

Random Assignment to Death *

Deborah Beim
Assistant Professor
Department of Political Science
Yale University
deborah.beim@yale.edu

Tom S. Clark
Charles Howard Candler Professor
Department of Political Science
Emory University
tom.clark@emory.edu

Benjamin E. Lauderdale
Professor
Department of Political Science
University College London
b.lauderdale@ucl.ac.uk

January 29, 2019

ABSTRACT

We use the quasi-random assignment of cases to three-judge panels on the US Courts of Appeals to assess the consistency of adjudication of death penalty appeals. We find clear evidence that panels apply different standards depending on whether a majority of the panel was appointed by Democratic or Republican presidents. Unlike previous work on panel effects in the US Courts of Appeals, we show that these effects persist to the end of the process of adjudication. Since the the early 1980s, across the US as a whole, the probability of ultimate execution has depended on the judges that were assigned to an inmate's first Court of Appeals case.

*Previous versions of this research have been presented at the 2015 Philadelphia Region American Politics Conference, 2017 European Political Science Association, and the 2017 American Political Science Association.

Capital punishment is the most punitive and irreversible form of judicial sanction. As a result, it is clear that its application must meet the very highest standards of fairness and justice. Indeed, it was out of a concern about unfair application of the death penalty that the US Supreme Court, in *Furman v. Georgia* (1972), struck down all death penalty statutes in the country. In a concurring opinion in that case, Justice Stewart wrote, “These death sentences are cruel and unusual in the same way that being struck by lightning is cruel and unusual. I simply conclude that the Eighth and Fourteenth Amendments cannot tolerate the infliction of a sentence of death under legal systems that permit this unique penalty to be so wantonly and so freakishly imposed.” When, just four years later, in *Gregg v. Georgia* (1976), the Supreme Court approved of newly written statutes governing sentencing procedures, the majority specifically argued that the new statutes addressed the problem of arbitrariness in death penalty sentencing.

There remains considerable disagreement as to whether these and subsequent reforms in capital sentencing indeed established a fair, even, and consistent application of the death penalty. As Baumgartner et al. (2018) argue, the spirit of the Court’s decisions validating new death penalty statutes in 1976 included the proposition that the imposition of the death penalty cannot be arbitrary or random. But recent research suggests juries and trials introduce wide variance in outcomes, including as a function of race (Anwar, Bayer, and Hjalmarsson 2012; Alesina and Ferrara 2014; Spurr 2002). Moreover, nearly all death penalty cases enter a long appeals process after the initial conviction and sentencing phase, which leads to a lengthy period of review and frequent success (see Gelman et al. 2004). Of the 6,000 individuals sentenced to death between 1973 and 1995, only 5% had been executed by 1995. Most of the remaining 95% of the sentences were either overturned or under continued appellate review. This appellate process was frequently characterized by claims that were very unlikely to prevail, and as a consequence Congress passed, in 1996, the Antiterrorism and Effective Death Penalty Act (AEDPA), which contained procedural hurdles including a

one-year statute of limitations period for seeking habeas corpus and severely restricted the ability to file a second or successive petition for a writ of habeas corpus, as well as making it more difficult to meet the standards for a writ.

We examine one aspect of this capital sentence appellate process: the US Courts of Appeals, which, due to the limited number of cases heard by the US Supreme Court, is usually the court of last resort for those who face execution. Generally speaking, death penalty cases reach the Courts of Appeals after the convicted defendant has exhausted all state-level appeals. Appellate decisions are almost always made by a panel of three judges selected from the pool of judges in the circuit.¹ These courts have the opportunity to correct for error and inconsistency that has occurred previously, but also potentially to introduce inconsistency themselves. Previous work has demonstrated that the existence of three judge panels reduces inconsistency in death penalty cases versus a hypothetical alternative of single-judge decisions in the Courts of Appeals (Beim and Kastellec 2014; Fischman 2015)(though see Sunstein et al. (2006), which argues that judges do not influence one another’s decision-making in death penalty cases). The deliberative process of collegial decision making on these courts may help promote a uniform standard across the cases heard in a circuit. However, even such influence is not necessarily sufficient to achieve a uniform standard across a circuit if judges’ pre-deliberation standards vary widely, because randomly assigned three-judge panels will frequently group judges with similar views together. The question of influence among the judges on a panel is intertwined with the question of how we measure which standards individual judges are inclined to apply, given that we only observe them making decisions in panels.

In this paper, we quantify the inconsistency that remains in the implementation of the

¹The US Courts of Appeals are arranged into 12 geographically-defined circuits. Currently sitting judges from the circuit responsible for the state seeking the death penalty are the primary pool for any case; however judges from US District Court, from other circuits of the U.S. Courts of Appeals, or retired judges from the circuit sometimes sit as one of the three judges “by designation”.

death penalty—*both in case outcomes and in actual executions*. Consistent with a generation of research on the relationship between judicial ideology and decision making, we show predictable variation in the standards judges apply to death penalty cases. Panels with a majority of Democratic-appointed judges grant relief more often; panels with a majority of Republican-appointed judges grant relief less often. Because of the effectively random manner in which death penalty cases are initially assigned to panels at the Court of Appeals, this level of appellate review creates variation in the standards that are applied at this critical stage in the post-conviction review of death sentences. We go further, though, and estimate the consequences of this heterogeneity on whether individuals are executed. While the effects for execution are somewhat smaller than for relief (probably due to the remaining oversight mechanisms of en banc review and the Supreme Court) we nonetheless find evidence that panel composition predicts execution probability. This finding is the first to our knowledge to systematically document the consequences of panel composition for the way in which criminal justice is ultimately administered to individual defendants in the highest stakes cases. This finding calls into question the extent to which the American legal system is meeting the standard that the Supreme Court has set out for the death penalty’s conformity to the Eight Amendment’s prohibition on cruel and unusual punishment.

IDENTIFICATION, DATA AND METHODS

We estimate the consequences of panel composition on decisions and executions by comparing cases that were assigned to different judges. Like much previous research, our study relies on the assumption that as-if random assignment of cases to judges holds for these cases. Internal assignment rules vary by circuit, but there is no evidence that cases are systematically assigned to judges in a way that relates to their merits or content.² Available evidence

²Not all circuits use computerized randomization (Hall 2010) or publicise their procedures. Levy (2017) writes that judges are randomly assigned in all death penalty cases in the Second, Third, Fourth, and

suggests that the assignment procedures are sufficient to prevent problematic imbalances in the strength of cases heard by different judges.³ In the domain of death penalty cases, previous research has indicated that Republican and Democratic appointees are equally likely to see defendants who won at the district court level (Beim and Kastellec 2014).

As-if-random assignment only plausibly holds for the first case involving a given death-row inmate before the Courts of Appeals, and only when comparing such cases heard in the same year in the same circuit. Some inmates have multiple cases heard over a period of years, but whether further cases are heard for the same inmate can be an outcome of the initial case, panel assignment is not independent across these cases, and the strength of such cases are unlikely to be independent either. Different circuits have different mixes of cases coming up from their constituent states and also different mixes of Democratic and Republican appointed judges. Within each circuit, the mix of cases and the mix of appointed judges potentially change over time. Further, the Supreme Court precedents that the Appeals courts apply are also changing over time. Taking these points together, among the set of first appearance cases considered at the same time by the same appellate court, the expected strength of the cases heard by any possible three-judge panel is the same. This means that we must analyze each Circuit separately and also adjust for the average “case strength” at any given moment in time in order to isolate the causal effect of the panel assignment.

There are many databases of death penalty appeals previously in existence, however none linked all the information we required for our analysis. We performed an over-inclusive

Eleventh Circuits. In the Fifth, Sixth, Seventh, Ninth, and Tenth Circuits, judges are randomly assigned but the assignment also takes into account concerns such as workload and the variety of judges with which an individual sits each year.

³It is not a problem for our identification strategy if panels themselves are not randomly constructed, so long as cases are assigned to those panels without respect to the facts of those cases. The most serious threat to our inferential strategy would be if cases are assigned in a way that causes some judges to get systematically weaker cases than other judges, because those judges would spuriously appear more conservative by virtue of more frequently denying relief.

Westlaw search for all cases that could be a case from a death row inmate before any circuit in the US Court of Appeals following the procedures in Fischman (2015) and Beim and Kastellec (2014). This procedure yielded more than 20,000 cases between 1983 and 2012. Each case was then read, individually, by either one of us or a member of our research team that comprised law students, graduate students in political science, and undergraduate students. Each case was assessed for whether it was a death penalty case. We retained each case decided by a three-judge panel (i.e., we exclude cases decided *en banc*) and recorded a number of pieces of information, which we describe below.⁴ As a result, our data includes those individuals who appeared with death penalty-related cases for the first time before the Courts of Appeals between 1 January 1983 and 31 December 2012.⁵

Because some states sentence far more people to death than others and circuits are organized geographically, cases are very unevenly distributed across circuits. The D.C. Circuit, and the First, Second and Third circuits (covering the mid-Atlantic through New England) yield too few cases for us to study. Therefore, we focus on the eight circuits—the Fourth through Eleventh, inclusive—which handle nearly all death penalty cases. The resulting data include 1,991 initial death penalty cases decided in the Court of Appeals between 1983 and 2012. For each case, we recorded the names of the judges on the panel, how each judge voted (whether to support any relief at all for the defendant), and whether the decision supported any relief at all for the defendant. We also recorded the name of the prisoner, the state of origin, procedural history (which other courts heard the case), the date of the panel’s decision, who the district court judge was, and how s/he voted.

We focus on two outcomes. The first outcome is the panel’s immediate decision to grant

⁴We then further examined all cases identified by a similar search decided between July 2, 1976 (when the Supreme Court decided *Gregg v. Georgia* 428 U.S. 153) and 12/31/1982, to see if any of our death penalty defendants had a case in the Court of Appeals prior to 1983. If they did, we excluded them from our data as their first appearance in our data set was not their first appearance at the Court of appeals.

⁵Our data exclude all cases in which a person was sentenced to death if that sentence was commuted prior to the appeal. Our data also exclude cases brought by next friends and cases in which the person on death row does not seek relief (such as cases about prison conditions).

or deny relief to a death-row prisoner. This allows us to focus on the reduced-form effects of inconsistency, knowing that suppressed dissents, bargaining, and other interjudge dynamics may influence the decisions that we observe. The second outcome of interest is whether the death-row prisoner is ultimately executed. We employ the list of executions maintained by the Death Penalty Information Center (<https://deathpenaltyinfo.org/>) current as of 2017, which we verified by cross-referencing against states' websites listing executions. We then matched each execution to an inmate appearing in our data.

Inmates who were granted relief are not necessarily removed from death row, and inmates granted relief may still ultimately be executed. For example, "relief" may take the form of a remand but the outcome of the remand may once again be the death penalty. Nevertheless, whether an inmate is granted relief in their initial appeals court case and whether they are ultimately executed are highly correlated outcomes. In most circuits, the relief denial in the initial panel decision is associated with a roughly 30 percentage point increase in ultimate execution rates.⁶

Our analysis focuses on a reduced form identification of the causal effect of having different panel compositions on grants of relief y_j and ultimate execution z_j in cases j . This approach has the advantage of generating comparable estimates of the causal effect of panel composition on both the relief and the execution outcomes. The latter effect is likely to run partly, but not entirely, through the immediate decision of the panel on whether to grant relief. However this is not the only causal pathway because the same panel is more likely to hear subsequent appeals from that defendant and the identities of the assigned judges may influence the subsequent oversight process via en banc and/or Supreme Court review. It is the very complexity of this subsequent process that makes the reduced form analysis

⁶Appendix Table A1 reports the proportion of inmates ultimately executed in each circuit (as of 2017) divided into those who were granted relief at their first appeal before the Court of Appeals and those who were not. The baseline rates at which inmates are granted relief and the rates at which they are executed vary over the period we study within and across circuits (see appendix Figure A1).

attractive: it allows us to estimate whether there is a causal effect of initial panel assignment on execution, regardless of the relative importance of the various causal pathways through which that effect could arise.

Our primary treatment variable is whether a panel has a majority of Democratic appointed judges $T_j = 0$ or a majority of Republican appointed judges $T_j = 1$. The key identifying assumption is that this treatment assignment is as-if-randomly assigned. As discussed above, this assumption is only plausible conditional on circuit and date. Therefore, we estimate linear regression models separately for each circuit, with no time effects and with year fixed effects. Year fixed effects provide the most credible causal estimates, as they allow for potentially different baseline rates of grant denial and ultimate execution for cases that first reached the Courts of Appeals in each individual year. We include the model with no time effects simply for descriptive purposes, it is not causally credible for the reasons described above. As a balance check, in the appendix we show that the state from which each case arose is not systematically associated with having a Republican majority, but only once we condition on circuit and year.

Our analysis only estimates variation in decision-making that is a function of the treatment variable. This means we cannot detect inconsistency that is associated with “within-party” variation in the standards judges apply. We are also estimating an average treatment effect in each circuit, but the treatment effect may itself vary over the period under study. These limitations cannot lead us to overestimate inconsistency in decision-making however, if judges vary within appointing party (which they surely do) or over time (which they likely do) that implies greater inconsistency in decisions than we estimate, albeit inconsistency that is not straightforwardly associated with whether there is a Democratic or Republican appointed majority on the panel.

Outcome	Year FE	Circuit								Avg. Effect	Total Effect	95% CI
		4	5	6	7	8	9	10	11			
Relief Denial	No	0.14 (0.04)	0.07 (0.04)	0.30 (0.07)	0.23 (0.14)	0.12 (0.07)	-0.01 (0.08)	0.03 (0.07)	0.16 (0.04)	0.12 (0.02)	232 (41)	152-313
	Yes	0.18 (0.05)	0.09 (0.04)	0.31 (0.07)	0.19 (0.15)	0.17 (0.08)	0.00 (0.10)	0.04 (0.08)	0.08 (0.04)	0.12 (0.02)	235 (44)	149-322
Execution	No	0.08 (0.06)	0.12 (0.05)	0.12 (0.07)	0.13 (0.16)	0.10 (0.09)	0.00 (0.07)	0.07 (0.08)	0.04 (0.05)	0.09 (0.02)	169 (48)	76-263
	Yes	0.02 (0.07)	0.14 (0.05)	0.17 (0.07)	-0.07 (0.16)	0.12 (0.10)	-0.12 (0.09)	0.03 (0.09)	0.05 (0.05)	0.07 (0.03)	143 (52)	42-244

Table 1: *Estimated treatment effects in each circuit for a change from a Democratic majority to a Republican majority panel, as differences in proportion of negative outcome for the inmate. The total effect is the implied difference in the total number of negative case outcomes over all circuits. Standard errors in parenthesis.*

RESULTS AND DISCUSSION

Table 1 shows the estimated effect of a change from a Democratic appointed panel majority to a Republican appointed panel majority on the probability of relief denial and execution in each of the circuits we consider. Republican majority panels are associated with higher relief denial rates and higher probability of ultimate execution in most circuits. In the fixed effects model the circuit-level point estimates vary from a 0.00 to a 0.31 higher probability of a relief denial with a Republican appointed majority versus a Democrat appointed majority panel. For execution, the estimates range from -0.12 to 0.17. Each of these circuit-specific estimates are relatively imprecise, but they are independent from one another and so the collective evidence they provide about the full set of circuits is much stronger. If we weight the circuit-level estimates by the number of cases decided in those circuits, we can construct an estimate of the average treatment effect of going from a Democratic appointed majority to a Republican appointed majority, for all death penalty cases in our data set. For relief denial, the average treatment effect is 0.12; for execution, it is 0.07. These average treatment effects of 12 and 7 percentage points correspond to 235 additional relief denials (95%: 149-322) and 143 additional executions (95%: 42-244) in a world where all panels had Republican

majorities versus a world where all panels had Democratic majorities. The p-value for the null hypothesis that there was no average treatment effect on switching between Democratic and Republican majority panels across all circuits is 0.003 for executions and negligible for relief denial.

It is important to reiterate that the individual-level treatment effects underlying these average treatment effects are both heterogenous across circuits and across time, and also that appeals would have faced different probabilities of having panels with Republican or Democratic majorities across circuits and across time. While we have made some minor functional form assumptions in order to form this estimate, there is no reason to expect those assumptions to bias the estimate in a particular direction or away from zero.

Because the death penalty entails such a high degree of punitiveness and irreversibility, its exercise requires the highest level of legal scrutiny to ensure its application does not violate individual rights. In the US, the Supreme Court has held that the constitutional prohibition on cruel and unusual punishment requires the death penalty not be administered randomly or in a way affected by factors that are orthogonal to the legal merits of the crime and defendant. Our analysis demonstrates two findings that call into question the extent to which the death penalty is administered to this standard. First, consistent with scores of research, we find that judges apply different standards which correlate with partisanship in deciding cases that are, in expectation, equally strong on the merits. We then go further to show a causal effect of the partisan composition of Appeals Court panels for the ultimate fate of the litigants coming before the court. If the identity of the judges influenced how panels decided cases but the institutions of judicial oversight remedied this variation before execution, one could argue that such variation is of limited normative concern. What we find instead is that being randomly assigned to differently composed panels has a causal effect on whether or not an individual lives or dies.

References

- Alesina, Alberto and Eliana La Ferrara (2014). “A Test of Racial Bias in Capital Sentencing”. *American Economic Review* 104.11, pp. 3397–3433.
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson (2012). “The Impact of Jury Race in Criminal Trials”. *Quarterly Journal of Economics* 127, pp. 1017–1055.
- Baumgartner, Frank et al. (2018). *Deadly Justice: A Statistical Portrait of the Death Penalty*. New York: Oxford University Press.
- Beim, Deborah and Jonathan P. Kastellec (2014). “The Interplay of Ideological Diversity, Dissents, and Discretionary Review in the Judicial Hierarchy: Evidence from Death Penalty Cases”. *Journal of Politics* 76.4.
- Fischman, Joshua B. (2015). “Interpreting Circuit Court Voting Patterns: A ‘Social Interactions’ Framework”. *Journal of Law, Economics, & Organization* 31.4, pp. 808–842.
- Gelman, Andrew et al. (2004). “A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States”. *Journal of Empirical Legal Studies* 1.2, pp. 209–261.
- Hall, Matthew (2010). “Randomness Reconsidered”. *Journal of Empirical Legal Studies* 7.3, pp. 574–589.
- Levy, Marin K. (2017). “Panel Assignment in the Federal Courts of Appeals”. *Cornell Law Review* 65.
- Spurr, Stephen J. (2002). “The future of capital punishment: determinants of the time from death sentence to execution”. *International Review of Law and Economics* 22.1.
- Sunstein, Cass R. et al. (2006). *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington, D.C.: Brookings Institution Press.

ONLINE APPENDIX

Circuit	Execution Rate				Difference
	Granted Relief	n	Denied Relief	n	
4th	33%	27	64%	217	31%
5th	40%	110	72%	490	32%
6th	10%	60	43%	108	33%
7th	0%	13	29%	65	29%
8th	28%	40	58%	148	30%
9th	17%	71	37%	81	20%
10th	20%	41	76%	118	56%
11th	20%	117	48%	285	28%

Table A1: *Proportion of inmates executed in each circuit, according to whether they are granted any relief at the initial appeal.*

Circuit	State	# Petitioners	Relief Denied	Petitioner Executed
4	MD	10	10	2
4	NC	90	79	30
4	SC	33	29	24
4	VA	111	99	91
5	LA	46	32	24
5	MS	42	27	16
5	TX	512	431	356
6	KY	17	15	2
6	MI	1	1	0
6	OH	118	73	44
6	TN	32	19	6
7	IL	53	47	9
7	IN	25	18	10
8	AR	50	35	22
8	IA	3	3	0
8	MO	119	97	71
8	ND	1	1	0
8	NE	13	10	3
8	SD	2	2	1
9	AK	1	0	0
9	AZ	52	30	26
9	CA	66	37	10
9	ID	6	2	1
9	MT	5	4	1
9	NV	9	2	1
9	OR	3	1	0
9	WA	10	5	3
10	CO	3	3	1
10	KS	1	0	0
10	NM	2	2	1
10	OK	145	107	92
10	UT	6	5	3
10	WY	2	1	1
11	AL	75	61	46
11	FL	212	157	68
11	GA	115	67	46
All		1991	1512	1011

Table A2: *Distribution of cases and outcomes by circuit and state.*

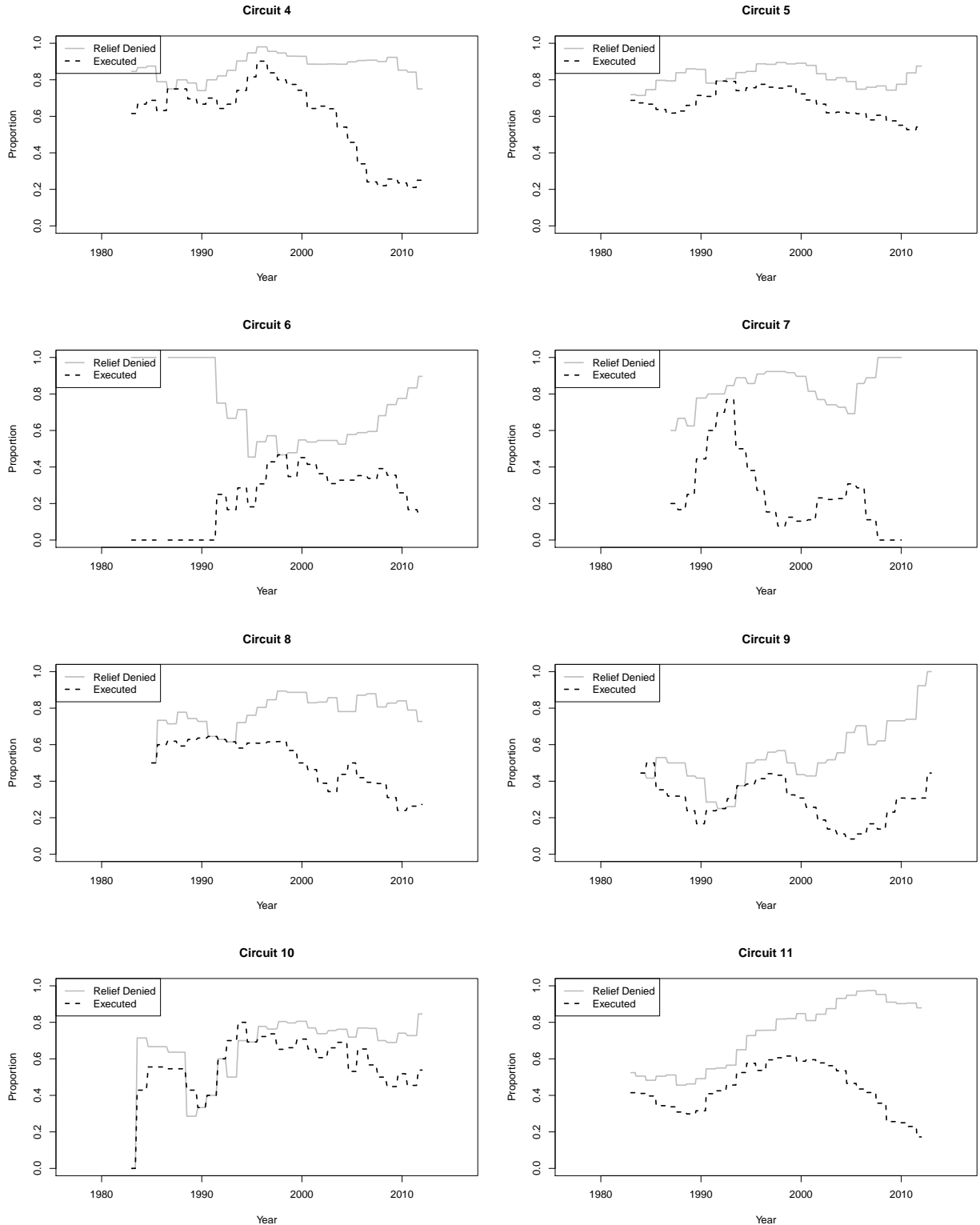


Figure A1: Five-year moving average for the rate at which death row inmates are denied relief by the Courts of Appeals, and at which they are ultimately executed, by year of initial Courts of Appeals case decision.

State Balance Check

The only variable that we can easily measure which is potentially associated with case strength and therefore might provide a balance check is the origin state of each case. Because it is state law and criminal procedures that are typically the subject of the case being heard by the US Court of Appeals, any imbalance where Republican majorities were more or less likely to hear cases from certain states, holding constant circuit and year, would be problematic for our analysis. In order to test whether such an imbalance exists, we ran the same circuit-specific regressions used in our main analysis—with the Republican majority treatment variable and including or excluding year fixed effects—to predict the origin state of each case. This generates a set of 36 p-values on the treatment variable, one for each state that appears in our data. If there is no association between state and treatment, we would expect these to follow a uniform distribution.

Using the year fixed effects analysis that forms our main estimates, we cannot reject the null of a uniform distribution for the treatment variable p-values: the Kolmogorov-Smirnov test p-value is 0.32. Conditional on circuit and year, there is no consistent evidence that Republican majorities were more likely to hear cases from particular states. The analysis without the year fixed effects fails this test, with a p-value of 0.01, which demonstrates the importance of conditioning on year to credibly identify a causal effect. We obtain similar results using a one-sample t-test versus a mean of 0.5: the p-values from the year fixed effects models have mean of 0.43 with $p = 0.20$ versus the null of 0.5, and the p-values from the models without fixed effects have mean 0.36 with $p = 0.01$ versus the null of 0.5.

The only circuit where any state has a p-value less than 0.05 in this balance check is the 4th circuit, which consists of NC, VA, SC and MD (WV is in the 4th circuit, but does not use the death penalty). Even conditional on year, NC cases are less likely to have Republican majorities than cases from the other three states, and VA cases are more likely to. However

the 4th circuit is not driving our main results. Our fixed effects estimate for the treatment effect of moving from a Democratic to Republican majority is somewhat higher in the 4th circuit than the average circuit for relief denials (0.18 vs 0.12), but somewhat lower than average for execution (0.02 vs 0.07). The 4th circuit is only responsible for 12% of the cases in the data and the national average effects barely change if it is excluded. Adding state fixed effects in addition to year fixed effects to the main models has no consequential effect on either the 4th circuit estimates or the national estimates for either outcome.