows that, for all functions $s(\cdot)$, the following inequalities hold

$$\begin{aligned}
E_1(t) \leq &\ \max\left\{\sup_{t\in t\in[\underline{t},\bar{t}]} 2\ \eta\ b_r(s(t)) + \eta^2\ ;\ \sup_{t\in t\in[\underline{t},\bar{t}]} 2\ \eta\ \breve{b}_r(s(t)) + \eta^2\right\} \\
\leq &\ \eta^2 + 2\ \eta\ \bar{B}_r, \\
E_2(t) \leq &\ \max\left\{\sup_{t\in t\in[\underline{t},\bar{t}]} 2\ \eta\ g_r(t) + \eta^2\ ;\ \sup_{t\in t\in[\underline{t},\bar{t}]} 2\ \eta\ \breve{g}_r(t) + \eta^2\right\} \\
\leq &\ \eta^2 + 2\ \eta\ \bar{G}_r, \\
E_3(t) \leq &\ 2\ \max\left\{\sup_{t\in t\in[\underline{t},\bar{t}]} \eta\left[b_r(s(t)) + g_r(t)\right] + \eta^2\ ;\ \sup_{t\in t\in[\underline{t},\bar{t}]} \eta\left[\breve{b}_r(s(t)) + \breve{g}_r(t)\right] + \eta^2\right\} \\
\leq &\ 2\ \eta^2 + 2\ \eta\left[\bar{B}_r + \bar{G}_r\right].
\end{aligned}$$

Since neither $\bar{B}_r$ nor $\bar{G}_r$ depends on $s(\cdot)$, condition (d) is satisfied, which concludes the proof of Proposition 3.

□

Table 16. Distance between trial and settlement density ratios

Group	Offer function	
	Main estimate	Naive estimate
1	20.8263	31.9346
2	39.4660	51.0874

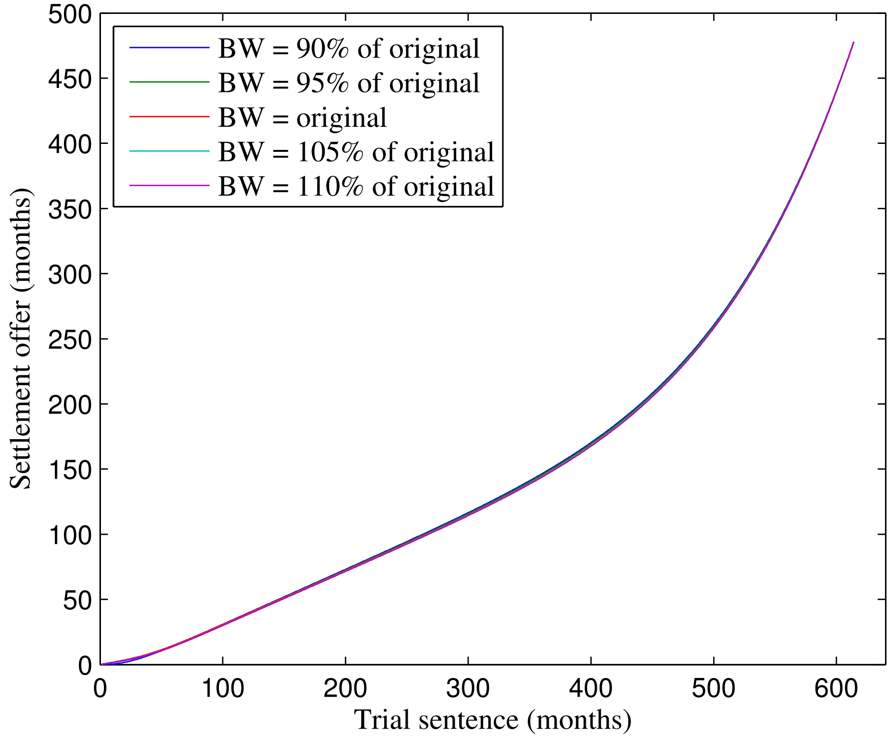
Notes: This table compares the objective function of the second stage of the offer function estimator (equation (5.14)) evaluated at two different estimated functions: The main estimate, which is the one employed throughout the main text (figure 3); and a naive estimate, obtained by connecting the extremes of the supports of trial sentences and settlement offers. This objective function serves as a distance between the density ratios of trial sentences and settlement offers. In the absence of sampling error, the distance should be zero at the true offer function. The table reports the distance multiplied by 100,000.

B.4. **Settlement offer estimator – Fit, precision and basic robustness analysis.** In this section, I discuss the fit and precision of the settlement offer function estimates. I also conduct a robustness analysis of these estimates, in which I consider alternative bandwidths for the kernel density estimators employed in the first stage.

B.4.1. *Fit.* I assess the fit of the settlement offer function estimates, based on the objective function in (5.14). This objective function serves as a distance between the density ratios of trial sentences and settlement offers. In the absence of sampling error, and if the model is correctly specified, this distance should be zero at the true settlement offer function. Column (1) of table 16 shows the distance obtained at the main estimated offer function, as reported in Section 6. To facilitate the interpretation of the distance, I multiply it by 100,000. As a basis for comparison, the table also shows the distance obtained by evaluating (5.14) at a naive offer function estimate—a line connecting the points $\left(\underline{\hat{t}}, \underline{\hat{s}}\right)$ and $\left(\hat{\bar{t}}, \hat{\bar{s}}\right)$, where $\underline{\hat{t}}$, $\underline{\hat{s}}$, $\hat{\bar{t}}$ and $\hat{\bar{s}}$ are the minimum trial sentence, the minimum settlement offer, the maximum trial sentence and the maximum settlement offer in the sample, respectively. For group one, relative to the naive estimate, my main estimate reduces the distance between the density ratios by roughly one third. For group two, the reduction is of approximately 20 percent.

B.4.2. *Robustness: Alternative bandwidths.* The empirical results presented in the main text employ kernel density estimates of trial sentences and settlement offers selected by Silverman's "rule-of-thumb" (Silverman, 1986). Table 11 in Appendix A.2 reports these bandwidths. Here I show estimates of the settlement offer function

Figure 6. Settlement offer function estimates with alternative bandwidths—Covariates group one



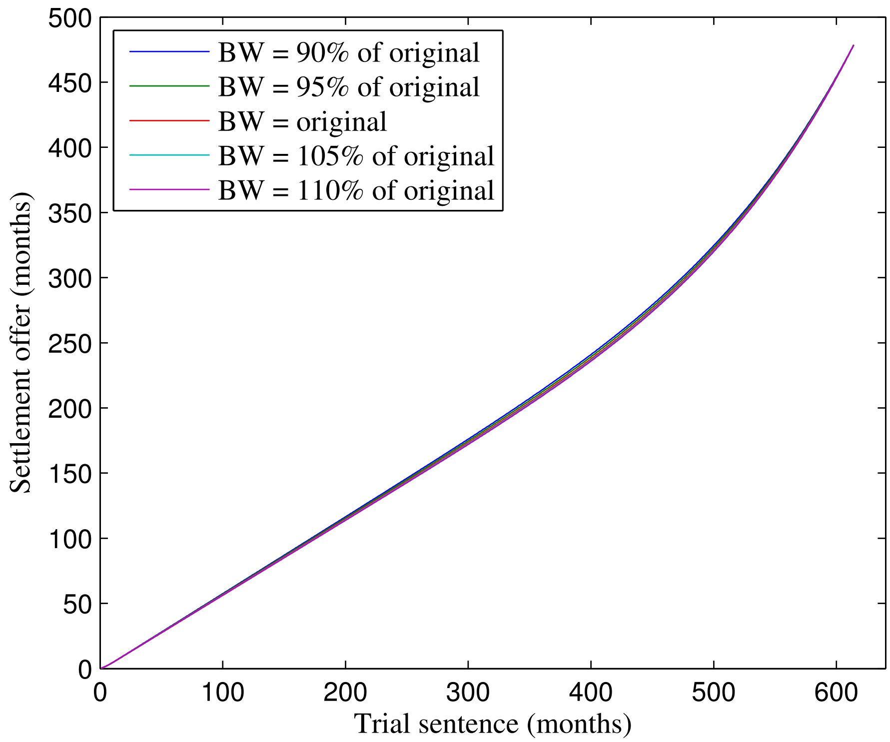
Note: This figure shows estimates of the settlement offer function of covariate group one based on alternative density estimates of the trial sentences and settlement offers. The alternative density estimates use bandwidths that are 90, 95, 105 and 110 percent of those employed in the main text.

obtained when I vary the bandwidths. Specifically, I consider the original rule-of-thumb bandwidth multiplied by 0.9, 0.95, 1.05 and 1.10. The results for groups one and two are reported in figures 6 and 7, respectively. For both groups, the offer functions estimated using the alternative bandwidths are essentially identical to those obtained with the original rule-of-thumb bandwidth.

B.4.3. *Precision.* The exercise above, which aims at verifying the sensibility of the estimated offer function to changes in the bandwidths of the kernel density estimates of trial sentences and settlement offers, serves also as an assessment of the stability of my estimation procedure. Even after a non-trivial variation of the bandwidths (plus or minus ten percent), the offer function estimator returns virtually the same results.

Another way of assessing the estimator's precision is by using the bootstrap samples employed in the construction of standard errors for the model parameters, as reported

Figure 7. Settlement offer function estimates with alternative bandwidths—Covariates group two



Note: This figure shows estimates of the settlement offer function of covariate group two based on alternative density estimates of the trial sentences and settlement offers. The alternative density estimates use bandwidths that are 90, 95, 105 and 110 percent of those employed in the main text.

in table 6 and discussed in Appendix A.2. I compute the standard deviation of the estimated offer function for each point of its domain. I then calculate the average standard deviation across the function's domain, weighting the point-wise standard deviations by the unconditional distribution of trial sentence densities, as reported in figure 14. For covariate group one, the average standard deviation is 25.74 months. For group two, it is 10.84 months.

B.5. **Extra empirical results.** This section reports results of the estimation of the basic model for covariates groups three to 12. In the interest of space, I do not show these results in the main text. They are largely consistent with the results for groups one and two, and point to a substantial variation across races in the outcomes of criminal cases. Notice that the covariate group classification in table 5 (Section 6) attributes even numbers to groups in which the defendant is African-American and odd numbers to groups in which the defendant is not. It is possible to assess the differences between African-American and non-African-American defendants by comparing the estimation results for subsequent odd and even-numbered groups. Groups seven and eight are the only groups for which the estimation results differ considerably from those in the main text.

Figure 8. Conditional trial sentence density estimates—lenient and harsh judges

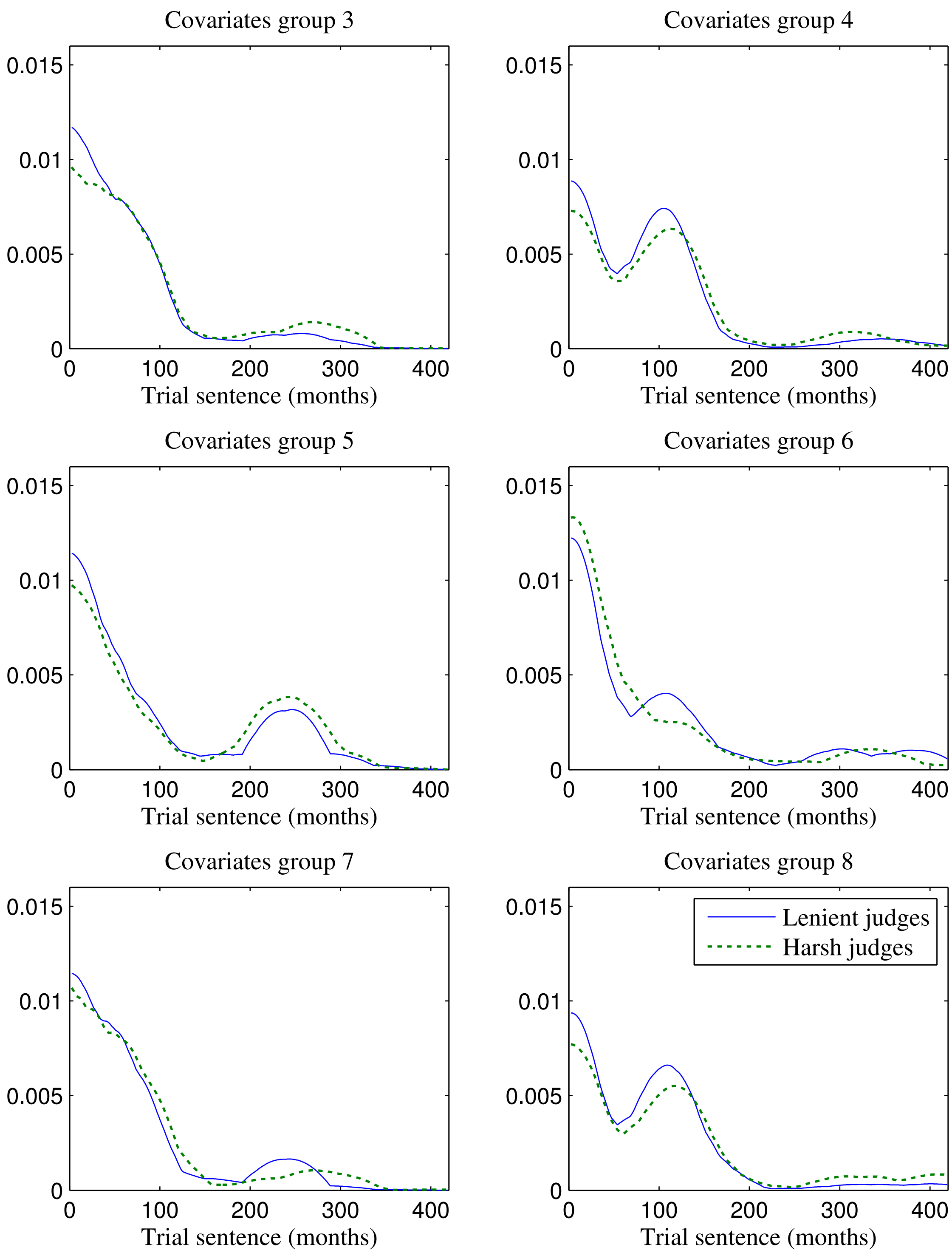
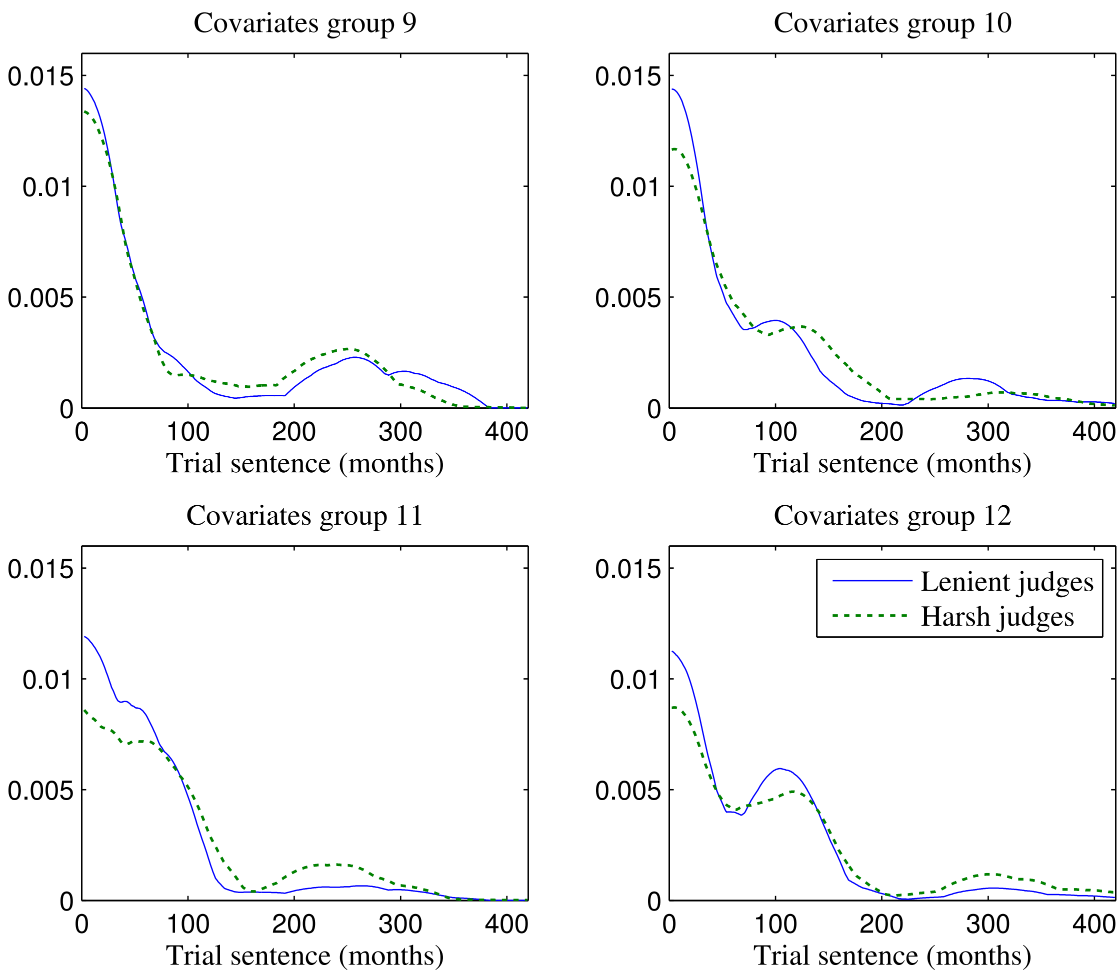


Table 17. Parameter Estimates by Covariate Group

Group	Parameters				
	$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\mu}$
3	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.64 (0.22)	1.00 (0.00)
4	0.03 (0.00)	0.00 (0.00)	0.00 (8.77)	1.43 (0.73)	1.00 (0.00)
5	0.02 (0.00)	0.00 (0.00)	0.00 (0.00)	0.67 (0.11)	1.00 (0.00)
6	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	1.06 (0.31)	1.00 (0.00)
7	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	2.43 (0.83)	1.00 (0.00)
8	0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	0.71 (0.25)	1.00 (0.00)
9	0.03 (0.01)	0.00 (0.00)	0.00 (0.00)	1.25 (0.28)	0.92 (0.04)
10	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	1.16 (0.38)	1.00 (0.00)
11	0.03 (0.00)	0.00 (0.00)	0.00 (0.00)	0.48 (0.08)	1.00 (0.00)
12	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.91 (0.39)	1.00 (0.00)

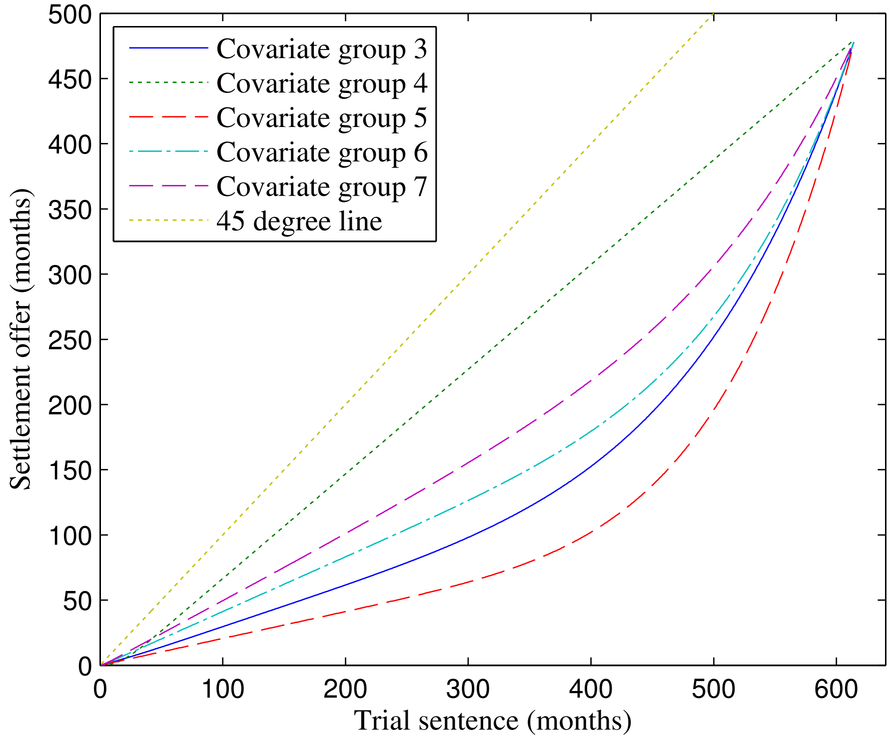
Notes: MLE estimates of the model parameters, conditional on covariates. See table 5 for a description of the covariate groups. Bootstrap standard errors in parenthesis.

FIGURE 9. Conditional trial sentence density estimates—lenient and harsh judges (cont.)



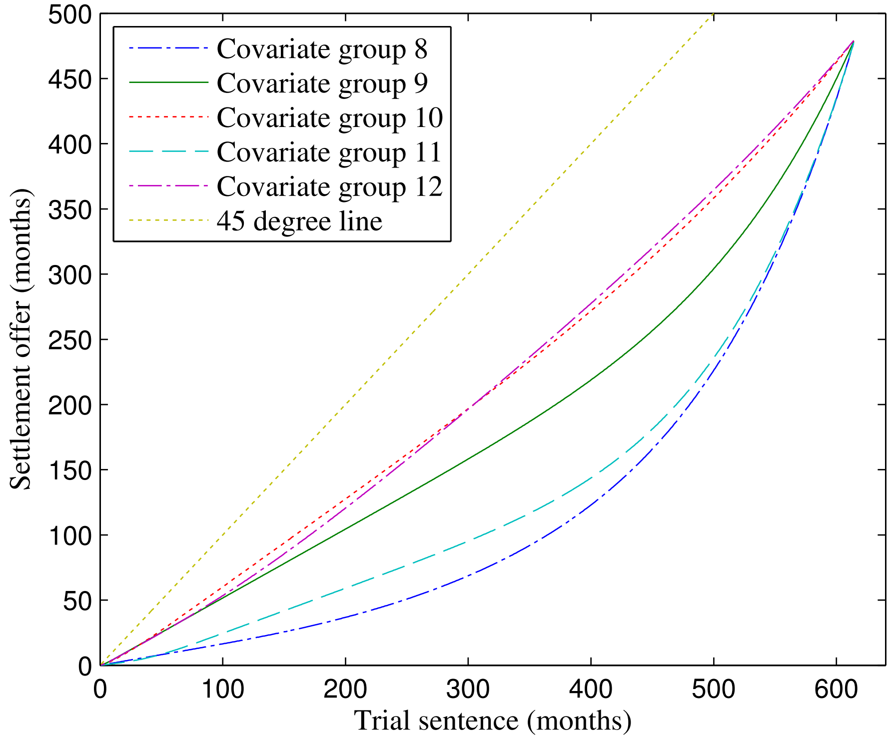
Note: Kernel density estimates of trial sentences assigned by lenient and harsh judges, conditional on covariates. See table 5 for a description of the covariate groups and Section 3 for details on the classification of judges.

FIGURE 10. Settlement offer function estimates—covariate groups 3 to 7



Note: Spline regression estimates of the prosecutor's settlement offer functions. See table 5 for a description of the covariate groups.

Figure 11. Settlement offer function estimates—covariate groups 8 to 12



Note: Spline regression estimates of the prosecutor's settlement offer functions. See table 5 for a description of the covariate groups.

FIGURE 12. Defendants' types distribution estimates

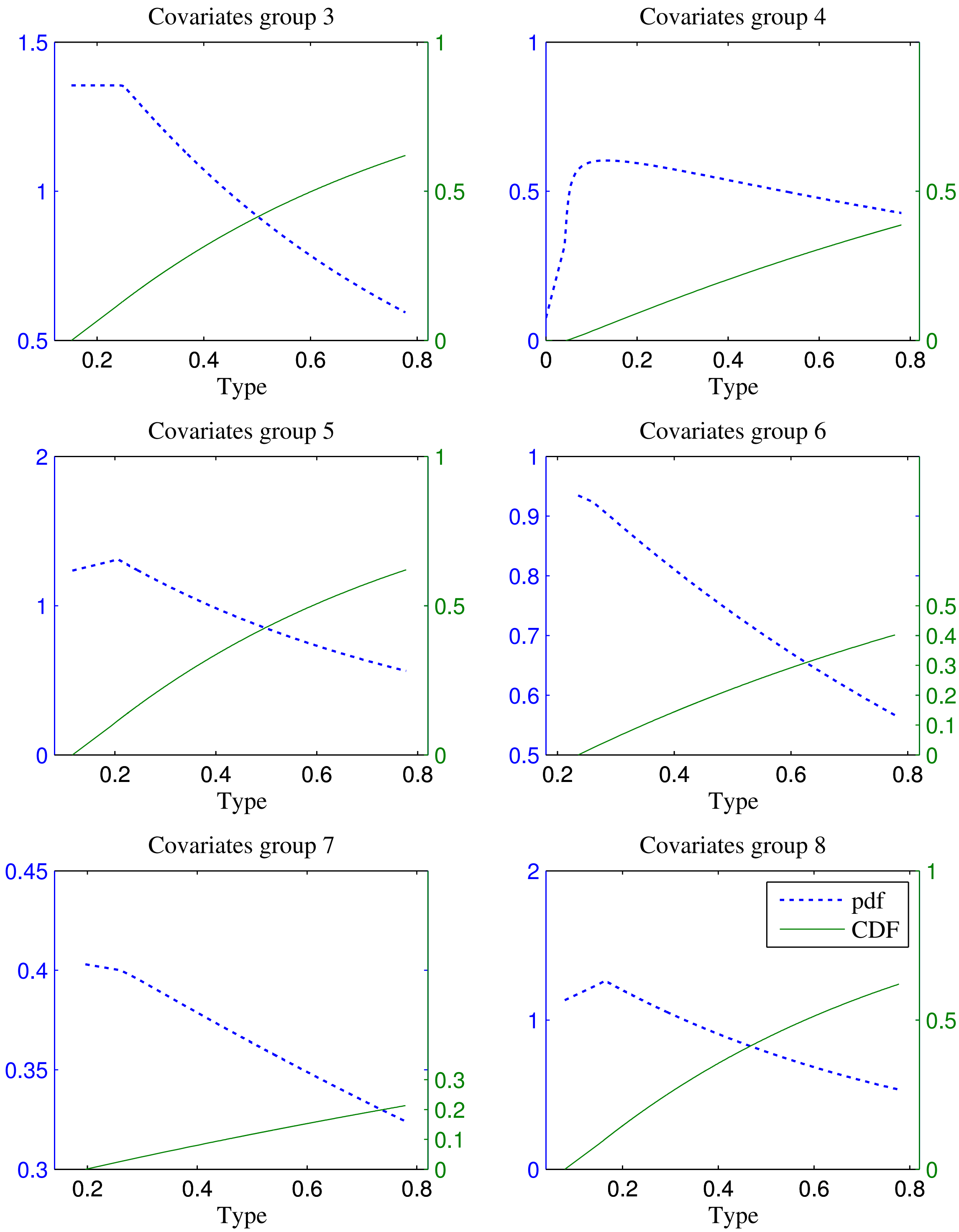
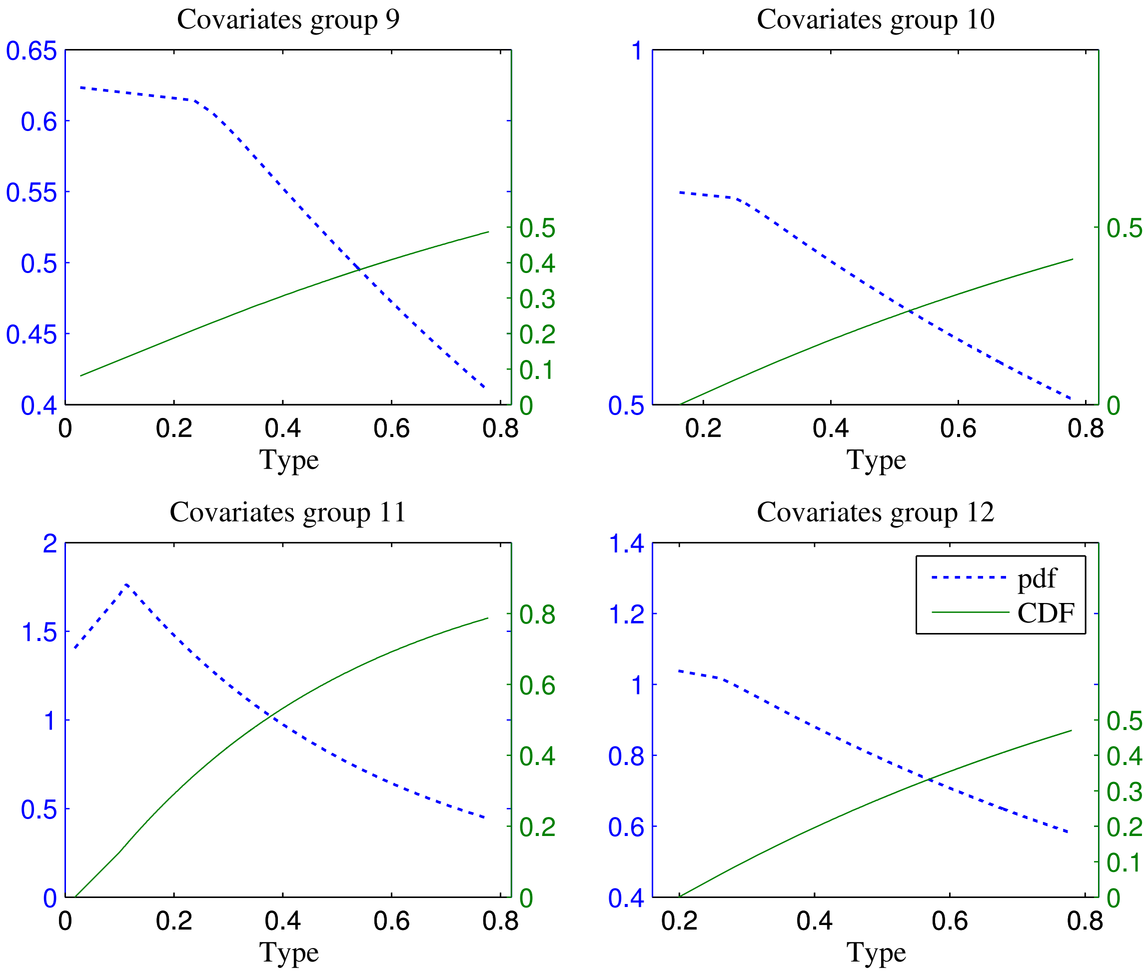


FIGURE 13. Defendants' types distribution estimates (cont.)



Note: Estimated pdf and CDF of the distribution of defendants' types (probabilities of conviction at trial), conditional on covariates. See table 5 for a description of the covariate groups. The distributions are only identified over part of their support.

Figure 14. Estimated unconditional distribution of trial sentences

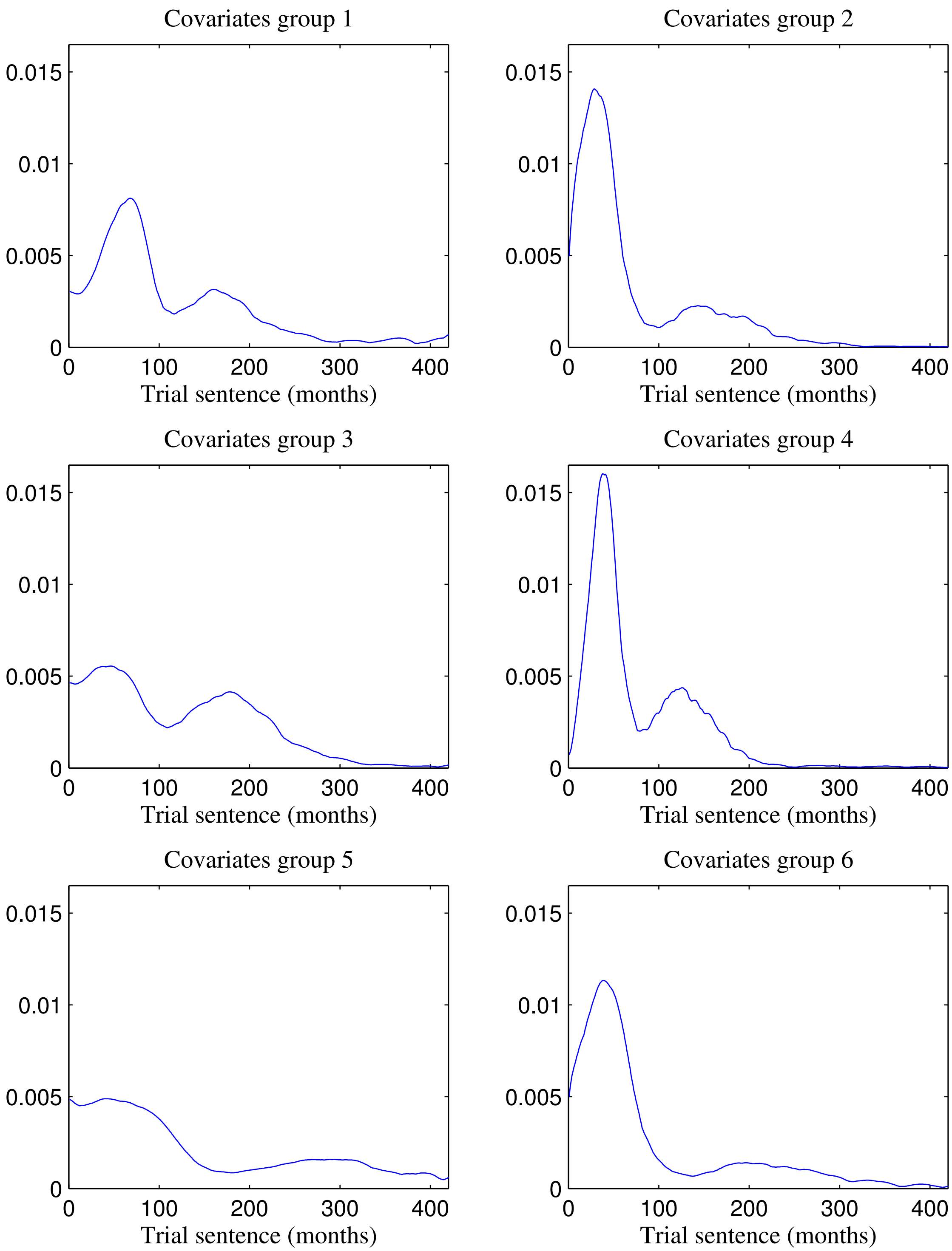
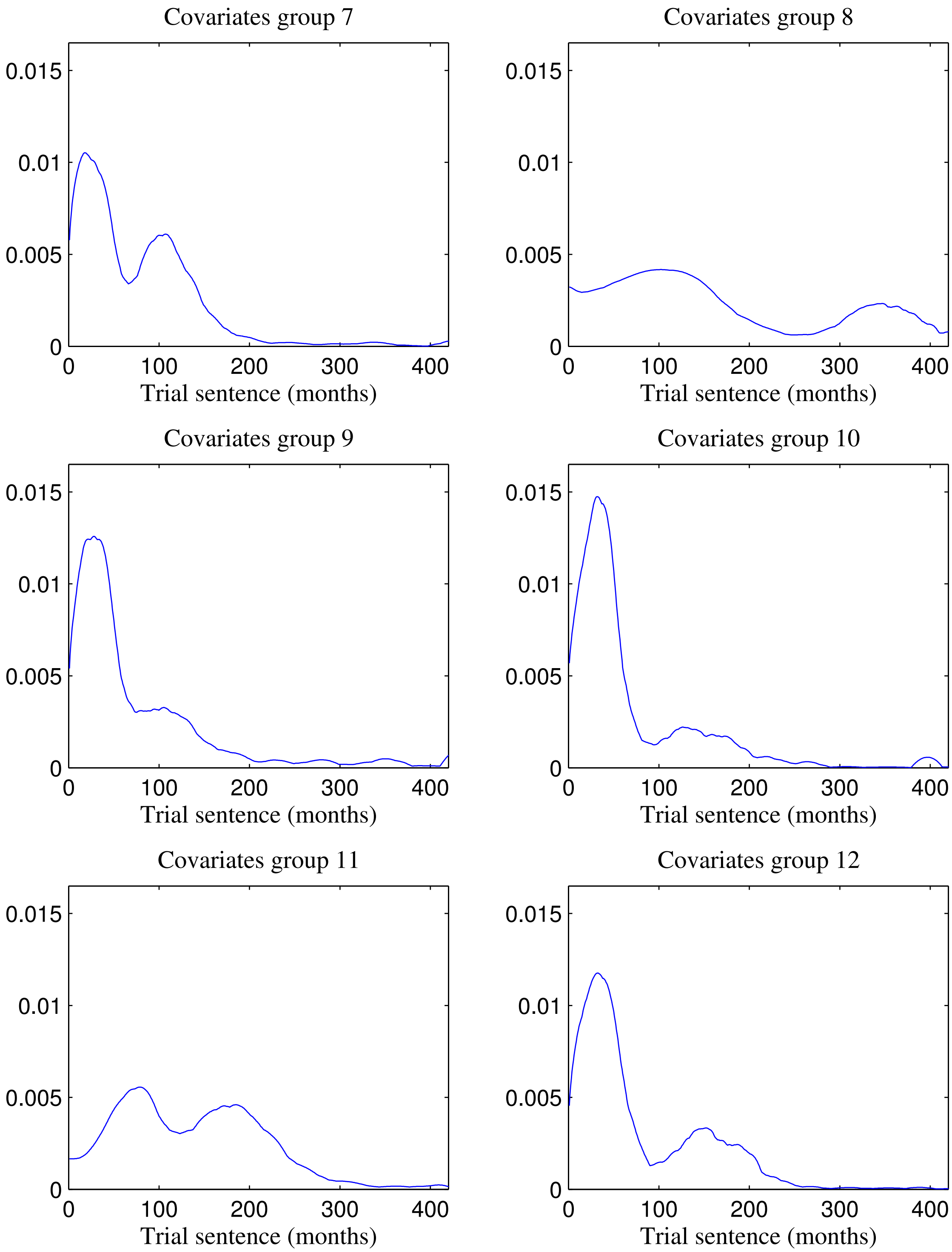


FIGURE 15. Estimated unconditional distribution of trial sentences (cont.)



Note: Distribution of trial sentences for all judges, unconditional on the case outcome. I obtain this distribution based on estimated model.

Figure 16. Estimated prosecutor's expected payoff, conditional on the trial sentence

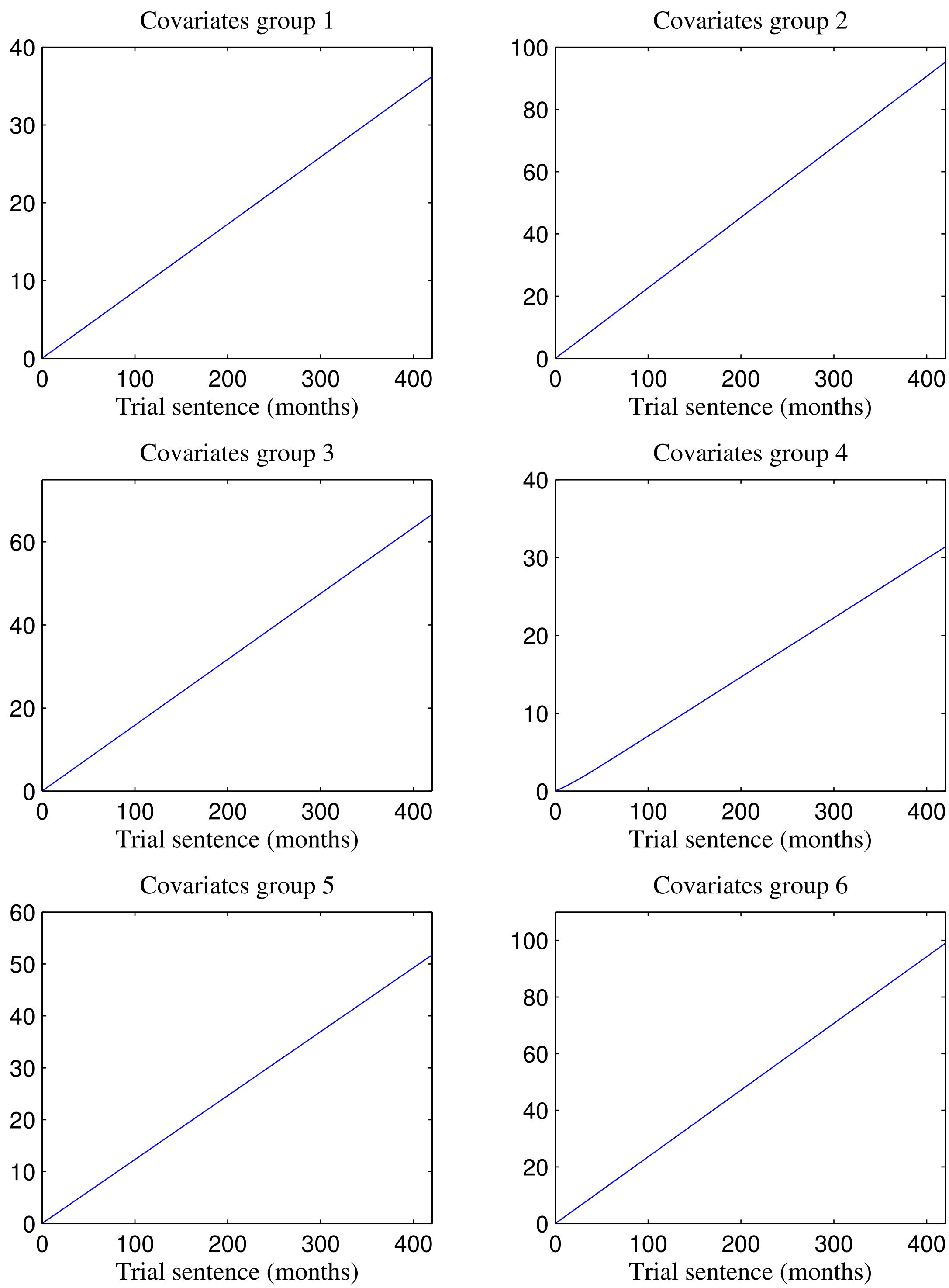
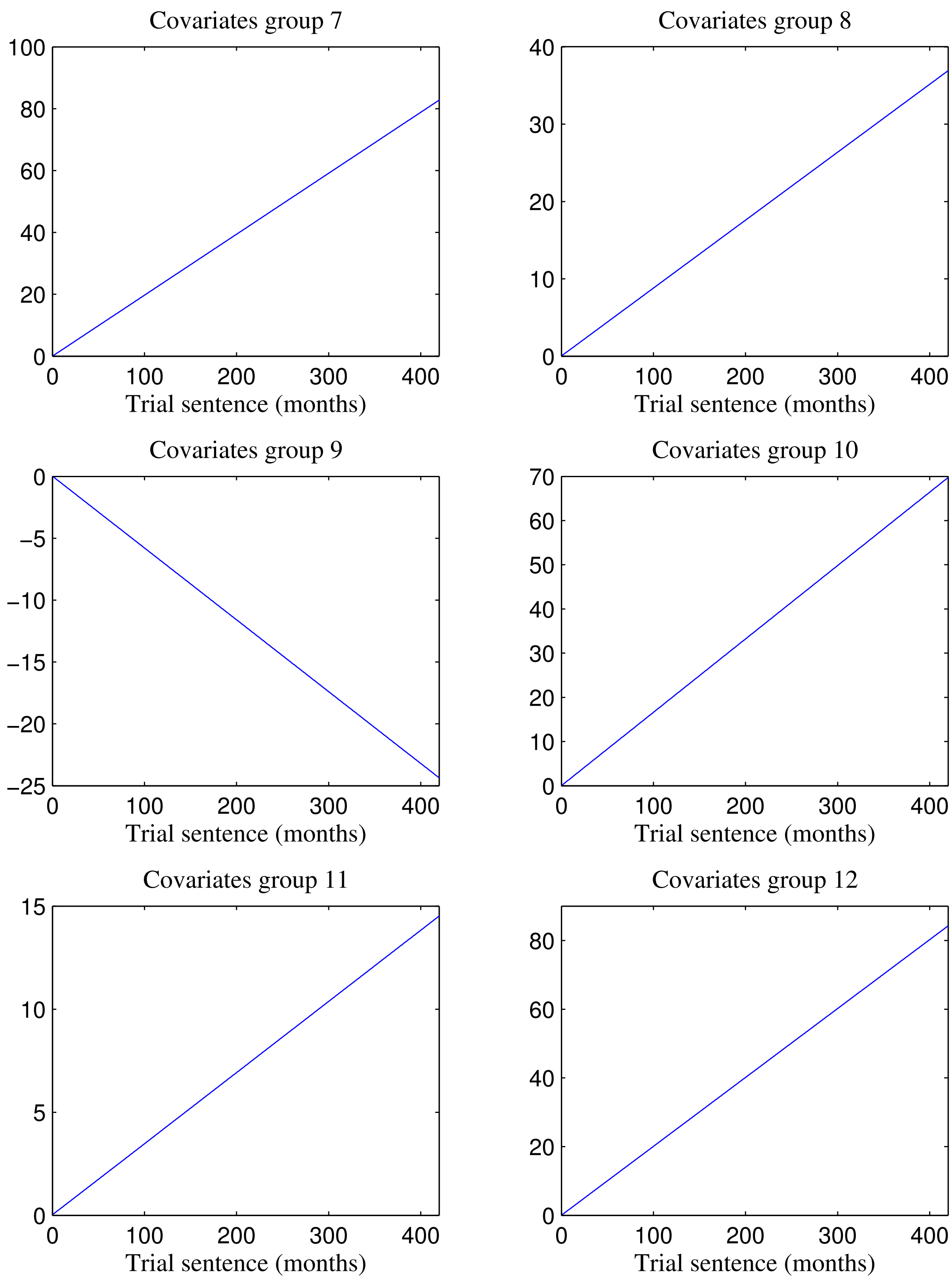


FIGURE 17. Estimated prosecutor's expected payoff, conditional on the trial sentence (cont.)



Prosecutor's payoff, as defined in the optimization problem (4.2), Section 4.

Table 18. Fitted values versus data—Distribution of outcomes

Group		Conviction		
		Any ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
3	Data	38.35%	34.60%	3.75%
	Model	36.66%	34.59%	2.08%
4	Data	44.11%	36.66%	7.45%
	Model	40.68%	36.64%	4.04%
5	Data	38.16%	35.11%	3.05%
	Model	35.55%	34.44%	1.10%
6	Data	40.32%	37.53%	2.79%
	Model	39.68%	37.64%	2.04%
7	Data	38.16%	35.11%	3.05%
	Model	37.53%	36.10%	1.43%
8	Data	40.32%	37.53%	2.79%
	Model	38.62%	37.60%	1.01%
9	Data	21.65%	19.37%	2.28%
	Model	21.59%	19.51%	2.08%
10	Data	29.43%	25.51%	3.92%
	Model	29.09%	25.83%	3.26%
11	Data	21.65%	19.37%	2.28%
	Model	20.91%	19.33%	1.58%
12	Data	29.43%	25.51%	3.92%
	Model	28.61%	25.02%	3.59%

Notes: $\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
See table 5 for a description of the covariate groups.

Table 19. Fitted values versus data—Sentences

Group		Average sentence, conditional on method of resolution†		
		All ($\Psi \in \{1, 2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
3	Data	52.43	46.79	104.45
	Model	54.29	45.64	198.33
4	Data	60.48	50.51	109.54
	Model	51.98	45.96	106.46
5	Data	53.53	47.65	121.25
	Model	52.76	45.55	277.60
6	Data	41.27	38.92	72.95
	Model	42.93	39.13	113.04
7	Data	47.35	42.96	98.01
	Model	43.43	41.21	99.55
8	Data	44.81	38.49	129.96
	Model	47.07	40.66	284.73
9	Data	47.76	41.13	104.00
	Model	46.15	42.25	82.86
10	Data	42.67	37.19	78.36
	Model	43.55	38.83	80.99
11	Data	44.68	41.35	72.95
	Model	55.21	41.45	223.86
12	Data	54.54	47.64	99.47
	Model	50.59	43.56	99.63

Notes: $\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

TABLE 20. Counterfactual results—Sentencing reform (probabilities of conviction and settlement)

Group		Conviction		
		Any	Settlement	Trial
	Current	36.66%	34.59%	2.08%
3	-20% trial sentence length	37.08%	35.49%	1.59%
	-10% incarceration cases	32.86%	31.02%	1.84%
	Current	40.68%	36.64%	4.04%
4	-20% trial sentence length	41.58%	38.21%	3.37%
	-10% incarceration cases	36.40%	32.56%	3.84%
	Current	35.55%	34.44%	1.10%
5	-20% trial sentence length	36.08%	35.43%	0.65%
	-10% incarceration cases	32.02%	31.23%	0.79%
	Current	39.68%	37.64%	2.04%
6	-20% trial sentence length	40.00%	37.95%	2.05%
	-10% incarceration cases	35.76%	33.70%	2.06%
	Current	37.53%	36.10%	1.43%
7	-20% trial sentence length	37.44%	36.18%	1.26%
	-10% incarceration cases	33.58%	32.36%	1.22%
	Current	38.62%	37.60%	1.01%
8	-20% trial sentence length	39.46%	38.81%	0.65%
	-10% incarceration cases	34.91%	34.07%	0.84%
	Current	21.59%	19.51%	2.08%
9	-20% trial sentence length	21.57%	19.56%	2.01%
	-10% incarceration cases	19.24%	17.36%	1.88%
	Current	29.09%	25.83%	3.26%
10	-20% trial sentence length	29.44%	26.64%	2.80%
	-10% incarceration cases	26.07%	23.09%	2.98%
	Current	20.91%	19.33%	1.58%
11	-20% trial sentence length	21.52%	20.55%	0.97%
	-10% incarceration cases	18.23%	17.07%	1.16%
	Current	28.61%	25.02%	3.59%
12	-20% trial sentence length	28.98%	25.92%	3.06%
	-10% incarceration cases	25.69%	22.53%	3.16%

Notes: This table reports the results of two counterfactual exercises. In the first one I reduce the length of the trial sentences of every case by 20 percent. In the second I set the incarceration sentences of all cases below the tenth percentile to zero. The current values are the ones fitted by the estimated model. See table 5 for a description of the covariate groups.

Table 21. Counterfactual results—Sentencing reform (sentences)

Group		Expected sentence†	
		Given $\Psi \in \{1, 2\}$	Unconditional
	Current	54.29	19.91
3	-20% trial sentence length	38.83	14.40
	-10% incarceration cases	59.38	19.51
	Current	51.98	21.14
4	-20% trial sentence length	38.79	16.13
	-10% incarceration cases	57.77	21.03
	Current	52.76	18.75
5	-20% trial sentence length	34.47	12.44
	-10% incarceration cases	56.46	18.08
	Current	42.93	17.03
6	-20% trial sentence length	33.85	13.54
	-10% incarceration cases	47.69	17.05
	Current	43.43	16.30
7	-20% trial sentence length	33.71	12.62
	-10% incarceration cases	47.80	16.05
	Current	47.07	18.18
8	-20% trial sentence length	31.42	12.40
	-10% incarceration cases	51.09	17.84
	Current	46.15	9.96
9	-20% trial sentence length	36.41	7.85
	-10% incarceration cases	51.18	9.85
	Current	43.55	12.67
10	-20% trial sentence length	33.45	9.85
	-10% incarceration cases	48.06	12.53
	Current	55.21	11.54
11	-20% trial sentence length	36.73	7.90
	-10% incarceration cases	58.95	10.75
	Current	50.59	14.47
12	-20% trial sentence length	38.57	11.18
	-10% incarceration cases	55.52	14.26

Notes: This table reports the results of two counterfactual exercises. In the first one I reduce the length of the trial sentences of every case by 20 percent. In the second I set the incarceration sentences of all cases below the tenth percentile to zero. The current values are the ones fitted by the estimated model. See table 5 for a description of the covariate groups.

$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.

† Measured in months.

Table 22. Counterfactual results—No plea bargaining

Group		Outcome	
		Probability of conviction	Expected sentence†
3	Current	36.66%	19.91
	No plea bargaining	[24.03% , 27.81%]	[34.05 , 39.29]
4	Current	40.68%	21.14
	No plea bargaining	[31.53% , 38.32%]	[24.98 , 30.36]
5	Current	35.55%	18.75
	No plea bargaining	[21.62% , 25.25%]	[40.41 , 46.82]
6	Current	39.68%	17.03
	No plea bargaining	[29.35% , 35.19%]	[27.87 , 33.45]
7	Current	37.53%	16.30
	No plea bargaining	[28.68% , 35.91%]	[23.64 , 29.41]
8	Current	38.62%	18.18
	No plea bargaining	[23.44% , 27.29%]	[42.63 , 49.48]
9	Current	21.59%	9.96
	No plea bargaining	[16.22% , 19.91%]	[13.84 , 16.74]
10	Current	29.09%	12.67
	No plea bargaining	[22.39% , 27.71%]	[15.70 , 18.90]
11	Current	20.91%	11.54
	No plea bargaining	[12.04% , 13.61%]	[19.44 , 21.65]
12	Current	28.61%	14.47
	No plea bargaining	[21.84% , 26.21%]	[18.24 , 21.71]

Notes: This table reports the results of forcing all cases to go to trial. Because the distribution of defendants' types is not identified over its entire support, I can only calculate bounds for the probability of conviction and the expected sentence. The current values are the ones fitted by the estimated model.
† Measured in months.
See table 5 for a description of the covariate groups.

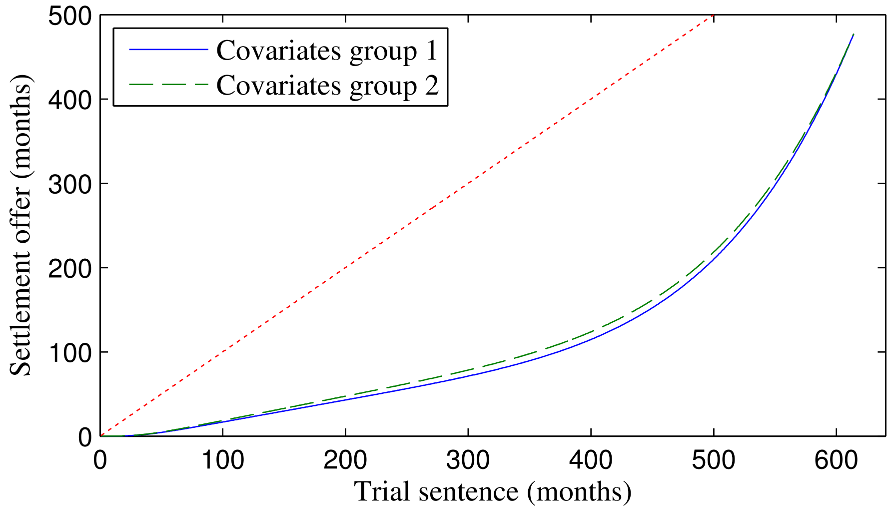
B.6. **Extension: Incorporating non-incarceration sentences.** In the analysis presented in the main text I disregard non-incarceration sentences, such as probation and community service. That is, I treat any case resulting in such a sentence as dropped by the prosecutor (if the case was settled) or as an acquittal (if the non-incarceration sentence was assigned at trial). Here I take an alternative approach. In addition to incarceration sentences, I consider probation and community service sentences, and I normalize the length of the non-incarceration sentences, so that their sample maximum is equivalent to the fifth percentile of the distribution of incarceration sentences.[57] Under this criterium, 61.31 percent of the cases of the sample are settled, 3.20 percent result in a trial acquittal and 4.12 percent result in a trial conviction. The remaining cases are dismissed. The settlement sentences have an average of 19.88 months and a standard deviation of 38.84 months. For the trial sentences, these statistics are 87.04 and 104.69, respectively.

I use this sample to implement the settlement offer function estimator described in the main text. Figure 18 shows the estimation results for covariate groups one and two. The estimated settlement offers are "more generous" than the ones reported in the main text—that is, they indicate that, given the same trial sentence, the prosecutor offers to settle for a shorter sentence. For example the settlement offers corresponding to a trial sentence of 200 months in figure 18 are 42.97 months for group one and 47.48 for group two. According to the estimators in the main text (Section 6, figure 3), these settlement offers are 72.18 and 115.00 months for groups one and two, respectively. The cross-race differences pointed out in the main text still hold. Conditional on the trial sentence, the prosecutors offer shorter sentences to non-African-American defendants than to African-American ones. But, as figure 18 makes clear, these differences become much less pronounced once non-incarceration sentences are considered.

In spite of the much larger number of observations employed, the offer function estimates shown in figure 18 have a worse fit to the data than the ones reported in the main text. As above, I use the objective function in (5.14) multiplied by 100,000 as a measure of fit. For the estimates incorporating non-incarceration sentences, this measure is 65.76 (group one) and 75.26 (group two). For the main estimates, as reported in table 16, the measure is 20.83 and 39.47 for groups one and two, respectively.

[57] I obtain results that are very similar to the ones presented below if, instead of the fifth, I use the third percentile of the incarceration sentences distribution in the normalization.

FIGURE 18. Settlement offer—Incorporating non-incarceration sentences



This figure shows the estimated offer functions for groups one and two using data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.

Table 23 presents estimates of the trial costs parameters and the auxiliary parameter μ. Qualitatively, most of these estimates are similar to the ones reported in the main text: Both $\hat{\alpha}_d$ and $\hat{\beta}_d$ are zero, indicating that defendants behave as if trials were essentially costless. $\hat{\alpha}_p$ is positive but very small: 2.08 and 0.27 for groups one and two, respectively. Similarly to the main text estimates, $\hat{\mu}$ is close to one for both groups. The estimates of $\hat{\beta}_p$ are the only ones that differ substantially from those in the main text. While the latter are roughly one for both groups, the former are much larger: 11.52 for group one and 4.98 for group two. The high prosecutor's costs implied by these estimates rationalize the large proportion of cases resulting in a settlement, once non-incarceration sentences are considered.

Figure 19 shows the estimated distributions of defendants' types for covariate groups one and two. Like in the main text, the identified range of the support comprises most of the unit line (from zero to roughly 0.8), and the estimated distributions suggest at least two modes: One at zero, and another greater than 0.8.

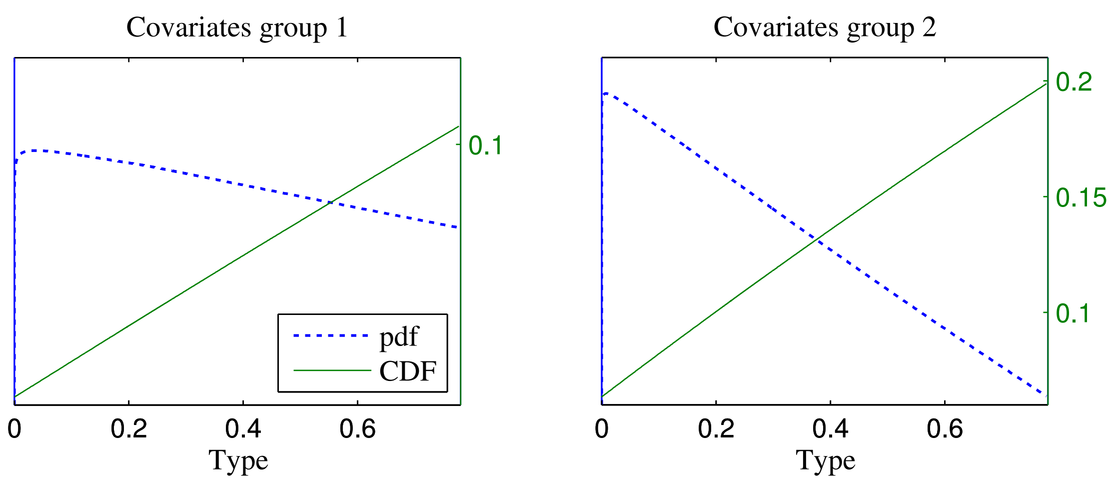
Table 24 shows the fit of the model to the data. The fit is not as good as that of the estimates in the main text. For both groups one and two, the model over-predicts the settlement probability and under-predicts the probability of conviction at trial.

Table 23. Parameter estimates by covariate group

Group	Parameters				
	$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\mu}$
	0.00	0.00	2.08	11.52	0.96
	0.00	0.00	0.27	4.98	0.94

Notes: MLE estimates of the model parameters, conditional on covariates. I obtain these estimates using data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
See table 5 for a description of the covariate groups.

Figure 19. Defendants' types distribution estimates



Note: Estimated pdf and CDF of the distribution of defendants' types (probabilities of conviction at trial), conditional on covariates. The distributions are only identified over part of their support. The estimates employ data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.

The model also over-predicts the expected sentence, conditional on a trial conviction, to a considerably larger extent than the main estimates.

Tables 25 and 26 present the results of the counterfactual analysis. Specifically, the former table contains the outcomes of the sentencing reform simulations, and the

TABLE 24. Fitted values versus data

Group		Conviction probability		
		Any ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	65.43%	61.31%	4.12%
	Model	65.23%	65.15%	0.07%
2	Data	62.71%	55.08%	7.63%
	Model	62.12%	61.96%	0.16%
Group		Average sentence, conditional on method of resolution†		
		All ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	33.05	27.02	122.85
	Model	26.70	26.37	329.11
2	Data	31.20	24.77	77.64
	Model	24.38	23.88	218.70

Notes: I obtain these estimates using data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

latter reports the effects of a ban on plea bargains. The results indicate that reducing mandatory minimum sentences by 20 percent would increase conviction rates by roughly three p.p., for group one, and four p.p., for group two—even stronger impacts than those suggested by the estimates in the main text. Like in the main text, the same intervention would reduce expected sentences by roughly 20 percent. Also similarly to the main text results, the assignment of non-incarceration sentences to all cases with sentences below the tenth percentile would greatly reduce the conviction rate (by roughly 15 percent for both groups) and leave the expected sentence essentially unchanged. Regarding the abolition of plea bargains, the results in table 26 suggest that such an intervention would not necessarily lead to a decrease in incarceration convictions, as, for both groups, the estimated upper bounds of the conviction probabilities in the scenario without plea bargaining are very similar to the current probabilities. According to these results, eliminating plea bargaining would also cause

TABLE 25. Counterfactual results—Sentencing reform

Group		Conviction probability		
		Any	Settlement	Trial
1	Current	65.23%	65.15%	0.07%
	-20% trial sentence length	68.17%	68.14%	0.03%
	-10% incarceration cases	54.18%	54.10%	0.08%
2	Current	62.12%	61.96%	0.16%
	-20% trial sentence length	66.21%	66.08%	0.13%
	-10% incarceration cases	52.46%	52.30%	0.16%
Group		Expected sentence†		
		$\Psi \in \{1, 2\}$	Unconditional on Ψ	
1	Current	26.70	17.42	
	-20% trial sentence length	16.91	11.53	
	-10% incarceration cases	33.27	18.03	
2	Current	24.38	15.15	
	-20% trial sentence length	17.19	11.38	
	-10% incarceration cases	30.29	15.89	

Notes: This table reports the results of two counterfactual exercises. In the first one I reduce the length of the trial sentences of every case by 20 percent. Such a scenario intends to simulate the lowering of mandatory minimum sentences for all types of cases. In the second exercise I set the incarceration sentences of all cases below the tenth percentile to zero, with the objective of capturing the broader assignment of alternative sentences to mild offenders. The current values are the ones fitted by the estimated model.
I obtain these estimates using data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

an increase of almost 400 percent in the expected sentences, indicating that defendants would be much worse off without the possibility of settling their cases. Thus the results of the ban on plea bargains differ somewhat from those presented in the main text, since the former suggest an unequivocal reduction in the conviction probability and a more modest increase in the expected sentences.

TABLE 26. Counterfactual results—No plea bargaining

Group		Outcome	
		Probability of conviction	Expected sentence†
1	Current	65.23%	17.42
	No plea bargaining	[51.37% , 65.88%]	[58.01 , 74.79]
2	Current	62.12%	15.15
	No plea bargaining	[48.94% , 62.09%]	[54.48 , 69.44]

Notes: This table reports the results of forcing all cases to go to trial. Because the distribution of defendants' types is not identified over its entire support, I can only calculate bounds for the probability of conviction and the expected sentence. The current values are the ones fitted by the estimated model.
I obtain these estimates using data on incarceration, probation and community service sentences. I normalize the non-incarceration sentences, so that their sample maximum equals the fifth percentile of the incarceration sentences distribution.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

As shown above, many of the estimation results obtained using non-incarceration sentences are similar to the ones presented in the main text—with a few notable differences concerning the prosecutor's trial costs and the effects of a ban on plea bargains on conviction rates. But, overall, the estimates incorporating non-incarceration sentences offer a substantially worse fit to the data than those reported in the body of the paper, which leads me to choose the latter as my preferred estimates.

B.7. **Extension: Relaxing the independence between T and Θ.** In the empirical model described in Section 5, I assume that the potential trial sentences and the defendants' types are independently distributed. Below, I relax this assumption and show that it is still possible to obtain partial identification of the model. In fact, the optimal settlement offer function is exactly identified and can be estimated by the procedure proposed in Section 5. Therefore, the estimates of settlement offer functions presented in the main text are robust to the dependence between defendants' types and potential trial sentences, in the way defined below.

Let $F(\cdot|T)$ be the distribution function of defendants' types, conditional on the potential trial sentence T. Assume that, for all $t \in [\underline{t}, \bar{t}]$, the function $F(\cdot|T)$ satisfies the technical assumptions outlined in Section 4. Denote the density and the hazard functions associated with $F(\cdot|T)$ by $f(\cdot|T)$ and $\lambda(\theta, t)$, respectively. Assume that

$\lambda(\theta, t)$ is differentiable in both its arguments, and that

$$\frac{\partial}{\partial\theta}\lambda(\theta, t) > 0 \quad \text{and} \quad \frac{\partial}{\partial t}\lambda(\theta, t) < \frac{1}{c_p + c_d} \tag{B.3}$$

for all $t \in [\underline{t}, \bar{t}]$ and $\theta \in (\underline{\theta}, \bar{\theta})$. The first inequality ensures that the hazard function associated with $F(\cdot|T)$ is increasing in θ. The second one limits the amount of mass $F(\cdot|T)$ can redistribute towards low defendants' types as the potential trial sentence increases.

Given a realization t of T, the equilibrium of the bargaining game can be found in the same way as in Section 4. It is characterized by (4.1) and by the first-order condition for the prosecutor, which is given by

$$\frac{t}{c_p + c_d} = \frac{f\left[\theta(s^*)|T = t\right]}{1 - F\left[\theta^*|T = t\right]}. \tag{B.4}$$

As before, I define the equilibrium settlement offer and defendant's threshold type as functions of t, and denote them by $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$, respectively. Using condition (B.3), and applying the implicit function theorem to (B.4), I have that $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$ are strictly increasing in t, and $\tilde{s}(\cdot)$ is strictly convex. The argument in Section 5 can then be easily adapted to show that, given the appropriate exogenous variation in the distribution of trial sentences, the function $\tilde{s}(\cdot)$ is identified. Notice that the estimator of $\tilde{s}(\cdot)$ proposed in the main text does not make direct use of the distribution of defendants' types. That means that it can be applied under the more general conditions described here, and the estimates of $\tilde{s}(\cdot)$ presented in Section 6 are still valid.

Although I am able to recover the optimal offer function after relaxing the independence assumption between T and Θ, the exact identification of the model's primitives does not hold. I now outline a strategy for the partial identification of such primitives. I begin by noticing that I can still recover $\tilde{\theta}(t)$ and the hazard function $\lambda(\tilde{\theta}(t), t)$ for all $t \in [\underline{t}, \bar{t}]$, up to the scalars c_d and c_p. I now strengthen condition (B.3), by assuming that

$$\frac{\partial}{\partial\theta}\lambda(\theta, t) > 0 \quad \text{and} \quad \frac{\partial}{\partial t}\lambda(\theta, t) < 0. \tag{B.5}$$

The second inequality in this condition implies that $F(\cdot|T)$ places more mass on high defendants' types as the potential trial sentence increases. A consequence of this inequality is that $\lambda(\theta, t) \leq \lambda(\theta, \tilde{\theta}^{-1}(\theta))$ for all $\tilde{\theta}(\underline{t}) \leq \theta \leq \tilde{\theta}(t)$. Therefore, I have that

$$F(\theta|T) = 1 - \exp\left(-\int_0^\theta \lambda(x,t)\,dx\right)$$
$$\leq 1 - \underline{\mu}\exp\left(-\int_{\tilde{\theta}(\underline{t})}^\theta \lambda(x, \tilde{\theta}^{-1}(x))\,dx\right)$$

for all $\tilde{\theta}(\underline{t}) \leq \theta \leq \tilde{\theta}(t)$, where $\underline{\mu} = \exp\left(-\int_0^{\tilde{\theta}(\underline{t})} \lambda(x,\underline{t})\,dx\right)$. Define the function

$$\ddot{F}(t) = 1 - \underline{\mu}\exp\left(-\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x, \tilde{\theta}^{-1}(x))\,dx\right).$$

I then have that $\ddot{F}(t) \geq F\left(\tilde{\theta}(t)|T = t\right)$ for all $t \in [\underline{t}, \bar{t}]$. Now consider the function

$$\ddot{g}(t) = \left[1 - \ddot{F}(t)\right]^{-1}\left[\frac{\partial}{\partial t}\tilde{s}(t)\right] b\left(\tilde{s}(t)|\Psi = 1\right) P[\Psi = 1].$$

Notice that

$$\ddot{g}(t) \geq \left[1 - F\left(\tilde{\theta}(t)\right)\right]^{-1}\left[\frac{\partial}{\partial t}\tilde{s}(t)\right] b\left(\tilde{s}(t)|\Psi = 1\right) P[\Psi = 1] = g(t)$$

for all $t \in [\underline{t}, \bar{t}]$, where the equality comes from (5.1), (5.4) and (5.5). The expression above allows me to write

$$1 - \int_{[\underline{t},\bar{t}]} \ddot{F}(t)\ddot{g}(t)\,dt \leq 1 - \int_{[\underline{t},\bar{t}]} F(\tilde{\theta}(t)|T = t)g(t)\,dt. \tag{B.6}$$

The expression on the right-hand side of (B.6) is the probability that a case is settled, conditional on $\Psi \neq 0$ (i.e., on it not being withdrawn by the prosecutor). That probability is observed by the econometrician. The expression on the left-hand side is known only up to the scalars c_d, c_p and $\underline{\mu}$. Thus, (B.6) establishes a non-linear bound for such scalars. If appropriate exogenous variation on the distribution of trial sentences is available, the equation may imply multiple bounds—which can be combined with the bounds for c_d described in footnote 31, in order to partially identify the model's primitives. Implementing such a strategy is an interesting topic for future research.

B.8. **Extension: Non-linear preferences.** A simplifying assumption of the model estimated in the main text is that both the defendant and the prosecutor's utilities are linear in the assigned sentence. Here I extend the model to allow for non-linearities in the defendant's utility function. Conceptually, this generalization has a small impact on the identification strategy, and I can still follow the estimation steps proposed in the main text. Most of the estimated primitives are very similar to the ones obtained using the simpler, linear utility model. Reassuringly, the results of the counterfactual analysis are also analogous to the ones presented in Section 7.

Below I specify the data-generating process for the extended model, followed by a brief description of the identification and estimation strategies. I then present the empirical results, stressing the similarities and differences between the outcomes of the extended and the basic models.

B.8.1. *Data-generating process.* For each case in the data, a prosecutor and a defendant bargain as follows: The prosecutor offers the defendant to settle for a sentence s. If the defendant rejects the offer, the case is brought to trial and the defendant is found guilty with probability Θ. This probability is draw from a distribution F that is twice-differentiable over the support $(\underline{\theta}, \bar{\theta}) \subseteq (0,1)$. The associated density f is strictly positive on $(\underline{\theta}, \bar{\theta})$ and is non-increasing in a neighborhood of $\bar{\theta}$. Only the defendant knows the realization of Θ at the beginning of the game, and henceforth this realization is denoted the defendant's type.

Let the random variable Z with support $\{l, h\}$ represent the judge responsible for the case. In the event of a conviction at trial, a sentence T is assigned. With probability $\nu(Z)$, T is equal to zero, which I interpret as a non-incarceration sentence. If T is different from zero, it is distributed according to the CDF $G(\cdot|Z)$, with support $t \in [\underline{t}, \bar{t}]$, where $\underline{t} > 0$. Let $g(\cdot|Z)$ be the associated density, and assume that $g(t|Z) > 0$ for all $t \in [\underline{t}, \bar{t}]$. The realization of T is common knowledge to the defendant and the prosecutor at the beginning of the game. Regardless of the trial outcome, the defendant and the prosecutor pay trial costs of $\alpha_d + \beta_d T$ and $\alpha_p + \beta_p T$, respectively.

The prosecutor's utility is linear on the assigned sentence. Specifically, the prosecutor's utility is given by: (i) s, if the defendant accepts the offer s; (ii) $T - \alpha_p - \beta_p T$, if the case results in a conviction at trial; and (iii) $-\alpha_p - \beta_p T$, if the case results in a trial acquittal. As for the defendant, any sentence $x \geq 0$ decreases her utility by $\frac{x^{1-\eta}}{1-\eta}$, where $\eta \in [0,1]$. Therefore the defendant's utility is: (i) $\frac{s^{1-\eta}}{1-\eta}$, if the case settles for s; (ii) $\frac{T^{1-\eta}}{1-\eta} - \alpha_d - \beta_d T$, in the event of a trial conviction; and (iii) $-\alpha_d - \beta_d T$, if the case results in an acquittal at trial.

Assume that none of the following objects varies with Z: $F, \alpha_d, \beta_d, \alpha_p, \beta_p$ and η. These objects, together with $\nu(Z)$ and $G(\cdot|Z)$, constitute the primitives of the structural model.

I solve the game by backward induction. Given a realization t of T, a defendant of type θ accepts an offer s if and only if $\frac{s^{1-\eta}}{1-\eta} \leq \theta \left(\frac{t^{1-\eta}}{1-\eta} \right) + \alpha_d + \beta_d t$. Thus, given s and

t, the defendant's strategy is characterized by the cutoff

$$\theta(s) = \frac{\frac{s^{1-\eta}}{1-\eta} - \alpha_d - \beta_d t}{\frac{t^{1-\eta}}{1-\eta}} \tag{B.7}$$

such that the defendant accepts the offer if and only if $\theta \geq \theta(s)$. For any t, the prosecutor then solves

$$\max_s \quad \{1 - F[\theta(s)]\} s + F[\theta(s)] \left\{ -\alpha_p - \beta_p t + t \frac{\int_{\underline{\theta}}^{\theta(s)} x f(x) dx}{F[\theta(s)]} \right\}.$$

Under the hypotheses above, the argument in Bebchuk (1984) still applies, and the optimal prosecutor's offer s^* satisfies $\theta(s^*) \in (\underline{\theta}, \bar{\theta})$. The prosecutor's first order condition is

$$\frac{f[\theta(s^*)]}{\{1 - F[\theta(s^*)]\}} = \frac{(s^*)^\eta \left(\frac{t^{1-\eta}}{1-\eta}\right)^2}{(s^* + \alpha_p + \beta_p t)\left(\frac{t^{1-\eta}}{1-\eta}\right) + t\left[\alpha_d + \beta_d t - \frac{(s^*)^{1-\eta}}{1-\eta}\right]}. \tag{B.8}$$

As in Section 4, define the equilibrium settlement offer and defendant's cutoff point as functions of the realized trial sentence t, and denote these functions by $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$, respectively. Assume that $\tilde{\theta}(\cdot)$ is strictly increasing, which implies that $\tilde{s}(\cdot)$ is also strictly increasing.[58] Also, assume that the $\tilde{s}(0) = 0$. The prosecutor's offer is then described by the random variable $S = \tilde{s}(T)$, which is zero with probability $\nu(Z)$ and positive otherwise. Conditional on being positive, S is distributed according to the CDF $B(\cdot|Z)$, which has support $[\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$, is continuous and has an associated density $b(\cdot|Z)$.

Let the random variable Ψ describe the method of resolution of the case: $\Psi = 0$ if the prosecutor drops the case or the parties agree to settle for a non-incarceration sentence; $\Psi = 1$ if the case is settles for an incarceration sentence; $\Psi = 2$ in the event of an incarceration conviction at trial; and $\Psi = 3$ if the defendant is found not guilty at trial. The observables for each case include the realizations ψ of Ψ and z of Z. Moreover, the realization t of T is observed if and only if $\Psi = 2$, while the realization s of S is observed if and only if $\Psi = 1$.

B.8.2. *Identification.* The identification strategy for the extended model is analogous to the one described in Section 4 for the basic model. In fact, the argument for the identification of the offer function $\tilde{s}(\cdot)$ is identical. Using the offer function and

[58]Below, in discussing the identification of the model, I present a set of sufficient conditions for $\tilde{\theta}(\cdot)$ to be strictly increasing, given the settlement offer function $\tilde{s}(\cdot)$.

(B.7), I recover the function $\tilde{\theta}(\cdot)$ for all $t \in [\underline{t}, \bar{t}]$, up to α_d, β_d and η. I can then use (B.8) to identify $\lambda(\cdot)$, the hazard function of $F(\cdot)$, up to $\alpha_d, \beta_d, \alpha_p, \beta_p$ and η. From the knowledge of $\lambda(\cdot)$, the identification of the model primitives follows the steps presented in Section 4 and Appendix A, except that (A.2) now defines a system of infinitely many equations and six variables: $\alpha_d, \beta_d, \alpha_p, \beta_p, \mu$ and η. By solving the system for these variables, I complete the identification of the model.

The argument above relies on the strict monotonicity of $\tilde{\theta}(\cdot)$. From (B.7), this condition holds if and only if

$$t^{\eta}\left[\frac{\tilde{s}'(t)}{\tilde{s}(t)^{\eta}} - \beta_d\right] > \frac{\frac{\tilde{s}(t)^{1-\eta}}{1-\eta} - \alpha_d - \beta_d t}{\frac{t^{1-\eta}}{1-\eta}} \tag{B.9}$$

for all $t \in [\underline{t}, \bar{t}]$. Moreover, by definition, $\tilde{\theta}(\cdot) \in (0,1)$, which is equivalent to

$$\frac{\tilde{s}(t)^{1-\eta}}{1-\eta} - \alpha_d - \beta_d t > 0 \tag{B.10}$$

$$\text{and} \quad \frac{\tilde{s}(t)^{1-\eta}}{1-\eta} - \alpha_d - \beta_d t - \frac{t^{1-\eta}}{1-\eta} < 0 \tag{B.11}$$

for all $t \in [\underline{t}, \bar{t}]$. Together, (B.9), (B.10) and (B.11) impose restrictions on the parameters α_d, β_d and η. These restrictions, which I explicitly consider in the estimation of the model, ensure that $\tilde{\theta}(\cdot)$ behaves as required by the identification strategy.

B.8.3. *Estimation.* The estimation of the extended model is similar to that of the basic model, as presented in the main text and in Appendix A. The estimation of the offer function is exactly the same, and, based on this function, I estimate the parameters that characterize the model primitives by maximum likelihood. But here there are six parameters to be estimated: α_p and β_p, which characterize the prosecutor's trial costs; α_d and β_d, which characterize the defendant's trial costs; η, which characterizes the defendant's utility function; and μ, which captures the behavior of the distribution of defendants' types for values of θ smaller than $\tilde{\theta}(\underline{t})$. As in the main text, I estimate these parameters separately across covariate groups. The estimation of the extended model is computationally intensive, relative to that of the basic model. The high computational costs result, in part, from the extra argument in the optimization procedure. Moreover, the non-linearity of the defendant's preferences complicates the numerical integrations employed in each evaluation of the likelihood function, defined in (A.5). For this reason, I am unable to compute standard errors for the model parameters using bootstrap methods.

TABLE 27. Parameter estimates for covariates group one and two—Extended model

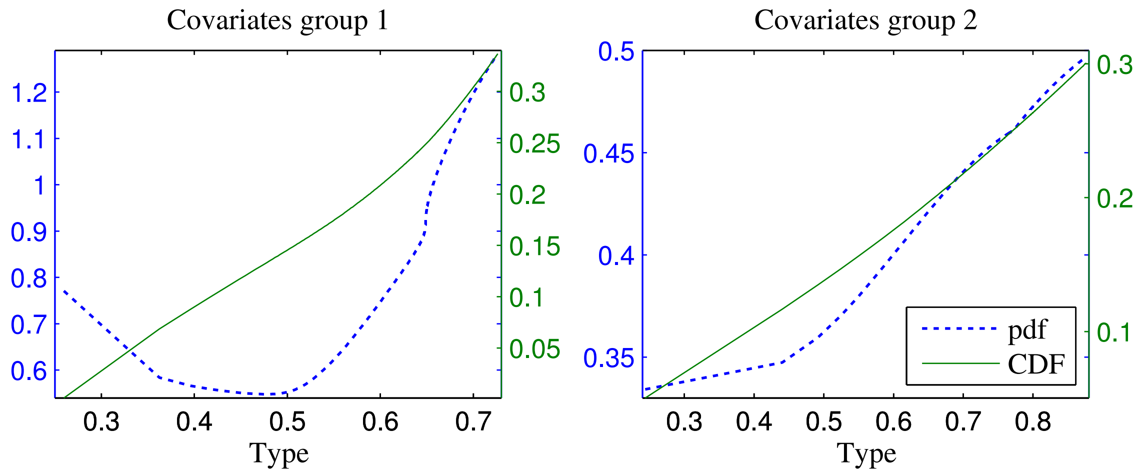
Group	Parameters					
	$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\eta}$	$\hat{\mu}$
1	2.2240	0.0030	0.0000	2.4280	0.8307	0.9986
2	0.3357	0.0009	0.0000	3.0442	0.5874	0.9496

Notes: MLE estimates of parameters of the extended model, conditional on covariates. The extended model allows for non-linearities in the defendant's utility, captured by the parameter η.
See table 5 for a description of the covariate groups.

B.8.4. *Estimation results.* Table 27 reports the estimation results. The defendant's and prosecutor's trial costs are measured in terms of utils and months, respectively. For both covariate groups, I estimate the coefficient η to be relatively high: 0.83 for group one and 0.59 for group two. Nevertheless, most of the other estimates resemble those obtained using the linear utility model. Specifically, all the parameters associated to the defendant's trial costs are very small, the intercept of the prosecutor's trial costs is zero and the auxiliary parameter $\hat{\mu}$ is close to one. The one notable difference between the estimates of the extended and the basic models is that I find β_p, the slope of the prosecutor's trial costs, to be substantially larger in the former. Using the extended model, I estimate β_p to be 2.43 for group one and 3.04 for group two. In the results reported in table 6 of the main text, $\hat{\beta}_p$ is equal to 0.97 and 1.06 for groups one and two, respectively.

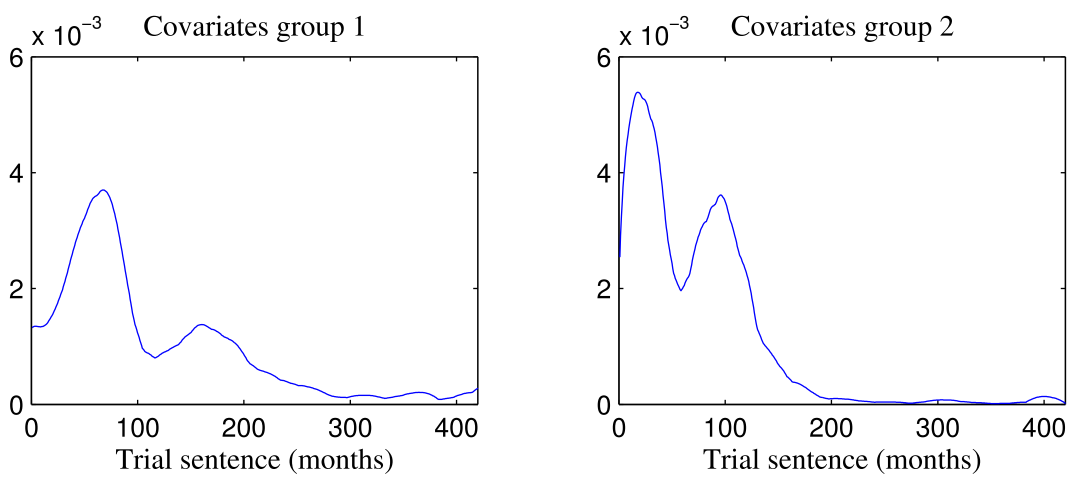
Figure 20 shows the estimated distributions of defendants' types. Remember that this distribution is only identified over the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$. For both groups, this interval comprises a large portion of the unit line. In the case of group one, $\tilde{\theta}(\underline{t})$ is 0.26 and $\tilde{\theta}(\bar{t})$ 0.73, approximately. For group two, $\tilde{\theta}(\underline{t})$ is 0.24 and $\tilde{\theta}(\bar{t})$ is 0.88. For group one, the estimated cumulative distribution function evaluated at 0.73 is 0.34, while, for group two, the estimated cumulative distribution evaluated at 0.88 is 0.30. Similarly to the basic model estimates, these numbers help rationalizing the differences between the estimated settlement offer functions of African-American and non-African-American defendants, which were pointed-out in the discussion of figure 3. Prosecutors offer non-African defendants to settle for relatively short sentences because, compared to their African-American counterparts, these defendants are less likely to be of a high type.

FIGURE 20. Defendants' types distribution estimates—Extended model



Note: Estimated pdf and CDF of the distribution of defendants' types (probabilities of conviction at trial), conditional on covariates. These estimates are based on the extended model, which allows for non-linearities in the defendant's utility function. The distributions are only identified over part of their support.

FIGURE 21. Estimated unconditional distribution of trial sentences—Extended model



Distribution of trial sentences for covariates groups one and two and all judges, unconditional on the case outcome. I obtain this distribution based on estimated extended model, which allows for non-linearities in the defendant's utility function.

The final step in the estimation of the model primitives is to obtain the full distribution of potential trial sentences for each covariate group—i.e., the distribution without conditioning on a trial conviction. Figure 21 shows the estimated distributions for both covariate groups.

Table 28. Fitted values versus data—Extended model

Group		Conviction probability		
		Any ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	38.35%	34.60%	3.75%
	Model	38.33%	34.55%	3.78%
2	Data	44.11%	36.66%	7.45%
	Model	43.47%	38.65%	4.81%

Group		Average sentence, conditional on method of resolution†		
		All ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	63.13	54.93	138.73
	Model	61.54	50.25	164.66
2	Data	48.14	41.51	80.73
	Model	44.65	40.56	77.50

Notes: Fitted values according to the extended model, which allows for non-linearities in the defendant's utility function.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

The fit of the extended model to the data, reported in Table 28, is better than that of the basic model. Like the latter, the extended model fits well the overall probability of conviction and the probabilities of settlement and conviction at trial. The extended model also fits well the overall average sentence and the average sentence, conditional on a plea bargain. The model over-estimates the average trial sentence for group one, but it does so to a lesser extent than the basic model. Using the standard deviation of the observed trial sentences as a reference, the extended model predicts the average trial sentence of group one to be 0.24 standard deviations longer than observed, as opposed to 0.59 standard deviations for the model in the main text. As it happens in the basic model, the over-prediction of the trial sentences has a small impact on the fit of the unconditional expected sentence.

B.8.5. *Counterfactual analysis.* Table 29 shows the effects of a twenty percent reduction in the length of potential trial sentences for all cases in the sample. The probabilities of incarceration by plea bargain and trial are in the top half of the table.

Table 29. Counterfactual results—Sentencing reform—Extended model

Group		Conviction probability		
		Any	Settlement	Trial
1	Current	38.33%	34.55%	3.78%
	-20% trial sentence length	38.44%	34.85%	3.59%
	-10% incarceration cases	34.44%	30.64%	3.57%
2	Current	43.47%	38.65%	4.81%
	-20% trial sentence length	43.55%	38.83%	4.73%
	-10% incarceration cases	39.08%	34.55%	4.53%

Group		Expected sentence†	
		$\Psi \in \{1,2\}$	Unconditional on Ψ
1	Current	61.54	23.59
	-20% trial sentence length	45.51	17.49
	-10% incarceration cases	68.50	23.43
2	Current	44.65	19.41
	-20% trial sentence length	35.31	15.38
	-10% incarceration cases	49.29	19.27

Notes: This table reports the results of two counterfactual exercises using the extended model, which allows for non-linearities in the defendant's utility function. In the first exercise I reduce the length of the trial sentences of every case by 20 percent. In the second one I set the incarceration sentences of all cases below the tenth percentile to zero. The current values are the ones fitted by the estimated model.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
† Measured in months.
See table 5 for a description of the covariate groups.

Similarly to the results using the basic model, I find that shorter potential trial sentences slightly raise the probability of conviction. The increases are of 0.29 percent for group one and 0.18 percent for group two. The impact on the expected length of the assigned sentences is on the bottom half of the table. For groups one and two, the intervention reduces the expected length by 25.86 and 20.76 percent, respectively. Again, the results are similar to those obtained using the basic model.

Table 29 also shows the results of setting the trial sentences below the tenth percentile to zero. This intervention reduces the total probability of conviction by roughly

Table 30. Counterfactual results—No plea bargaining—Extended model

Group		Outcome	
		Probability of conviction	Expected sentence†
1	Current	38.33%	23.59
	No plea bargaining	[28.64% , 36.76%]	[39.62 , 51.11]
2	Current	43.47%	19.41
	No plea bargaining	[38.14% , 43.43%]	[27.66 , 31.47]

Notes: This table reports the results of forcing all cases to go to trial. I use the extended model, which allows for non-linearities in the defendant's utility function. Because the distribution of defendants' types is not identified over its entire support, I can only calculate bounds for the probability of conviction and the expected sentence. The current values are the ones fitted by the estimated extended model.
† Measured in months.
See table 5 for a description of the covariate groups.

ten percent for each covariate group, but its impact on the expected sentence is relatively low: less than one percent. These results are, once more, consistent with the ones reported in the main text.

Table 30 presents the results of eliminating plea bargains. As in the basic model, this intervention reduces the proportion of cases resulting in a conviction. The differences between the current probabilities of conviction and the estimated lower bounds in the counterfactual scenario are 9.69 p.p. and 5.33 p.p. for groups one and two, respectively. Considering the upper bounds, the differences are 1.57 p.p. for group one and 0.04 p.p. for group two. The expected sentences increase by 67.95 to 116.66 percent for group one and 42.50 to 62.13 percent for group two, indicating that the defendants substantially benefit from their private information in the process of plea bargaining. These results are also analogous to those in the main text.

B.9. **Separate identification of the trial costs for the defendant and the prosecutor: an empirical illustration.** This sections illustrates the features in the data that drive the separate identification of the trial costs for the defendant and the prosecutor. As argued in Section 5, the trial costs for both agents jointly determine the probability of settlement while, given a trial sentence t, only the defendant's costs directly impact the equilibrium settlement offers. Specifically, the defendant's costs are equal to $\tilde{s}(t) - \tilde{\theta}(t)t$. Longer settlement offers, relative to the trial sentences, are thus associated with higher trial costs for the defendant. Taking $\tilde{s}(t)$ as given, as in the identification strategy described in Section 5, the only way of increasing the defendant's trial costs is by decreasing $\tilde{\theta}(t)$, which results in a lower probability of conviction, conditional on a trial. Here, to illustrate this argument, I report estimation results obtained from two modified data sets. In the first one, I artificially increase the observed settlement offers. In the second, I artificially increase the defendant's trial win rate. In both exercises, the estimates of the defendant's trial costs are greater than those reported in the main text.

B.9.1. *Longer settlement offers.* I consider data that are similar to those in the main text, except that I artificially increase the observed trial and settlement sentences.[59] Specifically, I increase each observed trial sentence by 36 months, and I separately estimate the model after increasing the observed settlement offers by 12 and 36 months. The results for covariate groups one and two are in Table 31. For both groups, increases in the length of settlement offers lead to the estimation of higher defendant's trial costs. Notice that adding constants to the trial sentences and settlement offers results in a translation of the offer function $\tilde{s}(\cdot)$. As a consequence, only $\hat{\alpha}_d$, the estimated intercept of the trial costs, increases. Notice also that, as the defendant's trial costs increase, the trial costs for the prosecutor decrease—which is expected, since the settlement probability remains constant.

B.9.2. *Lower probability of conviction, conditional on a trial.* Here, instead of changing the observed sentences, I artificially increase by 1000 the number of cases that result in an acquittal at trial. For covariate group one, such an increase raises the defendant's trial win rate from 57.18 to 93.26 percent. For group two, the raise is from

[59]The purpose of also increasing the trial sentences is to ensure that $\tilde{s}(t) < t$ for all $t \in [\underline{t}, \bar{t}]$, so that condition (ii) in footnote 31 of the main text holds. Conditions (i) to (iv) in that footnote are sufficient for $\tilde{\theta}(t)$ to be strictly increasing and bounded between zero and one, in accordance with the theoretical model. In the raw data, many observed trial sentences are far shorter than 12 months—the shortest increase in settlement offers considered here. In fact, the first quartile of the distribution of observed trial sentences is just five months.

Table 31. Parameter estimates with modified data—longer settlement sentences

Group	Δ settlement sentences	Parameters				
		$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\mu}$
1	+12 months	8.31	0.00	0.00	0.52	1.00
	+36 months	32.59	0.00	0.00	0.40	1.00
2	+12 months	1.56	0.00	0.00	0.44	1.00
	+36 months	25.90	0.00	0.00	0.08	1.00

Notes: MLE estimates of the model parameters, conditional on covariates. The estimates in this table are based on artificial data, in which the trial and settlement sentences are longer than in reality. Specifically, I add 36 months to each observed trial sentence. The table separately reports the estimates obtained after increasing the observed settlement sentences by 12 and 36 months.

Table 32. Parameter estimates with modified data—lower probabilities of conviction, conditional on a trial

Group	Parameters				
	$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\mu}$
1	0.03	0.14	37.79	0	1
2	0.02	0.25	4.89	0	1

Notes: MLE estimates of the model parameters, conditional on covariates. The estimates in this table are based on artificial data, in which I increase the number of cases resulting on a trial acquittal by 1000 for each group. To keep the settlement rate constant, I also increase the number of cases resulting in a successful plea bargain.

44.03 to 85.11 percent. I also increase the number of cases resulting in a successful plea bargain, so that the settlement rate remains constant. Table 32 contains the results for both groups. Consistently with the discussion above, the higher defendant's trial win rates lead to greater estimates of the defendant's trial costs.