Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges

Supplementary Material

Claire S.H. Lim*

This paper contains additional details of the data and the estimation of the model in our paper "Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges". We describe the details in the following order: (1) the main features of sentencing data and the aggregation procedure, (2) details of the judicial selection systems in Kansas, (3) data on political climate, (4) data on exit decisions, (5) an alternative specification for the reelection probability of appointed judges, and (6) the procedure of counterfactual experiments.

1 Sentencing Data and the Aggregation Procedure

In this section, we document the composition of the raw sentencing data, its major features, the aggregation procedure we used to generate the aggregate sentencing variable used in the paper, and robustness checks of the sentencing patterns with respect to various aggregation schemes.

1.1 Composition and Major Features of the Raw Data

The raw sentencing data contains rich information on each criminal case. The set of major variables that we use in our analysis are listed in Table 1. Under the Kansas Criminal Sentencing Guidelines

Table 1: Major Variables in the Sentencing Data

Variable Type	Variables
Basic Information	County, Sentencing Date, Sentencing Judge, Date of Conviction, Type of Counsel
Major Case Characteristics	Defendants' Criminal History, Name of Primary Offense of Conviction, Severity Level
Sentencing Outcome	Guideline Range Imposed, Type of Departure, prison sentencing/month

*Lim: Department of Economics, Cornell University, 404 Uris Hall, Ithaca, NY 14853 (e-mail: claire-lim@cornell.edu).

(Figure 1 on page 3), judges' discretion in a given case is determined by two case characteristics: *defendants' criminal history* and *severity level of primary offense*. Each felony case is classified into one of the 90 categories in the sentencing guidelines in Figure 1, based on the criminal history of defendants (9 categories: category A $\sim$ I) and the severity of primary offense (10 levels: level 1 $\sim$ 10). Table 2 shows examples of offenses that constitute each severity level.[1] As shown in Table

Table 2: Examples of Offenses in Each Severity Level

Severity Level	Offense
Level 1	Murder in the first degree - attempt Murder in the second degree - intentional Rape; sexual intercourse; no consent; overcome with force or fear Aggravated kidnapping
Level 2	Murder in the second degree - reckless Rape; knowingly misrepresenting sexual intercourse legally Aggravated criminal sodomy
Level 3	Voluntary manslaughter Aggravated robbery
Level 4	Aggravated battery; intentional, great bodily harm Involuntary manslaughter while under the influence of alcohol or drugs
Level 5	Involuntary manslaughter Battery Sexual exploitation of a child Theft; $100,000 or more
Level 6	Arson Aggravated assault on a law enforcement officer
Level 7	Aggravated assault Perjury
Level 8	Aggravated battery; reckless; bodily harm with deadly weapon
Level 9	Aggravated endangering a child Theft; at least $1,000 but less than $25,000 Burglary; motor vehicle, aircraft, or other means of conveyance
Level 10	Bigamy Incest Nonsupport of a child

2, serious offenses such as rape and murder, which are relatively more often publicized by media, belong to high severity levels (level 1$\sim$5).

For each of the 90 categories in Figure 1, the guideline specifies three numbers - minimum, standard, and maximum jail time. The judge can choose any jail time between the minimum and the maximum. Table 3 shows the overall distribution of cases in the raw sentencing data across severity level and the category of defendants' criminal history. The sentencing guideline and the case distribution show two features that are noteworthy.

[1] A complete manual for severity level classification of criminal offenses is available at http://www.accesskansas.org/ksc/2007desk.shtml.

Figure 1: Kansas Criminal Sentencing Guidelines

SENTENCING RANGE - NONDRUG OFFENSES

Category →	A	B	C	D	E	F	G	H	I
Severity Level ↓	3 + Person Felonies	2 Person Felonies	1 Person & 1 Nonperson Felonies	1 Person Felony	3 + Nonperson Felonies	2 Nonperson Felonies	1 Nonperson Felony	2 + Misdemeanor	1 Misdemeanor No Record
I	653 620 592	618 586 554	285 272 258	267 253 240	246 234 221	226 214 203	203 195 184	186 176 166	165 155 147
II	493 467 442	460 438 416	216 205 194	200 190 181	184 174 165	168 160 152	154 146 138	138 131 123	123 117 109
III	247 233 221	228 216 206	107 102 96	100 94 89	92 88 82	83 79 74	77 72 68	71 66 61	61 59 55
IV	172 162 154	162 154 144	75 71 68	69 66 62	64 60 57	59 56 52	52 50 47	48 45 42	43 41 38
V	136 130 122	128 120 114	60 57 53	55 52 50	51 49 46	47 44 41	43 41 38	38 36 34	34 32 31
VI	46 43 40	41 39 37	38 36 34	36 34 32	32 30 28	29 27 25	26 24 22	21 20 19	19 18 17
VII	34 32 30	31 29 27	29 27 25	26 24 22	23 21 19	19 18 17	17 16 15	14 13 12	13 12 11
VIII	23 21 19	20 19 18	19 18 17	17 16 15	15 14 13	13 12 11	11 10 9	11 10 9	9 8 7
IX	17 16 15	15 14 13	13 12 11	13 12 11	11 10 9	10 9 8	9 8 7	8 7 6	7 6 5
X	13 12 11	12 11 10	11 10 9	10 9 8	9 8 7	8 7 6	7 6 5	7 6 5	7 6 5

Note: The first, the second, and the third numbers in each category are minimum, standard, and maximum prison time specified by the law. The bright area in the upper-left part of the table is the category of crimes for which presumptive sentencing is imprisonment. The dark area in the lower-right part of the table is the category of crimes for which presumptive sentencing is probation.

Table 3: Distribution of Cases across Severity Levels and Defendants' Criminal History

Severity Level		Category of Defendants' Criminal History A	B	C	D	E	F	G	H	I	Total (Row)
I	Frequency	49	46	76	57	38	24	64	54	241	649
	Proportion (%)	0.09	0.09	0.14	0.11	0.07	0.04	0.12	0.10	0.45	1.20
II	Frequency	16	24	30	38	19	15	42	41	184	409
	Proportion (%)	0.03	0.04	0.06	0.07	0.04	0.03	0.08	0.08	0.34	0.76
III	Frequency	166	186	240	195	156	92	211	186	877	2,309
	Proportion (%)	0.31	0.34	0.44	0.36	0.29	0.17	0.39	0.34	1.62	4.28
IV	Frequency	44	51	68	70	32	38	68	64	245	780
	Proportion (%)	0.08	0.09	0.13	0.13	0.06	0.07	0.13	0.12	0.45	1.26
V	Frequency	272	289	374	264	223	153	337	445	1,590	3,947
	Proportion (%)	0.50	0.54	0.69	0.49	0.41	0.28	0.62	0.82	2.95	7.31
VI	Frequency	76	87	93	72	71	51	119	115	448	1,132
	Proportion (%)	0.14	0.16	0.17	0.13	0.13	0.09	0.22	0.21	0.83	2.10
VII	Frequency	677	787	1,332	798	1,130	686	1,246	1,183	3,141	10,980
	Proportion (%)	1.25	1.46	2.47	1.48	2.09	1.27	2.31	2.19	5.82	20.34
VIII	Frequency	387	506	1,040	363	1,677	654	1,071	873	2,000	8,571
	Proportion (%)	0.72	0.94	1.93	0.67	3.11	1.21	1.98	1.62	3.71	15.88
IX	Frequency	926	1,289	2,475	967	2,912	1,524	2,548	2,377	4,711	19,729
	Proportion (%)	1.72	2.39	4.59	1.79	5.39	2.82	4.72	4.40	8.73	36.55
X	Frequency	232	368	626	332	772	402	803	552	1,487	5,574
	Proportion (%)	0.43	0.68	1.16	0.62	1.43	0.74	1.49	1.02	2.75	10.33
Total (Column)	Frequency	2,845	3,633	6,354	3,156	7,030	3,639	6,509	5,890	14,924	53,980
	Proportion (%)	5.27	6.73	11.77	5.85	13.02	6.74	12.06	10.91	27.65	100

Note: The 90 cells made by the severity level and the criminal history in this table correspond to the 90 cells in the sentencing guidelines (Figure 1). The upper number in each cell shows the frequency of cases, and the lower number shows the proportion of cases in the cell among all the cases.

First, in the sentencing guidelines, there is a substantial degree of variation in the standard prison time (i.e., the prison time recommended by the law) across both categories of defendants' criminal history and the severity level. This feature implies that we should take the minimum and the maximum jail time specified in the guidelines into consideration in measuring a judges' sentencing harshness. That is, the measure of sentencing harshness should be *normalized* relative to the guidelines. If we use absolute (non-normalized) jail time to measure sentencing harshness, even a small degree of variation in the severity level of offenses in the pool of cases handled by each judge will result in inadequate variation in the measure of harshness. Hence, in the aggregation procedure described below, we use sentencing outcomes normalized relative to the guidelines.

Second, in Table 3, high-severity levels (level I-V) constitute approximately 15% of all cases. Additionally, the first four categories of defendants' criminal history (category A-D) constitute approximately 30% of all cases. Since severe crimes by criminals with lengthy histories constitute a relatively small proportion of cases, if we give equal weight to each case, the measure of sentencing harshness is likely to be driven by sentencing patterns for low-severity offenses. In reality, however, the type of offenses for which a sentencing decision becomes an important issue are of high severity. Therefore, to reflect the importance of each sentencing decision correctly, it is necessary to give large weight to high-severity offenses in measuring sentencing harshness. Specifically, we use *standard prison time* specified in the sentencing guidelines as the weight of each case.

Before describing the aggregation procedure in detail, we document additional major features of the raw sentencing data that lead to our design of the aggregation procedure:

(1) Discreteness of the jail time variable: While judges' discretion in sentencing has a continuous nature according to the law (given that they can choose any jail time between minimum and the maximum), the data on sentencing is almost discrete in that verdicts are concentrated on one of the three points - minimum, standard, and maximum jail time prescribed by the guidelines. Figure 2 shows the distribution of sentenced jail time for cases with severe crimes (severity level 1-5 out of 10 levels) when we normalize sentenced jail time at [0,1] interval. As the figure shows, there are strong concentrations at three different points - 0 (minimum), 0.5 (standard), and 1 (maximum).

The strong concentrations at these three points makes it difficult to use concepts such as quintile to measure sentencing harshness even though it may be a sensible choice in the abstract. More specifically, for high severity (severity level 1-5) cases, 0 (minimum sentencing) constitutes 45 percent, 0.5 (standard sentencing) constitutes 15 percent, and 1 (maximum sentencing) constitutes 18 percent of the cases. The rest of the data is sparsely spread. Because of this almost-discreteness of sentencing decisions, *it is more appropriate to regard sentencing as a discrete decision*.

(2) 'Guideline' variable in the raw data: There is also a (discrete) variable in the raw data named '*guideline*' (coded by the sentencing commission that collected the raw data) that classifies each

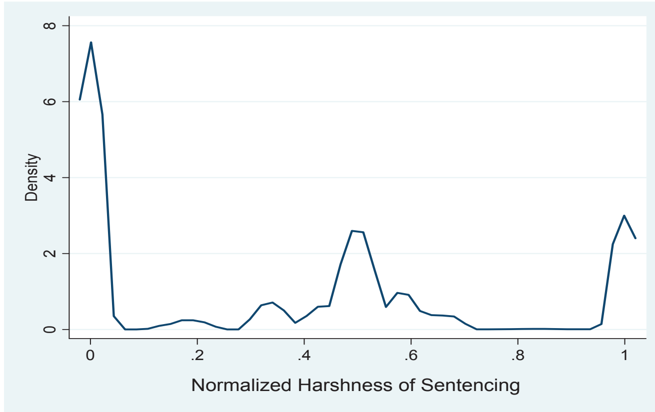


Figure 2: Distribution (Kernel Density Estimate) of the Normalized Jail Time

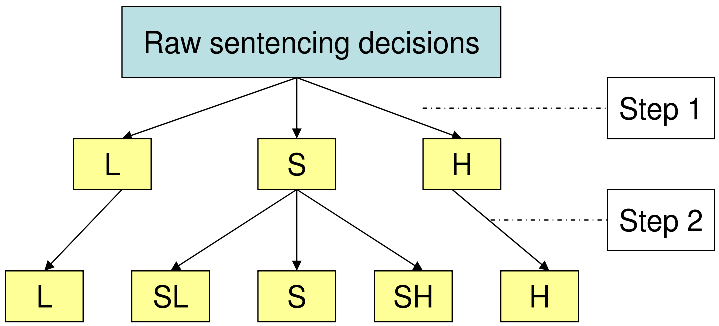
sentencing decision into one of the following categories: "standard", "mitigated", "aggravated", and "departure". Three categories, "standard", "mitigated", and "aggravated", of the 'guideline' variable roughly correspond to the standard, minimum, and maximum jail time prescribed by the sentencing guidelines. Additionally, "departure" category captures sentencing decisions that deviate from the range prescribed by the sentencing guidelines. The overall proportion of departure decision was small (around 5% of the whole cases). To avoid subjectivity in classifying sentencing decisions into categories, we use the 'guideline' variable provided by the sentencing commission in the aggregation procedure described below. We use decisions in "standard", "mitigated", and "aggravated" category of the 'guideline' variable as they are. For cases with "departure" decisions, there is a separate variable in the data that shows whether they were upward departure (sentencing above the maximum) or downward departure (sentencing below the minimum). Cases that resulted in upward (downward) departure are merged into cases with "aggravated" ("mitigated") decisions. Through this step, all sentencing decisions are classified into one of the three categories: mitigated, standard, and aggravated.

(3) Discrete-to-discrete aggregation with weights: Given the discrete nature of the sentencing variable, the appropriate aggregation scheme should be one that maps discrete sentencing decisions in about 87 cases to one discrete choice for each judge-period, giving each case a different weight based on its importance (severity). We use *weighed mode* as the aggregated measure, where the weight is the standard prison time for each case specified by the sentencing guidelines. We describe the aggregation procedure in greater detail in Section 1.2.

1.2 Aggregation Procedure of Sentencing Data

The aggregation is done in two steps. In the first step, discrete sentencing decisions in the three categories - mitigated, standard, and aggravated, described above - in on average 87 decisions for each judge-period are aggregated into one the three decisions - Lenient (L), Standard (S), and Harsh (H). In the second step, we divide the Standard (S) category in the first step into three sub-categories: Standard-harsh (SH), Standard (S), and Standard-lenient (SL). Hence, the two-step procedure results in five categories.

Figure 3: Aggregation of Sentencing Decisions



(1) First Step:

We weight the *frequency* of each of mitigated, standard, and aggravated decision with the standard prison time in the guidelines. Let us consider the following example (Table 4). Suppose that a judge makes decisions in six cases A, B, C, D, E, and F in a period as follows: A-mitigated, B-standard, C-aggravated, D-mitigated, E-standard, and F-mitigated. Further, suppose that the primary offense and the defendant's criminal history in each case yields the standard prison time of 9, 66, 160, 43, 130, or 12 months, respectively (based on the sentencing guidelines in Figure 1). In aggregate, "mitigated", "standard", and "aggravated" decisions receive a total score of 64, 196,

Table 4: Example – Aggregation of Sentencing Decisions (the first step)

Case	Severity Level	Category of Criminal History	Weight (Standard Prison Time)	Sentencing		
				mitigated	standard	aggravated
A	IX	F	9	√		
B	IV	D	66		√	
C	II	F	160			√
D	VI	A	43	√		
E	V	A	130		√	
F	VII	I	12	√		
Total Score				64	196	160
				Decision : S (Standard)		

Note: The table of sentencing guidelines on page 3 yields the standard prison time used as the weight for each case.

and 160 months, respectively. If the "mitigated" decision gets the highest total score, we classify the aggregated decision of the judge-period as Lenient (L). If the "aggravated" decision gets the highest score, we classify the aggregated decision as Harsh (H). If the "standard" decision gets the highest score, we classify the aggregated decision as Standard (S). In the example, the "standard" decision has the highest score. Therefore, the sentencing outcome in the period is classified as Standard in this first step.

Following this first step of the aggregation scheme leads to the distribution that is highly concentrated on Standard (S) decision. In the first-stage of classification, the Standard category constitutes more than 70% of the aggregated decisions.

(2) Second Step: The purpose of the second step is to further divide the Standard category into three sub-categories in order to more finely capture the variation in judges' sentencing decisions. If the aggregation in the first step results in classification into H or L, no further classification occurs. If the first step resulted in S, we conduct further classification giving weights only to the high-severity (severity I-V) cases.[2] Table 5 illustrates the second step with the example considered above. In the example, cases B, C, and E belong to the high severity level. Hence, these three cases are counted in the second step of the aggregation. In this particular case, S is still the category that receives the highest score in the second step. Hence, the final result of aggregation is S. If L or H receives the highest score in the second step, the final classification result would be SL or SH, respectively.

Table 5: Example – Aggregation of Sentencing Decisions (the second step)

Case	Severity Level	Category of Criminal History	High Severity	Weight	Sentencing		
					mitigated	standard	aggravated
A	IX	F	No	9	√		
B	**IV**	D	**Yes**	**66**		√	
C	**II**	F	**Yes**	**160**			√
D	VI	A	No	43	√		
E	**V**	A	**Yes**	**130**		√	
F	VII	I	No	12	√		
Total Score					0	196	160
					Decision : S (Standard)		

1.3 Robustness of the Major Sentencing Patterns

In this section, we document the robustness of the major sentencing patterns with respect to alternative aggregation procedures. The two major sentencing patterns with which we check the robust-

[2] There is a natural reason to give special weight to high-severity level cases: High severity cases have significantly more variation in sentencing outcomes than low severity cases do. Specifically, less than half of sentencing decisions for high severity cases are "standard" decisions (in the 'guideline' variable), while 69% of sentencing decisions for low severity cases are "standard" decisions. Hence, high severity cases are not only socially more important, but they are also the cases that convey more information about variation across judges.

ness are as follows: 1) there is a substantial difference between sentencing patterns in conservative districts and liberal districts when judges are elected, while there is little difference between conservative and liberal districts when judges are appointed; 2) Republican judges are not harsher than Democrats when judges are elected.

In checking the robustness of these two patterns, we try three alternative measures. We describe the procedures by which the alternative measures are constructed, and we document the major patterns.

1.3.1 Alternative Measure A: aggregation from 5 decisions to 5 decisions

For the first alternative measure we try ("alternative measure A"), the outcome of the aggregation is five categories, as in the case of the baseline measure we used in our main analysis. The main difference between alternative measure A and the baseline measure is in the processing of case-level decisions. For alternative measure A, we classify each case-level decision into five categories, while we used three categories (mitigated, standard, and aggravated) for case-level decisions in constructing the baseline measure.

The aggregation is completed in two steps. In the first step, we normalize sentencing harshness on a [0,1] scale, relative to the minimum and the maximum jail time in the sentencing guidelines. Then, we classify the sentencing outcome in each case to five intervals: [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), and [0.8, 1.0]. Decisions in each of these five intervals are labeled as *L*, *SL*, *S*, *SH*, and *H*. For each judge-period, we choose the weighted mode of all cases, using the standard prison time for each case as the weight. This step aggregates on average 87 decisions in each judge-period to one decision in one of the five categories. This first step is similar to the first step of the aggregation procedure for the baseline measure, introduced on page 6, except that we use five categories instead of three categories in the first step.

If the first step resulted in *L*, *SL*, *SH*, or *H*, then no further classification occurs. If the first step resulted in *S*, then we divide the category *S* into three subcategories, *SH*, *S*, and *SL*, in the second step. We give weights only to categories of crimes for which presumptive sentencing is imprisonment. (These categories constitute the bright area in the upper-left part of the sentencing guideline on page 6.) Then, if the second-step classification of category *S* results in *SH* or *H*, then the final outcome of aggregation becomes *SH*. If the second-step classification of category *S* results in *SL* or *L*, then the final outcome of aggregation becomes *SL*.

Figure 4 shows the difference between conservative and liberal districts for appointed and elected judges. As in the case of the baseline measure, the difference between conservative and liberal districts is substantially larger when judges are elected, compared to the case in which judges are appointed. Figure 5 compares sentencing decisions by Democrats and Republicans under the two systems. The pattern that the figure shows is similar to Figure 3 in the main text of the paper in that elected Republicans do not exhibit harsher sentencing compared with elected Democrats.

Figure 4: Sentencing Patterns based on Alternative Measure A - across selection systems and political orientations

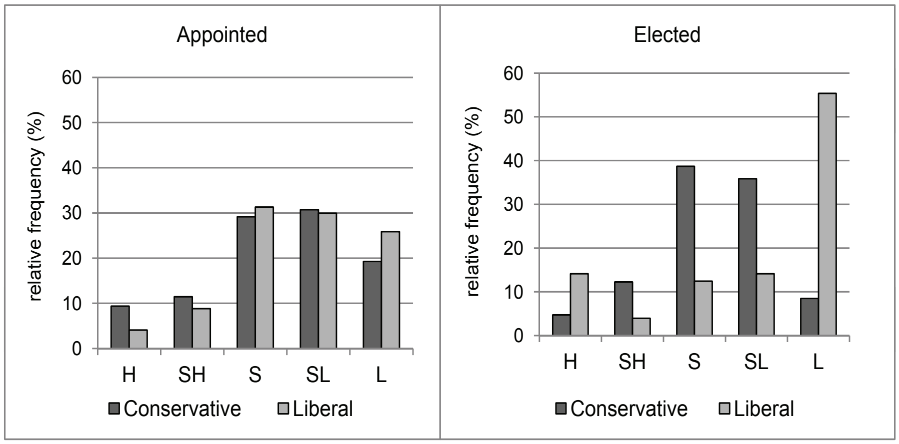
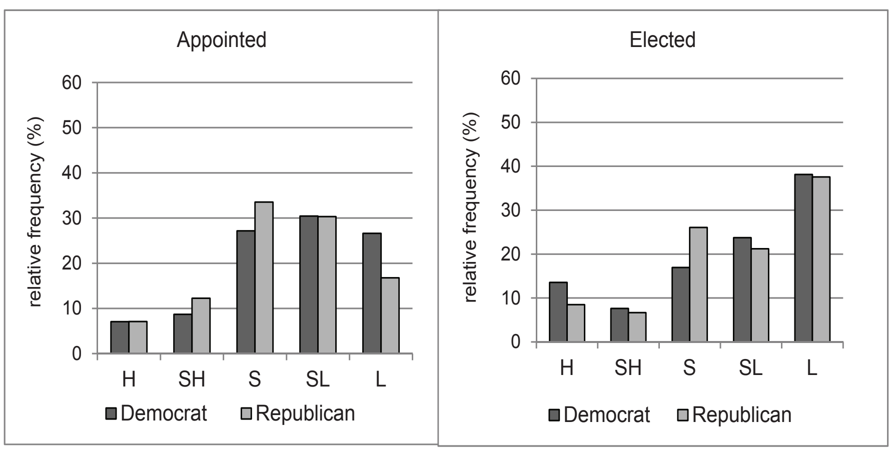


Figure 5: Sentencing Patterns based on Alternative Measure A - across selection systems and parties



1.3.2 Alternative Measure B: Aggregation from 5 decisions to 5 decisions

The second alternative measure ("alternative measure B") that we consider is similar to alternative measure A considered above. This measure is also constructed in two steps. The first step in constructing this measure is identical to the first step in constructing alternative measure A. Additionally, if the first step results in *L*, *SL*, *SH*, or *H*, no further classification occurs. If the first step results in *S*, we divide the category *S* into three subcategories, *SH*, *S*, and *SL*, by giving weights only to the categories of cases for which presumptive sentencing is imprisonment. If the second step yields *L* for decisions in a judge-period classified as *S* in the first step, the final outcome of the aggregation becomes *SL*. If the second step yields *H* for decisions in a judge-period classified as *S* in the first step, the final outcome of the aggregation becomes *SH*. If the second step results in *SL*, *S*, or *SH*, for decisions in a judge-period classified as *S* in the first step, then the final outcome of the aggregation becomes *S*. In brief, construction of alternative measure B differs from that of alternative measure A in that the subcategories *SL* and *SH* in the second step results in *S* for the final outcome for alternative measure B, which is not the case of alternative measure A.

Figure 6 shows the sentencing patterns in conservative and liberal districts for appointed and elected judges, based on alternative measure B. Figure 7 shows the sentencing patterns by Democrats and Republicans for the two selection systems. The sentencing patterns shown in the two figures are almost identical to the patterns shown in Figure 4 and Figure 5 based on alternative measure A.

Figure 6: Sentencing Patterns based on Alternative Measure B - across political orientations and selection systems

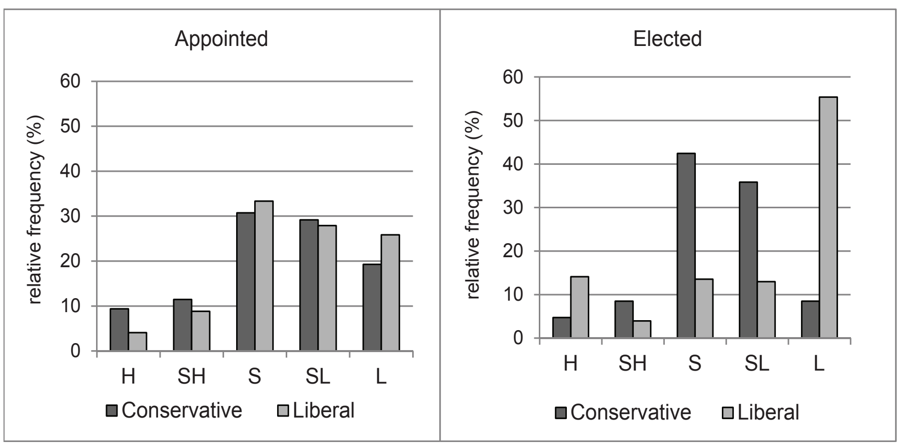
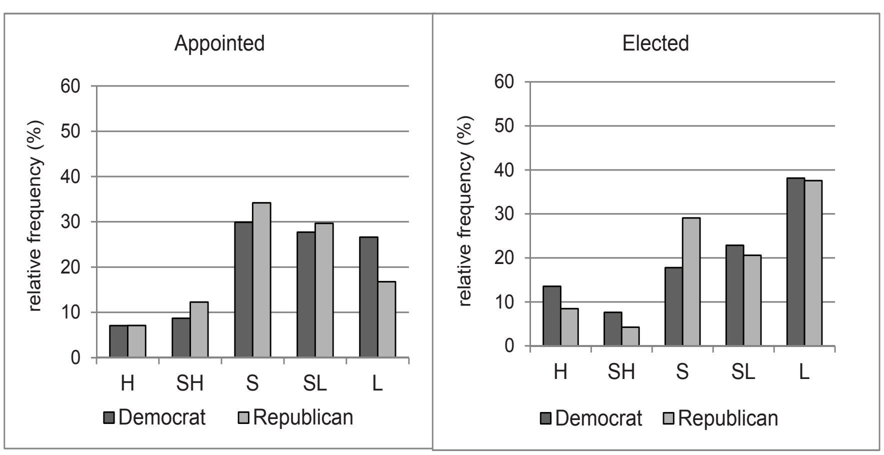


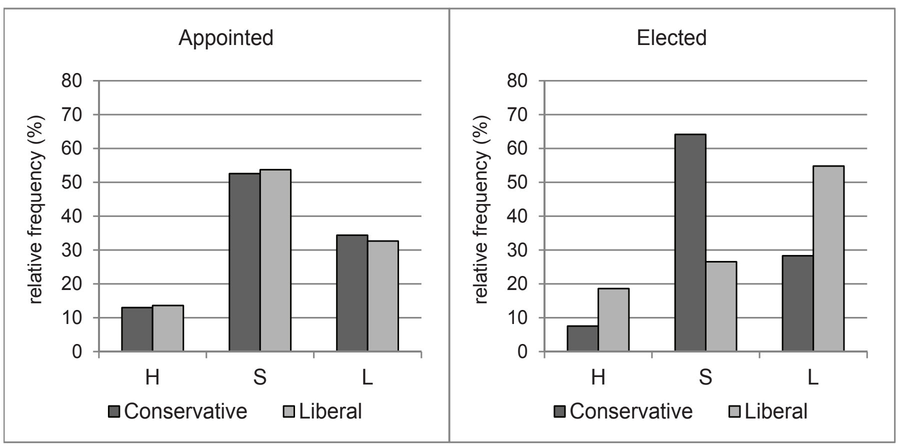
Figure 7: Sentencing Patterns based on Alternative Measure B - across parties and selection systems



1.3.3 Alternative Measure C: Aggregation from 3 decisions to 3 decisions

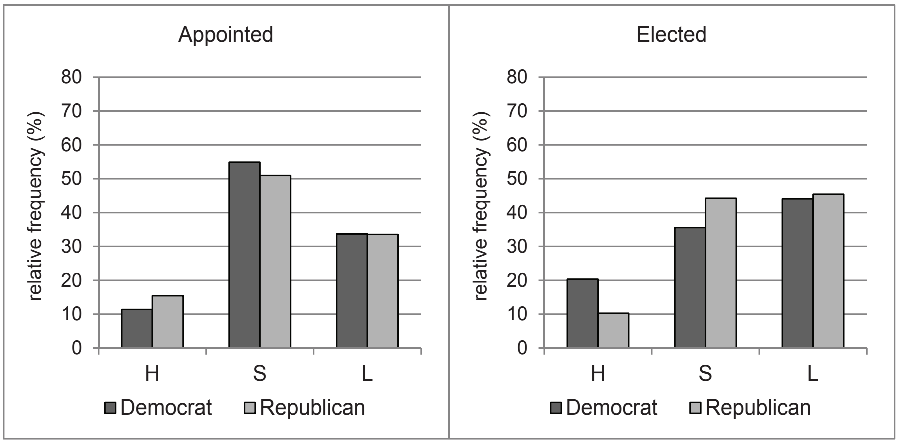
For the third alternative measure ("alternative measure C") that we consider, we use only the categories of cases for which presumptive sentencing is imprisonment. The aggregation procedure consists of only one step, and the final outcome of the aggregation belongs to one of *three* categories: H, S, or L. In contrast to alternative measures A and B, the sentencing decision in each criminal case is first classified into one of the three categories: mitigated, standard, or aggravated. (This is similar to the aggregation procedure that gave the baseline measure). Then, we aggregate sentencing decisions in each judge-period into one of the three categories - H, S, or L, using standard prison time as the weight. (This part of the procedure is almost identical to the first step of the baseline aggregation procedure described in Section 1.2.) The difference from the baseline measure is that we use only the cases for which presumptive sentencing is imprisonment. Figure 8 and 9 again show the robustness of the major sentencing patterns.

Figure 8: Sentencing Patterns based on Alternative Measure C - across political orientations and selection systems



The three alternative measures that we documented in this section show that the major sentencing patterns that were introduced in the main text of the paper are invariant to the aggregation procedures. In the next section, we document the history and socio-economic characteristics of judicial selection systems in Kansas that are related to Section 3 of the main text of the paper.

Figure 9: Sentencing Patterns based on Alternative Measure C - across parties and selection systems



2 Details of the Judicial Selection Systems in Kansas

2.1 History

In this section, we describe the history of the two selection systems in Kansas.[3] Until the middle of the 20th century, Kansas elected all judges and justices for its state court. In the year 1958, they amended the constitution to appoint justices for the state supreme court. In 1972, they amended the constitution to allow for an appointment system for district court judges. Then, in the 1974 general election, there was a question on the ballot asking voters in each district whether to use appointment or election for their district court judges. This was the origin of the co-existence of the two systems in the state. Selection systems are prescribed by Article 3 of the Kansas Constitution.[4]

2.2 Relationship between the judicial selection systems and socio-economic characteristics

In Section II.A of the main paper, we described the overall similarity of districts that belong to the two systems, in terms of major social and political characteristics. In this section, we further investigate socio-economic characteristics of the judicial districts under the two systems. We focus on the following variables: income, crime rate, industrial characteristics, and the level of education. We investigate the relationships at the county-level.[5] We conduct the analysis at two time

[3] A similar description of the history of the two systems in Kansas can be found in the American Judicature Society's web site on judicial selection systems (http://www.judicialselection.us/judicial_selection).

[4] See the following web page for details: http://www.kslib.info/constitution/art3.html

[5] Even though the operating unit of the system is judicial district, using county-level data helps us to have a large number of observations, which makes it easier to detect any systematic differences between the two systems.

points, the 1970s and the 1990s, for the following reasons: we chose the 1970s because it was the period when the appointment system was adopted; we chose the 1990s to see whether there are correlations between socio-economic characteristics and the systems that did not exist in the 1970s but evolved later. The data source is the City and County Data Book in 1977 and 2000, by the U.S. Census Bureau. In the logit regressions we show below (Table 7 and Table 9), the dependent variable is the system, a dummy variable that takes value 1 when a county belongs to the system of appointment and yes-or-no vote, and takes value 0 when a county belongs to the system of competitive election.

For the 1970s, we focus on the following four variables: per capita income, crime rate per 1,000 population, percentage of population in farming, and percentage of employment in manufacturing. Table 6 shows the descriptive statistics, and Table 7 shows the result of the logit regression. (We did not include education-related variables, because they are not available for Kansas in the 1970s. As for income, we include per capita income rather than median income, because median income was available only for family income, not for individual income.) None of the variables have a statistically significant effect on the probability that a county adopts the system of appointment.

Table 6: Descriptive Statistics: County-level Socio-economic Characteristics in 1970s

variable	year	mean	std. dev.	min	max
per capita income	1974	4717.4	945.29	3415	7420
crime rate (per 1,000 population)	1977	20.62	18.07	.25	97.66
employment in manufacturing (%)	1970	10.00	7.74	.8	32.1
farming population (%)	1970	23.26	11.91	.39	51.54

Table 7: Logit Regression of Systems on Socio-economic Characteristics in 1970s

variable	coefficient	std. err.	z	$P >\| z \|$
constant	.0034	1.8122	0.00	0.998
per capita income	-.0003	.0003	-1.47	0.141
crime rate	.0197	.0180	1.09	0.276
employment in manufacturing	.0705	.0382	1.84	0.065
percentage of farming population	.0306	.0276	1.11	0.267

For the 1990s, we focus on the following variables: median income, crime rate per 1,000 population, percentage of population in farming, percentage of population with high school education or higher, and percentage of population with bachelor's degree or higher. For the 1990s, we do not include percentage of employment in manufacturing, because the variable is not available for the majority of counties in Kansas for this period. Descriptive statistics are in Table 8. Table 9 shows the results of the logit regression. No variables have a coefficient estimate that is statistically significant at the 5% level. Only the coefficient of the crime rate is statistically significantly related to the system at the 10% level. Moreover, even for the crime rates, the magnitude of the coefficient is fairly small. In Table 10, we also document the result of a t-test (comparison of mean crime rates between the two systems). The magnitude of overall difference between the two systems in mean crime rates is much smaller than that of variance within the systems.

Table 8: Descriptive Statistics: County-level Socio-economic Characteristics in 1990s

variable	year	mean	std. dev.	min	max
median income	1997	33389.20	5158.91	23604	59870
crime rate (per 1,000 population)	1997	26.86	20.44	.75	105.68
farming population (%)	1990	10.70	6.15	.1	27.1
education: high school or higher (%)	1990	77.67	4.52	67.3	92.9
education: college or higher (%)	1990	14.58	5.09	8.1	40.5

Table 9: Logit Regression of Systems on Socio-economic Characteristics in 1990s

variable	coefficient	std. err.	z	$P >\| z \|$
constant	-9.005166	4.776122	-1.89	0.059
median income	.0000647	.0000504	1.28	0.199
crime rate	.0269492	.0145273	1.86	0.064
farming population	.0727795	.0510192	1.43	0.154
high school or higher	.0746148	.0695045	1.07	0.283
college or higher	-.0317852	.0710383	-0.45	0.655

Table 10: Two Sample T-test with Unequal Variances for Crime Rate

Group	Obs	Mean	Std. Error	Std. Dev	95 % Confidence Interval
Election	53	23.98	2.60	18.94	[18.76, 29.20]
Appointment	52	29.80	3.00	21.66	[23.77, 35.83]
Combined	105	26.86	1.99	20.44	[22.91, 30.82]
Difference		-5.81	3.97		[-13.69, 2.07]

Difference = mean (election) - mean (appointment)

H_0 : difference $= 0$, t-value = -1.46, $\Pr\{|T| > |t|\}$ =0.1466

The result of the logit regressions shown above alleviates the concern for the possibility that differences in sentencing decisions between the systems may have been caused by the unobserved heterogeneities of the judicial districts.

In the next section, we describe how the political climate, which captures the stochastic aspect of voters' party preferences, is coded.

3 Political Climate

As stated in the main paper, the political climate is one of the three states - 'favorable to Republican', 'neutral', or 'favorable to Democrat'. The measure is based on each judicial district's normalized vote share of Democrats in presidential and gubernatorial elections. We separately construct the state-of-the-district variables from presidential vote shares and gubernatorial vote shares. This is because the meaning of the state-level Republican and Democratic parties can differ from the meaning of the national ones. However, we keep the frequencies of the three states ('favorable to Republican', 'neutral', and 'favorable to Democrat') consistent across the presidential elections and gubernatorial elections. In our data, judges face the three states 'favorable to Republican', 'neutral', and 'favorable to Democrat' for 30.1%, 47.2%, and 22.7% of the time, respectively.

The relationship between the classification of the political climate and the district-level Democratic vote share in presidential election years is described in Table 11. The 248 observations in Table 11 are from 8 presidential elections and 31 judicial districts in Kansas from 1976 to 2004. The table shows asymmetry of classification, yielding relatively small frequencies of the state

Table 11: Classification of Political Climate – presidential election years

Political Climate	Frequency	Normalized Democratic Vote Share (%)			
		mean	std. dev.	minimum	maximum
favorable to Republican	85	30.0	3.9	18.4	33.3
neutral	117	39.7	3.6	33.5	45.6
favorable to Democrat	46	52.9	6.9	46.1	72.8

'favorable to Democrat'. Since the distribution of district-level Democratic vote share is right-skewed, equally dividing the three states based on frequencies would yield a disproportionately long interval of vote share being classified as the state 'favorable to Democrat'. The political climate variable not only means the relative preference of voters, but it also has a meaning in terms of the absolute level of vote share. Additionally, the classification in Table 11 is balanced given the overall shape of the vote share distribution. The classification of political climate in gubernatorial election years is summarized in Table 12. The 248 observations in the table are based on 8 gubernatorial elections and 31 judicial districts in Kansas from 1978 to 2006. The rationale behind the

Table 12: Classification of Political Climate – gubernatorial election years

Political Climate	Frequency	Normalized Democratic Vote Share (%)			
		mean	std. dev.	minimum	maximum
favorable to Republican	108	33.6	9.3	16.2	46.5
neutral	102	52.1	3.1	46.5	57.0
favorable to Democrat	38	63.7	6.7	57.1	80.6

classification using gubernatorial election years is similar to the one for presidential election years. We summarize the relative frequency of the political climates that judges face in conservative and

Table 13: Relative Frequency of Political Climate that Judges face (%)

Political Climate	Appointed		Elected		Overall
	Conservative	Liberal	Conservative	Liberal	
favorable to Republican	41.70	17.51	60.20	16.83	30.05
neutral	50.87	33.95	38.80	57.23	47.24
favorable to Democrat	7.43	48.54	1.00	25.94	22.71

liberal districts under the two systems in Table 13. In the next section, we describe the details of the exit decisions in the data.

4 Exit Decisions

As described in the main paper, a judge makes an exit decision at the end of each period. In our data, we have 1541 observations of exit decisions and other modes of exit. We show the overall distribution of exit decisions in two different situations in Table 14 and Table 15: (a) when the seat is not up for reelection (i.e., when a judge is in the first period of a term), and (b) when the seat is up for reelection (when a judge is in the second period of a term). The two other modes of termination - death and promotion - in the table are not counted as voluntary exit in our estimation.

5 An Alternative Specification for Appointed Judges

In the main text of the paper, we assumed that appointed judges are reelected with probability 1. In this section, we introduce an alternative specification of the reelection probability of appointed judges. Since we do not have any observation of defeat, the probit model (which we used for

Table 14: Exit Decisions and Other Modes of Termination - when the seat is not up for reelection

	Appointed		Elected	
	Frequency	Proportion(%)	Frequency	Proportion(%)
Voluntary Exit	18	4.49	9	2.42
Staying	377	94.01	358	96.24
Death	0	0.00	1	0.27
Promotion	6	1.50	4	1.08

Table 15: Exit Decisions and Other Modes of Termination - when the seat is up for reelection

	Appointed		Elected	
	Frequency	Proportion(%)	Frequency	Proportion(%)
Voluntary Exit	13	3.00	28	8.38
Running	420	96.77	302	90.42
Death	0	0.00	2	0.60
Promotion	1	0.23	2	0.60

elected judges) is not feasible for appointed judges. Hence, we use a probabilistic voting model in which we identify the reelection probability function with the distribution of the vote share. We specify the model, discuss identification, describe the data on vote share, and show the results.

5.1 Model

When appointed judges run for reelection, they do not face challengers. Voters in the district take a yes-or-no vote for the incumbent. The probabilistic voting model consists of three elements: voter utility from observable characteristics and sentencing decisions of the incumbent, individual voters' idiosyncratic taste shocks, and district-level taste shocks.[6] A voter votes for the incumbent when the total of the three utility components is larger than zero. Or, equivalently (and for ease of exposition), a voter votes for the incumbent when the sum of two components - utility from observables of the incumbent and his (voter's) idiosyncratic taste shock - exceeds a district-level threshold, which is also a random variable. That is, voter j in the district of judge i at period t casts a yes-vote if

$$h(XR_{it}) + \varepsilon_{jt} \geq \eta_{Ait},$$

[6]For papers describing the probabilistic voting model and its empirical application, see the following: Lindbeck, A., and J. Weibull (1987): "Balanced-budget Redistribution as Political Equilibrium," *Public Choice*, 52, and Strömberg, D. (2008), "How the Electoral College Influences Campaigns and Policy: The Probability of Being Florida", *American Economic Review*, 98-3.

where XR_{it} is a state vector (a bundle of observables and sentencing decisions of incumbents), $h(XR_{it})$ is voters' utility from XR_{it}, ε_{jt} is voter j's idiosyncratic taste shock, η_{Ait} is district-level taste shock, and ε_{jt} and η_{Ait} follow normal distribution, $\varepsilon_{jt} \sim N(0,1)$ and $\eta_{Ait} \sim N(0, \sigma_A^2)$. The specification of the function $h(\cdot)$ is identical to that of the latent variable $g(\cdot)$ we used for elected judges.

For a realization of district-level taste shock η_{Ait}, the vote share of the incumbent is

$$1 - \Phi(-h(XR_{it}) + \eta_{Ait}) = \Phi(h(XR_{it}) - \eta_{Ait}),$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution. Additionally, the *ex-ante* reelection probability of a judge with state vector XR_{it} (before realization of η_{Ait}) is

$$\begin{aligned} \text{reelection probability} &= \Pr\left\{\Phi(h(XR_{it}) - \eta_{Ait}) \geq \frac{1}{2}\right\} \\ &= \Phi\left(\frac{h(XR_{it})}{\sigma_A}\right). \end{aligned}$$

Remark: The above mathematical relation between distribution of vote share and reelection probability hinges on the fact that voters always have two fixed options (yes or no for the incumbent). We cannot apply a probabilistic voting model to elected judges, since an elected judge may often face no challengers if he is strong. That is, we cannot derive the above relation between vote share and reelection probability for elected judges.

5.2 Identification

Parameters of the probabilistic voting model are identified from the variation of the share of yes-votes across time and districts. Since we observe only the proportion of voters who voted yes, *not* individual voters' utility from incumbents, the parameters of voter utility from incumbents ($h(XR_{it})$) are identified *only up to scale*. Hence, we normalize the variance of individual voters' taste shock to 1. Then, parameters of $h(\cdot)$ capture the relationship between variation in XR_{it} and variation in the share of yes-votes. Variation in vote share not explained by variation in XR_{it} is attributed to district-level taste shock η_{Ait}.

5.3 Data: Distribution of Yes-vote Share

Since an appointed judge loses in reelection when the yes-vote share is below 50%, the reelection probability function is determined by the overall frequency that the yes-vote share falls under (or close to) 50% and the variation in observable variables. In this section, we document the overall distribution of the yes-vote share and its relationship to key observables (sentencing decision and political climate).

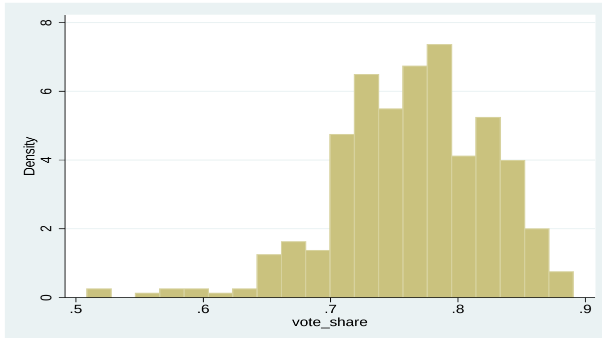


Figure 10: Distribution of the Yes-vote Share of Appointed Judges (All Sample)

Statistics	All Sample	By Overall Sentencing			By Political Climate		
		Low	Middle	High	Favorable	Neutral	Unfavorable
Mean	76.52	75.52	76.34	75.29	75.81	77.03	76.49
Std. Dev.	5.94	5.92	4.70	3.81	6.08	5.81	5.97
Minimum	50.86	51.33	63.46	65.75	50.86	51.33	58.14
Maximum	89.04	88.29	84.58	83.48	88.29	89.04	85.51
10th percentile	69.64	70.64	71.33	70.46	67.93	70.96	68.84
25th percentile	72.77	72.43	73.56	73.52	71.82	73.95	72.19
50th percentile	76.91	75.10	76.75	75.96	76.10	77.70	76.76
75th percentile	80.96	79.35	79.07	77.09	80.42	80.39	81.88
90th percentile	83.87	81.41	82.45	78.73	83.48	84.13	84.06

Table 16: Summary Statistics of the Yes-vote Share (%) of Appointed Judges

Figure 10 and the second column ('All Sample') of Table 16 show the overall distribution of the yes-vote share for the whole sample of reelection of appointed judges and its summary statistics, respectively. The mean of the distribution is 76.52%, the standard deviation is 5.94%, and the 10th percentile is 69.64%. These summary statistics show that there is very little variation in the yes-vote share, and appointed judges are extremely safe most of the time.

5.3.1 By Sentencing Decision

In this section, we document the overall distribution of yes-vote share of appointed judges by sentencing decisions in the term preceding the reelection. For simplicity of exposition, we categorize the sentencing decisions in a term (two periods) as follows: (a) If the pair of sentencing decisions in the term is one of the following six combinations – (L,L), (SL,L), (S,L), (SL,SL), (S,SL), or (SH,L) – we classify the overall sentencing as "Low", (b) if it is one of the following three combinations

– (S,S), (SH,SL), or (H,L) – we classify the overall sentencing as “Middle”, (c) if it is one of the following six combinations – (H,H), (H,SH), (H,S), (SH,SH), (S,SH), or (H,SL) – we classify the overall sentencing as “High”. (This classification is summarized in Table 17.) The third, fourth,

Table 17: Three categories of the Combinations of Sentencing Decisions

Category	Combination of Sentencing Decisions
Low	(L,L), (SL,L), (S,L), (SL,SL), (S,SL), and (SH,L)
Middle	(S,S), (SH,SL), and (H,L)
High	(H,H), (H,SH), (H,S), (SH,SH), (S,SH), and (H,SL)

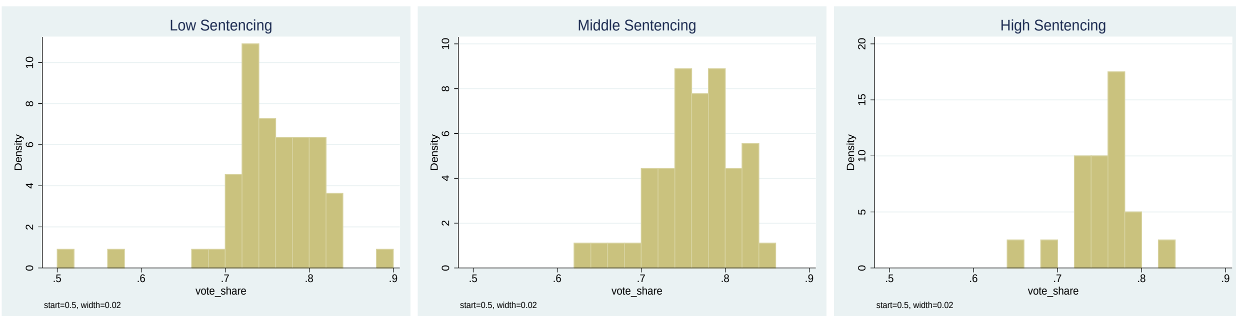


Figure 11: Distribution of the Yes-vote Share of Appointed Judges by Overall Sentencing

and fifth columns of Table 16 and the histograms in Figure 11 show the summary statistics and the overall distribution of the yes-vote share of appointed judges by sentencing decisions in the term preceding the election. In all three categories, the mean is around 75%, the standard deviation is around 4∼6%, and the 10th percentile is above 70%.

5.3.2 By Political Climate

The last three columns of Table 16 and the histograms in Figure 12 show summary statistics and the distribution of the yes-vote share of appointed judges under three different conditions of political climate: (a) when political climate is unfavorable to the party (i.e., when a judge was initially appointed by a Republican governor and the current political climate is favorable to Democrat, or vice versa) (b) when political climate is neutral, (c) when political climate is favorable to the party (when a judge was initially appointed by a Republican governor and the political climate is favorable to Republicans, or vice versa). Under all three conditions of political climate, the mean yes-vote share is above 75%, and the standard deviation is around 6%. Under all three conditions, the 10th percentile is around 70%.

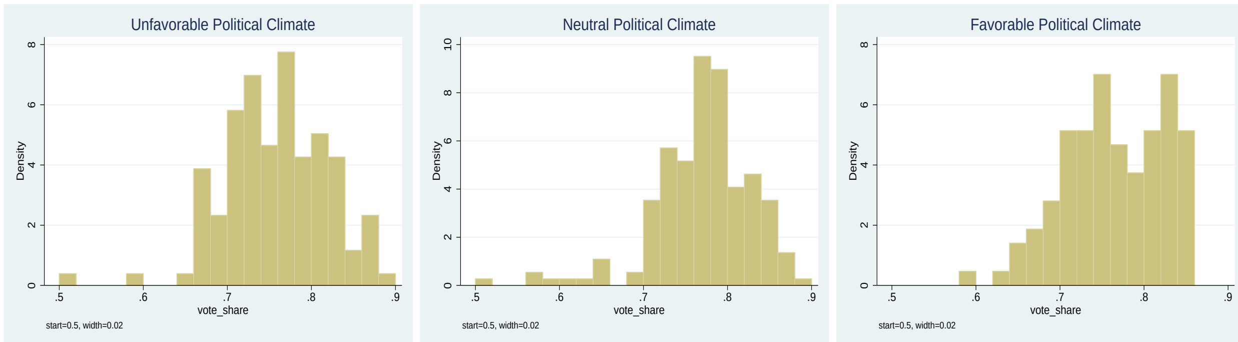


Figure 12: Distribution of the Yes-vote Share of Appointed Judges by Political Climate

5.4 Discussion

The overall distribution of the yes-vote share of appointed judges shown in the previous section implies that appointed judges are extremely safe. *A priori*, having no observations of failure of appointed judges may *not* necessarily imply that appointed judges are free from reelection concerns. It may very well be the case that appointed judges adjust their decisions just sufficiently to be reelected. But, distribution of the yes-vote share of appointed judges generated from such a situation would normally have more observations of the yes-vote share in a relatively low range (i.e., 50∼60% of vote share) than our data shows. Overall, the distribution of the yes-vote share in our data shows the mean (around 70%) well above the threshold of reelection (50%), with small standard deviation. Hence, it is reasonable to consider that appointed judges are reelected with probability 1. In the next section, we show that this assumption in the model in the main text is consistent with the estimation result of an alternative specification in which reelection probability of appointed judges is estimated with the probabilistic voting model specified above.

5.5 Estimation Result

The parameter estimates of the probabilistic voting model specified above and their standard errors are in Table 18.[7] The specification of $h(XR_{it})$ for appointed judges, used on page 18, is identical to that of $g(XR_{it})$ for elected judges (specified in the appendix of the paper), and the definition of each parameter among ψ's is identical to its counterpart among ϕ's.

The second column of Table 19 shows the summary statistics of the vote share simulated from the estimated model parameters, and Figure 13 shows its overall distribution. The estimated model has good performance in predicting the key summary statistics of the overall vote share. Addition-

[7]The parameters were estimated along with other parameters from the baseline model.

[8]Since there is only very little variation in the vote share that is related to the covariates, the coefficient estimates naturally have large standard errors. This feature is another reason why it is better to set the reelection probability of appointed judges at 1 (as in the main text of the paper) rather than to estimate it.

Table 18: Parameter Estimates - Reelection Probability of Appointed Judges

Parameter	Component of the Model	Estimate	Std. Error [8]
ψ_1	Constant	1.3211	0.7631
$\tilde{\psi}_{DC}$	Scale - Democrat, conservative	0.2526	0.4301
$\tilde{\psi}_{DL}$	Scale - Democrat, liberal	0.3252	0.4249
$\tilde{\psi}_{RC}$	Scale - Republican, conservative	0.2510	0.4336
$\tilde{\psi}_{RL}$	Scale - Republican, liberal	0.3151	0.4285
$\hat{x}_C$	Bliss point - conservative districts	0.9795	0.8316
$\hat{x}_L$	Bliss point - liberal districts	0.2387	0.5491
σ_f	Common scale parameter	0.6512	0.7279
ψ_3	$I[Noncrime_i]$	-0.4893	0.7592
ψ_4	Age_{it}	-0.0018	0.0010
ψ_5	$Tenure_{it}$	-0.0107	0.0035
ψ_6	$I[SOD=1]*I[Party_i=D]$	-0.0147	0.0396
ψ_7	$I[SOD=2]*I[Party_i=D]$	0.0328	0.0424
ψ_8	$I[SOD=3]*I[Party_i=D]$	-0.0207	0.0463
ψ_9	$I[SOD=1]*I[Party_i=R]$	0.0598	0.0408
ψ_{10}	$I[SOD=3]*I[Party_i=R]$	-0.0358	0.0571
σ_A	Std. Dev of the Taste Shock η_{Ait}	0.1782	0.0059

ally, the overall distribution of the vote share predicted from the estimated model, in Figure 13, is similar to the empirical observation. (It covers the range from around 50% to 90% with slight left-skewness.)

The last column of Table 19 shows the summary statistics of the reelection probability predicted from the estimated model. This clearly shows that there is extremely small variation in the reelection probability of appointed judges, and *the whole distribution lies between 99% and 100% reelection probability*. Therefore, we can conclude that it is a reasonable approximation to consider that appointed judges are reelected with almost probability 1 irrespective of their sentencing behavior.

6 Procedures of Counterfactual Experiments

In this section, we describe how counterfactual experiments in Section VI.A of the main paper are conducted. In both counterfactual experiments, we use parameter values of the model that are estimated in the main analysis. The exact procedure of counterfactual experiments is as follows.

Step 1 (value function calculation): As in the estimation procedure, we solve a dynamic programming problem by backward induction, using the parameters of the model. That is, we compute the present discounted value of each decision from the last period and proceed backward.

Table 19: Vote Share and Reelection Probability of Appointed Judges from the Estimated Model

Statistics	Predicted Vote Share (%)	Predicted Reelection Probability (%)
Mean	76.39	99.99351
Std. Dev.	5.85	0.01385
10th percentile	68.72	99.98407
25th percentile	72.69	99.99390
50th percentile	76.75	99.99789
75th percentile	80.53	99.99937
90th percentile	83.68	99.99980

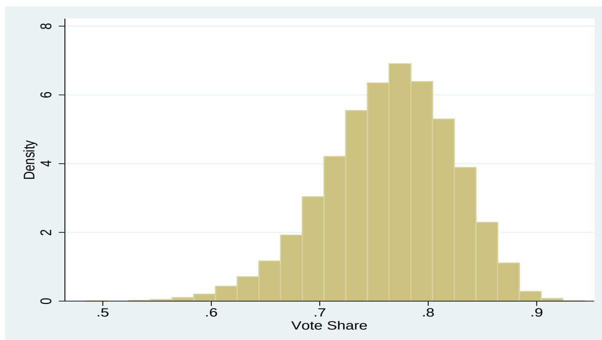


Figure 13: Predicted Distribution of the Yes-vote Share of Appointed Judges

The difference between the estimation procedure and the counterfactual experiments is that we use a hypothetical reelection probability function in simulation (a) (life-tenure) and hypothetical preference distribution in simulation (b). That is, in simulation (a), we replace the actual reelection probability function with the hypothetical "reelection probability=1". In simulation (b) where appointed judges face competitive elections, we replace elected judges' preference distribution with that of appointed judges.

Step 2 (drawing initial conditions): We set the distribution of initial conditions of individual judges (entry age, pre-entry work experience, party, etc.) at the empirical distribution in the data. We draw 10,000 judges from this distribution.

Step 3 (simulation of decisions): We simulate the decision of each judge with initial conditions drawn from Step 2, from the initial period to the period after exit. From the initial period, we move forward simulating each judge's decision and random components that affect decisions (e.g., taste shocks, political climate, reelection uncertainty, etc.).

Step 4 (aggregation): Aggregate decisions simulated in Step 3.