Electoral Institutions and the Performance of Public Officials: Evidence from State Supreme Courts

Elliott Ash and W. Bentley MacLeod*

November 1, 2016

Abstract

This paper exploits the variation in how U.S. state supreme court judges are appointed and retained to measure the effect of these changes upon performance using a panel of all judicial opinions written between 1947 and 1994. We find evidence of both incentive and selection effects due to electoral procedures. Election-year politics reduces the output of judges in non-partisan elections, but not in partisan elections or uncontested elections. Moving from non-partisan elections to uncontested elections causes incumbent judges to improve work quality, while moving from partisan to uncontested elections has no effect on this choice. Judges selected by technocratic merit commissions produce higher-quality work than either partisan-elected judges or non-partisan-elected judges. These results are consistent with the view that technocratic merit commissions have better information about judge quality than voters, and that political bias can reduce the quality of officials selected.

*Elliott Ash: etash@princeton.edu; Center for Study of Democratic Politics, Woodrow Wilson School of Public Affairs, Princeton University; Department of Economics, University of Warwick. W. Bentley MacLeod: bentley.macleod@columbia.edu; Department of Economics, Columbia University. We thank Yisehak Abraham, Ankeet Ball, Josh Brown, Josh Burton, Matthew Buck, Eammonn Campbell, Daniel Deibler, Seth Fromer, Gohar Harutyunyan, Archan Hazra, Montague Hung, Dong Hyeun, Mithun Kamath, James Kim, Michael Kurish, Jennifer Kutsunai, Steven Lau, Sharon Liao, Sarah MacDougall, Justin McNamee, Sourabh Mishra, Brendan Moore, Arielle Napoli, Karen Orchansky, Bryn Paslawski, Olga Peshko, Quinton Robbins, Ricardo Rogriguez, Xiaofeng Shi, Carol Shou, Alex Swift, Holly Toczko, Tom Verderame, Sam Waters, Sophie Wilkowske, John Yang, Geoffrey Zee, Fred Zhu, and Jon Zytnick for their meticulous help in assembling data and other research assistance. We thank Daniel Chen, Tom Clark, John Ferejohn, Sanford Gordon, Chris, Hanretty, Jon Kastellec, Lewis Kornhauser, Eric Posner and the participants at Princeton University Conference on Bureaucrats, SIOE meetings at Harvard Law School, NYU Law and Economics Workshop, and Conference on Empirical Legal Studies in Europe for helpful comments. Columbia University's Program for Economic Research, Columbia Law School, and the National Science Foundation Grant SES-1260875 provided financial support for this research.

1 Introduction

The decisions of public officials have a large impact upon our daily lives, yet they typically face weak incentives for good performance. In the case of judges, weak incentives are a design feature justified by the need to have them issue unbiased decisions in the public interest.[1] This is particularly important for appellate courts in common-law systems, such as the United States, where judicial decisions have the power of law. Moreover, there is now a significant literature suggesting that effective courts and judges are important ingredients for economic growth (La Porta et al. (1998), Glaeser et al. (2001), and Djankov et al. (2003)).[2]

In this paper we focus upon the methods by which judges are selected and retained, and ask how changes in these methods affect the quality of judicial decisions. The U.S. states have experimented with a variety of procedures, including partisan elections (judges affiliated with a political party), non-partisan elections (competitive, but no party affiliation on the ballot), and the "Missouri Plan" (merit-based selection by expert commission, with uncontested retention elections). Using this variation, matched with a database containing all the decisions in state supreme courts from 1947 to 1994, we measure the effect of changes in electoral institutions on the performance of state supreme court judges.

In the absence of explicit pecuniary rewards, what motivates good performance for public officials? One possibility is the intrinsic reward from doing a good job, which may arise when organizations have a clear "mission" that can motivate its members (Wilson, 1989). Dewatripont et al. (1999) study the design of rewards in the presence of mission, and in particular highlight the role of "professionalism" in encouraging task-specific skills, which in turn increase the professional's intrinsic incentive for work (see White (1959) and Wilensky (1964)). Francois (2000) uses this idea to build a theory of non-profit firms.

Maskin and Tirole (2004) introduce a theoretical model of policy decision-making by public officials, showing that the optimal choice of institution (removable "politician" or unaccountable "judge") depends upon the extent to which the preferences of citizens are important. Election pressure can cause officials to modify their decisions to reflect the interests of the electorate. Alesina and Tabellini (2007, 2008) extend this analysis to consider

[1] See for example Epstein et al. (2013). The optimality of this policy follows directly from the Holmström and Milgrom (1991) multi-tasking model when there are no good measures of decision quality.

[2] The state court websites provide up-to-date information on how appellate courts work in each state. See `https://www.ohiobar.org/forpublic/resources/lawyoucanuse/pages/lawyoucanuse-448.aspx` for a concise explanation. See Landes and Posner (1980) on the role of judges in common law systems. The work of has been influential in documenting how variation in legal systems, including the effectiveness of judicial decision making, can affect economic performance. Klerman and Mahoney (2005) provide additional historical evidence linking increases in judicial independence with increased economic performance.

the nature of the task at hand. Tenured "bureaucrats" are preferred for technical tasks, where organizations can evolve a mission to carry out these tasks well. In contrast, elected "politicians" are more sensitive to the preferences of citizens, and hence to the distributional concerns of the median voter. The choice of one or the other system is a practical question that depends upon the empirical magnitude of these factors.

A first-order question is, simply, whether or not there is a difference in policy choices between elected and appointed officials. For example, Besley and Coate (2003) show that electricity prices are lower in states with elected (versus appointed) utility regulators, consistent with elected regulators being more responsive to consumers. In the case of judges, there is now a significant body of evidence showing that stronger electoral pressures are associated with harsher criminal sentences (Huber and Gordon, 2004; Gordon and Huber, 2007; Lim, 2013; Berdejo and Yuchtman, 2013; Iaryczower et al., 2013; Canes-Wrone et al., 2014; Park, 2014).

More generally, a large literature documents significant impacts of political motivations on judge decision-making. Epstein et al. (2013) find that decisions by federal judges (although they are appointed and have tenure) tend to reflect the ideological leanings of the president that appointed them. Shepherd (2009) and Canes-Wrone et al. (2010) find that retention elections cause state appellate judges to vary their decisions in politically sensitive cases. Besley and Payne (2013) find that elected judges are more likely to support anti-discrimination law. This evidence broadly supports the view that selection and retention procedures have an impact on the policy choices made by judges.

A separate question is what evidence we have on the preferences of judges, and whether or not, as in Francois (2000) or Alesina and Tabellini (2007), they have a "mission" that motivates them to provide high work quality. An example of this type of evidence is Dal Bo et al. (2013), who find in a field experiment that higher wage offers for low-skill public-sector jobs results in a more self-motivated workforce. In this paper, we examine the trade off between policy choices and mission. Does the choice between a "politician" and "bureaucrat" entail a tradeoff between responsiveness to the electorate and the execution of a mission to make high quality legal decisions?

Providing evidence on this question requires first that we measure judicial performance. The job of a appellate court judge is to clarify the law, so well-crafted decisions will guide and constrain future judges.[3] That will be reflected in the number of citations to a decision.

[3]See Gennaioli and Shleifer (2007), Gennaioli and Shleifer (2008), and Baker and Mezzetti (2012) for formal models of the common-law process.

Therefore Choi et al. (2010) suggest using citations to measure judge work quality. Using data for 1998 through 2000, Choi et al. (2010) find no evidence of a cross-sectional relationship between judicial quality and the way state appellate judges are selected (partisan election, non-partisan election, merit commission, and governor appointment).

The data used for this paper come from Ash and MacLeod (2015), who introduce a dataset constructed from the written opinions of all state supreme court judges for the years 1947 through 1994. The major task of judging (writing judicial opinions) did not change much over this period, and we have a complete record of judge work output – the text of the decisions that they have authored. In Ash and MacLeod (2015) we use exogenous variation in time pressure to measure its effect upon decision quality. The results suggest that when judges have more time, they write longer, more influential decisions, consistent with the hypothesis that they have a "mission" for high-quality work.

This paper extends the analysis in a number of directions. First, we explore the extent to which decision quality varies by judge, rather than by non-judge factors such as caseload or type of case.[4] To that end we refine the performance measures used in Choi et al. (2010) and Ash and MacLeod (2015). The volume of text written presents a reasonable measure of work output, while the number of endorsing citations by later judges presents a reasonable measure of work quality. For both output and quality, we condition on features of the case portfolio that may affect length/citations but are outside the authoring judge's control. We show that there are systematic differences in performance between judges that allow a consistent ranking over a number of years.

Using these measures we then estimate the incentive and selection effects of changes in the procedures for appointing and retaining judges. We focus on three major systems: partisan elections, where judges are explicitly affiliated with a political party on the ballot; non-partisan systems, where there is a vote but party affiliation is not listed; and the "Missouri Plan," a merit-based system where judges are initially appointed by a commission of experts and then face uncontested retention elections. We observe changes in a judge's performance as an election becomes more imminent, as well as in response to changes in the appointment system. We measure the effect upon performance of changing from partisan elections to the merit system, and changing from non-partisan elections to the merit system.

[4]Judges do not write decisions alone; they manage law clerks who are responsible for doing research and in some cases writing initial drafts of opinions. Hence, if quality varies consistently as a function of the judge, part of this variation is due to their managerial skills. This is related to the active literature showing that managers make a difference (see Bloom and Van Reenen 2010). We also know from Coviello et al's (2014) work on Italian courts that the way judges organize their workload has an impact upon their performance.

We find that being up for election is associated with a reduction in judge output under non-partisan elections. There is no election-year effect on performance under partisan elections or under uncontested elections. These contrasting results make sense in light of the evidence in Lim and Snyder (2015), who have data on bar association evaluations of judges up for election. They find that these evaluations have no effect in partisan elections, since voters mainly follow party preferences. However, in non-partisan-elections, these evaluations do have an effect on voter choices. This helps explain why judges in a non-partisan system spend time campaigning: to affect voter choices.

Next, we find that judges appointed under a partisan system do not change output or work quality when the re-appointment process is changed to uncontested elections. However, we find that when a judge begins service under a non-partisan system that is changed to uncontested re-appointment, quality rises. Again, this contrasting result is consistent with Lim and Snyder's (2015) hypothesis that partisan elections place small re-election pressures upon judges; therefore, moving to an uncontested system does not make much of a difference. But non-partisan elections impose strong election pressures, so moving to uncontested elections gives the judges more time to spend on judging (rather than campaigning).

From the election-year results, we already know that there is a reduction in output in the period before an election under non-partisan systems, thus mechanically the movement to an uncontested system might be expected to increase output. The fact that we do not observe an increase in output suggests that judges are using the reduction in the time constraint to increase quality instead. This is consistent with Ash and MacLeod (2015), who find that reducing time pressure increases opinion quality, which in turn implies that judges have an intrinsic mission to produce high-quality decisions.

Finally, we explore how changes in appointment systems affect the quality of selected judges. We do this by comparing the performance of judges on the same court at the same time, but selected under different procedures, with controls for time and age effects. We have evidence that a change in the appointment system from electoral selection to merit selection has no effect upon output, but a positive effect upon opinion quality. This evidence is consistent with the intuitive presumption that state governments decide to change their judge selection system in order to enhance performance. These governments have chosen more "bureaucratic" appointment systems for judges with the (correct) expectation of an increase in the importance of mission and reduction of the bias and distractions of electoral politics.

The rest of the paper is organized as follows. Section 2 provides an institutional back-

ground on state supreme court selection and retention. Section 3 introduces a model of the selection and incentive effects of judicial elections. Section 4 discusses the issue of measuring judge performance. Sections 5, 6, and 7 report respectively the results on election year politics, the effect of the retention system, and the effect of the appointment system. Section 8 provides a concluding discussion.

2 Background

This section provides relevant background for the theoretical and empirical analysis. First, Subsection 2.1 describes the electoral institutions that provide our treatment variation. Subsection 2.2 provides an overview of our data sources.

2.1 Institutions

Our institutional setting is the set of state supreme courts, also known as state courts of last resort. As described in greater detail in Ash and MacLeod (2015), these courts serve as the state judiciary's analogue to the U.S. Supreme Court. These courts have authority to review laws produced by state legislatures and decisions produced by lower state courts, thus state supreme court judges are some of the most powerful officials in state government.

As with other appellate courts, a judge writes an opinion explaining the court's decision. The job of a supreme court judge does not change much over the course of the career, and it does not vary across states.

While the work tasks are the same, the rules for selecting and retaining appellate judges vary across states and over time. These rules are listed in Table 1, with rule changes indicated by cell borders. These changes are used in our empirical section to identify the incentive and selection effects of changing electoral systems.

The first system, partisan elections, is used for both selection of new judges and retention of incumbent judges. For these elections, judges are members of a political party, Republican or Democrat. They must win a primary election for their party before running in a general election, where their political affiliation is labeled on the ballot. Incumbent judges rarely face a credible challenge in the primary, but in the general election they usually face a challenger from the opposing political party.

The second system, non-partisan elections, are also used for both selection and retention. In this system there are competitive elections, but there are no primaries and party affiliations

Table 1: Judicial Selection and Retention Systems

State (Years)	Selection	Retention	State (Years)	Selection	Retention
Alaska	Merit	Uncontested	New Hampshire	Governor	Tenure
Alabama	Partisan	Partisan	North Carolina	Partisan	Partisan
Arkansas	Partisan	Partisan	North Dakota	Non-Partisan	Non-Partisan
Arizona (-1974)	Non-Partisan	Non-Partisan	Nebraska (-1962)	Partisan	Partisan
Arizona (1975-)	Merit	Uncontested	Nebraska (1963-)	Merit	Uncontested
California	Governor	Uncontested	New Jersey	Governor	Tenure
Colorado (-1966)	Partisan	Partisan	New Mexico (-1988)	Partisan	Partisan
Colorado (1967-)	Merit	Uncontested	New Mexico (1989-)	Partisan	Uncontested
Connecticut	Governor	Governor	Nevada	Non-Partisan	Non-Partisan
Delaware	Governor	Governor	New York (-1976)	Partisan	Partisan
Florida (-1971)	Partisan	Partisan	New York (1977-)	Governor	Governor
Florida (1972-1976)	Non-Partisan	Non-Partisan	Ohio	Partisan	Non-Partisan
Florida (1977-)	Merit	Uncontested	Oklahoma (-1967)	Partisan	Partisan
Georgia (-1984)	Partisan	Partisan	Oklahoma (1968-)	Merit	Uncontested
Georgia (1985-)	Non-Partisan	Non-Partisan	Oregon	Non-Partisan	Non-Partisan
Iowa (-1962)	Partisan	Partisan	Pennsylvania (-1968)	Partisan	No Retention
Iowa (1963-)	Merit	Uncontested	Pennsylvania (1969-)	Partisan	Uncontested
Idaho	Non-Partisan	Non-Partisan	Rhode Island	Governor	Tenure
Illinois (-1964)	Partisan	Partisan	South Carolina	Legislature	Legislature
Illinois (1965-)	Partisan	Uncontested	South Dakota (-1980)	Non-Partisan	Non-Partisan
Indiana (-1970)	Partisan	Partisan	South Dakota (1981-)	Merit	Uncontested
Indiana (1971-)	Merit	Uncontested	Tennessee (-1971)	Partisan	Partisan
Kansas (-1958)	Partisan	Partisan	Tennessee (1972-1977)	Merit	Uncontested
Kansas (1959-)	Merit	Uncontested	Tennessee (1978-)	Partisan	Partisan
Kentucky (-1975)	Partisan	Partisan	Texas	Partisan	Partisan
Kentucky (1976-)	Non-Partisan	Non-Partisan	Utah (-1951)	Partisan	Partisan
Louisiana	Partisan	Partisan	Utah (1952-1985)	Non-Partisan	Non-Partisan
Maine	Governor	Governor	Utah (1986-)	Merit	Uncontested
Maryland (-1976)	Non-Partisan	Non-Partisan	Vermont (-1971)	Legislature	Legislature
Maryland (1977-)	Merit	Uncontested	Vermont (1972-)	Governor	Legislature
Massachusetts	Governor	Tenure	Virginia	Legislature	Legislature
Michigan	Partisan	Non-Partisan	Washington	Non-Partisan	Non-Partisan
Minnesota	Non-Partisan	Non-Partisan	Wisconsin	Non-Partisan	Non-Partisan
Missouri	Merit	Uncontested	West Virginia	Partisan	Partisan
Mississippi	Partisan	Partisan	Wyoming (-1972)	Non-Partisan	Non-Partisan
Montana	Non-Partisan	Non-Partisan	Wyoming (1973-)	Merit	Uncontested

Notes. This table lists the elections systems for state supreme court judges observed in our data. Election-system reforms indicated by cell borders.

are not on the ballot. There are generally two candidates, an incumbent and a challenger, but the incumbent is not identified as such.

The third major system is merit selection with uncontested retention elections, also known as the Missouri Plan. In this system, judges are nominated by a commission of experts – senior attorneys and retired judges – and confirmed by the governor. Incumbent judges face an up-or-down retention vote with no challenger. This system is designed to be more meritocratic, and to impose weaker political incentives, than electoral selection.

In a fourth hybrid system, judges are initially selected through partisan elections but thereafter face uncontested retention elections. California has governor appointment but uncontested retention elections. The other states either have some combination of governor or legislative appointment, both for initial selection and for period retention. In Massachusetts, New Hampshire, New Jersey, and Rhode Island, judges have lifelong tenure.

In Ohio and Michigan, judicial elections are difficult to classify within the partisan/non-partisan dichotomy because they have partisan primaries and nomination processes, but the political party is not on the ballot in general elections. Following Nelson et al. (2013), we classify these states as partisan selection and non-partisan retention. Alternative coding, or leaving them out of the analysis, does not change our results.

Hanssen (2004) provides evidence and discussion of the political motivations behind reforms to these regimes. He shows that the systems tended to become less politicized during times of political transition in the state's electorate. For example, in the South, when states began transitioning from majority Democrat to majority Republican, state governments tended to enact laws giving judges stronger tenure. This suggests that the reforms were not a response to the quality of the judges, but rather to external political factors outside the judges' control.

2.2 Data Overview

The data-set used for the empirical analysis is an extension of that used in Ash and MacLeod (2015). It merges information on judge biographies, state-level court institutions, and published judicial opinions. These data allow panel estimates on the effects of court institutions on judge performance.

There are 1,628 state supreme court judges in our data. Table 2 reports summary statistics on the characteristics of judges working in one of the three selection systems discussed in Section 2.1. For many of the variables, the systems are comparable. Relative to the partisan judges, the non-partisan and merit judges are more likely to be female. Merit judges are

Table 2: Summary Statistics on Judge Characteristics by Selection System

	Partisan Elections		Non-Partisan Elections		Merit Selection	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Background						
Start Age	53.6969	8.8354	52.8235	8.3454	52.1143	7.8792
Female	0.0305	0.1721	0.0663	0.2491	0.0616	0.2410
Top School	0.0973	0.2966	0.1040	0.3058	0.1081	0.3114
Previous Experience						
Private Practice	0.6862	0.4644	0.8141	0.3897	0.6645	0.4737
Judiciary	0.6082	0.4886	0.5630	0.4969	0.6818	0.4673
Politics	0.2818	0.4503	0.2269	0.4197	0.1118	0.3162
Academia	0.0879	0.2834	0.1076	0.3105	0.1060	0.3088
Partisan Affiliation						
Republican	0.5517	0.4675				
Democrat	0.4483	0.4412				
Career Length	11.8401	8.6620	13.6512	9.6248	13.2120	7.0797
How Ended						
Retired	0.3293	0.4703	0.3896	0.4883	0.3825	0.4871
Resigned	0.1111	0.3145	0.2098	0.4077	0.0553	0.2291
Died in Office	0.1070	0.3094	0.1144	0.3188	0.0553	0.2291
Lost Election	0.0650	0.2468	0.0409	0.1983	0.0138	0.1170
Impeached	0.0054	0.0735	0.0000	0.0000	0.0046	0.0679
Judges	738		367		217	

Notes. Biographical information by judge election system. Observation is a judge. *Start Age* is judge age upon joining the court. *Female* is a dummy for being female. *Top School* means the judge attended law school at Yale, Harvard, Columbia, Stanford, or Chicago. The *Previous Experience* items equal one if the judge has previous experience in the respective area. *Republican* is a dummy for being Republican, *Democrat* for being Democrat. *Career Length* is number of years working on the court, conditional on having left the court before 2014. The *How Ended* items equal one if the judge's state supreme court judgeship ended for this reason.

the most likely to have judicial experience, while partisan judges are the most likely to have political experience. Non-partisan and merit judges have longer career lengths. Merit judges are the least likely to lose re-election.

Our performance measures were constructed from published state supreme court opinions for the years 1947 through 1994, obtained (along with some annotated meta-data) from bloomberglaw.com. The full sample includes 1,024,261 cases. We drop opinions that do not have a named author (per curium decisions), resulting in a sample of 404,928 majority opinions. This is an average 47.2 cases per judge per biennium.

3 Mechanisms

In practice, democratic institutions are designed to achieve a variety of goals (see for example Besley and Coate 1997). As we discussed in the introduction, there is a rich literature on how appointment and retention systems shape the preferences of judges. Our goal here is a more narrow "labor economics" question, namely, how do these systems affect work incentives and the quality of selected judges? The purpose of this section is to build a model that makes some predictions that can be answered with the available data, and provide an organizing framework to discuss the empirical evidence. In particular we would like to focus on the role of voter information on outcomes. This focus is particularly interesting in the context of judicial elections because, as Lim and Snyder (2015) observe, it is not clear that voters have a great deal of information about judges.

The model is based on Condorcet's Condorcet (1785) jury theorem that views voting as an information revelation problem (see also Young, 1988). The model provides a simple framework that is sufficiently rich to capture the effect of election rules upon the selection of and incentives for judges.[5] It is assumed that each voter has a noisy measure of judge quality that is used to make their decision between candidates. In addition, they care about the political views of judges, the strength of which is measured by the bias in favor of judges from their preferred party.

More precisely, suppose that there is an opening for a judge for which there are two candidates, A and B. One of these could be an incumbent, but we abstract from this and suppose that each judge j has a quality level q_j drawn from a normal distribution: $q_j \sim N(0,1), j \in \{A, B\}$. It is assumed that these draws are uncorrelated, though different

[5]See Ashworth and de Mesquita (2008) and Ashworth et al. (2015) for more sophisticated versions of this class of models.

jurisdictions may have different distributions. The socially desirable outcome is to choose the most able judge, though a judge's political views may bias this decision.

The remaining subsections analyze how differences in information on judge candidates may influence the expected quality q_j of the judge selected, as well as the judge's performance once he is in office. Subsection 3.1 introduces a merit selection baseline where the better judge is always selected. Subsection 3.2 considers the consequences of electoral selection, where voters do not have perfect information, and may be biased by politics. Subsection 3.3 looks at the effects on an incumbent judge of electoral campaign demands.

3.1 Merit Selection and Governor Appointment

The salient feature of merit selection is a committee that looks carefully at each potential candidate. For simplicity suppose that q_j is observable to members of the commission.[6] The merit commission is assumed to be able to communicate its finding clearly to the governor, who in turn will select the more able candidate. Thus, the expected quality of a judge under a merit system is the first order statistic:

$$\bar{q}^M = \mathbb{E}\left\{\max\left\{q_A, q_B\right\}\right\} = \frac{1}{\sqrt{\pi}} > 0.$$

If the expected ability of a randomly chosen candidate is 0, then selecting the better one from a pool of two judges results in positive expected quality. Increasing the size of the pool would simply increase the expected quality of the appointed judge; it is the same logic as Condorcet's (1785) jury theorem.

We can compare this to an appointment system where political bias enters. As a matter of convention we suppose that the governor (and later the representative voter) prefers Judge A. We can model this as a bias b and suppose that Judge A is chosen if and only if:

$$q_A + b \geq q_B. \tag{3.1}$$

Let $I\left(q_A, q_B, b\right) = 1$ if (3.1) holds and zero otherwise. Let

$$\bar{q}^G\left(b\right) = \mathbb{E}\left\{q_A I\left(q_A, q_B, b\right) + \left(1 - I\left(q_A, q_B, b\right)\right) q_B\right\}. \tag{3.2}$$

In the appendix we show:

[6] This assumption can be relaxed and does not substantially affect the results.

Proposition 1. *The average quality of judges chosen under an unbiased merit panel is higher than that under governor appointment with bias:* $\bar{q}^M = \bar{q}^G(0) > \bar{q}^G(b), b \neq 0$. *The difference in quality rises as the level of political bias increases (*$\bar{q}^G(b)$ *is strictly decreasing in b).*

This rather intuitive result illustrates the cost associated with bias. In the absence of any bias the best candidate is chosen. However, preference for one or the other candidate can lead to the less able individual being chosen in some cases.

3.2 Selection of Judges by Election

Next we consider the effect on quality of selecting judges by election using a representative voter. This is modeled by supposing that the quality of information held by the electorate is lower than that of the merit panel. Suppose that the representative voter gets a signal of judge $j's$ quality:

$$s_j = q_j + \epsilon_j$$

where ϵ_j is normally distributed with mean zero and variance σ_j^2. The precision is defined by $\rho_j = 1/\sigma_j^2$. The representative voter observes the two signals and then assesses the relative quality of the judges.

We distinguish partisan and non-partisan electoral systems by introducing bias b. As a matter of convention suppose that judge A comes from the same party as the representative voter, where b represents the voter's utility weight on partisan affiliation. In a non-partisan system $b = 0$, while a partisan system is characterized by $b > 0$.

After observing s_j, the voter's posterior distribution on q_j is normal with mean

$$\mathbb{E}\{q_j|s_j\} = \pi_j s_j$$

and precision $1 + \rho_j$, where $\pi_j = \frac{\rho_j}{1+\rho_j}$ is the weight assigned to s_j.The representative voter selects Judge A if and only if

$$\pi_A s_A + b \geq \pi_B s_B.$$

As the bias in favor of a judge from the same party increases, the probability that Judge A is selected increases. This can be understood as reducing the competitiveness of the election. The expected quality of a judge selected under an electoral system with bias b is defined by:

$$\bar{q}^E(b) = \mathbb{E}\{q_A I(\pi_A s_A, \pi_B s_B, b) + q_B(1 - I(\pi_A s_A, \pi_B s_B, b))\}. \tag{3.3}$$

In the appendix we show:

Proposition 2. *When voters do not perfectly observe judge quality, the average quality of elected judges is lower than that of merit-selected judges:*

$$q^M \geq \bar{q}^G(b) > \bar{q}^E(b).$$

Average judge quality falls with the strength of political bias, and therefore quality with partisan elections is lower than that with non-partisan elections: $\bar{q}^E(b)$ falls with b.

As in the previous case, bias reduces the effectiveness of the electoral system. Bias does not mean that the better candidate cannot win, only that the standard is higher. It is an empirical question as to whether or not the bias in the system provides an insurmountable barrier for more qualified candidates. As we previewed in the introduction, we find that on average merit systems do indeed select better judges than partisan systems.

3.3 Campaign Incentives

We now extend the model to explore the effect of electoral incentives upon judicial performance. Suppose that when a judge allocates time to campaigning this enhances the quality of the signal observed by voters. We then proceed by assuming that judges have a normal level of effort for their work, given by $\bar{y}_A$ and $\bar{y}_B$ for the incumbent A and the challenger B, respectively. In an election year the judges may divert some of this effort to election-year politics. While B is a challenger and is not sitting on the court, for simplicity we assume he faces the same decision problem as the incumbent A. This approximates the situation where B is a judge on another court – the state's intermediate appellate court, for example.

Thus in an election year it is assumed that the individuals supply y_A and y_B to their jobs, resulting in election year effort:

$$e_A = \bar{y}_A - y_A \geq 0,$$

$$e_B = \bar{y}_B - y_B \geq 0.$$

The consequence is that the representative voter chooses judge A over judge B if and only if

$$\pi_A(s_A + e_A) + b \geq \pi_B(s_B + e_B).$$

The probability of A winning is:

$$p_A\left(e_A, e_B | q_A, q_B\right) = \mathbb{E}\left\{I\left(\pi_A(s_A + e_A), \pi_B(s_B + e_B), b\right) | q_A, q_B\right\}.$$

Correspondingly, define $p_B\left(e_A, e_B | q_A, q_B\right) = 1 - p_A\left(e_A, e_B | q_A, q_B\right)$.

We suppose that candidate j has preferences:

$$U_j = Bp_j\left(e_A, e_B | q_A, q_B\right) - C\left(e_j\right),$$

where B is the intrinsic value from winning the election and $C_j(e_j) = C_j\left(\bar{y}_j - y_j\right)$ is the utility cost of campaign effort. The campaigning cost $C\left(e\right)$ is assumed to be twice differentiable in e and satisfies $C_j\left(0\right) = C_j'\left(0\right) = 0, C_j'' > 0$. This guarantees an interior solution.

In our data we can observe the output of sitting judges, and hence both $\bar{y}_A$, the output before an election year, and y_A, the output in an election year, are observable. Consider first an uncontested elections (the "Missouri Plan"), in which judges do not face a challenger. This can be understood in the model notation as $e_B = 0$; the challenger sets zero campaign effort. The incumbent judge A sets e_A accordingly.[7]

Next, we consider the equilibrium when there is an active challenger (details in the appendix). If we suppose that the problem is symmetric, then the judges have the same quality precision: $\rho_A = \rho_B$ and have $e_A = e_B$. The first-order conditions for effort in this case are given by:

$$C_j'\left(e_j\right) = \sqrt{\frac{\rho}{2}}\phi\left(\sqrt{\frac{\rho}{2}}\left(q_A - q_B\right) + b\left(\frac{1+\rho}{\sqrt{2\rho}}\right)\right). \tag{3.4}$$

where $\phi(\cdot)$ is the standard normal pdf. Since $\phi\left(x\right)$ achieves its maximum value at $x = 0$, we see that effort is highest when:

$$\left(q_A - q_B\right) + b\left(\frac{1+\rho}{\rho}\right) = 0. \tag{3.5}$$

These observations can be summarized as follows:

Proposition 3. *When voters have the same quality of information regarding candidates, the candidates choose the same level of campaign effort. Moreover, the amount of effort is highest in the most competitive races - when (3.5) is small. In particular, campaign effort*

[7]The only caveat is for judges who feel they may not get re-elected for whatever reason (for example, bad press from a high-profile case). Thus, there may be some judges who do exert effort, in which case e_A may be positive. There is never any reason to observe a negative effort level.

decreases with the bias b. This means that campaigns reduce judging effort more (put more effort into campaigning) under non-partisan elections than under partisan elections.

In the appendix we prove that an equilibrium to the campaign effort game exists and that the effort of Judge A is greater than candidate B if and only if the electorate has a better measure of Judge A's quality.

This proposition has the following implications in our data. First, uncontested elections are the least competitive and have the weakest electoral incentives. Among the electoral systems, they should have a smaller effect on judging effort than partisan elections or non-partisan elections. Second, if non-partisan elections have less bias then they are more competitive than partisan elections. Therefore non-partisan elections should have a larger negative effect on judging effort than partisan elections.

4 Measuring Judge Performance

In this section we discuss the problem of measuring judicial performance. We focus on two simple metrics for judge performance, output and quality. Previous researchers, such as Epstein et al. (2013) and Choi et al. (2010), use total number of opinions and case citations as measures of output and quality, respectively. This approach faces the challenge that the difficulty and importance of a ruling varies from case to case, for reasons outside a judge's control. This in turn implies that simple citation and output measures are likely to be noisy. This problem is exacerbated when trying to compare judges across different states (as done in Choi et al. (2010)), since not only lower-court characteristics but also a range of court and state factors might affect the number of citations to a decision.

We find that using simple output and quality measures results in noisy measures that do not clearly and persistently distinguish the performance between judges. One reason for this is that a "judge" is not really a single individual, but a team of individuals, that include clerks and secretarial staff. Judges do select the clerks that are working for them, and hence our measures can be seen as composites that depend upon both the judge's legal skill when researching, reasoning, and writing, as well as managerial skill when selecting and directing clerks. As we know from Bloom et al. (2012), management quality varies across firms, and there are systematic relationships between management quality and firm performance.

In our data we cannot directly disentangle managerial skill from legal skill. However, we can ask if there is a judge fixed effect – namely, are there some judges who are consistently better than others? If we can measure these differences, then we can ask whether the choice of

selection system affects decision quality. In Ash and MacLeod (2015) we constructed yearly measures of performance for a range of variables. We found significant effects, but these measures are quite noisy and faced a multiple-comparison issue. In this section we discuss an improvement to those measures that increases the signal-to-noise ratio. This is achieved first by using two-year rather than one-year means. This helps by, for example, averaging more over the clerks chosen by a judge. Second, we control for case characteristics. Some cases, such as death penalty cases, simply get much more attention, and thus adjustment need to be made to tease out judge effects. We provide some graphical evidence that indeed our new measures are more successful in capturing variation in judge performance.

4.1 Conditioning on Case Characteristics

An important issue in using judicial decisions (and citations to those decisions) to study judge performance is that citations to an opinion are the product of non-judge case factors as well as the choices of the authoring judge. In particular, some types of cases are just more important than others. For example, cases that review the constitutionality of statutes are probably relatively important. In addition, judges have some discretion over the types of cases they are chosen to author opinions for. If we want to compare the quality of judges working on the same court at the same time, we need to try to account for these non-judge factors.

Assume that the number of citations to a decision i, Q_i, are determined by case factors C_i and judge factors q^j_{ist}. We are interested in conditioning on C_i. We use the full range of dummy variables for the area of law of a case, as well as the related industries of a case. These are coded for each case by Bloomberg staff attorneys, and there may be up to three legal areas and three related industrial sectors for any particular case. Summary tabulations for the most frequent legal areas and industrial sectors are reported in table 3. The case characteristics vector C_i includes a dummy variable for each area and sector, equaling one if case i has been assigned to that area or sector.

We estimate

$$Q_{ist} = \alpha_s + \alpha_t + C_i'\beta + \tilde{q}_{ist}$$

where we have also included state effects α_s and time effects α_t in addition to case characteristics C_i. Using the estimates $\hat{\alpha}$ and $\hat{\beta}$, we take the residuals vector

$$q^j_{ist} = Q_{ist} - \hat{\alpha}_s - \hat{\alpha}_t - C_i'\hat{\beta}.$$

Table 3: Summary Statistics on Area of Law and Related Industries

Area of Law	Freq.	Percent	Related Industrial Sector	Freq.	Percent
Criminal Law	191810	21.85	Real Estate	28527	13.64
Civil Procedure	74757	8.52	Law Enforcement	10758	5.14
Evidence	66377	7.56	Automobiles	10206	4.88
Torts	57915	6.6	Insurance	9158	4.38
Damages & Remedies	45073	5.14	Tax	8509	4.07
Contracts	40888	4.66	Construction & Engineering	6332	3.03
Real Property	36408	4.15	Workers' Compensation	5397	2.58
Constitutional Law	34038	3.88	Banking	4917	2.35
Family Law	32191	3.67	Legal & Compliance Services	4682	2.24
Workers' Compensation	22955	2.62	Automobile Insurance	4124	1.97
Insurance Law	19375	2.21	Property Management	4108	1.96
Administrative Law	18264	2.08	Transportation	3890	1.86
Wills, Trusts & Estates	18179	2.07	Child Welfare	3689	1.76
Tax & Accounting	16978	1.93	Employment Services	3679	1.76
Employment Law	14601	1.66	Health & Medical	3478	1.66
Habeas Corpus	13426	1.53	Oil & Gas	3189	1.52
Appellate Procedure	13140	1.5	Railroads	2777	1.33
Professional Responsibility	12052	1.37	Hospitals	2719	1.3
Motor Vehicles & Traffic Law	9644	1.1	Education	2586	1.24
Land Use Planning & Zoning	9122	1.04	Trucking	2097	1
Government	8942	1.02	Bridges & Roads	1751	0.84
Mortgages & Liens	7531	0.86	Agriculture & Farming	1729	0.83
Landlord & Tenant	5499	0.63	Mortgage Lending	1680	0.8
Construction Law	4997	0.57	Manufacturing	1612	0.77
Elections & Politics	4972	0.57	Real Estate Agents & Brokers	1573	0.75
Eminent Domain	4943	0.56	Unions	1485	0.71
Labor Law	4790	0.55	Financial Services	1469	0.7
Government Employees	4773	0.54	Judiciary	1448	0.69
Debtor Creditor	4260	0.49	Politics	1336	0.64
Employee Benefits	4208	0.48	Teachers	1300	0.62
Medical Malpractice	4113	0.47	Medical Procedures	1273	0.61
Personal Property	3994	0.46	Public Works	1223	0.58
Corporate Law	3958	0.45	Life Insurance & Annuities	1155	0.55
Negotiable Instruments	3843	0.44	Apartment Leasing	1127	0.54
Education Law	3803	0.43	Mining & Natural Resources	1115	0.53
Banking & Finance	3380	0.39	Drug Trafficking	1105	0.53
Alcohol & Beverage	3213	0.37	Sewer & Water	990	0.47
Civil Rights	3138	0.36	Electric	985	0.47
Health Law	2950	0.34	Water & Sewer	972	0.46
Transportation Law	2839	0.32	Physicians	966	0.46
Partnerships	2333	0.27	Firearms & Weapons	962	0.46
Natural Resources	2301	0.26	Motorcycles	919	0.44
Legal Malpractice	2285	0.26	Water	904	0.43
Products Liability	2280	0.26	Food & Beverage	888	0.42
Alternative Dispute Resolution	2144	0.24	Commercial Real Estate	883	0.42
Communications & Media	2048	0.23	Property & Casualty Insurance	854	0.41
Environmental Law	1857	0.21	Administration	837	0.4

This is a vector of case quality measures for each case i, purged of variation due to area of law, the industrial sector related to the case, the state of the decision, and time.

Similarly, for a case of length Y_{ist}, we construct a purged length measure as

$$y^j_{ist} = Y_{ist} - \hat{\alpha}_s - \hat{\alpha}_t - C_i'\hat{\beta}.$$

Next we construct judge-specific metrics of quality and output over time. For each two-year period (biennium) t, the quality of judge j is represented as

$$Q_{jst} = \frac{1}{|q_{ist}|} \sum_i q^j_{ist}$$

the average adjusted quality of his cases. Output is defined as

$$Y_{jst} = \sum_i y^j_{ist}$$

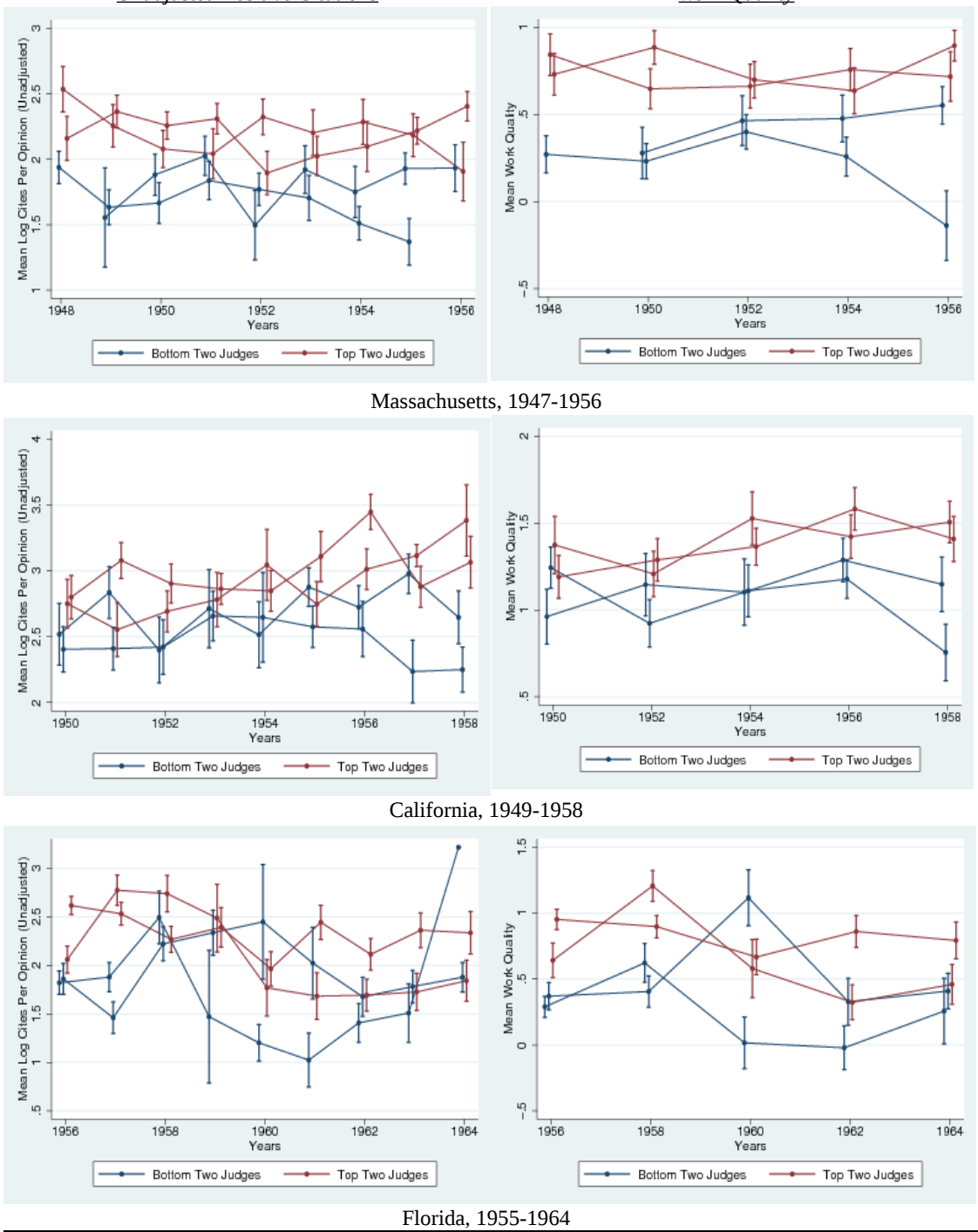
the summed adjusted length of his cases. Because we cannot easily interpret the units, in the regression tables we standardize the outcome variables to mean zero and standard deviation one.

4.2 Distinguishing Judge Quality Over Time

This section shows that our two-year-average measures of case-adjusted quality demonstrate durable differences across judges in the quality of their work product. Our measures are mechanically less noisy due to the the smoothing over two years, and the conditioning upon case characteristics. The issue this is whether or not they capture variations in judicial performance. We do this by looking at the top- and bottom-ranked judges by state for a selection of ten-year periods.

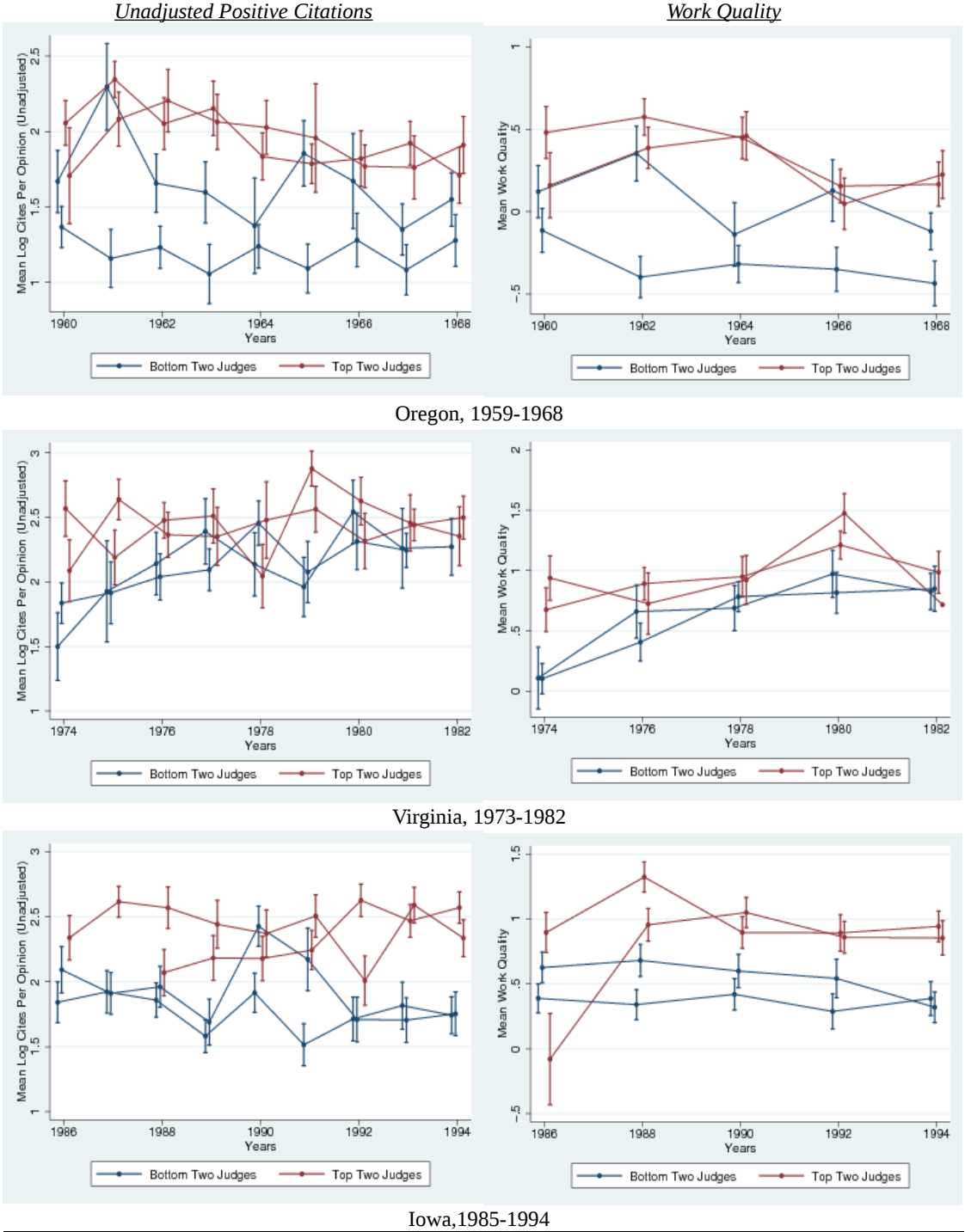
Figures 1 and 2 illustrates our quality measure (right-hand side), and its comparison with the raw number of citations (left-hand side), for the top and bottom two judges, in six selected states. The error spikes give the standard error of the mean for the set of cases in that period. On the right hand side for each figure, one can clearly see that the bottom two judges have consistently lower citation rates per opinion than the upper two judges. More precisely, this suggests that the lower citations of one judge relative to another is not a chance event, but one that is associated with durable judge characteristics. In the following sections, we explore the extent to which these characteristics are related to the appointment

Figure 1: Work Quality Distinctions Between Judges



Raw data for Opinion Quality (positive citations per opinion, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 2: Work Quality Distinction Between Judges (cont.)



Raw data for Opinion Quality (positive citations per opinion, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Table 4: Quality and Output as Predictors of Bar Association Evaluations of Quality

	Logit Estimate for Effect on "Good Judge" Designation		
	(1)	(2)	(3)
Output	0.154		-0.0771
	(1.046)		(1.100)
Quality		1.059**	1.076**
		(0.363)	(0.112)
State Fixed Effects	X	X	X
Year Fixed Effects	X	X	X

N= 51 judge-bienniums for set of judges in Pennsylvania, Texas, and Washington for the years 1987 through 1994. Outcome is an indicator for being a "good" judge has defined in Lim and Snyder (2015). Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01.

and retention system.

In the appendix we include similar figures for number of opinions and total output. In the case of number of lead opinions written, judges write almost identical numbers for each two-year period. This reflects that judges are generally assigned cases through docket rotation and have little control over the number of opinions written. This means that number of opinions is not a good performance measure. In the case of total output (total number of words written), the rankings, trends, and separation are similar to those reported in Figures 1 and 2 for citations.

4.3 Relation to Bar Association Evaluations of Quality

To further validate our outcome variables as judge-specific measures of performance, we explore the extent to which they are correlated with another performance measure previously used in the literature, the quality ratings issued by state bar associations. We were able to merge our data for a small number of judges with the data on evaluations provided by Lim and Snyder (2015). We then estimated a logit model with state and year fixed effects to see whether our quality and output measures are predictive of the bar association evaluations of a "good judge," as coded by the authors. Table 4 shows that quality, but not output, is a strong predictor of the bar association qualifications evaluation. An important caveat is that this difference is only identified off of three states where there were both good and not-good supreme court judges during the time overlap (1987 through 1994).

This result supports the hypothesis that our measures have a judge specific component. The extent to which this component varies over time is hypothesized to depend upon the appointment and retention system. The fact that output has no effect, implies that the citation measure we construct is positively correlated with how a state bar evaluates a judge. The extent to which voters are sensitive to this information is measured by looking at the effect that merit appointment systems have upon selection.

5 Effect of Being Up For Election

This section examines how judges change their behavior over time in response to the election cycle. Ash and MacLeod (2015) show that contested elections reduce judicial performance. We add to that analysis by distinguishing between partisan and non-partisan elections and using our new performance measures. In theory, if judges wish to be re-elected then they should put effort into election-year politics. This in turn would lead to a reduction in output on the court. The question is whether bias matters for the electoral effect, as implied by the model in this paper.

5.1 Empirical Strategy

The empirical strategy for examining the effects of electoral demands on judicial behavior is to exploit the staggered election cycle for identification of stronger electoral incentives. The election schedule is arbitrarily assigned by history, so it is reasonable to assume that the schedule is uncorrelated with other institutional or socioeconomic factors that might affect individual judge performance. For this analysis we used data provide by Kritzer (2011), supplemented by new data collection and checking by a team of research assistants.

The electoral cycle is represented in our regressions as a vector of dummy variables E_{jst}, which equals one if judge j in state s is up for election at biennium t. The vector includes separate indicators for partisan, non-partisan, and uncontested retention elections. The dummy variable is coded as a one regardless of whether the judge actually ran for election – this is intended to avoid endogeneity problems from the judge's choice whether to actually run.

One possible source of bias in this analysis comes from time-invariant characteristics of individual judges. Some judges may have higher or lower performance than others on average due to unobservable characteristics, and they may be up for election more often or less often for any number of reasons. To deal with this possibility, we include a full set of judge-specific

fixed effects. Therefore any estimated election coefficients are relative to a judge's personal average.

A second major source of bias comes from the time-varying changes in the court work environment which may be correlated with the electoral schedule. For example, there may be campaigning demands during election years on all judges – not just those up for election – if they are asked to assist fellow members of their political party. To deal with this possibility, we include a full set of state-year fixed effects. Therefore any estimated election coefficients are also relative to the court average in each year. This means they effectively compare judges sitting on the same court, working at the same time, but who are in different stages of the electoral cycle.

Our preferred specification is

$$y_{ist} = \text{JUDGE}_i + \text{STATE}_s \times \text{YEAR}_t + E'_{ist}\rho + \epsilon_{ist} \tag{5.1}$$

where JUDGE_i is a judge fixed effect, $\text{STATE}_s \times \text{TIME}_t$ is a state-year fixed effect for each s and biennium t, and E_{ist} includes the election-year treatments. Standard errors are clustered by judge, since the electoral treatment varies within judge over time (although our main results are statistically significant with clustering by state). Note that this gives the average output deviation for the two years before an election. The one-year effect is larger and more statistically significant.

5.2 Results

The coefficient estimates from Equation (5.1) are reported in Table 5. The top panel has results for output, while the bottom panel has results for quality.

The different columns experiment with different specifications for the fixed effects and controls. Looking first at the bottom panel, one can see that there are no effects of the electoral cycle on case quality. In the case of output, we still don't see effects for partisan and uncontested elections. However, across all the specifications, we see significant negative effects on output from non-partisan elections.

The negative output effect in Column 1, with only state fixed effects, means that when there are more judges up for election, that is associated with reduced court-wide output. Since the estimated coefficient is almost the same as the other columns, this suggests that the judge-specific negative effects of being up for election have an aggregate negative effect on the court as a whole (rather than being compensated by other judges who are not up for

Table 5: Effect of Being Up For Election

	Effect of Election Year				
	(1)	(2)	(3)	(4)	(5)
Output					
Non-Partisan Election Year	-0.0829*	-0.0868**	-0.118**	-0.0870**	-0.118*
	(0.0331)	(0.0283)	(0.0289)	(0.0279)	(0.0560)
Partisan Election Year	-0.0325	-0.0117	-0.0223	0.00344	-0.0223
	(0.0280)	(0.0231)	(0.0241)	(0.0241)	(0.0237)
Uncontested Election Year	0.0181	0.0368	0.0529	0.0761+	0.0529
	(0.0436)	(0.0392)	(0.0391)	(0.0390)	(0.0511)
Quality					
Non-Partisan Election Year	0.0416+	-0.0104	-0.0115	-0.0149	-0.0115
	(0.0213)	(0.0208)	(0.0200)	(0.0199)	(0.0226)
Partisan Election Year	-0.0376+	-0.0234	-0.0170	-0.0162	-0.0170
	(0.0218)	(0.0214)	(0.0193)	(0.0193)	(0.0214)
Uncontested Election Year	0.0405+	0.0258	0.00524	0.00325	0.00524
	(0.0241)	(0.0225)	(0.0205)	(0.0206)	(0.0131)
Year Fixed Effects	X	X			
State Fixed Effects	X				
Judge Fixed Effects		X	X	X	X
State-Year Fixed Effects			X	X	X
Experience Controls				X	
Clustering by State					X

N= 7,999 judge bienniums. Standard errors clustered by judge in parentheses (except column 5, where errors are clustered by state). + p < .1, * p < .05, ** p < .01. Treatment variable is a dummy equaling one for the bienniums where a judge is facing reelection.

election).

5.3 Discussion

The election-cycle results suggest that non-partisan elections reduce output, while partisan elections and uncontested elections have no effect. Since uncontested elections are non-competitive by design, the null effect here is unsurprising (see Proposition 3).[8]

The more interesting distinction is between partisan elections (no electoral effect) and non-partisan elections (negative electoral effect). These results are consistent with the idea from Lim and Snyder (2015) that non-partisan elections are more competitive than partisan elections and require that judges spend time campaigning. Since voters tend to follow party lines, partisan elections are not as competitive as non-partisan elections.

The method for controlling for case characteristics (described in Section 4.1) plays an important role in the electoral-cycle results. The introduction of case controls for area of law and related industries results in the negative effect of Partisan elections on election year output going to zero. In the case of non-partisan systems, the negative effect is robust and does not change with the introduction of these controls.

6 Effect of Judge Retention Process

This section reports the results on how changing the system for judge retention affects the performance of sitting judges. Subsection 3.3 discusses the model mechanism for the effects of retention elections on incumbent judge behavior. More competitive elections result in more campaigning, which will reduce effort spent on judging. We examine this issue using judge fixed effects and institutional reforms to the retention system.

6.1 Empirical Strategy

Identification comes from discrete changes in the rules for retaining state supreme court judges. Eight states moved from partisan retention to uncontested retention elections: Colorado, Illinois, Iowa, Indiana, Kansas, Nebraska, New Mexico, and Oklahoma. Six states

[8]Note that since the mid-1990s, third-party funding for negative advertising in Missouri Plan elections has increased significantly. Our results may not extend to more recent years (our panel ends in 1994). This is an important area for future research.

moved from non-partisan retention to uncontested retention: Arizona, Florida, Maryland, South Dakota, Utah, and Wyoming.[9]

The regression framework is a standard differences-in-differences approach based on Bertrand et al. (2004). To control for time-invariant judge characteristics that may be correlated with the retention system in various states, we include judge fixed effects. To control for national trends in performance, we include year fixed effects. To control for pre-existing state trends in performance that may be confounded with the reforms, we include state-specific linear trends.

As in Ash and MacLeod (2015), we measure effects in a ten-year window around the reforms. The regressions include an indicator equaling one for the baseline time window of ten years before and ten years after a change to the retention system. The treatment variable is a dummy for the ten years after the change. Thus, with the inclusion of the judge fixed effects, the estimates can be interpreted as the average difference in within-judge performance for the ten years after the policy change relative to the ten years before the policy change. In a handful of states, we shrank the time window if the reform occurred close to the beginning or end of the sample.[10] In the appendix we include a table using other time windows; our main result on nonpartisan judges and quality is a somewhat lagged effect that is statistically significant with an effect window of at least eight years.

Formally, we estimate

$$y_{ist} = \text{YEAR}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \bar{R}'_{st}\bar{\rho} + R'_{st}\rho + X'_{ist}\beta + \epsilon_{ist} \tag{6.1}$$

where YEAR_t is a fixed effect for the two-year period t, JUDGE_i is a judge fixed effect, and $\text{STATE}_s \times t$ is a state-level linear time trend for state s. The term $\bar{R}_{st}$ is a vector of indicators equaling one for the baseline time windows of ten years before and ten years after each of the retention reforms. R_{st} is a vector of treatment indicators for the ten years after each rule change (with ρ measuring the corresponding causal effects of interest). X_{ist} includes

[9]There was an additional reform undertaken by four statesL partisan retention elections to non-partisan retention elections. This includes Florida, Georgia, Kentucky, and Utah. Because Florida and Kentucky had other important contemporaneous reforms, and because Utah's reform happened at the very beginning of our sample, we do not include this treatment in our main analysis. For the other treatments, there were other types of judicial reforms occurring around the same time. See Appendix B.1 for more details and robustness checks.

[10]These reforms are mostly enacted by voters through ballot referendums administered in November and officially going into effect the subsequent January. In these cases the dummy variable would turn on in the year following the vote. In cases where the policy is effective in the first half of the year, it is coded as turning on in that year. Note that Florida changes from partisan to non-partisan and then to uncontested elections. Our results do not change substantially if Florida is left out of the analysis.

Table 6: Effect of Changing the Retention System on Incumbent Judge Performance

	Effect of Change in Retention Rule			
	(1)	(2)	(3)	(4)
Output				
Non-Partisan to Uncontested	0.0617	0.0700	0.0249	0.00538
	(0.116)	(0.130)	(0.128)	(0.101)
Partisan to Uncontested	0.0124	0.00443	-0.00615	0.0166
	(0.0674)	(0.0868)	(0.0906)	(0.0748)
Quality				
Non-Partisan to Uncontested	0.176+	0.178*	0.196*	0.193**
	(0.0881)	(0.0876)	(0.0814)	(0.0646)
Partisan to Uncontested	0.0123	-0.00190	-0.0199	-0.0386
	(0.0777)	(0.0787)	(0.0746)	(0.0740)
Year Fixed Effects	X	X	X	X
State Time Trends		X	X	X
Judge Fixed Effects	X	X	X	X
Experience Controls			X	X
State Policy Controls				X

N= 7,380 judge bienniums. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Experience controls include a full range of dummies for years of experience. State policy controls includes other appointment-process changes, mandatory-retirement changes, changes in the number of judges, and log state government expenditures on the judicial branch.

other state and judge controls, namely judge experience and variables for other court-related policies. Standard errors are clustered by state.

6.2 Results

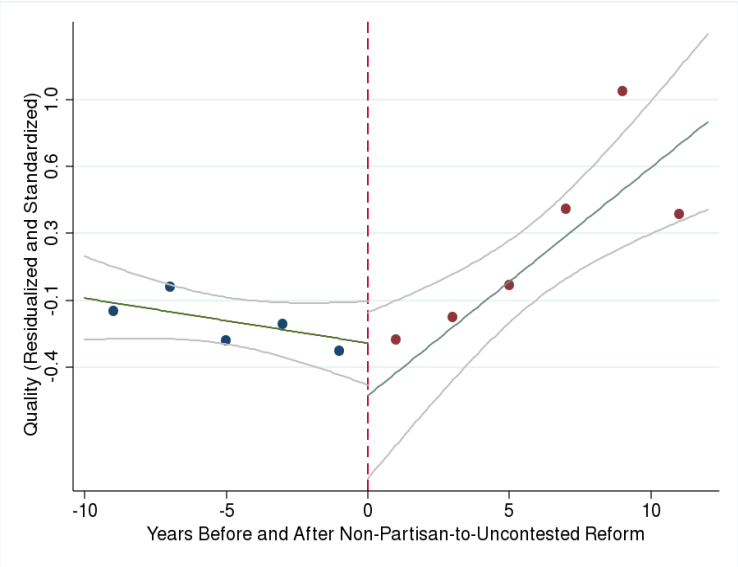
Table 6 reports our estimates for ρ from Equation 6.1. The top panel gives the effect on output, while the bottom panel gives the effect on quality. The four columns report different fixed effects specifications.

First, we see no effect of the rule changes on average judge output. However, we see a positive and statistically significant effect on judge work quality when moving from a non-partisan system to uncontested elections. This result is insensitive to the addition of state time trends, judge experience controls, and a range of state-level policy controls. There is no quality effect for partisan elections.

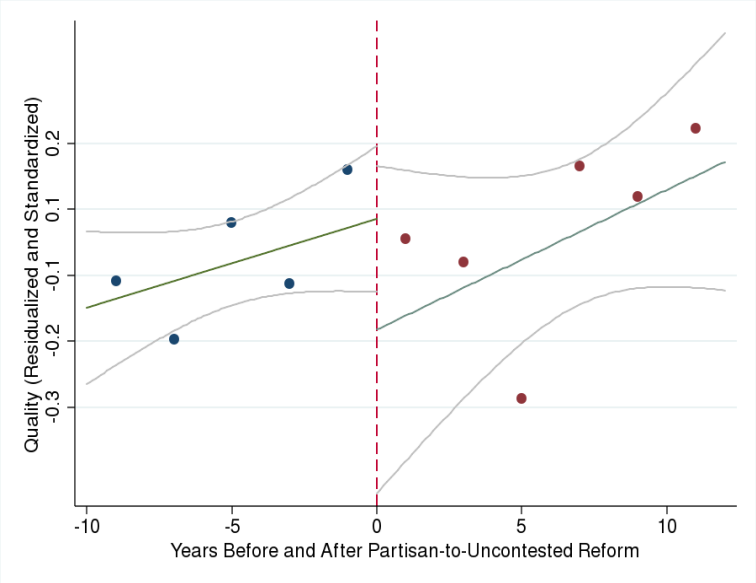
Figure 3 illustrates the dynamics of these effects in an event-study framework. These

Figure 3: Judge Work Quality Before and After Retention System Change

(a) Non-Partisan-to-Uncontested Reform



(b) Partisan-to-Uncontested Reform



N= 7,999 judge bienniums. Residualized judge work quality plotted by years before and after electoral reforms. Panel (a): Effect of moving from non-partisan to uncontested elections. Panel (b): Effect of moving from partisan to uncontested elections.

graphs plot the residualized work quality of incumbent judges for bienniums before and after the reform. They illustrate a positive change in work quality for non-partisan judges, although there is a six-year delay before the effect is observed. In the appendix we show that the effect is statistically significant with an effect window of at least 8 years. This delayed effect is likely due to the relatively long term lengths for supreme court judges. The minimum term length for these treated states is six years, the maximum is ten years, and the average is 7.8 years. Therefore one would expect a delay in the incentive effect until the next election, which could be as long as ten years.

6.3 Discussion

These results provide additional evidence that electoral incentives have an important impact on the quality of work produced by appellate court judges. In particular, the partisanship of elections matters. When taking non-partisan judges and giving them tenure, there is an increase in work quality. This is consistent with the idea from Ash and MacLeod (2015) that these judges have an intrinsic motivation to produce high-quality work, and that with weaker electoral incentives they will spend less time campaigning and more time writing good decisions.

On the other hand, when taking partisan judges and giving them tenure, there is no effect on output or quality. This is consistent with the idea from Lim and Snyder (2015) and from the model that partisan elections are less competitive than non-partisan elections due to political bias. Because there was little electoral pressure in the first place, eliminating those elections does not have a big impact on the way partisan judges spend their time.

7 Effect of the Selection Process on Judge Quality

In this section we investigate how changes to the procedure to select judges affects the quality of chosen judges. This analysis is motivated by Proposition 2. Selection mechanisms that use better information about candidates or have less bias should, all else equal, select better candidates on average.

A priori, there is no reason to suppose that a judge chosen by the Missouri Plan faces less bias than in, say, a non-partisan election. However, the intent of using a merit commission is to create a pool of better qualified judges. Similarly, political parties have an incentive to choose qualified judges that are consistent with the party's views. Hence, it is an empirical question whether or not the judges chosen by the Missouri Plan or by a partisan election

system are of higher or lower quality that those selected under a non-partisan system. What the theory illustrates is that the presence of bias reduces quality, while more precise signals increase quality.

7.1 Empirical Strategy

This subsection describes the empirical strategy for measuring the effects on judge quality of different judge selection systems. The source of identification is the set of reforms to the judicial selection systems, listed in Table 1. [11] Seven states moved from partisan selection to merit selection: Colorado, Iowa, Indiana, Kansas, Nebraska, and Oklahoma.[12] Five states moved from non-partisan selection to merit selection: Arizona, Maryland, South Dakota, Utah, and Wyoming.[13] The goal is to compare the performance of judges selected before these reforms to the performance of judges selected after these reforms.

We control for time-varying state-specific factors by including a full set of state-year (interacted) fixed effects. This specification effectively compares the performance of judges sitting on the same court at the same time, but selected under different regimes. We carry out some robustness checks to ensure that timing issues, such as the age of the judge, do not explain our results. We do this by including a full set of dummies for years of judge experience. This means that any estimates are made relative to other judges of the same experience level.

Second, the regressions include a full set of dummies for the judge's starting year. This set of controls complements the years of experience, with the goal of controlling for cohort-specific effects on performance. For example, judges beginning in the 1970s may be systematically better than judges beginning in the 1980s, due to changes in the economy. These indicators control for national variation in the market for judges as a function of time.

Third, the treatment indicators are active only for years where there are at least two judges selected from each system working on the court during that year. This is done to make a clean comparison that is not biased by outlier pre-reform judges who remain on the

[11] As mentioned previously, there was an additional reform where four states changed from partisan selection to non-partisan selection. Because of the small sample of states, we do not include this treatment in our main analysis. The coefficient on the output and quality of non-partisan judges relative to partisan judges is positive but not statistically different from zero ($p = .46$).

[12] Tennessee moved to merit selection in 1972, but moved back to partisan selection in 1978. Only one judge was selected by the merit process so it is not included in this analysis.

[13] Florida also moved from non-partisan to merit, but it is not included in this section because it had changed from partisan to non-partisan elections five years prior; therefore there were judges selected under all three systems in Florida and must be treated as a special case. Including Florida – with implicitly it being a part of both treatments – does not change the results.

bench long after the other pre-reform judges. Note that this two-judge-minimum results in Utah and Wyoming (non-partisan to merit reforms) being dropped from the baseline analysis, since these states did not have years in the data where at least two judges from each system. These two states are included in Column 5 of Table 7, which reports the estimates without the two-judge minimum. This specification uses all years with at least one judge per system, and the results are similar.

In the strictest specification, we include a large set of covariates for judge personal characteristics. These include a full range of dummies for judge starting age. We have indicators for judge gender, political party, and whether they have previous experience as judge, politician, private practice, or law teaching.

The estimating equation for performance variable y_{ist} for judge i in state s at year t is

$$y_{ist} = \text{STATE}_s \times \text{YEAR}_t + X'_{ist}\beta + S'_{ist}\rho + \epsilon_{ist} \tag{7.1}$$

where $\text{STATE}_s \times \text{YEAR}_t$ includes the state-year fixed effects, X_{ist} includes the controls for experience, starting year, etc., and S_{ist} includes the treatment indicators equaling one for judges selected under the post-reform system. Standard errors are clustered by state.

Given the inclusion of the fixed effects, the coefficients ρ procure the average difference in performance between judges selected under the new system and judges selected under the old system, controlling for other time-varying state-level factors, for years of experience, and for cohort effects.[14] To the extent that we observe effects on quality in the non-partisan-to-merit reform, those effects are on top of the within-judge incentive effects estimated in the previous section.

7.2 Results

Table 7 reports the estimates from Equation (7.1). Again, the top panel looks at output while the bottom panel looks at quality.

First, there is no selection effect on output. Judges selected by merit commissions write about the same number of words as elected judges, after controlling for case characteristics.

However, there are important effects on work quality. Across the specifications, we see that merit-selected judges write more highly cited opinions than judges selected by election

[14]Note that in the electoral selection systems, the judges may be initially appointed by the governor to fill a vacant seat, rather than being initially selected through a competitive electoral process. We still code the appointed judges as being selected under the electoral system – since the predecessor's choice whether to step down is endogenous to the system.

Table 7: Effect of Judicial Selection System on Judge Work Quality

	Difference between Merit and Elected Judges				
	(1)	(2)	(3)	(4)	(5)
Output					
Merit, Relative to Non-Partisan	0.171	-0.228	-0.105	-0.138	0.273
	(0.256)	(0.356)	(0.300)	(0.298)	(0.189)
Merit, Relative to Partisan	-0.142	0.224	0.228	0.372	0.0302
	(0.178)	(0.177)	(0.221)	(0.232)	(0.264)
Quality					
Merit, Relative to Non-Partisan	0.178**	0.268**	0.193**	0.214**	0.112+
	(0.0582)	(0.0809)	(0.0716)	(0.0726)	(0.0635)
Merit, Relative to Partisan	0.198+	0.242**	0.140**	0.107+	0.167*
	(0.0993)	(0.0424)	(0.0515)	(0.0572)	(0.0704)
Top School				-0.0208	
				(0.0377)	
Local School				-0.0295	
				(0.0273)	
Judge Experience				-0.00237	
				(0.0236)	
Political Experience				0.0597*	
				(0.0236)	
Private-Practice Experience				-0.00987	
				(0.0380)	
Academic Experience				0.0571	
				(0.0556)	
Female				0.0208	
				(0.0489)	
Year Fixed Effects	X				
State Fixed Effects	X				
State-Year Fixed Effects		X	X	X	X
Experience Controls			X	X	X
Starting Year Controls			X	X	X
Judge Characteristics				X	
All Years					X

N= 7,999 judge-bienniums. Estimates of the average difference between judges selected under a new system, relative to to judges selected under the old system, limited to years in which there are at least two judges on the court selected from each system (except column 5, which includes all years). Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01.

(either partisan or non-partisan). Column 3 shows this effect is not driven by experience or starting year. Column 4 shows that this effect is robust to including a large number of controls for judge personal characteristics that may be correlated with work quality. Finally, Column 5 shows that our choice to focus on years with at least two judges from each system is not necessary to observe the effect.

The results on merit selection suggest that merit commissions select better judges than elections. This is consistent with the model's notion that merit commissions have more information about judge quality than voters.

8 Conclusion

The goal of this paper has been to contribute some evidence regarding the hypothesis that the choice between a "politician" and "bureaucrat" entails a tradeoff between a sensitivity to the desires of the electorate and the execution of the mission to make high quality legal decisions (Maskin and Tirole, 2004; Alesina and Tabellini, 2007, 2008). A substantial body of work documents that public officials respond to the preferences of the electorate (Ashworth, 2012). We complement this literature with evidence on the other side of the balance sheet: how electoral pressures interact with a judge's mission to provide high work quality. Sitting judges increase the quality of their work in response to a weakening of re-election pressure. Moreover, judges selected by a technocratic merit commission are of higher quality. The results are detailed in Table 8.

For incumbent judges, we find that non-partisan systems reduce output in election years, but not partisan elections or uncontested elections. This is consistent with a simple model in which campaign effort takes time away from judging, and that political bias makes partisan elections less competitive. Moving from non-partisan to uncontested elections increases case quality, consistent with the notion that non-partisan contested elections are more demanding of a judge's time than uncontested elections. There is no within-judge effect of moving from partisan to uncontested elections, reflecting that non-partisan elections are most competitive – due to less bias than partisan elections. Finally, the merit-based selection process selects better judges than the election systems. These results are consistent with a selection model where better-informed experts can choose higher-quality officials than voters on average.

Our evidence is broadly in line with the early rational-choice approaches of Downs (1957) and Ferejohn (1986), in which voters use their information to make the best decisions they can, conditional upon their policy preferences. But more information is not always better;

Table 8: Summary of Results

	Effect on Output	Effect on Quality
Electoral Cycle Effects		
Partisan Election Year	↓	~
Non-Partisan Election Year	~	~
Uncontested Election Year	~	~
Retention Reform Effects		
Non-Partisan to Uncontested	~	↑
Partisan to Uncontested	~	~
Selection Process Effects		
Merit, Relative to Non-Partisan	~	↑
Merit, Relative to Partisan	~	↑

Summary of results. The left-most column indicates the treatment, and the other column headers indicate the outcome measure. Arrows indicate a positive or negative effect on judge performance. A tilde (~) indicates no effect. See text for details.

more information on candidate quality can improve performance (see Pande, 2011), but more information on political affiliation can reduce performance.

Should all states immediately move to a merit system with uncontested retention elections? Our evidence would certainly strengthen arguments to do so. But there are other criteria besides judicial citations for ranking courts, and ballot referenda for the merit plan have failed many times. There may be many other social impacts of these courts, but at present there aren't data-driven ways to measure them. There is an ongoing debate on which is the superior system (e.g. Pozen, 2010); the fact that states continue to experiment with different systems suggests that it is not clear which system is optimal. If a single system were clearly optimal, then we would have expected the market to have moved in that direction quickly, consistent with Posner's (1987) view that legal institutions move in the direction of efficient exchange.

The fact that we do find a pattern of effects predicted by our simple model helps explain why there is experimentation. While the results are consistent with merit commissions selecting better judges, judging is not a purely technical activity. The political views of judges color the ideological content of their decisions (see Epstein et al., 2013), which may explain why many jurisdictions prefer to give voters a clear signal of the political views of judges. Optimizing states would change systems only if it led to an improvement; hence at any point in time there should be only small variation across states (as Choi et al. (2010) find).[15]

Finally, our results highlight the fact that the American legal system is neither simple nor static. It is a complex, dynamic system consisting of a number of interlocking ingredients. Our study focuses upon one of the most important and influential ingredients of this system: the U.S. state supreme court judges who rule on all aspects of private law, including contract, tort, and property law. Our evidence is consistent with the hypothesis that these judges are professionals who are interested in enhancing the quality of the law. Hence, we have observed many states moving away from partisan political processes for selection toward non-partisan and merit-based processes. These more "bureaucratic" systems have selected better judges and imposed incentives more aligned with the mission of increasing the quality of American law.

[15]The issue here is similar to the well-known Roy model in economics. One cannot use cross-sectional data to measure the impact of a location on wages due to the self-selection by workers; similarly, one cannot use cross-sectional data to measure the impact of selection systems on judge performance due to self-selection of judges across states. See Heckman and Honore (1990) for details.

References

Alesina, A. and Tabellini, G. (2007). Bureaucrats or politicians? part i: A single policy task. *The American Economic Review*, 97(1):169–179.

Alesina, A. and Tabellini, G. (2008). Bureaucrats or politicians? part ii: Multiple policy tasks. *Journal of Public Economics*, 92(3-4):426–447.

Ash, E. and MacLeod, W. B. (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *Journal of Law and Economics*, 58(4).

Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science*, 15(1):183–201.

Ashworth, S. and de Mesquita, E. B. (2008). Electoral selection, strategic challenger entry, and the incumbency advantage. *Journal of Politics*, 70(4):1006–1025.

Ashworth, S., de Mesquita, E. B., and Friedenberg, A. (2015). Accountability and information in elections. mimeo, University of Chicago.

Baker, S. and Mezzetti, C. (2012). A theory of rational jurisprudence. *Journal of Political Economy*, 120(3):pp.513–551.

Berdejo, C. and Yuchtman, N. (2013). Crime, punishment, and politics: an analysis of political cycles in criminal sentencing. *Review of Economics and Statistics*, 95(3):741–756.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Besley, T. and Coate, S. (1997). An economic model of representative democracy. *The Quarterly Journal of Economics*, 112(1):pp. 85–114.

Besley, T. and Coate, S. (2003). Elected versus appointed regulators: Theory and evidence. *Journal of the European Economic Association*, 1(5):1176–1206.

Besley, T. and Payne, A. (2013). Implementation of anti-discrimination policy: Does judicial selection matter? *American Law and Economics Review*, 15(1):212–251.

Bloom, N., Genakos, C., Sadun, R., and Reenen, J. V. (2012). Management practices across firms and countries. NBER Working Papers 17850, National Bureau of Economic Research, Inc.

Bloom, N. and Van Reenen, J. (2010). Human resource managment and productivity. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics, Volume 4*, volume 4, chapter 19. Elsevier.

Canes-Wrone, B., Clark, T. S., and Kelly, J. P. (2014). Judicial selection and death penalty decisions. *American Political Science Review*, 108:23–39.

Canes-Wrone, B., Clark, T. S., and Park, J.-K. (2010). Judicial independence and retention

elections. *Journal of Law, Economics & Organization*, 28(2):211.

Choi, S. J., Gulati, G. M., and Posner, E. A. (2010). Professionals or politicians: The uncertain empirical case for an elected rather than appointed judiciary. *Journal of Law, Economics, and Organization*, 26(2):290.

Condorcet, M. d. (1785). Essay on the application of analysis to the probability of majority decisions.

Coviello, D., Ichino, A., and Persico, N. (2014). Time allocation and task juggling. *American Economic Review*, 104(2):609–23.

Dal Bo, E., Finan, F., and Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *Quarterly Journal of Economics*, 128(3):1169–1218.

Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part ii: Application to missions and accountability of government agencies. *Review of Economic Studies*, 66(1):pp.199–217.

Djankov, S., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2003). Courts. *Quarterly Journal of Economics*, 118(2):453–517.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):pp. 135–150.

Epstein, L., Landes, W. M., and Posner, R. A. (2013). *The Behavior of Federal Judges*. Harvard University Press.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1-3):5–25.

Francois, P. (2000). Public service motivation as an argument for government provision. *Journal of Public Economics*, 78(3):275–299.

Gennaioli, N. and Shleifer, A. (2007). The evolution of common law. *Journal of Political Economy*, 115(1):43–68.

Gennaioli, N. and Shleifer, A. (2008). Judicial fact discretion. *Journal of Legal Studies*, 37(1):1–35.

Glaeser, E. L., Johnson, S., and Shleifer, A. (2001). Coase versus the coasians. *Quarterly Journal of Economics*, 116(3):853–99.

Gordon, S. and Huber, G. (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science*, 2(2):107–138.

Hanssen, F. (2004). Is there a politically optimal level of judicial independence? *The American Economic Review*, 94(3):712–729.

Heckman, J. J. and Honore, B. E. (1990). The empirical content of the roy model. *Econometrica*, 58(5):pp.1121–1149.

Holmström, B. and Milgrom, P. (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7:24–52.

Huber, G. A. and Gordon, S. C. (2004). Accountability and coercion: Is justice blind when it runs for office? *American Journal of Political Science*, 48(2):247–263. Times Cited: 79 Huber, GA Gordon, SC Huber, Gregory/A-5950-2012 Huber, Gregory/0000-0001-6804-8148 79.

Iaryczower, M., Lewis, G., and Shum, M. (2013). To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics*, 97:230–244.

Klerman, D. M. and Mahoney, P. G. (2005). The value of judicial independence: Evidence from eighteenth century England. *American Law and Economics Review*, 7(1):1.

Kritzer, H. M. (2011). Competitiveness in state supreme court elections, 1946–2009. *Journal of Empirical Legal Studies*, 8(2):237–259.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6):1113–55.

Landes, W. M. and Posner, R. A. (1980). Legal change, judicial behavior, and the diversity jurisdiction. *The Journal of Legal Studies*, 9(2):pp.367–386.

Lim, C. H. S. (2013). Preferences and incentives of appointed and elected public officials: Evidence from state trial court judges. *American Economics Review*.

Lim, C. H. S. and Snyder, J. M. (2015). Is more information always better?party cues and candidate quality in u.s. judicial elections. *Journal of Public Economics*.

Maskin, E. and Tirole, J. (2004). The politician and the judge: Accountability in government. *The American Economic Review*, 94(4):1034–1054.

Nelson, M. J., Caufield, R. P., and Martin, A. D. (2013). Oh, mi: A note on empirical examinations of judicial elections. *State Politics & Policy Quarterly*, page 1532440013503838.

Pande, R. (2011). Can informed voters enforce better governance? experiments in low-income democracies. *Annual Review of Economics*, 3(1):215–237.

Park, K. H. (2014). *Judicial Elections and Discrimination in Criminal Sentencing*. PhD thesis, Harris School, University of Chicago, Chicago, IL.

Posner, R. A. (1987). The law and economics movement. *The American Economic Review*, 77(2):pp.1–13.

Pozen, D. (2010). Judicial elections as popular constitutionalism. *Columbia Law Review*, 110:2047–2134.

Shepherd, J. M. (2009). The influence of retention politics on judges' voting. *The Journal of Legal Studies*, 38(1):169–206.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*.

Wilensky, H. L. (1964). The professionalization of everyone? *American Journal of Sociology*, 70(2):pp.137–158.

Wilson, J. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books (AZ), New York, NY.

Young, H. P. (1988). Condorcet's theory of voting. *The American Political Science Review*, 82(4):pp. 1231–1244.

A Model Appendix

This appendix enumerates the proofs for the major theoretical results from Section 3. Subsection A.1 formalizes the effects of bias and noise on the quality of selected judges. A.2 formalizes the role of bias and noise in campaign incentives for judge effort.

A.1 Effect of bias and noise on judge quality

Let ϕ, Φ respectively denote the standard normal's probability density and cumulative distribution functions. The expected quality of judges selected by the governor, expression 3.2, can be written as:

$$\begin{aligned} \bar{q}^G(b) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(q_A + (q_B - q_A)\, I\left(q_A, q_B, b\right)\right)\phi(q_B)\,\phi(q_A)\,dq_B dq_A. \\ &= \int_{-\infty}^{\infty}\left(q_A + \int_{q_A+b}^{\infty}(q_B - q_A)\,\phi(q_B)\,dq_B\right)\phi(q_A)\,dq_A \end{aligned} \tag{A.1}$$

Clearly $\bar{q}^M = \bar{q}^G(0)$. Notice that:

$$\begin{aligned} \frac{d\bar{q}^G(b)}{db} &= \int_{-\infty}^{\infty}\left(-b\phi(q_A + b)\right)\phi(q_A)\,dq_A \\ &= -b\int_{-\infty}^{\infty}\phi(q_A + b)\,\phi(q_A)\,dq_A \\ &= -\frac{1}{\sqrt{\pi}} b \exp(-\frac{b^2}{4}) < 0 \end{aligned} \tag{A.2}$$

The second derivative is:

$$\frac{d^2\bar{q}^G(b)}{db^2} = \frac{(b^2-2)}{2\sqrt{\pi}}\exp(-\frac{b^2}{4}) < 0,$$

where the strict inequalities assume $b \in (0, \sqrt{2})$. A small amount of bias has a small negative effect on quality, that gets larger with b. This proves Proposition 1.

Next we consider the expected quality with elections. In this case the expected payoff is over q_A and q_B, with selection determined by the signals:

$$\bar{q}^E(b) = \int\int\left(q_A + (q_A - q_B)\Pr\left[s_B > s_A + b | q_A, q_B\right]\right)\phi(q_A)\,\phi(q_B)\,dq_A dq_B.$$

Notice that $(q_A - q_B)\, I\left(q_A, q_B, b\right) > (q_A - q_B)\Pr\left[s_B > s_A + b | q_A, q_B\right]$ and hence we have im-

mediately that $\bar{q}^G(b) > \bar{q}^E(b)$. Also since

$$\frac{d\Pr[s_B > s_A + b | q_A, q_B]}{db} < 0$$

for all q_A, q_B, we have that expected ability of judges falls with b. This implies Proposition 2.

A.2 Effect of campaign incentives on effort

We can write the signals observed by the voters as:

$$\begin{aligned} s_j &= m_j + r_j \epsilon_j, \\ &= \pi_j(x_j + e_j) + \pi_j \sigma_j \epsilon_j, \end{aligned}$$

where ϵ_j follows a standard normal distribution. Let us compute:

$$\Pr[m_A + r_A\epsilon_A + b \geq m_B + r_B\epsilon_B].$$

The inequality can be rewritten as:

$$m_A + b + e_A - e_B - m_B \geq r_B\epsilon_B - r_A\epsilon_A = \sqrt{r_B^2 + r_A^2}\epsilon,$$

where ϵ is a standard normal distribution. Hence, we have:

$$\Pr[m_A + b + e_A - e_B + r_A\epsilon_A \geq m_B + r_B\epsilon_B] = \Phi\left(\frac{m_A + b - m_B}{\sqrt{r_B^2 + r_A^2}}\right).$$

In our case we have

$$m_j = \frac{\rho_j}{1+\rho_j}(x_j + e_j)$$

and

$$\begin{aligned} r_j &= \frac{\rho_j}{1+\rho_j}\sigma_j \\ &= \frac{\sqrt{\rho_j}}{1+\rho_j}. \end{aligned}$$

Taking the effort of the other judge as given, the first order condition for a judge defines an optimal effort choice:

$$C_j'(e_j) = B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{m_A+b-m_B}{\sqrt{r_B^2+r_A^2}}\right). \tag{A.3}$$

$$= B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{\frac{\rho_A}{1+\rho_A}(q_A+e_A)+b-\frac{\rho_B}{1+\rho_B}(q_B+e_B)}{\sqrt{r_B^2+r_A^2}}\right). \tag{A.4}$$

Observe that if $\pi_A = \pi_B$, both judges choose the same level of effort, and this has *no effect* on the probability of winning – it is a negative sum game.

Assumption Effort costs are *strongly convex given* $\rho_j, i \in \{A,B\}$ if for every $x \in \Re$ the solution to the following equation is unique:

$$C_j'(e) = B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{\pi_j e}{\sqrt{r_B^2+r_A^2}}+x\right), i \in \{A,B\}.$$

Such functions exist because $\phi > 0$ and ϕ', ϕ'' are bounded, and $C_j(0) = C_j'(0) = 0, C_j'' > 0$. More generally, given any function $C(e)$ satisfying $C(0) = C'(0) = 0, C'' > 0$, and precisions ρ_j for $j \in A, B$, one can choose $\gamma_j > 0$ sufficiently large that this condition holds for $C_j(e) = \gamma_j C(e)$.

Proposition 4. *If effort costs are strongly convex* given $\rho_j, i \in \{A,B\}$ *then there exists a Nash equilibrium in campaign effort. Judge A chooses more effort than Judge B ($e_A > e_B$) if and only if the quality of information regarding Judge A is higher ($\pi_A > \pi_B$).*

Proof. Notice that the maximum effort possible for judge j is:

$$C_j'\left(e_j^{max}\right) = \frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi(0).$$

Let $m = \max\{\pi_A e_A^{max}, \pi_B e_B^{max}\}$ and define the function:

$$h : [-m,m] \to [-m,m]$$

by:

$$h(x) = \frac{\rho_A}{1+\rho_A}e_A(x) - \frac{\rho_B}{1+\rho_B}e_B(x)$$

where:

$$C_j'\left(e_j(x)\right) = B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{\frac{\rho_A}{1+\rho_A}q_A + b - \frac{\rho_B}{1+\rho_B}q_B + x}{\sqrt{r_B^2+r_A^2}}\right).$$

Strong convexity ensures that $e_j(x)$ is a uniquely defined continuous function of x that maximizes the payoff of judge j given the effort of the other judge. Hence $h(x)$ is continuous, and by Brower's fixed point theorem we have the existence of x^* such that $h(x^*) = x^*$, which is in turn by construction a Nash equilibrium, where:

$$\begin{aligned} C_j'\left(e_j^*\right) &= B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{\frac{\rho_A}{1+\rho_A}q_A + b - \frac{\rho_B}{1+\rho_B}q_B + x^*}{\sqrt{r_B^2+r_A^2}}\right), \\ &= B\frac{\pi_j}{\sqrt{r_B^2+r_A^2}}\phi\left(\frac{\frac{\rho_A}{1+\rho_A}(q_A+e_A^*) + b - \frac{\rho_B}{1+\rho_B}(q_B+e_B^*)}{\sqrt{r_B^2+r_A^2}}\right). \end{aligned}$$

. □

B Empirical Appendix

This appendix includes some further notes on the data and the institutional reforms, as well as further regression specifications.

B.1 Notes on Institutional Reforms

This section provides some notes on the institutional reforms. The key point is that there were often coterminous reforms, such as the introduction of an intermediate appellate court. To deal with this we ran all the regressions while leaving one state out. None of the results were substantially changed in these checks. Note that these coterminous reforms only threaten identification in the analysis of retention-process reforms. When we look at the electoral cycle and when we look at selection effects, we are holding court-specific incentives constant.

Colorado instituted an intermediate appellate court in 1971, four years after the election reform. Changing Colorado to a four year window does not change the results. Florida moved from partisan to non-partisan elections in 1972, then moved from non-partisan to merit-uncontested in 1977. In the retention-process regressions we treat these as separate reforms with five-year effect windows. Removing Florida from the regressions does not change the results.

At the same time that Illinois changed from partisan retention to uncontested retention (November 1962), the state also increased judge term lengths from nine years to ten years. However, the term-lengths change went into effect in January 1963, two years before the election reform went into effect.

At the same time it moved from partisan to merit-uncontested, Indiana increased term lengths from six years to ten years.

Kentucky instituted an intermediate appellate court at the same time that it moved from partisan to non-partisan elections.

The Maryland governor began selecting new appointees by merit commission beginning in 1971. When it moved from non-partisan retention to uncontested retention, the term length was reduced from 15 years to 10 years.

Oklahoma instituted an intermediate appellate court at the same time it moved from partisan to merit-uncontested.

In 1973, South Dakota increased its term length from six years to eight years, eight years before the non-partisan to merit-uncontested reform.

Tennessee moved from partisan to merit-uncontested in 1972, then moved back to partisan elections in 1975. It is not included in the analysis.

Utah instituted an intermediate appellate court in 1988, two years after the reform from non-partisan to merit-uncontested.

B.2 Appendix Tables

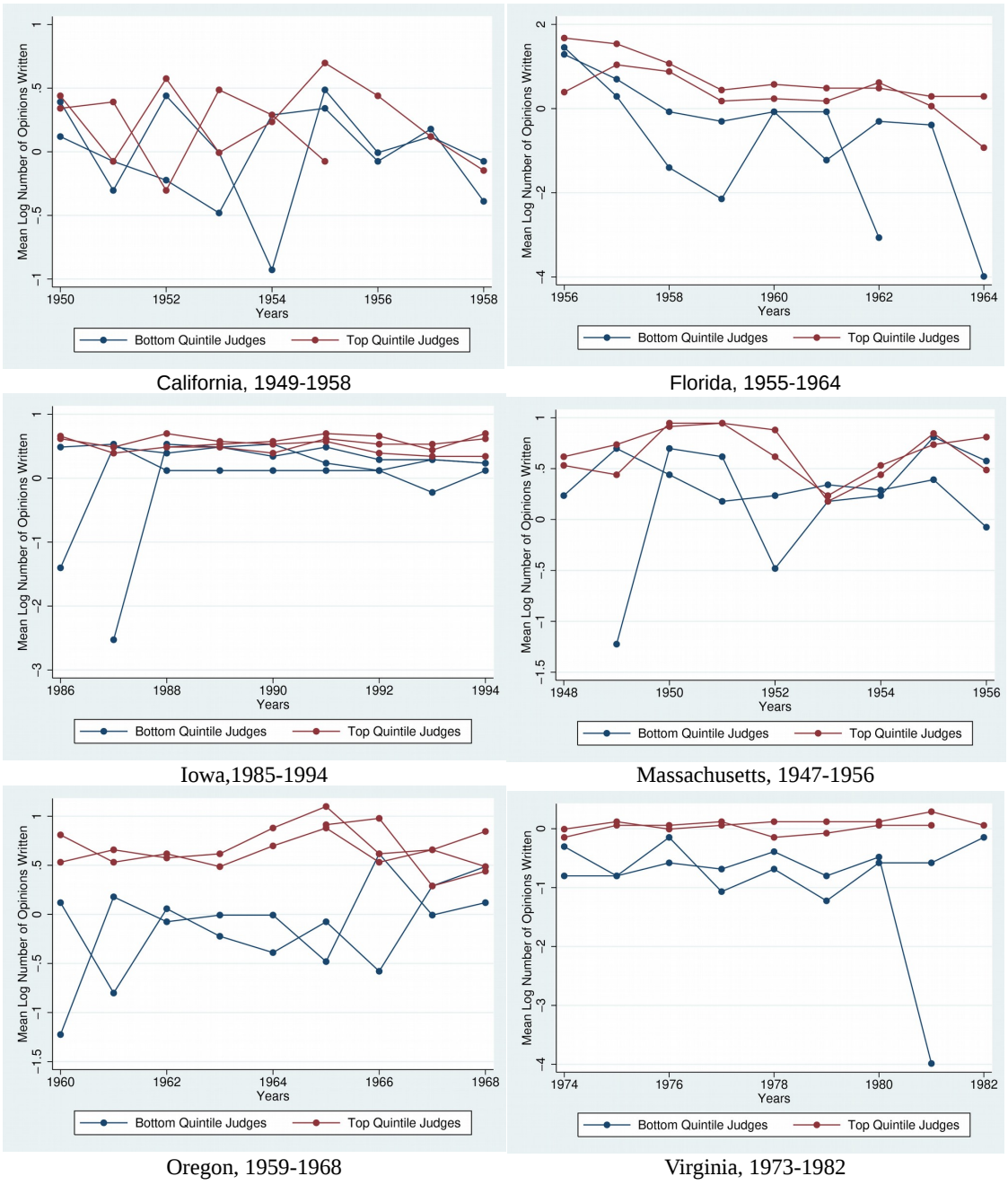
This section provides additional tables and empirical results.

First we look at additional outcome variables for the set of judges analyzed in Subsection 4.2. Figures 4, 5, and 6 provide judge-specific plots for three potential outcome variables to measure output. Respectively, they report the number of opinions written annually, the number of words written annually, and "work output" the number of words written per two-year period after residualizing on case characteristics. As discussed in the text, work output provides the most consistent distinctions between judges of these measures.

Next, Table B1 illustrates the dynamic effect of the election schedule on work output. There is no significant effect in the partisan and uncontested systems, but a clear drop before the election in the non-partisan system.

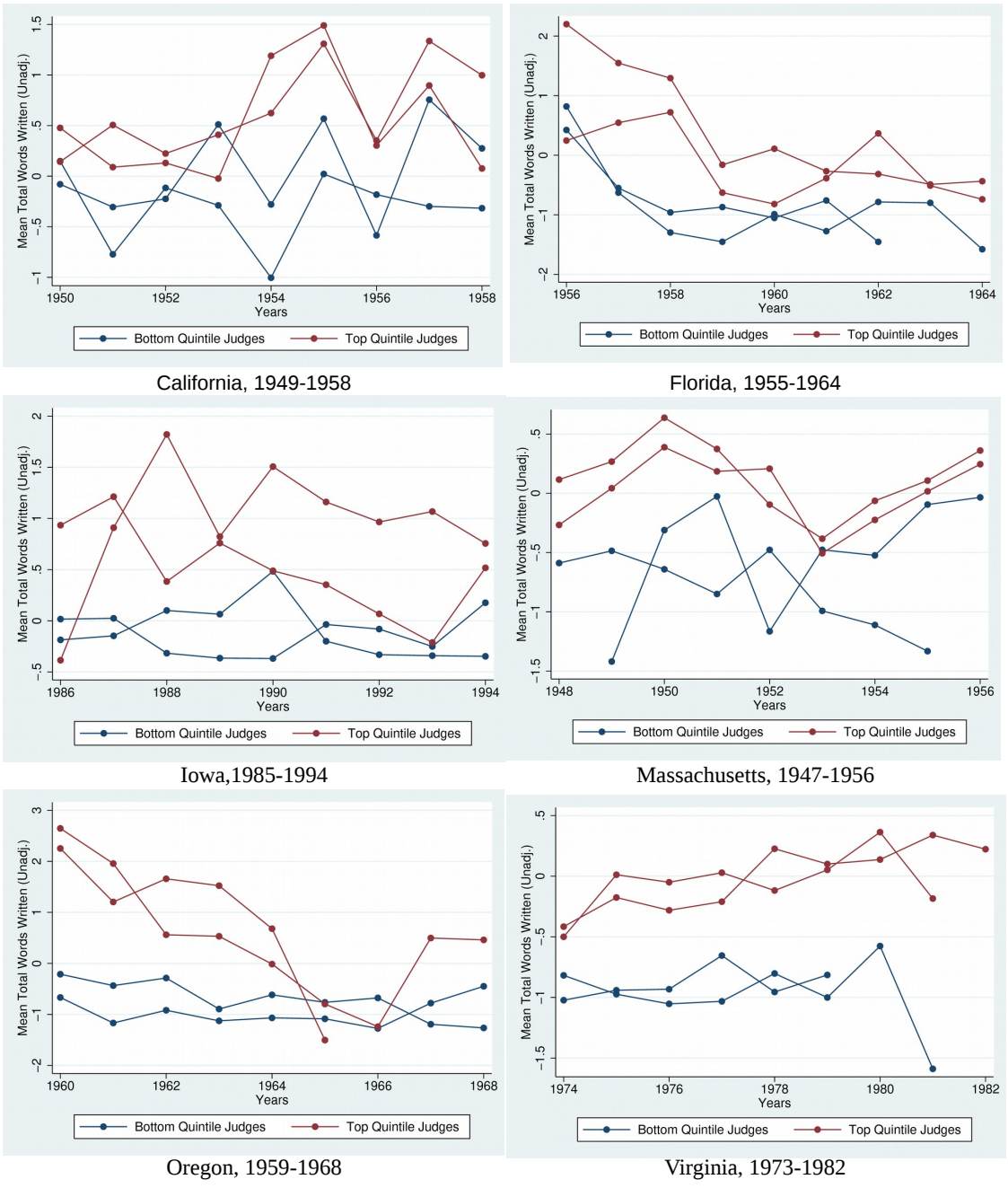
Table B2 illustrates the importance of the treatment window length for the retention-rule effect changes. While the coefficient on quality for non-partisan-to-uncontested is positive for a four year treatment, it is not statistically significant. With at least 8 years, there is a

Figure 4: Number of Opinions Distinction Between Judges



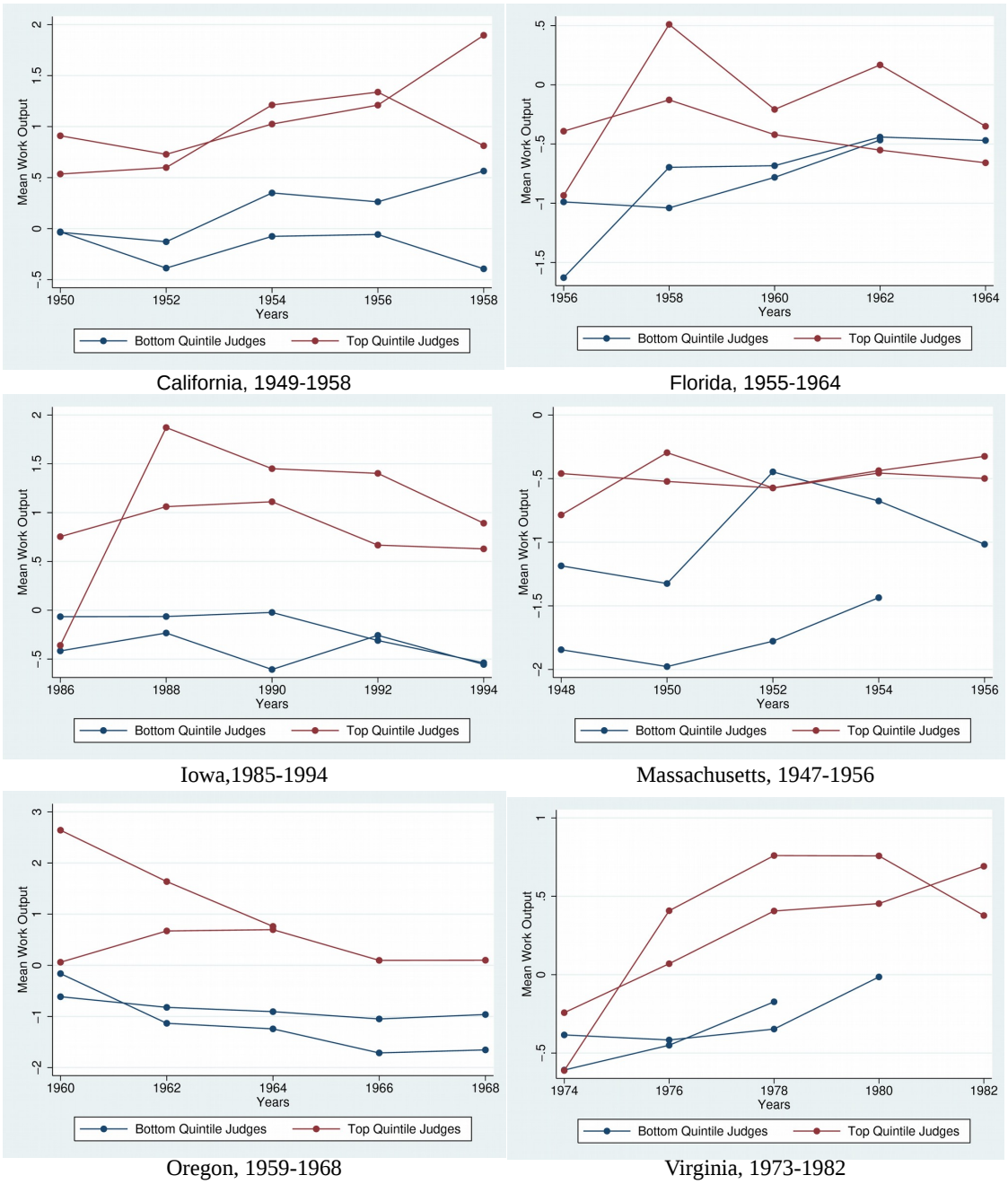
Raw data for log number of opinions plotted by year, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 5: Raw Words Written Between Judges



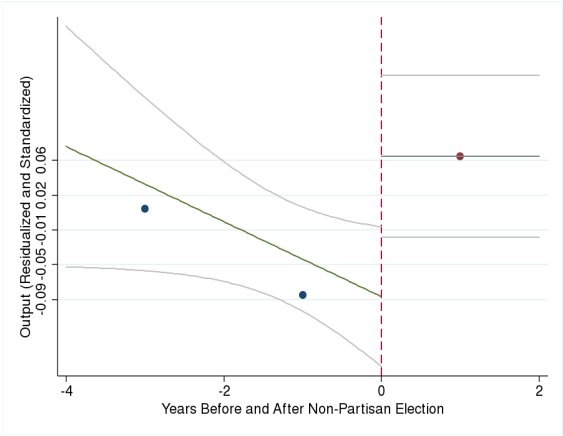
Raw data for number of words written plotted by year, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

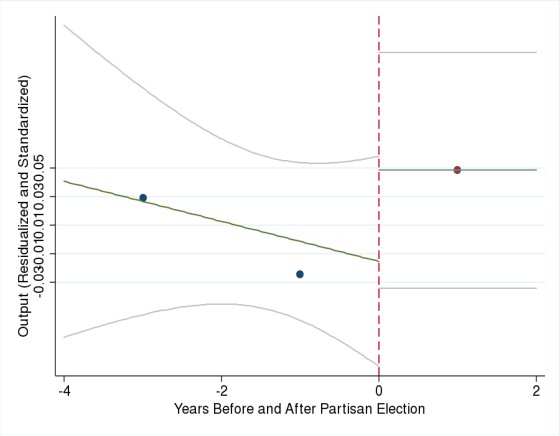
Figure 6: Work Output Distinction Between Judges

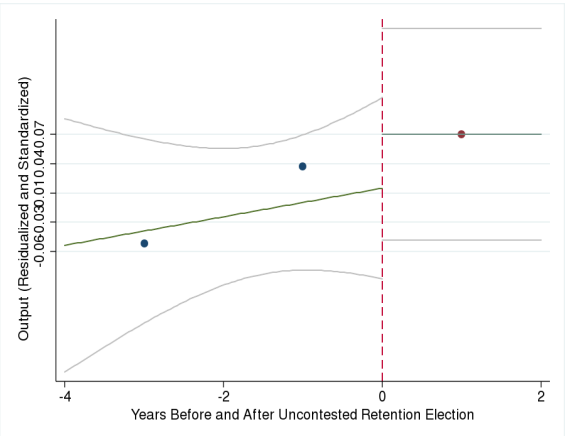


Raw data for Work Output (number of words written, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Table B1: Dynamic Effect of Election Cycle on Work Output







N= 7,999 judge bienniums.

Table B2: Treatment Window Size for Retention Rule Changes

	Effect of Change in Retention Rule					
	(1)	(2)	(3)	(4)	(5)	(6)
Output						
Non-Partisan to Uncontested	-0.125	-0.0606	0.0411	0.140	0.134	0.102
	(0.0989)	(0.117)	(0.114)	(0.121)	(0.121)	(0.169)
Partisan to Uncontested	0.0539	0.00438	-0.00301	0.0106	-0.0665	-0.0650
	(0.0955)	(0.113)	(0.0768)	(0.0786)	(0.0985)	(0.114)
Quality						
Non-Partisan to Uncontested	0.104	0.173*	0.211**	0.219**	0.199*	0.238*
	(0.0770)	(0.0812)	(0.0689)	(0.0712)	(0.0975)	(0.117)
Partisan to Uncontested	-0.0590	-0.0747	-0.0470	-0.0308	-0.0769	-0.0814
	(0.0515)	(0.0681)	(0.0817)	(0.0702)	(0.0619)	(0.0743)
Treatment Window Size	4	8	10	12	14	∞
Year Fixed Effects	X	X	X	X	X	X
State Time Trends	X	X	X	X	X	X
Judge Fixed Effects	X	X	X	X	X	X
State Policy Controls	X	X	X	X	X	X
Experience Controls	X	X	X	X	X	X

N= 7,380 judge bienniums. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01.

Table B3: Effect on Discretionary Opinions Output

	Effect on Discretionary Opinions Output				
	(1)	(2)	(3)	(4)	(5)
Election-Cycle Effect					
Non-Partisan Election Year	-0.00756	-0.0380	-0.0637*	-0.0741*	-0.0726*
	(0.0300)	(0.0287)	(0.0317)	(0.0329)	(0.0337)
Partisan Election Year	-0.101*	-0.0986**	-0.150**	-0.115**	-0.112**
	(0.0424)	(0.0314)	(0.0564)	(0.0429)	(0.0407)
Uncontested Election Year	0.0375	0.0640	0.185*	0.126*	0.133*
	(0.0618)	(0.0488)	(0.0727)	(0.0575)	(0.0582)
Retention System Effect					
Non-Partisan to Uncontested	-0.118	-0.0355	-0.0432	0.204	0.200
	(0.215)	(0.185)	(0.199)	(0.204)	(0.212)
Partisan to Uncontested	-0.159+	-0.134	-0.135	-0.0367	-0.0596
	(0.0926)	(0.100)	(0.0954)	(0.0595)	(0.0655)
Selection Effect					
Merit, Relative to Non-Partisan	-0.0811	0.261	0.185	0.331+	-0.445+
	(0.132)	(0.178)	(0.187)	(0.190)	(0.244)
Merit, Relative to Partisan	-0.261	-0.583+	-0.602*	-0.605+	0.108
	(0.168)	(0.332)	(0.264)	(0.335)	(0.200)

See notes for previous tables. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01.

statistically significant effect, extending to including all years in the window.

Finally, Table B3 shows the effects of our three treatments on the output of discretionary opinions. This is the volume of text in discretionary opinions, after adjusting for legal topics as done with the majority opinions analysis. We don't have citations to discretionary opinions, so we only look at output. In each row, Column 4 gives the preferred specification.

The results can be summarized as follows. Both partisan and non-partisan elections reduce the output of discretionary opinions when up for election. In uncontested elections, judges actually write more discretionary opinions. There is no effect of the retention rule change on the volume of discretionary opinions written. The selection effect is sensitive to the specification. In the preferred specification (Column 4), it looks like merit judges write more discretionary opinions than non-partisan judges, but less than partisan judges.