INTRINSIC MOTIVATION IN PUBLIC SERVICE:
THEORY AND EVIDENCE FROM STATE SUPREME COURTS

Elliott Ash
W. Bentley MacLeod

Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts
Elliott Ash and W. Bentley MacLeod
NBER Working Paper No. 20664
November 2014, Revised June 2015
JEL No. J3,J33,J44,J45

ABSTRACT

This paper provides a theoretical and empirical analysis of the intrinsic preferences of state appellate court judges. We construct a panel data set using published decisions from state supreme court cases merged with institutional and biographical information on all (1,636) state supreme court judges for the 50 states of the United States from 1947 to 1994. We estimate the effects of changes in judge employment conditions on a number of measures of judicial performance. The results are consistent with the hypothesis that judges are intrinsically motivated to provide high-quality decisions, and that at the margin they prefer quality over quantity. When judges face less time pressure, they write more well-researched opinions that are cited more often by later judges. When judges are up for election then performance falls, suggesting that election politics take time away from judging work – rather than providing an incentive for good performance. These effects are strongest when judges have more discretion to select their case portfolio, consistent with psychological theories that posit a negative effect of contingency on motivation (e.g. Deci, 1971).

Elliott Ash
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
eta2103@columbia.edu

W. Bentley MacLeod
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
and NBER
bentley.macleod@columbia.edu

1 Introduction

This paper provides a theoretical and empirical study of the intrinsic motives of state supreme court judges. We have constructed a panel data set that matches institutional and judge biographical information with published decisions for all state high courts in the United States from 1947 until 1994. This data is matched with information on institutional changes in judicial employment conditions, which either increased or decreased the constraints faced by judges when doing their work. The within-judge estimates of the behavioral responses to these changes are consistent with the hypothesis that judges are intrinsically motivated to choose higher-quality opinions when given more time, and that they intrinsically prefer quality over quantity at the margin.

Epstein et al. (2013) observe that judges are paid on fixed salaries to reduce pecuniary motivation that may bias their decision-making. This naturally leads to the question of the impact that such low-powered incentives have on judicial performance, and the extent to which judges are motivated to provide high-quality decisions. Second, there is the controversial issue of how judges should be retained and in particular whether they should take an active part in fund-raising for re-election (see Spottswood, 2007). Using a within-judge design that measures the dynamic effect of electoral demands, we find that judges reduce effort in the year they are up for electoral review, suggesting that they are spending time campaigning rather than judging.

The presence of intrinsic motivation is consistent with White's (1959) theory of *professionalism*. According to White, professionals are individuals who become personally invested in their skills and are motivated by the challenge of doing a job well. If judges have a preference for working on influential cases, then constraining their choices can be demotivating and lead to lower performance. As shown in Section 3, this prediction follows from a simple model of a utility-maximizing judge who allocates time across cases, outside activities, and leisure. It is also consistent with work on contingency and motivation developed in the psychology literature (Deci, 1971), and with recent work in economics on how monetary incentives can crowd out intrinsic incentives (Frey and Oberholzer-Gee, 1997; Gneezy and Rustichini, 2000).[1]

The job of an appellate court judge is to review trial cases and ensure that the law

---

[1] The career concerns literature (e.g. Dewatripont et al. 1999b, Francois 2000, and Prendergast 2007) has shown that the conclusions of the simple intrinsic motivation model developed here can also be derived with standard preferences in a more complex dynamic setting. However, concerns about future promotion have limited importance in our setting, where promotion to a higher court (the U.S. Supreme Court) is extremely rare, and almost all judges keep their job until retirement unless they lose election.

is properly applied. For empirical purposes this is an attractive setting because the basic functions of the job have not changed much in the last 200 years. Litigants appeal cases where they feel there has been an error. The judges choose which cases to hear and then write decisions regarding the merits of each case. From these written decisions one can build a number of performance measures, such as the number of decisions written in a year, the length of a decision, and how often a decision is cited by later judges.[2] Section 7 reports estimates of the effect of changes in employment conditions on these performance measures using judge fixed effects.

Much of the previous empirical work on appellate judging has focused on federal judges. In terms of the work requirements, state appellate judging and federal appellate judging are very similar, the main difference being that state judges mostly interpret state law rather than federal law. Our results therefore have relatively strong external validity for predicting the effects of institutional reforms on federal judges. The advantage of studying state judges consists in the substantial variation across states and over time in the rules for appointing, compensating, and retaining judges. This allows for a more credible identification of the effect that employment conditions have on performance than is possible with federal judges.

To organize the data and interpret the results, Section 3 extends the legal-realist model in Epstein et al. (2013) to allow for intrinsic rewards from high-quality decision-making.[3] In the model, judges prefer working on important cases that can influence the law in the future. This additional ingredient is sufficient to reproduce Deci's (1971) result in the judicial setting – namely, giving a judge more discretion over the cases they hear increases effort in cases they consider important, while reducing effort in less important cases. Besides providing micro foundations for how extrinsic incentives might crowd out intrinsic motivation, the model leads the empirical inquiry toward institutional reforms that give judges more time to write decisions and more discretion over their work environment.

The first change examined is the introduction of intermediate appellate courts (IAC's). This has the effect of reducing the case load of sitting supreme court judges. We measure the impact of the rule change by comparing the performance of individual judges before and after the introduction of an IAC. These judges could respond in various ways to the reduced workload, including working less (choosing more leisure) or working more intensely on the cases still on the docket. The results show that judges respond to the introduction of an IAC by focusing on the more interesting cases that are cited more often by later judges.

---

[2]See Choi et al. (2008) for a discussion of how to measure judicial performance.

[3]Recent work by economists in a legal-realist tradition includes Glaeser et al. (2001), Gennaioli and Shleifer (2007), and Baker and Mezzetti (2012).

With more time from the reduced caseload, the judges write longer and more well-researched opinions, suggesting an intrinsic motivation for their work. Moreover, this effect is strongest when judges have more discretion over what cases to select for review.

Next is the question of term length – the number of years a judge serves in between elections. An increase in term length reduces time and incentive pressures arising from the retention process. If judicial elections incentivize good performance, then increasing term lengths should reduce performance on average because elections play a weaker role in the judge's decision-making. If elections take time, for example due to electoral campaigning, then increasing term lengths (reducing electoral incentives) may improve performance due to the operation of intrinsic motivation. The results support the latter idea: judges respond to a term length increase with higher-quality judgments and no decline in output as measured by the number of cases or total number of words written.

Another implication of the time allocation model is that if judges are engaged in outside activities that provide income, then increasing income from judging should reduce time pressure and thus be associated with more and/or higher-quality output. In the data, wage increases have a positive effect on performance, but only in the states where judges have full discretion over selecting cases for review. In states with less discretion over their case portfolio, there is no effect of wages on performance. This suggests the importance of control over the work environment in the operation of intrinsic incentives.

Finally we measure the effect of the election process by comparing the performance of a judge in the year they are up for re-election to other years of their tenure. There are two types of judicial elections. In contested systems, judges face challengers and therefore face a real chance of being removed from office. Under uncontested elections, only the worst judges are removed by vote and electoral demands are much weaker. In contested systems, there is a decrease in both the number of decisions and the quality of those decisions during election years, while in uncontested elections there is no effect.

In summary, the results are consistent with the hypothesis that judges have an intrinsic motivation to write high quality opinions that interacts with extrinsic incentives to determine the allocation of time across judging and other activities. The results are also consistent with the contention that judicial elections interfere with good judging, rather than reward it.

The rest of the paper is organized as follows. The next section discusses the relevant literature. Section 3 introduces a model of judicial behavior, while the institutional context is discussed in Section 4. Section 5 discusses the data. The empirical strategy is outlined in Section 6 and the results are reported in Section 7. Section 8 provides a concluding

discussion.

2 Literature Background

This paper contributes to two distinct literatures. The first is the literature on public sector employment, where the concern is that use of low-powered incentives may result in less efficient provision of government services (see Francois and Vlassopoulos, 2008). The second is the large literature on judicial behavior, much of it in political science, that is concerned with understanding how judges make decisions.[4] These are discussed in turn.

2.1 Public Sector Incentives

A generic feature of compensation contracts for individuals who are being asked to provide expert opinion is that compensation is insensitive to the opinion provided. The cost for these individuals of changing their recommendation is low, so even small rewards for a particular position can lead to a large distortion in decision-making.[5] In 1991 the National Academies commissioned a report on how to improve compensation policy in government (Milkovich and Wigdor, 1991), concluding that the distortion costs of performance pay outweighed any benefits from increased performance. This report contributed to a large literature, beginning with Kerr (1975), on the dysfunctional behavior that incentive pay can create.

The challenge for economics is to understand performance in professions where incentive pay is weak. Wilensky (1964) explains this as a consequence of "professionalism" where norms of behavior evolve that create intrinsic incentives – individuals work hard to provide performance that is evaluated by their peers as high quality. Kreps (1997) introduced this idea into economics, leading to a number of interesting theoretical developments, including Dewatripont et al. (1999a,b), Besley and Ghatak (2005), and Benabou and Tirole (2002, 2003, 2006). In Dewatripont et al. (1999a,b), the effort incentive stems from career concerns – being rewarded with a better job in future periods. In Benabou and Tirole (2003), extrinsic incentives crowd out intrinsic motivations because they signal that the principal does not trust the agent.

Francois (2000), Prendergast (2007), Alesina and Tabellini (2007, 2008), and Delfgaauw

---

[4]See Epstein and Knight (2013) and in particular Epstein et al. (2013) who have a comprehensive review of the literature.

[5]This result is so well known in incentive theory that it is not typically presented as a result. Variants of the result can be traced to D'Aspremont and Gerard-Varet (1979).

and Dur (2008) apply these ideas to incentives for public officials. These papers are concerned with how to design public employment contracts under a number of assumptions about the self-motivation of agents. The corresponding empirical inquiry is to measure the extent to which intrinsic incentives lead public servants to modify their behavior as a function of the environment.

Beginning with Deci (1971), a substantial body of work has used experiments with college students to tease out the effect of intrinsic incentives. Economists have extended this work both in the lab (Gneezy and Rustichini, 2000) and in surveys (Lacetera and Macis, 2010). These papers show that explicit performance pay can crowd out intrinsic motivation. Gneezy et al. (2011) provide a review of this growing literature.

Dal Bo et al. (2013) conduct a field experiment in which they randomize salaries for public sector job offers in Mexico. Higher compensation is associated with increased quality of the applicant pool and better self-motivation among the hired workers. Besley and Coate (2003) find that appointed regulators are more sensitive to the issues that policy makers care about, while elected regulators are more pro consumer. These papers illustrate the effect of selection on intrinsic incentives, a central issue in the political science literature on judging (Knight and Epstein, 1996).

This paper is concerned with the effects of changes in the environment on acting officials, rather than on the effects of selection. Examples of this approach in looking at public sector employees include Mas (2006); he studies the effect of arbitration decisions on police performance and finds that favorable outcomes lead to better police performance. Dal Bó and Rossi (2011) find that longer term limits lead to higher effort by Argentine legislators, while Banerjee et al. (2014) find that reforms that reduced managerial autonomy in India police stations reduced police effectiveness.

2.2   Judicial Behavior

One strand of the large literature on judges asks how compensation affects the ability of judges selected. Klerman and Mahoney (2005) find that in eighteenth-century England, the passage of statutes giving judges higher salaries was associated with abnormally high returns on stock equity. Baker (2008) instruments for differences in federal judge salaries by comparing them to law-partner salaries in the judge's home state. He finds that these salary differences are not associated with judicial effort, as measured by tendency to dissent in controversial cases and time between case filing and case disposition. Choi et al. (2009) use a cross-section of state supreme court judges to measure the effect of salaries on the

quality of judges selected, and they find only mixed evidence on whether judges with higher salaries perform better, as measured by the number of opinions written and number of case citations.

A second literature explores the effect of election politics on judge decision-making. Tabarrok and Helland (1999) find that electoral incentives increase tort awards, while Gordon and Huber (2007) find that they lead to harsher criminal sentences. Shepherd (2009a,b) uses a cross-section of state supreme court decisions to show that judicial voting reflects the political preferences of the authority responsible for reappointment (the governor or the state legislature), and that judges favor litigants affiliated with that authority. In a separate cross-section of state supreme court decisions, Choi et al. (2010) find that elected judges write more opinions while appointed judges write more heavily cited opinions.

Lim (2013) compares the behavior of elected versus appointed state court trial judges and finds that as a group appointed judges are more homogeneous and tend to make less harsh sentencing decisions. Iaryczower et al. (2013) extend this work to allow for common values and dispersed information among judges, coming to similar conclusions.

Section 3's model of judicial behavior follows the legal-realist approach pioneered by Landes and Posner (1980) and supposes that judges allocate time and effort consistent with well-defined preferences. Recent work by economists in a legal-realist tradition includes Glaeser et al. (2001), Gennaioli and Shleifer (2007), and Baker and Mezzetti (2012). Epstein et al. (2013) introduce a rational-choice model of judging that explicitly allows for judges to enjoy allocating time to their work. The empirical difficulty is that this intrinsic work satisfaction cannot directly be observed. Our contribution is to formalize how intrinsic motivation would influence the behavioral responses to observed variation in employment conditions.

Epstein et al. (2013) use their model to motivate an empirical study of federal appellate courts. The focus here is on state appellate courts, where there is greater variety of treatments across states and across time. Subject to the caveats in Bertrand et al. (2004), one can view the U.S. states as a collection of "natural experiments," where a quasi-experimental approach can be taken to estimate the causal effects of changes in employment conditions on the behavior of individual judges (see Angrist and Pischke, 2009).

7

3 Judicial Preferences

This section extends the model of Epstein et al. (2013, pp. 25-50) to allow for an interaction between shocks to the environment and how judges allocate their time over cases and outside activities. This model provides a framework for interpreting the behavioral effects of exogenous shocks to the judicial employment contract.

Epstein et al.'s (2013) model begins with the observation that the judge's problem, like most incentive problems, is to allocate time across activities. In this case, those activities are leisure, judging, and "outside activities." Non-judging activities refers to outside work that increases (pecuniary and/or non-pecuniary) income. Moonlighting as a private-sector arbitrator or providing legal representation is rare for appellate judges, and generally forbidden by rules of judicial conduct. The key example of an outside activity for our purposes is campaigning for re-election. But as discussed in Epstein et al. (2013), outside activities may also include other political activities such as fund-raising for party affiliates, writing books and journal articles, or guest lecturing at law schools.

A judge must allocate time between leisure T_L, judging T_J, and outside activities T_A, subject to the constraint:

$$T_L + T_J + T_A \leq \bar{T}. \tag{3.1}$$

where $\bar{T}$ is the time available for the period in question (week, month, or year, for example). Let $\vec{T} = \{T_L, T_J, T_A\}$ denote the vector of time allocations. Suppose that the utility function takes a standard Cobb-Douglas form:

$$U\left(\vec{T}, \vec{\alpha}, \vec{\beta}\right) = U_L\left(T_L, \alpha_L\right)^{\beta_L} U_J\left(T_J, \alpha_J\right)^{\beta_J} U_A\left(T_A, \alpha_A\right)^{\beta_A}$$

where the vector of judge preferences $\vec{\beta} = \{\beta_L, \beta_J, \beta_A\} \in (0,1)^3$, includes the elasticity of leisure β_L, the elasticity from the intrinsic value of work on cases β_J, and the elasticity for outside activities β_A.[6] The parameters $\vec{\alpha} = \{\alpha_L, \alpha_J, \alpha_A\}$ are used to parametrize environmental features that change the rewards or time demands of judges. The parameters $\vec{\beta}$ can vary across judges but are assumed to be fixed within-judge over time. In contrast, the parameters $\vec{\alpha}$ can vary across courts and across time.

Suppose that preferences have the standard properties:

Assumption Judicial utility, $U_i\left(T_i, \alpha_i\right) \geq 0$, for activity $i \in \{L, J, A\}$ is continuous in T_i

---

[6]The elasticity is given by $\frac{U_i}{U} \frac{\partial U}{\partial U_i} = \beta_i$. Given that $\beta_i \in (0,1)$, this implies that preferences are strictly concave in the time choice.

and α_i. Holding α_i fixed, U_i is strictly increasing and concave for $T_i \geq 0$.

Further, suppose that the parameter values are such that at the optimum $T_i > 0$, to focus on the interesting cases. The fact that U_i are concave in T_i and the log function is strictly concave implies that the monotonic transformation of utility, $u = \log(U)$ gives us an equivalent representation of preferences in a linear form that is strictly concave in $\vec{T}$:

$$u\left(\vec{T}, \vec{\alpha}, \vec{\beta}\right) = \beta_L u_L(T_L, \alpha_L) + \beta_J u_J(T_J, \alpha_J) + \beta_A u_A(T_A, \alpha_A).$$

The judge's time allocation is the solution to:

$$\max_{\vec{T} \in \Re_+^3} \; u\left(\vec{T}, \vec{\alpha}, \vec{\beta}\right), \tag{3.2}$$

subject to the time constraint (3.1).

A key measure of judicial preference is the extent to which utility varies with time allocated to an activity. Accordingly, when utility is time-differentiable, define the *time elasticity of activity i* by:

$$\epsilon_i(T_i, \alpha_i) = \frac{\beta_i T_i}{U_i(T_i, \alpha_i)} \frac{\partial U_i(T_i, \alpha_i)}{\partial T_i} > 0. \tag{3.3}$$

The assumptions made regarding utility imply that a unique solution to the time allocation problem exists. In the appendix we provide a proof to the following proposition:

Proposition 1. *For $\beta_i > 0$ ($\forall i$), there exists a unique solution to the judge optimization problem, $\vec{T}^*\left(\vec{\alpha}, \vec{\beta}\right) > 0$. Moreover, if the solution is twice differentiable in $\alpha_i, \forall i$, then $\frac{\partial T_i^*}{\partial \alpha_i} > 0$ and $\frac{\partial T_j^*}{\partial \alpha_i} < 0$ for $j \neq i$ if $\frac{\partial \epsilon_i}{\partial \alpha_i} > 0$. The inequalities are reversed if $\frac{\partial \epsilon_i}{\partial \alpha_i} < 0$.*

One can gain a bit more insight into these conditions by defining the *price of time*, given by the Lagrange multiplier, μ, for the time constraint (3.1). The first order conditions for time allocation imply:

$$\text{value of time}_i = \mu T_i^* = \epsilon_i(T_i^*, \alpha_i), \forall i \in \{L, J, A\}. \tag{3.4}$$

Without loss of generality, normalize the value of leisure so that

$$U_L(T_L) = T_L,$$

and set $\beta_L = 1$, which implies $\epsilon_L = 1$. From (3.4), we have that $\mu = 1/T_L^*$, and hence:

$$T_i^* = T_L^* \times \epsilon_i \left(T_i^*, \alpha_i\right), i \in \{J, A\}. \tag{3.5}$$

Thus, the time allocated to judging and outside activities depends on the elasticity of preferences with respect to time. An immediate implication is:

Corollary 2. *An increase in the judge's elasticity for activity i, β_i, increases time allocated to that activity.*

This result highlights the fact that without some way to measure $\vec{\beta}$, one cannot measure the importance of intrinsic incentives across judges because observed variation in choices may be due to variation in preferences, and not due to variations in employment conditions. Therefore, we hold $\vec{\beta}$ fixed and ask how changes in employment conditions $\vec{\alpha}$ might affect judge behavior.

3.1   Preference for Outside Activities

Consider first the allocation of time to outside activities that add to the pecuniary returns from judging. Outside activities are interpreted in general terms; they include any activity that has a cost or benefit that is not affected by decision-making in the court. The main examples are electoral campaigning, writing books, and other activities that have future returns. We cannot directly measure these activities, but we may consider their implications when they are monetized and enter via the current period's pecuniary returns.

Let the preference for outside activities be given by:

$$U_A \left(T_A, \alpha_A\right) = I_0 + I_A \left(T_A, \alpha_A\right) \geq 0.$$

The term I_0 is the base income from employment as an appellate judge, while $I_A \left(T_A, \alpha_A\right)$ represents all other forms of non-leisure rewards, including career concerns from keeping one's job. Suppose that the parameter α_A represents changes in the electoral system. When α_A is larger, then the judge faces greater competition for re-election. If one does not increase effort, then increased electoral competition leads to a fall in future compensation:

$$\frac{\partial I_A}{\partial T_A} > 0, \frac{\partial I_A}{\partial \alpha_A} < 0.$$

Next suppose that an increase in electoral competition leads to an increase in the marginal

return from outside activities:

$$\frac{\partial^2 I_A}{\partial T_A \partial \alpha_A} > 0.$$

The effect on the elasticity of time is thus:

$$\frac{\partial \epsilon_A (T_A, \alpha_A)}{\partial \alpha_A} = \frac{\beta_A T_A}{I_0 + I_A(\cdot)} \left(\frac{\partial^2 I_A}{\partial T_A \partial \alpha_A} - \frac{1}{I_0 + I_A(\cdot)} \frac{\partial I_A}{\partial T_A} \frac{\partial I_A}{\partial \alpha_A} \right) > 0,$$

which gives, from Proposition 1:

Proposition 3. *An increase in electoral competition results in more time on outside activities and less time allocated to judging.*

The model also makes a clear prediction regarding the effect of pecuniary rewards on judging. Viewing the income from judging, I_0, as a parameter we have:

$$\frac{\partial \epsilon_A (T_A, I_0)}{\partial I_0} = \frac{\beta_A T_A}{I_0 + I_A(\cdot)} \frac{\partial I_A}{\partial T_A} \left(-\frac{1}{I_0 + I_A(\cdot)} \right) < 0.$$

Proposition 4. *An increase in judicial income reduces time allocated to outside activities, and increased time allocated to judging.*

This result requires judges to have some ability to allocate time to outside activities. If $T_A = 0$, then changing judicial income will have no effect on time allocated to judging.

3.2 Preference for Judging

This section considers the sub-problem of allocating time T_J to different cases. This in turn allows us to derive the indirect utility from judging, denoted by $U(T_J, N_J, d_J)$, where $N_J > 0$ is the total caseload, and $d_J \in [0, 1]$ is the amount of discretion. The discretion parameter is important because in some states supreme courts are required to review cases with certain characteristics, such as death penalty cases. The parameter d_J denotes the fraction of the case load over which the judge has discretion to hear the case or not. With a larger d_J, she will have more control over her case load, which in turn will have implications for the quality of her decisions.

Specifically, there is a continuum of cases indexed by their complexity/legal significance $\gamma \in [0, \infty)$, with probability distribution given by $f(\gamma)$. The index γ allows one to explore how variations in the cost of time affect the cases a judge chooses to hear and how much time is allocated to each case. The number of cases is given by N_J.

Time spent on case γ is given by $t(\gamma)$. If the judge had no interest in doing her job, then she would set $T_J = 0$, and allocate the rest of her time to leisure or outside activities. In practice, judges likely have some intrinsic desire to do a good job. White (1959) calls this *professionalism,* and supposes that it arises from the development of skills that the judge enjoys using. Our goal is to measure the empirical significance of such preferences. The rest of this section details a model that has a rich set of predictions regarding how time constraints affect the choices of a professional. The main results are outlined here with the details provided in the appendix.

Professionalism can consist of two elements. First, it is assumed that the judge receives utility $V_0 \geq 0$ simply from being a judge – we call this the *prestige* from judging. Second, for each case she hears, she obtains utility $V(t(\gamma), \gamma)$, representing the intrinsic reward a judge obtains from hearing a case of importance $\gamma \geq 0$.

Normalize this function so that $V(0, \gamma) = -c < 0$ for all γ. This is the minimum time cost from hearing a case. Also suppose $V_t(0,0) = 0$, and for $\gamma, t > 0$,

$$
\begin{aligned}
\frac{\partial V}{\partial \gamma} &> 0 \\
\frac{\partial^2 V}{\partial \gamma \partial t} &> 0 \\
\frac{\partial^2 V}{\partial t \partial t} &< 0.
\end{aligned}
$$

This captures the idea that the judge gets more intrinsic utility from cases with higher γ.[7] Finally, we suppose the judge always gets positive utility from cases with sufficiently high value. This is made more precise below. For the cases over which the judge has discretion, let $h(\gamma) \in [0,1]$ be the fraction of these cases that she hears. Given these assumptions, the value from the discretionary allocation of time to judging is given by:

$$
\begin{aligned}
V_J(h(\cdot), t(\cdot), d_J, N_J) &= V_0 + d_J N_J \int_0^\infty V(t(\gamma), \gamma) f(\gamma) h(\gamma) d\gamma \\
&\quad + (1 - d_J) N_J \int_0^\infty V(t(\gamma), \gamma) f(\gamma) d\gamma.
\end{aligned}
$$

---

[7]An example of such a function is $V(t, \gamma) = t^r \gamma - c, r \in (0,1)$. Notice that these assumptions imply $V(t,0) = -c$ for all t.

Given this, the judge solves the following time allocation problem:

$$U_J(T_J, d_J, N_J) = \max_{h(\cdot), t(\cdot)} V_J(h(\cdot), t(\cdot), d_J, N_J) \tag{3.6}$$

subject to:

$$d_J N_J \int_0^\infty t(\gamma) h(\gamma) f(\gamma) d\gamma + (1 - d_J) N_J \int_0^\infty t(\gamma) f(\gamma) d\gamma \leq T_J. \tag{3.7}$$

Consider first the case of the unmotivated judge for which $\beta_J = 0$. In that case $T_J = 0$, and the judge only allocates time between leisure and outside activities. For such a judge variation in working conditions that do not affect their time allocation will have no affect on judicial performance.

The more interesting case is when they do care and $\beta_J \in (0, 1)$. Let μ_J be the Lagrange multiplier for (3.7). Using (3.4) we have:

$$\mu_J^* = \frac{\partial U_J}{\partial T_J} = \mu^* \frac{U_J}{\beta_J}. \tag{3.8}$$

The marginal value of time for the judging sub-problem is equal to the judge's marginal value of time, times the utility from judging, divided by the elasticity for judging.

The Effect of Discretion

The solution to the judging sub-problem can be conveniently expressed in terms of the cost of time. The amount of time allocated to a case of quality γ depends only on the value of time in the judging sub-problem:

$$V_t(t^*(\mu_J, \gamma), \gamma) = \mu_J.$$

This expression implies that time, and thereby opinion quality, is *increasing* with the difficulty of the case: $t_\gamma^*(\mu_J, \gamma) > 0$.

An increase in discretion reduces the time constraint, and hence μ_J. Thus there is an increase in time per case holding T_J fixed. The choice of cases to hear is given by the *net value* of case γ:

$$\Delta(\mu_J, \gamma) = V(t^*(\mu_J, \gamma), \gamma) - \mu_J(t^*(\mu_J, d_J)).$$

This value is increasing in γ and decreasing in μ_J. For $\gamma = 0$ it is $-c$, and therefore negative.

We assume that for for γ sufficiently large this is positive. Thus there is a unique $\gamma^* \left(\mu_J \right)$ such that $\Delta \left(\mu_J, \gamma^* \left(\mu_J \right) \right) = 0$. For cases $\gamma > \gamma^* \left(\mu_J \right)$, the judge gets positive utility from hearing these cases, and so sets $h \left(\gamma \right) = 1$. For $\gamma < \gamma^* \left(\mu_J \right)$, the judge prefers not to hear these cases, and thus only hears the fraction required, $\left(1 - d_J \right)$. By the envelope theorem, the net value is decreasing in μ_J. Therefore an increase in the value of time μ_J leads to fewer cases being heard and less time allocated to the cases that are heard. These observations are summarized as follows:

Proposition 5. *An increase in discretion reduces the elasticity of the utility from judging $\left(\frac{\partial \epsilon_J \left(T_J, d_J \right)}{\partial d_J} < 0 \right)$, and hence less total time is allocated to judging, with fewer cases being heard.*

This result illustrates the countervailing properties of discretion – it tilts the balance in favor of more important cases. Since these cases provide more utility, they have an effect similar to V_0, and thus lead to a reduction in total time in judging.

The effect upon time per case is ambiguous. The reason is that from expression 3.8, we see that that the cost of time in judging solves depends both on the overall cost of time μ^* and on utility U. An increase in discretion lowers the cost of time but raises utility, so the net effect is ambiguous. This will be an effect that we will measure in the data.

The model also helps explain the role of discretion as a matter of public policy. If there are cases that are deemed to be in the public interest, then a mandatory review might be warranted. Otherwise, they may not be heard if considered to be of marginal legal significance. On the other hand, allowing discretion may result in more time allocated to cases with greater legal significance if the time constraint is relaxed sufficiently.

The Effect of Case Load

Finally, consider the effect of changes in the case load. Rewrite the variable utility component for judging as:

$$\begin{aligned} V \left(T_J, d_J, N_J \right) &= U \left(T_J, d_J, N_J \right) - V_0, \\ &= N_J V \left(\frac{T_J}{N_J}, d_J, 1 \right). \end{aligned}$$

This implies that the optimal marginal value of time in the judging sub-problem can be written as $\mu_J^* \left(\frac{T_J}{N_J}, d_J \right)$, which is decreasing in both the time and discretion arguments.

The appendix shows that the effect of case load on the preference for judging is given by:

$$\frac{\partial U_J}{\partial N_J} = V\left(\frac{T_J}{N_J}, d_J\right) - V_{T_J}\left(\frac{T_J}{N_J}, d_J\right)\frac{T_J}{N_J} > 0. \tag{3.9}$$

This expression is positive because an increase in case load results in more high-value cases; the judge can always reallocate effort from the low-value to the high-value cases to increase total utility. The effect on the elasticity with respect to case load is given by:

$$\frac{\partial^2 \epsilon_J\left(T_J, d_J, N_J\right)}{\partial T_J \partial N_J} = \frac{\beta_J T_J}{U_J}\left(\frac{\partial^2 U_J}{\partial T_J \partial N_J} - \frac{1}{U_J}\frac{\partial U_J}{\partial T_J}\frac{\partial U_J}{\partial N_J}\right).$$

The first term in parentheses is positive – increasing the case load always increases the marginal value of time – while (3.9) implies that the second term is negative in sign. The relative importance of the two terms is determined by prestige, V_0. Thus we have:

Proposition 6. *An increase in the case load* (N_J) *increases the time allocated to judging* T_J *if the prestige from judging is sufficiently high, otherwise it lowers total time on judging.*

In summary, judges have to decide which cases to hear, and how to allocate time across these cases. Given that pecuniary returns do not vary with these decisions, the allocation of time depends on their personal preferences for their work, or their professionalism. As professionals, judges may spend more time on the more important cases. When judges have discretion, then the theory predicts that they adjust their workload on two margins – which cases to hear, and how much time to allocate to the cases that they do hear. The theory provides little guidance regarding the size of these effects – that requires empirical evidence.

4 Institutional Background

State supreme courts operate as the state judiciary's analogue to the U.S. Supreme Court, where judges rule on questions of state law rather than federal law. These questions arise in cases appealed from lower state courts. The state supreme court consists of a panel of between five and nine judges, who decide together how to rule on appeals.

A case begins when a plaintiff files a lawsuit or a prosecutor indicts a criminal. At trial, facts are litigated and a judge/jury gives a verdict, which the losing party can appeal. If the state has an intermediate appeals court, they will then take the case and may affirm, reverse, or modify the trial verdict. After this intermediate court's decision (or after the

trial decision when the state does not have an intermediate appellate court), the ruling can be appealed to the state supreme court.

If the supreme court accepts a case for review, the panel of judges will rehear the case at oral argument and review the submitted briefs for legal error. Each judge votes whether to affirm or reverse the lower decision. One of the majority judges writes an opinion explaining the decision. In rare cases, the state supreme court ruling is appealed to the U.S. Supreme Court.

This is the institutional context for the present study of judicial incentives. Importantly, the job of a supreme court judge does not change much over the course of a career. A judge in his first year of work has essentially the same task as a judge in his last. While the state supreme court judge's job of reviewing cases and establishing precedent is similar to that of a U.S. Supreme Court justice, there are important institutional differences between state supreme courts and the U.S. Supreme Court that provide attractive opportunities for the empirical study of how employment conditions affect judicial behavior.

[TABLE 1.1 HERE]

This paper identifies the causal relationship between employment conditions and judge performance using institutional changes at state supreme courts. The list of treatments is described in Table 1.1. These include the addition of an intermediate appellate court, changes in term length between elections, changes in judge salaries, and the staggered judicial electoral cycle. Some of the reforms to intermediate appellate courts and term lengths were not included due to co-occurrence of other major reforms. The following reforms are not the focus of the analysis but are included as control variables: establishment of a court administrative office, change in the judge retention system, increase in the number of judges, and establishment of a mandatory retirement age.

One of the most common reforms is the establishment of an intermediate appellate court (IAC). Intermediate appellate courts significantly filter the set of cases that supreme court judges have to review. When an IAC is operating, supreme court judges have a lot of help in reviewing cases and have more discretion in whether to accept cases for review. We expect that the introduction of an IAC will increase the time and discretion available to judges, so they can devote more time to what they care about. If judges care about their work product, they will spend more time on the opinions that remain on the docket. In terms of the model notation, this reform can be seen as reducing N_J, the number of cases they must hear, and/or increasing d_J, the level of discretion in case selection. As discussed in more

detail below, some states gave full discretion to the supreme court after establishing an IAC, while other states gave only limited discretion. This will turn out to matter empirically.

[FIGURE 1.1 HERE]

Because compensation is not contingent on performance, the retention process is probably the strongest incentive system facing state appellate judges. In most states, judges have to be reappointed by the governor or by voters every few years. There is substantial variation in the lengths of terms of office for state appellate judges, as shown in Figure 1.1. Moreover, there are discrete changes in the judge term length. Section 7.2 reports estimates of the effect on performance of these changes. In the sample of years, while ten states in total changed judge term lengths, there were only five states without significant coterminous reforms to judge employment conditions.[8] Increasing term lengths reduces the frequency with which a judge faces re-election, so the effects of election on a judge's time allocation will be reduced. Decreasing term lengths should have the opposite effect. In the model, decreasing electoral incentives corresponds to a decrease in the return to outside (campaigning) activities α_0, which by Proposition 1 should increase time spent on cases and be reflected in higher quantity and/or quality of output.

[FIGURE 1.2 HERE]

We measure the effects of IAC establishments and term length changes in a panel framework using judge fixed effects. Because there is such a long time period, we measure effects in a window around the reforms. While the results are not particularly sensitive to the time window chosen (as illustrated in the figures below), we settled on a ten-year window as reasonably balanced given our purposes. This ten-year-window specification accommodates the average career length of state supreme court judges – having a longer effect window would give too much weight to the handful of judges who work on the court for many years before and after the reform. The regressions include an indicator equaling one for the baseline time window of ten years before and ten years after an IAC establishment or a term length change. 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. This setup is illustrated in Figure 1.2. In a handful

---

[8]The changes are as follows: Hawaii, seven to ten years; Montana, six to eight; South Dakota, six to eight; Louisiana, fourteen to ten; Vermont, two to six.

of states, we shrank the time window if the reform occurred close to the end of the sample (ND and NE) or if there was a conflict with another reform (HI). Where reforms occurred within 5 years of each other, they were not included in the analysis.

[FIGURE 1.3 HERE]

[FIGURE 1.4 HERE]

[FIGURE 1.5 HERE]

Next we have monthly data on individual judge salaries between 1974 and 1994. This analysis add to existing findings, such as Choi et al. (2009) and Baker (2008), by measuring the within-judge effect in a panel data framework. Proposition 4 suggests that increasing base income I_0 would reduce incentives to engage in outside activities and therefore may increase time spent on judging work. To get an idea of the scope of variation in salaries, Figure 1.3 shows the average real judge salary over time for the sample of judges. The negative trend early in the period is due to pay increases not keeping up with inflation. Figure 1.4 provides a histogram of the number of discrete salary increases by year in the data. Figure 1.5 shows the average size, in proportional terms, of the judicial salary increases that occurred in that year (if one did occur). As in Mas (2006), these discrete compensation changes may result in measurable performance changes.

[TABLE 1.2 HERE]

To look more closely at judicial election incentives, we exploit the staggered electoral cycle. Like U.S. Senators, nontenured state supreme court judges face election on a staggered basis, where a subset of judges are up for election in any particular election year. One can compare the performance of judges who are up for election to their colleagues who are not up for election. If judges choose high performance to impress voters or the party organization, then being up for election should improve judge performance. If instead judges have an intrinsic motivation to choose high performance and elections take up their time, then being up for election should reduce judge performance. Or elections may have no effect. Helpfully, there are both contested election systems (where elections are competitive and incumbents face challengers) and uncontested election systems (where incumbent judges face an up-or-down retention vote and only the worst judges are removed from office). Whatever their

effect, elections should have a weaker or non-existent effect in uncontested systems.[9] Table 1.2 lists the states by their retention system.

The judge retention system also matters for the analysis of salary increases. An alternative model by Besley (2004) suggests that increasing salaries may affect performance by increasing the value of retention, which would cause judges to change their performance to influence their election. If Besley's model is correct, one would observe an incentive effect of salaries only in the competitive retention systems, and not under uncontested elections or tenure.

[TABLE 1.3 HERE]

As emphasized in the theory, the judge's discretion to select cases for review is crucial to the operation of extrinsic incentives and intrinsic motivation. The central idea is that with discretionary case selection, intrinsic motivation should play a more important role in performance because judges have more control over their work environment. One can divide state supreme courts into three categories, depending on whether they have full discretion in case selection, partial discretion (that is, some cases are mandatory), or fully mandatory review (that is, all cases must be reviewed to some degree).[10] These states are listed in Table 1.3. A line of papers in the political science literature have shown that while judges have some level of discretion under mandatory review, it nonetheless has substantial compositional effects on the state supreme court caseload (e.g. Eisenberg and Miller, 2009).[11]

---

[9]There are two types of contested system. In partisan contested elections, incumbent judges face a challenger, with party affiliations on the ballot. In non-partisan contested elections, incumbent judges face a challenger, but party affiliations are not on the ballot. We examine the differences between these two systems in separate work. Three other judge retention systems have a relatively small number of states. In governor retention, the governor decides whether a judge should be re-appointed at the end of his/her term. With legislative retention, the state legislature decides by majority vote whether a judge should be re-appointed. With life tenure, judges cannot be removed except by impeachment.

[10]The U.S. Supreme Court has fully discretionary review.

[11]In practice, the appellate review standards are relatively complex. Splitting the states into three categories required simplification and some subjective coding decisions. That is why we report aggregate results without regard to the discretion level. A second procedural rule that did not turn out to matter was the procedure for assigning cases to judges. At state supreme courts, discretionary assignment by the chief justice (the rule at the U.S. Supreme Court) is the minority rule followed in just 15 states. In 13 states, cases are randomly assigned to authoring judges. In the remaining 22 states, cases are assigned on a rotating system, with cases arbitrarily assigned to judges based on their order on the docket. Christensen et al. (2012) show some differences in case assignment characteristics across systems. In their sample, for random assignment and rotating assignment, case characteristics and judge characteristics are only negligibly correlated. In the electoral cycle results, however, we find that the number of opinions still goes down for judges up for election relative to their colleagues, meaning that true rotation/randomness is not occurring. It turns out that looking only at the states with rotating/random assignment does not affect our results.

We do not have variation over time within state in the rules for case selection. Our analysis seeks to study their importance by examining the relative magnitude and significance of the other treatments depending on the case selection rule. In particular, the establishment of an IAC should have a larger treatment effect under discretionary case selection, as this gives the supreme court more scope to reduce their caseload. In terms of the model, this can be thought of as a greater increase in the parameter d_J due to establishing an IAC under full discretion relative to limited discretion. Similarly, for courts with higher review discretion, treatments that reduce the return to outside activities (α_A) or increase base income (I_0) might have a larger positive effect on judging effort (relative to courts with lower review discretion) since the return to judging time (relative to leisure time) is greater in those courts.

5   Data

One contribution of this project is the assembly of a new integrated data set on state appellate courts. At present, there is extensive data on federal court judges (e.g., Epstein et al., 2013), but no existing comprehensive panel data on state courts. Existing studies, such as Landes and Posner (1980), Shepherd (2009b), Tabarrok and Helland (1999), Hall and Bonneau (2006), Gordon and Huber (2007), Kritzer (2011, 2015), Lim (2013), and Iaryczower et al. (2013), ask different questions and/or use different, shorter time periods. Our data set has a much longer time period and is more comprehensive. The State Court Data Project had four years of data (1995-1998), and did not have data on how often cases were cited by later judges. Choi et al's (2009, 2010) data included three years of cases (1998-2000). These short time frames only allow cross-sectional studies of judge behavior. Since the data set spans 48 years of data (1947-1994) it allows us to use a within-judge identification strategy.

Three data sets have been constructed with three types of data: 1) judge characteristics, 2) institutional variables, and 3) judicial output measures. These data are discussed in the next three subsections.

5.1   Judge Characteristics

The data on the characteristics of individual judges is collected from a range of sources. The key sources include state court web sites, judge obituaries, and Marquis Who's Who. Items that were unavailable from these sources were obtained through interviews of state court administration staff.

Table 2.1 presents summary means and standard deviations for a collection of judge variables in the sample. State supreme court judges are hired in middle age: 53.6 years old on average. They work as judges for an average of 12 years, are overwhelmingly male, and most of them resign or retire (rather than earn a promotion, die in office, or lose election). Because promotion (defined in this table as moving to a federal appeals court or to a governorship) is so rare, career concerns are likely a limited source of incentive pressure for these judges.

5.2 Institutional Variables

The key changes in institutional treatment variables, as described in Section 4 above, are listed in Tables 1.1 and 1.2. Summary statistics for these variables by state and year are listed in Table 2.2. The discrete rule changes are represented in the data as dummy variables that equal one for the years after the law change and zero for the years before. In almost all cases, reforms are 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. For annual salaries, we use the weighted average across months. The term length changes are combined in a single variable, where a term length decrease is represented as a negative one.[12]

[TABLE 2.2 HERE]

The institutional variables were collected from a range of sources. Most of the judge salary data were obtained from the National Center for State Courts (NCSC), which administers an annual survey of state judge salaries. These data were error-checked by research assistants. Inflation adjustments are made with BLS data on prices for regions or MSA's, when applicable. The appellate review standards were also collected from the NCSC. The data on discrete rule changes were collected from previous papers on judge election rules, from the web site judicialselection.us, and from inspection of legislation and constitutional amendments. The records on election events were collected from a range of sources, with most of the records obtained from Kritzer (2015).

---

[12]Treating these changes symmetrically is a strong assumption, but most of the estimates are similar if we treat them as separate reforms. Statistical significance is lower due to the smaller number of experiments.

5.3 Judicial Performance

The data set contains judicial performance measures constructed for judicial opinions. Here we use performance measures suggested by Landes and Posner (1980), Choi et al. (2010), and Epstein et al. (2013). These include the number of opinions written, the length of opinions, the amount of research put into opinions, and the number of subsequent citations to a judge's opinions. A detailed description of how the data were collected and constructed is included in Appendix B.

The judicial performance data are constructed from the universe of opinions published by state supreme courts between 1947 and 1994. The full sample includes 1,025,461 cases. Many of these cases are summary orders – certiorari denials, habeas corpus denials, and other brief orders that do not require a full written opinion. These orders are just a few sentences long and are rarely if ever cited by future judges. Many states in the sample do not publish these types of orders. Our interest is in the professionally authored legal precedents that explain the ruling for future judges, so summary orders are excluded from the empirical analysis. Specifically, we focus on published majority opinions that are seven or more sentences in length – orders with six or fewer sentences are removed. This step shrinks the sample to 496,099 majority opinions. Next, since the goal is to study the behavior of individual judges across time, unauthored (per curiam) opinions are removed, as well as the small number of opinions written by non-supreme-court judges, such as magistrates, commissioners, and other special sitting judges.[13] This step shrinks the sample to 387,905 majority opinions (plus attached discretionary opinions) written by judges with biographical information. This results in 184.7 cases per state per year and 25 cases per judge per year on average.

The list of performance variables, along with summary statistics, are presented in Tables 2.3 through 2.6. At the state supreme court level, whether to accept a case for review is often discretionary, so if judges accept more cases for review they are taking on more work. Whether to write a discretionary opinion – a concurrence or a dissent – is always up to the judge's discretion and involves willingly taking on more work.

Second, effort statistics are constructed from the raw text of a judge's opinions. An appellate judge's output is his writing; a rough measure of increased effort would be increased

---

[13]The treatment variables are uncorrelated with the number and proportion of per curiam opinions, with the exception of the establishment of the intermediate appellate court. The proportion of per curiam opinions goes down after an IAC is established. This likely reflects that per curiam opinions are on average less important than authored opinions, and that after an IAC is established the court reviews fewer less important cases.

language output that is measured by the total number of words written during a time period, as well as a basic opinion length measure – the average number of words per majority opinion written. In addition, the amount of research a judge engages in is measured by the average length of the Table of Cases, which is the number of previous authorities cited in her opinions.

When using opinion length and case-law research as effort measures, it's important to note that much of the raw labor inputs into opinion writing and research are provided by supreme court clerks. If one observes changes in output in response to rule changes, that effect may be due in part to changes in how the judge manages her clerks. We interviewed state supreme court court staff, and learned that the number of clerks per judge remained relatively constant over the time period of the study. Moreover, the processes of clerk selection and retention were relatively stable and do not seem to be correlated with changes in other judicial institutions.

The third measure is the number of citations to a judge's opinions by other judges. Bloomberg Law staff attorneys have categorized citations as positive, distinguishing, or negative. A positive cite is a clear signal that a decision is found useful by a future judge. A distinguishing cite means that part of the ruling is useful, but needs to be clarified – so this is perhaps a weaker signal of opinion quality. The significance of a negative citation is more problematic; the most intuitive interpretation is that a negative cite means a judge made the wrong decision. On the other hand, negative cites could mean that a judge is being more creative in his judging and allowing for more experimentation in lawmaking. A final possibility is that negative cites are just another signal of an opinion's influence relative to other opinions, and therefore could serve as an additional quality measure.

The Bloomberg citation analysis features allow the construction of more fine-tuned data on judicial citations, and includes information about whether a case is discussed by the future court (rather than cited without comment), or whether it is directly quoted by the citing court. These measures can be understood as more direct signals that the citing court finds the opinion useful. The Out-of-State Cites measure includes positive cites from out-of-state courts; as noted by Choi et al. (2010) among others, this is perhaps the best quality measure because the cited case serves as persuasive rather than binding precedent. For all of these citation measures, however, an important caveat is that the number of cites is a joint measure of both the importance of a case and the effort of the judge.[14]

Citations are an interesting outcome variable because they reflect case quality, which

---

[14]For the citations measures, versions using only citations within ten years of the publication of the opinion (rather than all citations through 2012) did not qualitatively affect the results.

judges may be intrinsically motivated to produce when other time constraints are reduced. Citation counts would not be useful on this point if they could easily be manipulated by judges independently of case quality. Another problem would be if judges monitored their citations and derived utility from citation counts (rather than case quality itself).

[FIGURE 2.1 HERE]

There are a number of reasons why this may not be a serious concern. First, the vast majority of citations to state supreme court opinions are from other courts (mostly lower courts in the same state, but also courts in other states), meaning that the judge cannot meaningfully influence citation counts either from citing his own cases or from influencing his colleagues to do so. Second, during this time period it was difficult to monitor one's citations, since they would have to be looked up manually in citation digests – large tomes that span multiple shelves in law libraries. Moreover, many citations occur long after an opinion is written (and presumably forgotten). Figure 2.1 shows the distribution of the delay between a case and its citations – the median delay is ten years. Recalling that the average career length in the sample is 12 years, this shows that a majority of citations occur after a judge leaves the bench. Citations are therefore unlikely to be a source of expected rewards for the authoring judge.

[TABLE 2.3 HERE]

[TABLE 2.4 HERE]

Table 2.3 provides summary statistics for case-level variables. Besides the performance-related variables just described, this table includes a handful of opinion-level variables that are not used in the empirical analysis but provide context for the type of work performed by state supreme court judges. They include the case outcome – affirm, reverse, remand, or modify – and the main areas of law for each case – civil, criminal, administrative, or constitutional. The first pair of columns gives statistics for the full sample of cases, while the second pair of columns gives statistics for the pruned sample of authored legal opinions, which is the sample of opinions used in the empirical analysis. As expected, the average opinion in the pruned sample is longer, more well-researched, more well-cited, and has more discretionary opinions attached. This reflects that the less important summary orders have been excluded. Table 2.4 gives summary correlations within case for the set of performance measures used in the empirical analysis. Notably, all of the measures are positively correlated

within opinion, some strongly so. In particular, the fact that negative cites are strongly correlated with positive cites would support the use of negative cites as an additional signal of opinion quality.

It's tempting to draw causal inferences from these correlations; after all, a well-researched opinion is likely to be more well-cited due to its higher-quality research. But there are other unobserved qualities of opinions that also influence the number of citations, such as the clarity of the legal reasoning or the novelty of the legal issues presented, which are likely also correlated with the length of the table of cases. While it is an interesting open empirical question what case features determine the number of citations, we do not have independent exogenous variation in particular features of opinions and therefore cannot procure credible estimates of those effects. Therefore, the empirical analysis uses each performance variable separately and only as a dependent variable.

[TABLE 2.5 HERE]

Table 2.5 aggregates the data by judge-year. This level of aggregation is used in the results reported below because it allows the use of a judge fixed effect and treats a year of work by an individual judge as the unit of observation. If case-level data were used in the regressions, then the number of opinions written would skew a judge's weighting in the estimates. Also, it makes sense for judges who work many years to count more than judges who work just a few years. Finally, using years rather than months or quarters is helpful for avoiding problems associated with seasonal variation in performance, for example due to vacation time.

The statistical levels observed in the data are not especially relevant to the empirical analysis. The coefficient estimates are derived from log specifications and can therefore be interpreted as proportional changes due to the treatments. That said, one might note that each of these judges is responsible for a large corpus of output in any given year. The average annual output is 63,831 words, the length of a short novel. On average, a judge's opinions for a year are used 43 times by judges in other jurisdictions (Total Out-of-State Cites), illustrating that state supreme court judges are influential figures that can play an important role in the broader legal system.

[TABLE 2.6 HERE]

Finally, Table 2.6 reports summary correlations for the performance variables using the judge-year data set. As with the case-level correlations reported in Table 2.4, the judge-year

correlations indicate that the performance measures are mostly correlated within judge over time. The multiple measures of output and quality provide multiple signals of the amount of time a judge spends on his opinions. The important exception in this table is the number of majority opinions, which is negatively correlated with most of the quality measures. This suggests that judges face a tradeoff between quantity and quality. These correlations may reflect that judges with larger caseloads have to sacrifice on quality, or it may just be a case composition effect where judges with larger caseloads also tend to work on less important cases. The goal of our empirical work is to discriminate between these types of explanations using panel data.

6   Empirical Strategy

The core of our empirical approach is to exploit *within-judge* variation in performance. More precisely, consider the following setup with three types of data. First, judge characteristics ,X_i, include observables such as age, experience, and education, as well as unobservables such as ability and preferences (These would include $\vec{\beta}$ from the model). Second, employment conditions Z_{it} generally vary within state over time but can also vary by judge within state (due to a staggered electoral cycle, for example). These variables include compensation, rules for appointment and retention, term length, etc. Third, judge performance outcomes Y_{it} are constructed from the sample of judicial opinions as described in Section 5.3.

The goal is to compare judge performance under conditions Z^A and Z^B. The causal effect of potential outcome A compared to B is given by:

$$Y_{it}^A - Y_{it}^B = F\left(Z^A, X_i, t\right) - F\left(Z^B, X_i, t\right)$$

where $F(\cdot)$ describes the outcome as a function of the treatment (Z), judge characteristics (X_i), and time (t). As Holland (1986) observes, measuring the causal effect is impossible because it requires comparing different treatments to the same judges administered at the same time.

Previous papers on judicial employment conditions have taken two different approaches to solving this identification problem. The structural approach, best-known from the industrial organization literature (e.g., Angrist and Pischke, 2010), assumes that the model of judge behavior is known and that only parameter values are unknown. Two leading applications of this approach to state supreme court judges are Lim (2013) and Iaryczower et al. (2013). These papers illuminate the relative importance of different mechanisms assuming that those

mechanisms exist. However, a causal interpretation relies on the correctness of a model that cannot in general be proven to be correct.

A second approach is to estimate cross-sectional effects conditional on observables. The two leading applications of this approach to state supreme court judges are Choi et al. (2009) and Choi et al. (2010). This approach assumes that X_i captures all the relevant characteristics of a judge. Then whenever $X_i = X_j$,

$$F\left(Z, X_i, t\right) = F\left(Z, X_j, t\right), \forall Z.$$

If the treatment Z is assigned to different judges, then the causal effect is given by:

$$Y_{it}^A - Y_{jt}^B = F\left(Z^A, X_i, t\right) - F\left(Z^B, X_j, t\right)$$

and no time variation is needed. The problem with this approach is that judges and courts may have unobserved characteristics that vary systematically by state ($X_i \neq X_j$). If so, correlation between performance and employment conditions are not likely to be causal links, but the result of differences between the judges selected.

Our solution to the identification problem is to use the panel structure in the data. Changes in state laws determining judicial employment conditions over time are viewed as "natural experiments," measuring the changes in judicial performance in response to changes in employment conditions. Formally, we measure:

$$Y_{it}^A - Y_{it-1}^B = \sum_{(i,t) \in T(A,B)} \frac{F\left(Z^A, X_i, t\right) - F\left(Z^B, X_i, t-1\right)}{\#T\left(A, B\right)},$$

where $T\left(A, B\right)$ is the set of all the judges i and periods t where employment rule Z^A prevailed in period t and employment rule Z^B prevailed in period $t-1$. By construction, we measure *within-judge* effects of the treatment. The benefit of this approach is that many of the features of the court and judge that are likely to be important, such as the number of judges or judge ability, are held fixed.

There is still the problem of secular changes from one period to the next. We follow Bertrand et al. (2004) and address this issue with the inclusion of state time trends and time dummies. Under relatively weak conditions one can correctly identify the effect of a change at the state level on individuals in that state. A number of studies have used this approach in a law-and-economics context, including Miles (2000), Autor et al. (2004), Autor

et al. (2006), MacLeod and Nakavachara (2007), Currie and MacLeod (2008), Carvell et al. (2012) and Avraham et al. (2013).

The econometric specification is a linear model estimated by ordinary least squares. Records are indexed by ist, representing judge i, state s, and year t.[15] The set of judge characteristics, described in Section 5.1, is represented by X_i. The vector of treatment variables, described in Section 5.2, is represented by Z_{ist}; what is included varies by regression and is described in more detail in the results section. The performance measure is denoted by Y_{ist}, as described in Section 5.3. It is constructed from the sample of opinions written by judge i working in state s during year t. The outcome variables in the regressions are $y_{ist} = \log(1 + Y_{ist})$; coefficients can therefore be interpreted as proportional changes due to reforms.[16]

The estimated equation is given by:

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + Z'_{ist}\rho + \epsilon_{ist}$$

where TIME_t is a fixed effect for year 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 year fixed effect allows for arbitrary national trends in the performance variable. The judge fixed effect controls for time-invariant state-level and judge-level characteristics. The state time trends control for preexisting trends in the performance variable that may be confounded with changes in state laws.[17]Again following Bertrand et al. (2004), standard errors are clustered at the state level, allowing for heteroskedasticity and for arbitrary correlation of the error term within state across judges and across time. This is sensible because unobserved shocks to performance are likely correlated within the same court.[18]

---

[15]More precisely, it is *court s*, since Oklahoma and Texas each have two courts of last resort.

[16]The variable $1 + Y_{ist}$ is used to account for zeros in the data; the means of the variables are mostly far from zero. The results are robust to using levels or the inverse hyperbolic sine rather than logs of the dependent variable. The adjusted R^2 is usually higher in the logs specification than in levels.

[17]Our results are robust to the inclusion of a full set of dummies for years of judge experience, meaning that the effects are not generated by mechanical changes in judge human capital. The results are robust to removing the first and last years of each judge's career, meaning that they are not generated by outliers related to different case compositions for younger/older judges, or for judges transitioning between positions. We have done the regressions with all the opinions (not just those with more than seven sentences), which strengthens our results. Finally we have done regressions using state-year data (rather than judge-year data) that includes unauthored (per curiam) opinions and those authored by special judges, with substantially similar results.

[18]The statistical tests are robust to using two-way clustering by state and year. See Cameron et al. (2011).

7 Results

This section presents four sets of empirical results. First, Section 7.1 reports results on establishing an intermediate appellate court. The results on extending judge term lengths are in Section 7.2. Results on increasing judge salaries are presented in Section 7.3. Section 7.4 reports the effect of being up for election.

7.1 Effect of an Intermediate Appellate Court

First consider the effect of introducing an intermediate appellate court (IAC). IAC judges take over much of the caseload and and give the supreme court judges more discretion in whether to accept cases for review. Reducing the caseload should mechanically reduce time demands for the supreme court judges, and hence these regressions provide a robustness check.

A concern is that the introduction of the IAC may be correlated with judge performance. These courts were established because the supreme court was overworked, and anecdotal evidence suggests that the supreme court judges actively lobbied for the lower court.[19] We can measure the effect of the treatment on the judges in these states but cannot make strong external validity claims about the effects of an IAC in states that did not choose to establish one. What we have done is checked carefully whether there were other significant reforms to the court system at the same time as an IAC being established. We have not considered any of those reforms in our treatments (although including them does not substantially change the results).[20] A detailed history of these reforms is available from the authors upon request.

As discussed in Section 6, our basic regression model is given by

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \bar{Z}'_{st}\bar{\rho} + Z'_{st}\rho + \epsilon_{ist} \tag{7.1}$$

where TIME_t is a fixed effect for year t, JUDGE_i is a judge fixed effect, and $\text{STATE}_s \times t$

---

[19]For example, the Massachusetts judiciary web site states: "The Supreme Judicial Court's appellate caseload had greatly expanded through the late 1950s and 1960s. Expansion was fueled in part by a huge increase in criminal appeals... Supreme Judicial Court Chief Justice Joseph Tauro, with considerable support and assistance from SJC Clerk John E. Powers, leaders of the Legislative and Executive Branches, and the state's bar associations, succeeded in getting an intermediate appellate court established in 1972." Besides Massachusetts, we found historical evidence of this justification for Florida, Arizona, New Mexico, Maryland, North Carolina, Colorado, Iowa, Kansas, Wisconsin, Arkansas, Hawaii, Alaska, Idaho, Connecticut, Minnesota, and North Dakota.

[20]We have not excluded reforms that occurred coterminously with salary increases (Florida, New Mexico, Washington, Colorado, Kansas) or establishment of a mandatory retirement age (Florida, Massachusetts).

is a state-level linear time trend for state s. The term $\bar{Z}_{st}$ is a vector of indicators equaling one for the baseline time windows of ten years before and ten years after each of the policy changes discussed in Section 4. Z_{st} is a vector of treatment indicators for the ten years after each rule change.[21] Thus, with the inclusion of the judge fixed effects, the estimates for the elements of ρ 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. Let ρ_{IAC} be the component of ρ corresponding to the IAC treatment, with estimate $\hat{\rho}_{IAC}$.

The level of discretion given to state supreme court judges in selecting cases for review should affect how they reallocate their time in response to changes in employment conditions. Therefore, the effects of the treatments are interacted with the level of discretion given in each state. The second regression model is

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \bar{Z}'_{st}\bar{\rho} + Z'_{st}\rho + \gamma^{LD} Z'_{st}\rho^{LD} + \epsilon_{ist} \qquad (7.2)$$

where γ^{LD} is a dummy equaling one in the states with limited discretion – whether partial discretion or mandatory review, as described in Section 4 and listed in Table 1.3. In this specification, the coefficients of $\hat{\rho}$ measure the baseline effect of establishing an IAC in the full-discretion system, while the coefficients of $\hat{\rho}^{LD}$ give the additional interacted effect under limited discretion.

[TABLE 3.1 HERE]

The estimated effects of establishing an IAC are reported in Table 3.1. Each row corresponds to a different outcome variable, listed in the leftmost column. The items in Columns 1 through 3 give the estimated $\hat{\rho}_{IAC}$ from Equation (7.1), where subsequent columns add additional fixed effects and/or trends. Column 1 estimates the IAC effect by comparing across states within years, basically giving the average difference in supreme court judge performance between states that have an intermediate appellate court and those that don't. Column 2 gives the within-state effect, which compares how the supreme court as a whole did after establishing the IAC, with both sitting judges and newly arrived judges. Column 3 gives the within-judge effect, looking only at the average treatment effect on sitting judges at the time of the rule change. The estimates in Columns 4a and 4b are derived from Equation (7.2), where Column 4a includes the baseline effect $\hat{\rho}_{IAC}$ and Column 4b includes the

---

[21]As discussed in Section 4 and illustrated in Figure 1, the time window is smaller for several of the reforms.

interacted effect $\hat{\rho}_{IAC}^{LD}$. For this table, we focus on Columns 3, 4a, and 4b, which give the within-judge effect of establishing an IAC.

Overall the number of opinions decreases after the reform, reflecting a decreased workload (Column 3, first row). Columns 4a and Column 4b show that the effect is focused in the full-discretion states, with a statistically negative interaction effect under limited discretion which fully cancels out the effect. This is consistent with the idea that lower discretion constrains the ability of judges to reduce their caseload in response to the reform.

In Columns 3 and 4a, the estimated coefficient on total words written is much smaller in magnitude than that on number of opinions, and in the aggregate it is statistically insignificant. The reason can be seen in the row on majority opinion length – after the establishment of an IAC, state supreme court judges are using their extra time to write longer, more detailed majority opinions. This effect is almost totally focused in the full-discretion states, with limited-discretion states (Column 4b) having a significantly smaller effect on majority opinion length. Correspondingly, the cases are more well-researched (Length of Table of Cases), an effect which is also strongest (though not significantly so) under full discretion.

The longer, more well-researched opinions are also of higher quality – but only under full discretion. The coefficients on positive cites, discussion cites, quote cites, and out-of-state cites are all significantly positive in the aggregate (Column 3), and they are even larger under full discretion. The relative effect in limited-discretion states is weaker (that is, negative) for all of the citation measures, with statistical significance on most measures. Finally, the increase in distinguishing and negative cites may reflect a greater level of judicial independence and experimentation, which is consistent with their having more time and energy in opinion-writing.

Next we analyze the dynamics of the judge response to establishing an IAC by estimating

$$y_{ist} = \text{JUDGE}_i + \sum_{\tau=-6}^{-1} \rho^\tau Z_{st}^\tau + \sum_{\tau=1}^{6} \rho^\tau Z_{st}^\tau + \epsilon_{ist}$$

where JUDGE_i is a judge fixed effect and the indicator variables $Z_{st}^\tau, \tau \in \{-6, ..., -1, 1, ..., 6\}$ are the leads and lags from the IAC reform year. In particular, given IAC reform year T_s in state s, Z_{st}^τ equals one for judges working in state s at year $T + \tau$. The sample is restricted to treated judges, and to the six years before and after the IAC reform. This means the coefficients can be interpreted as the average within-judge log deviations in the performance variable relative to the year the reform was enacted.

[FIGURE 3.1 HERE]

31

This regression is run separately in states with full discretion and limited discretion over case selection after the IAC reform. As illustrated in Figure 3.1, the IAC had a large effect in the states with full discretion (left column of graphs) and a more muted effect in the states with limited discretion (right column of graphs). Under full discretion, we see clear and immediate effects on the number of opinions, caselaw research, and citations. Under limited discretion, we only see a performance improvement on the caselaw research margin. This is consistent with the importance of discretion in the effectiveness of an IAC in reducing the caseload.

The effects on per-opinion performance measures in Table 3.1 are potentially from two sources. First, an IAC increases judge discretion over case selection, so they may be selecting a set of cases that are more interesting on average. Second, judges may be putting more effort into the cases left over. The average length, research, and cites might increase without increases in judge effort just because the set of cases is different. To study this, we look at how the performance measures change on a fixed number of the most important cases in a judge's portfolio, with the idea that higher performance on these cases is due to higher effort in response to having more time. Specifically, a set of judge-year performance data are constructed using averages from the five lowest-quality opinions and five highest-quality opinions published for each judge, ranked by the number of positive citations. This necessitates the exclusion of a handful of judge-years with fewer than ten opinions. If there is a change only in the bottom-quality cases, that suggests the IAC effect consists solely of a change in the composition of cases. If there is a change in the top-quality cases as well, that suggests the judges are using their extra time to put more work into important cases.

[TABLE 3.2 HERE]

The effect of establishing an intermediate appellate court on the bottom five and top five cases by quality are reported in Table 3.2. As with Table 3.2, we report the interacted effect with limited discretion. As expected, the bottom 5 cases (Column 1a) show large increases in most quality measures, reflecting that the least important cases are no longer being accepted for review and the tail of the distribution is being cut off. In terms of the model, this is consistent with the model's prediction that judges rule only on the more important cases first. As seen in Column 1b, these effects on the least-important cases are significantly weaker under limited discretion. These judges have less control over their caseload, and hence they cannot filter out the least-important cases.

The effects on the top 5 cases (Column 2a) are weaker, with mostly zeros but a statistically positive effect on average majority opinion length and the average length of the table of

cases. This suggests that with the extra time from a reduced caseload, judges are spending more time on their most important cases, writing longer and more well-researched majority opinions. That extra work does not translate into more citations, however. Again, discretion seems to matter, with limited discretion killing the effect on opinion length.

Overall, these results are consistent with the hypothesis that judges have an intrinsic incentive to do good work, particularly on the cases they find more interesting. When given more discretion to select cases, judges choose the more interesting cases that have a stronger impact on the law. When their workload is reduced due to a smaller caseload, they spend more time working on the cases remaining on the docket.

7.2 Effects of Term Length Changes

This section discusses the estimated effects of changes in term lengths. With a longer term of office, judges face weaker electoral incentives because they have to face election less often. Therefore increasing the term of office should result in judges spending more time on what they care about.

There were changes in supreme court term length in ten states in the sample, but the treatment is independently identified in only five states due to the co-occurrence of other reforms in five of the states. The changes are as follows: Hawaii, seven to ten years; Montana, six to eight years; South Dakota, six to eight years; Louisiana, fourteen to ten years; and Vermont, two to six years. Of these states, at the time of the reform Hawaii had governor retention, Vermont had legislative retention, and Montana, South Dakota, and Louisiana had contested elections.[22] The effects have similar magnitude but are more precise if the five other states are included while controlling for the co-occurring reforms.

[TABLE 4 HERE]

The results reported in Table 4 come from estimating Equations (7.1) and (7.2), as done with IAC's in the previous section. The table includes estimates for ρ^{TL}, where TL stands for a term length change. Column 1 gives the across-state effect, Column 2 the within-state effect, and Column 3 the within-judge effect. Columns 4a and 4b present the interacted treatment by discretion over case selection.

---

[22]If one interacts the effect with the retention system, There were similar positive effects on case-law research and quality in governor retention and contested election systems. The term length effect is actually negative overall in the legislative retention system (Vermont).

Unlike the IAC results, including judge or state fixed effects has a significant impact on the term lengths effects. In Column 1, we see that the states that increased term lengths have much lower scores on opinion quality measures than other states on average. Once one controls for a state fixed effect (Column 2), that disparity disappears. This is consistent with significant variation in measured judge quality across states that is controlled with the addition of judge or state fixed effects.

As can be seen in Column 3, when judges get more time due to changes in term length, they respond with increases in opinion quality. While there are no significant effects on the number of opinions and opinion length, there is an increase in several of the opinion quality measures. In particular, increased term length results in significant improvements in distinguishing cites, discussed cites, and quoted cites.

Looking to the analysis of discretion over case selection, similar patterns to that in the IAC are observed. The positive effects of increasing term lengths are strongest under full discretion, and weaker under limited discretion. The positive effect on caselaw research becomes statistically significant for full-discretion states. The relative effects on opinion quality under limited discretion are statistically negative for caselaw research, discuss cites, and quote cites. These estimates are consistent with the hypotheses that longer terms make intrinsic incentives more powerful, and that judges invest more in the task of opinion writing when work discretion allows.

7.3   Effects of Judge Salary Changes

This section discusses the effect of judge salary on judge performance. This analysis is motivated by Proposition 4, which states that increasing base income I_0 should reduce time spent on outside activities and thereby increase time spent on judging. The outcome y_{ist} for judge i in state s at year t is modeled as

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \rho W_{st} + X'_{st}\beta + \epsilon_{ist} \qquad (7.3)$$

where TIME_t is a fixed effect for year t, JUDGE_i is a judge fixed effect, $\text{STATE}_s \times t$ is a state-level linear time trend for state s, W_{st} is the log real annual salary paid to judges in state s at year t, and ϵ_{ist} is an error term. The set of control variables X_{st} includes treatment dummies for all of the institutional reforms described in Section 4. This regression effectively compares deviations from the detrended mean log salary to deviations from the detrended outcome variable. Outcome variables are in logs, so coefficients can be interpreted as the

predicted percent change in the outcome variable for a one percent increase in real salary.

As done with intermediate appellate courts and term lengths, we estimate a separate specification where salary changes are interacted with a dummy for limited discretion. This specification is given by

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \rho W_{st} + \rho^{LD} \gamma^{LD} W_{st} + X'_{st}\beta + \epsilon_{ist} \qquad (7.4)$$

where everything is the same as (7.3) except for the interacted salary term. As before, γ^{LD} is a dummy equaling one in the states with limited discretion, meaning that $\hat{\rho}^{LD}$ gives the relative salary effect in limited-discretion states. The estimated wage elasticity of performance for full-discretion states is given by $\hat{\rho}$; for limited discretion states, $\hat{\rho} + \hat{\rho}^{LD}$.

[TABLE 5.1 HERE]

Table 5.1 reports the salary results. Column 1, with only year fixed effects, is the closest specification to that used in Choi et al. (2009). Column 2 gives the within-state effect which would include the effect of salary increases on sitting judges and the quality of new judges. For analyzing only the incentive effect, we look at Column 3.

Column 1 shows that there are large and significant relationships between wages and the outcome variables when comparing judges across states. In states with higher salaries, judge tend to write fewer opinions, but they are longer and have more citations. However, these effects are zero when controlling for state fixed effects (Column 2) or judge fixed effects (Column 3). This is consistent with selection bias in the cross-sectional regression (Column 1). Due to other judge-level and court-level factors, those judges paid higher salaries are also the ones who write fewer, better opinions. Recall that this regression includes the other treatment variables, highlighting that there are other unobserved variables driving the effect.

However, we can see in Columns 4a and 4b that when accounting for discretionary review, there are positive effects of salary changes. In particular, both within-state and within-judge, increases in log salary in full-discretion states are associated with increases in positive cites, discuss cites, and quoted cites per opinion. The interacted effects on these measures for limited discretion are negative – and as large or possibly larger in magnitude than the positive effects for the full-discretion states.

The effects on time use of relieving time pressure from outside activities is stronger when a judge has discretion over using his time. One way to interpret this result is that time spent on judging is less valuable under limited discretion. They have less intrinsic motivation

for their job due to having less control over their work environment, so the leisure effect is stronger than the intrinsic-motivation effect.

Besley (2004) provides another potential explanation of these results. A positive effect of salary on performance could arise from an incentive effect where a larger salary increases the value of being retained for future judge terms. Because retention is more valuable, the judge works harder to be retained, and performance increases on average. If this is the key explanation, wages would increase performance in states where judges undergo a contested retention process, but should have a weaker or no effect in states where judges have tenure. To consider this hypothesis, we look at salary effects separately by the contestedness of the retention process. As listed in Table 1.2, "contested retention" includes contested elections, governor retention, and legislative retention. "Uncontested retention" includes uncontested elections and life tenure.

Formally, we define the dummy variable δ^C to equal one in those states with contested elections, governor retention, or legislative retention. The dummy variable δ^U equals one under uncontested elections and life tenure. The estimated equation is:

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + W'_{st}\rho + X'_{st}\beta + \epsilon_{ist} \tag{7.5}$$

where:
$$W'_{st}\rho = \rho^C \delta^C W_{st} + \rho^{CLD} \delta^C \gamma^{LD} W_{st} + \rho^U \delta^U W_{st} + \rho^{ULD} \delta^U \gamma^{ULD} W_{st}.$$

The coefficients $\hat{\rho}^C$ and $\hat{\rho}^U$ give the baseline salary-performance elasticities under contested and uncontested retention systems, respectively. The coefficients $\hat{\rho}^{CLD}$ and $\hat{\rho}^{ULD}$ give the respective interacted effects with limited discretion.

[TABLE 5.2 HERE]

In Table 5.2, Columns 1a, 1b, 2a, and 2b respectively report the estimates for $\hat{\rho}^C$, $\hat{\rho}^{CLD}$, $\hat{\rho}^U$, and $\hat{\rho}^{ULD}$. It is apparent that there is a striking similarity between contested and uncontested retention systems. Especially for the effects on positive cites, discuss cites, and quoted cites, the estimates are very similar. If anything, the effects are slightly stronger under weak retention incentives. Moreover, the interacted salary effects with limited discretion are strikingly similar across the retention systems.

These results contradict the hypothesis that higher performance in response to higher pay is due to an increased value of being retained for future terms of office, but consistent with the simple time allocation model. A third alternative explanation would be a behavioral reciprocity effect (e.g. Akerlof, 1982). While the evidence cannot convincingly reject a

reciprocity model, it's worth noting that most of the empirical studies on reciprocity shows that the effect is ephemeral, dying out within days or even hours (e.g. Gneezy and List, 2006; Bellemare and Shearer, 2009).

7.4   Electoral Cycle Effects

To further study the effects of retention incentives on judge behavior, consider how judges change their behavior over time in response to the election cycle. In particular, the staggered election cycle allows one to compare judges sitting on the same court who are not up for election with those who face election in the current period. The election schedule is arbitrarily assigned by history, so one can reasonably assume that it is uncorrelated with other institutional or socioeconomic factors that would affect individual judge performance. If judges who are up for election have less free time than judges who are not up for election, then this will crowd out effort in judging. This effect should be weakest in the uncontested electoral system where judges have de facto tenure.

This effect is estimated with:

$$y_{ist} = \text{JUDGE}_i + \text{STATE}_s \times \text{TIME}_t + \rho^C Z_{ist}^C + \rho^U Z_{ist}^U + \epsilon_{ist} \tag{7.6}$$

where JUDGE_i is a judge fixed effect,and $\text{STATE}_s \times \text{TIME}_t$ is a state-year fixed effect for each state s and year t. The terms Z_{ist}^C and Z_{ist}^U are indicators that equal one when judge i is up for election at year t, where C stands for contested elections and U stands for uncontested elections. Note that the dummy variable is coded as a one regardless of whether the judge actually ran for election – this is needed to avoid endogeneity problems from the judge's choice whether to actually run.

[TABLE 6 HERE]

Table 6 reports the results on the performance effect of being up for election. There are three "a-b" column pairs, with the "a" columns reporting $\hat{\rho}^C$ and the "b" columns reporting $\hat{\rho}^U$. Columns 1a and 1b include state fixed effects and state trends; Columns 2a and 2b include judge fixed effects and state trends; Columns 3a and 3b (the specification from Equation 7.6), include judge fixed effects and state-year fixed effects. While Columns 3a and 3b are preferred for measuring the causal effect on judges of being up for election, Columns 1a and 1b are useful because they summarize the average effect on the whole court when more

judges are up for election in a particular year, relative to trend (although these coefficients may be biased by other election-year factors unrelated to judicial elections).

Columns 1a and 1b include only state-year fixed effects. In contested elections, when more judges are up for election, that is associated with fewer opinions that are of lower quality. In uncontested elections, there is an opposite effect – when more judges are up for election, there are more opinions and they may be higher-quality on average. This is strong initial evidence of the differences in the incentive effects between these electoral systems. Contested elections seem to take time away from the judges, with an aggregate negative effect on output. Importantly, this means that any within-judge effects of elections are not compensated for by other judges. We don't have an explanation for why there would be more opinions in election years for uncontested systems; this is likely due to other non-election factors.

Next consider the effects in Column 3a, which give the effect of being up for election in a contested system relative to other judges on the court. With less time for judging, how do they reorganize their time? First, output is significantly reduced – they write fewer majority opinions and fewer discretionary opinions, and the opinions are shorter in length They are directly reducing the amount of work done. On the other hand, the coefficients on caselaw research and on out-of-state cites are zero. The negative coefficients on the other opinion quality measures are smaller in magnitude than the effects on opinion output. This suggests that that the judges reduce the number of opinions in order to maintain the quality of those opinions.

Finally, Column 3b reports results on uncontested elections. In contrast to Column 3a, there are no negative effects of being up for election. If anything, there are positive effects on performance of being up for an uncontested election. These results are consistent with the notion that these judges have de facto tenure and do not need to reduce their time allocation to judging based on the electoral cycle.

Finally we look at the dynamic within-judge effect of elections by estimating

$$y_{ist} = \text{JUDGE}_i + \sum_{\tau=-2}^{2} \rho^\tau Z_{ist}^\tau + \epsilon_{ist}$$

where JUDGE_i is a judge fixed effect and the indicator variables $Z_{ist}^\tau, \tau \in \{-2, -1, 0, 1, 2\}$ are the leads and lags from a judge election year. Given nearby election year T_{ist} for judge i working at year t, Z_{ist}^τ equals one for year $T + \tau$. The sample is restricted to states with elections, and only includes the last four years and first two years of a judge's term (when

38

the term length is longer than six). This means the coefficients can be interpreted as the average within-judge log deviations in the performance variable relative to the fourth-to-last year ($t = T - 4$) of a judge's term.

[FIGURE 3.2 HERE]

This regression is run separately in states with contested and uncontested elections, and performance trends are plotted separately in Figure 3.2. Consistent with the regression results, contested elections have a large effect on the number of majority and discretionary opinions, and only a small effect on the table of cases length and on citations. In the uncontested elections, meanwhile, there is no consistent trend in performance due to elections.

These trend estimates further support the hypothesis that strong electoral incentives impose time constraints, and that in response to those constraints judges reduce their time spent on judging. This behavioral response mostly consists of reducing the number of opinions written, rather than reducing opinion quality. These results are consistent with the hypothesis that judges care more about the quality rather than the quantity of their opinions.

8 Discussion

The goal of this paper has been to measure the causal effect of changes in employment conditions on judicial behavior. Given that judges have low-powered incentives that do not explicitly link pay to performance, a standard agency model would predict that reducing incentive pressure would have no effect on behavior. The evidence is inconsistent with that hypothesis; the reduction of time pressure is associated with longer, better-researched opinions that are cited more often by later judges.

When time pressure increases, judges prefer to reduce the number of opinions written rather than compromise on quality. These findings are consistent with the view that judges are professionals who care about the quality of their work, and that at the margin they prefer to maintain high quality at the cost of lower quantity. This is consistent with a model in which judges have an intrinsic desire to write high quality appellate opinions, tempered by the cost of their time. Moreover, discretion over case selection – that is, the level of control a judge has over his work material – appears to contribute to intrinsic motivation, consistent with the early work of Deci (1971), and more recent theory on intrinsic motivation (Benabou and Tirole, 2003; Prendergast, 2008; Rebitzer and Taylor, 2011).

State supreme court judges are a highly selected group of individuals, concluding a significant career in the law by serving in one of the most prestigious legal positions in their state. The evidence supports the (perhaps unsurprising) claim that state supreme court judges are motivated by professionalism (White, 1959; Wilensky, 1964). Because there are limited prospects for promotion for state supreme court judges, the results are likely due to an intrinsic motivation to good work, or a concern for their professional legacy, rather than career concerns (Dewatripont et al., 1999b; Francois, 2000; Prendergast, 2007). Extrinsic incentives that impose time pressure reduce judge performance by crowding out intrinsic motivation to do a good job, consistent with previous empirical work in other settings (Gneezy et al., 2011).

By demonstrating that judges care about their opinion output, this paper's findings strengthen previous empirical studies using features of judicial opinions to measure judicial performance (e.g., Choi et al., 2010; Epstein et al., 2013). The results support the view that both institutional rules and judge preferences are important inputs into a well-functioning legal system, which might partly explain the significant cross-country variation in the quality of legal systems (Djankov et al., 2003).

The recent work in behavioral economics has demonstrated a great deal of variation in individual preferences. In particular, this work has shown that the standard agency model's assumption that agents are motivated only by pecuniary returns is a (useful) simplification.[23] It remains an open question the extent to which one could design institutions that select individuals with particular social preferences, and in particular a preference to act in the public interest.[24] The evidence here suggests that state appellate court judges are one group of officials where institutions do matter for selection of preferences.

References

Akerlof, G. A. (1982). Labor exchange as partial gift exchange. *Quarterly Journal of Economics 97*(4), 543–69.

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

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

---

[23]See Fehr and Schmidt (1999) for a synthetic discussion of social preferences.

[24]See Lazear (1989) for an interesting paper that discusses how firms may choose employees with particular preferences.

Angrist, J. and J.-S. Pischke (2009). *Mainly Harmless Econometrics*. Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010, Spring). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Autor, D. H., J. J. Donohue, and S. J. Schwab (2004, May). The employment consequences of wrongful-discharge laws: Large, small, or none at all? *American Economic Review 94*(2), 440–446.

Autor, D. H., J. J. Donohue, and S. J. Schwab (2006, May). The costs of wrongful-discharge laws. *Review of Economics and Statistcs 88*(2), 211–231.

Avraham, R., L. S. Dafny, and M. M. Schanzenbach (2013). The impact of tort reform on employer-sponsored health insurance premiums. *Journal of Law Economics & Organization 29*.

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

Baker, S. A. (2008). Should we pay federal circuit judges more? *BUL Rev. 88*, 63.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2014). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. Technical report.

Bellemare, C. and B. Shearer (2009). Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior 67*(1), 233–244.

Benabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915.

Benabou, R. and J. Tirole (2003, JUL). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Benabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

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

Besley, T. (2004). Paying politicians: theory and evidence. *Journal of the European Economic Association 2*(2-3), 193–215.

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

Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *The*

American economic review 95(3), 616–636.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2).

Carvell, D., J. Currie, and W. B. MacLeod (2012, Spring). Accidental death and the rule of joint and several liability. *Rand Journal of Economics 43*(1), 51–77.

Choi, S., M. Gulati, and E. Posner (2008). Judicial evaluations and information forcing: Ranking state high courts and their judges. *Duke LJ 58*, 1313.

Choi, S. J., G. M. Gulati, and E. A. Posner (2009). Are judges overpaid-a skeptical response to the judicial salary debate. *Journal of Legal Analysis 1*, 47.

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

Christensen, R. K., J. Szmer, and J. M. Stritch (2012). Race and gender bias in three administrative contexts: Impact on work assignments in state supreme courts. *Journal of Public Administration Research and Theory.*

Currie, J. and W. B. MacLeod (2008, May). First do no harm? tort reform and birth outcomes. *Quarterly Journal of Economics 123*(2), 795–830.

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

Dal Bó, E. and M. Rossi (2011). Term length and the effort of politicians. *The Review of Economic Studies 78*(4), 1237–1263.

D'Aspremont, C. and L.-A. Gerard-Varet (1979, February). Incentives and incomplete information. *Journal of Public Economics 11*(1), 25–45.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic montivation. *Journal of Personality and Social Psychology 18*(1), 105–115.

Delfgaauw, J. and R. Dur (2008). Incentives and workers' motivation in the public sector. *The Economic Journal 118*(525), pp. 171–191.

Dewatripont, M., I. Jewitt, and J. Tirole (1999a). The economics of career concerns, part i: Comparing information structures. *REStudies 66*(1), 183–98.

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

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2003). Courts. *Quarterly*

Journal of Economics 118(2), 453–517.

Eisenberg, T. and G. P. Miller (2009). Reversal, dissent, and variability in state supreme courts: The centrality of jurisdictional source. *Boston University Law Review 89*, 1451–1504.

Epstein, L. and J. Knight (2013). Reconsidering judicial preferences. *Annual Review of Political Science 16*(1), 11–31.

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

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–68.

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

Francois, P. and M. Vlassopoulos (2008, February). Pro-social motivation and the delivery of social services. *CESifo Economic Studies 54*(1), 22–56.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review 87*(4), 746–755.

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

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

Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica 74*(5), 1365–1384.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives 25*(4), 191–209.

Gneezy, U. and A. Rustichini (2000, Aug). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*(3), 791–810.

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

Hall, M. and C. Bonneau (2006, January). Does quality matter? challengers in state supreme court elections. *American Journal of Political Science 50*(1), 20–33.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

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

244.

Kerr, S. (1975, December). On the folly of rewarding A, while hoping for B. *Academy of Management Journal 18*(4), 769–783.

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

Knight, J. and L. Epstein (1996, NOV). The norm of stare decisis. *American Journal oOf Political Science 40*(4), 1018–1035.

Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *The American Economic Review 87*(2), pp. 359–364.

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

Kritzer, H. M. (2015). *Justices on the Ballot: Continuity and Change in State Supreme Court Elections.* Cambridge University Press.

Lacetera, N. and M. Macis (2010). Do all material incentives for pro-social activities backfire? the response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology 31*(4), 738 – 748.

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

Lazear, E. P. (1989, December). Pay equality and industrial politics. *Journal of Political Economy 97*, 561–580.

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

MacLeod, W. B. and V. Nakavachara (2007, June). Legal default rules: the case of wrongful discharge laws. *Economic Journal 117*, F1–F62.

Mas, A. (2006, AUG). Pay, reference points, and police performance. *Quarterly Journal of Economics 121*(3), 783–821.

Miles, T. J. (2000, Apr). Common law exceptions to employment at will and us labor markets. *Journal of Law Economics & Organization 16*(1), 74–101.

Milkovich, G. T. and A. K. Wigdor (1991). *Pay for Performance: Evaluating Performance and Appraisal Merit Pay.* Washington, D.C., U.S.A.: National Academy Press.

Prendergast, C. (2007, March). The motivation and bias of bureaucrats. *American Economic Review 97*(1), 180–196.

Prendergast, C. (2008, May). Intrinsic motivation and incentives. *The American Economic Review 98*(2), 201–205.

Rebitzer, J. B. and L. J. Taylor (2011). Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 4*, Volume 4, Chapter 8. Elsevier.

Shepherd, J. (2009a). Are appointed judges strategic too? *Duke Law Journal 58*(7), 1589–1626.

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

Spottswood, M. (2007). Free speech and due process problems in the regulation and financing of judicial election campaigns. *Nw. UL Rev. 101*, 331.

Tabarrok, A. and E. Helland (1999). Court politics: The political economy of tort awards. *Journal of Law and Economics 42*(1), pp. 157–188.

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.

A Theory Appendix: Solution to Judge Optimization Problem

A.1 General Case

We consider first a general time allocation problem and then apply the solution to the judge's problem. There are n activities, indexed by $i \in \{1, .., n\}$. The time allocated to each activity is given by T_i, with vector representation $\vec{T} = \{T_1, ..., T_n\}$. Let $\bar{T} > 0$ be the total time available. The gain from each activity is represented by $U_i(T_i, \alpha_i)^{\beta_i}$, with the corresponding payoff function given by:

$$u\left(\vec{T}, \vec{\alpha}\right) = \sum_{i=1}^{n} \beta_i \log\left(U_i\left(T_i, \alpha_i\right)\right), \tag{A.1}$$

$$= \sum_{i=1}^{n} \beta_i u_i\left(T_i, \alpha_i\right). \tag{A.2}$$

Note that since U_i is (weakly) concave in T_i, then u_i is strictly concave in T_i (a strictly concave function of a concave function is strictly concave). The sum of concave functions is

concave, so $u(\cdot)$ is concave in $\vec{T}$. The agent faces the following time constraints:

$$\sum_{i=1}^{n} T_i \ \leq \ \bar{T} \tag{A.3}$$

$$T_i \ \geq \ 0, \forall i. \tag{A.4}$$

The objective is to allocate time across activities to maximize the payoff function subject to these constraints. Since $U_i(0, \alpha_i) = 0$, the non-negativity constraints are not binding and we do not need to consider them.

Existence and uniqueness follow from the standard concave optimization assumptions over a compact, convex set. We can characterize the optimum using the first order conditions for the Lagrangian:

$$L = u\left(\vec{T}, \vec{\alpha}\right) + \mu\left(\bar{T} - \sum_{i=1}^{n} T_i\right).$$

Let $\vec{T}^*(\vec{\alpha})$ denote the optimum, and let $\mu^*(\vec{\alpha})$ be the associated Lagrange multiplier. The first-order condition for T_i is given by

$$0 \ = \ L_{T_i} = \beta_i \frac{\partial u_i(T_i^*, \alpha_i)}{\partial T_i} - \mu^*. \tag{A.5}$$

Using the expression for the elasticity we have:

$$\epsilon_i(T_i, \alpha_i) = T_i^* \mu^*.$$

Next we work out the comparative static conditions. Let $J = \nabla u$ be the Jacobian for u and H be the corresponding Hessian matrix ($H_{ij} = \frac{\partial^2 u}{\partial T_i \partial T_j}$), both evaluated at $\vec{T}^*$. The first order conditions can be written in matrix form:

$$\begin{bmatrix} J - \mu^* \vec{1} \\ \bar{T} - \vec{1}^T \vec{T}^* \end{bmatrix} = \vec{0},$$

where $\vec{1}$ is a vector of ones and T denotes the transpose. We take derivatives with respect

to α_1 (the other cases are similar – this is easier to write), which gives us:

$$
\begin{bmatrix} H & ,-\vec{1} \\ -\vec{1}^T, & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \vec{T}^*}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{bmatrix} + \begin{bmatrix} \beta_1 \frac{\partial^2 u_1(T_1^*, \alpha_1)}{\partial T_1 \partial \alpha_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0}. \tag{A.6}
$$

Multiply each side by the vector of effects:

$$
\begin{bmatrix} \frac{\partial \vec{T}^*}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{bmatrix}^T \begin{bmatrix} H & -\vec{1} \\ -\vec{1}^T & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \vec{T}^*}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{bmatrix} + \begin{bmatrix} \frac{\partial \vec{T}^*}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{bmatrix}^T \begin{bmatrix} \beta \frac{\partial^2 u_1(T_1^*, \alpha_1)}{\partial T_1 \partial \alpha_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0}. \tag{A.7}
$$

Since the time constraint is always binding we have $\sum_{i=1}^n \frac{\partial T_i}{\partial \alpha_i} = \frac{\partial \bar{T}}{\partial \alpha_i} = 0$ and (A.7) implies:

$$
\frac{\partial \vec{T}^*}{\partial \alpha_1}^T H \frac{\partial \vec{T}^*}{\partial \alpha_1} = -\beta_1 \frac{\partial^2 u_1(T_1^*, \alpha_1)}{\partial T_1 \partial \alpha_1} \times \frac{\partial T_1^*}{\partial \alpha_1}. \tag{A.8}
$$

The term on the left-hand side is negative since the objective function is strictly concave, and hence H is negative definite. Since $T_i^* > 0$ then for a general activity i we conclude:

$$
\begin{aligned}
\text{sign}(\frac{\partial T_i^*}{\partial \alpha_i}) &= -\text{sign}\left(\frac{\partial \vec{T}^*}{\partial \alpha_i}^T H \frac{\partial \vec{T}^*}{\partial \alpha_i} \right) \times \text{sign}\left(\beta_i \frac{\partial^2 u_i(T_i^*, \alpha_i)}{\partial T_i \partial \alpha_i} \right) \\
&= \text{sign}\left(\beta_i \frac{\partial^2 u_i(T_i^*, \alpha_i)}{\partial T_i \partial \alpha_i} \right) \\
&= \text{sign}\left(\frac{\partial \epsilon_i(T_i^*, \alpha_i)}{\partial \alpha_i} \right).
\end{aligned}
$$

If $\text{sign} \frac{\partial \epsilon_i(T_i^*, \alpha_i)}{\partial \alpha_i} > 0$, then $\text{sign}(\frac{\partial T_i^*}{\partial \alpha_i}) > 0$. The binding time constraint implies that for some $j \neq i$ we have $\frac{\partial T_j^*}{\partial \alpha_i} < 0$. From (A.5) and the strict concavity of u_j in T_j, we have $\frac{\partial \mu^*}{\partial \alpha_i} > 0$. This implies that $\frac{\partial T_j^*}{\partial \alpha_i} < 0$ for all $j \neq i$ (as u_j is strictly concave in T_j for all j). A similar argument occurs if the sign is negative. This completes the proof of Proposition 1.

A.2 Time Allocation to Cases Sub-problem

The judging sub-problem can be viewed as finding the indirect utility function:

$$U_J\left(T_J, d_J, N_J\right) = \max_{h(\cdot), t(\cdot)} V_J\left(h\left(\cdot\right), t\left(\cdot\right), d_J, N_J\right) \tag{A.9}$$

subject to:

$$d_J N_J \int_0^\infty t\left(\gamma\right) h\left(\gamma\right) f\left(\gamma\right) d\gamma + \left(1 - d_J\right) N_J \int_0^\infty t\left(\gamma\right) f\left(\gamma\right) d\gamma \leq T_J. \tag{A.10}$$

A.2.1 First Order Conditions

Let μ_J be the multiplier for the time constraint. We can solve the problem by working out the first order conditions given the multiplier, and then choosing the multiplier to ensure the time constraint is satisfied. In this way μ_J represents the price of time in the judging sub-problem.

The Lagrangian can be written as:

$$\begin{aligned} L = \ & V_0 + \\ & d_J N_J \int_0^\infty \left[V\left(t\left(\gamma\right), \gamma\right) - \mu_J t\left(\gamma\right) h\left(\gamma\right)\right] f\left(\gamma\right) d\gamma \\ & + \left(1 - d_J\right) N_J \int_0^\infty \left[V\left(t\left(\gamma\right), \gamma\right) - \mu_J t\left(\gamma\right)\right] f\left(\gamma\right) d\gamma \\ & + \mu_J T_J \end{aligned} \tag{A.11}$$

Next, the first order condition for the allocation of time to a case of difficulty γ is given by:

$$V_t\left(t^*\left(\gamma, \mu_J\right), \gamma\right) = \mu_J. \tag{A.12}$$

Notice that the time allocated per case does not vary directly with discretion d_J nor with the case load N_J, only with the value of time (which in turn is affected by d_J).

We can see how time varies with the quality of a case since:

$$\frac{\partial t^*}{\partial \gamma} = -\frac{\frac{\partial^2 V\left(t^*\left(\gamma, \mu_J\right), \gamma\right)}{\partial t \partial \gamma}}{\frac{\partial^2 V\left(t^*\left(\gamma, \mu_J\right), \gamma\right)}{\partial t \partial t}} > 0.$$

Hence the time spent on cases increases with the importance of a case. Moreover, we also have :

$$\frac{\partial t^*\left(\gamma, \mu_J\right)}{\partial \mu_J} < 0.$$

48

As the price of time rises, the time spent per case falls.

The next issue is the choice of cases to hear. The amount of time spent on a case depends only on the quality of the case and the value of time. Hence we can write the net value of a case with quality γ given optimal time t as:

$$\Delta\left(\gamma,\mu_J\right) = V\left(t^*\left(\gamma,\mu_J\right),\gamma\right) - \mu_J\left(t^*\left(\gamma,\mu_J\right)\right).$$

Now the Lagrangian can be written as:

$$
\begin{aligned}
L &= d_J N_J \int_0^\infty \Delta\left(\gamma, t^*\left(\gamma,\mu_J\right),\mu_J\right) h\left(\gamma\right) f\left(\gamma\right) d\gamma \\
&+ \left(1 - d_J\right) N_J \int_0^\infty \Delta\left(\gamma, t^*\left(\gamma,\mu_J\right),\mu_J\right) f\left(\gamma\right) d\gamma \\
&+ \mu_J T_J.
\end{aligned}
$$

Notice that the Lagrangian is linear in $h\left(\gamma\right) \in [0,1]$, implying that:

1. $h^*\left(\gamma,\mu_J\right) = 0$ if $\Delta\left(\gamma,\mu_J\right) < 0$.

2. $h^*\left(\gamma,\mu_J\right) = 1$ if $\Delta\left(\gamma,\mu_J\right) > 0$.

Let us consider the case where the judge chooses to hear a subset of the cases over which she has discretion. Since $t^*\left(\gamma,\mu_J\right)$ is increasing in γ and $\Delta\left(0,\mu_J\right) = -c$ there is a unique $\gamma^*\left(\mu_J\right)$ such that:

$$\Delta\left(\gamma^*\left(\mu_J\right),\mu_J\right) = 0,$$

in which case the judge chooses to hear the harder cases with $\gamma \geq \gamma^*\left(\mu_J\right)$. What is interesting is that the cases she chooses are only a function of the value of time μ_J. Applying the envelope theorem to A.12 and the implicit function theorem to the equation above we get:

$$\frac{\partial\gamma^*}{\partial\mu_J} = \frac{t^*\left(\gamma^*,\mu_J\right)}{V_\gamma(t^*\left(\gamma^*,\mu_J\right),\gamma^*)} > 0.$$

As time becomes more scarce/expensive, the judge hears fewer cases (chooses a higher threshold γ^*).

The total time spent on judging given the time cost μ_J is given by:

$$
T_J\left(\mu_J, d_J, N_J\right) = N_J \int_0^\infty t^*\left(\gamma,\mu_J\right) f\left(\gamma\right) d\gamma \tag{A.13}
$$

$$
-d_J N_J \int_0^{\gamma^*\left(\mu_J\right)} t^*\left(\gamma,\mu_J\right) f\left(\gamma\right) d\gamma. \tag{A.14}
$$

Note that

$$\frac{\partial T_J}{\partial d_J} = -N_J \int_0^{\gamma^*(\mu_J)} t^*(\gamma, \mu_J) f(\gamma) \, d\gamma < 0, \tag{A.15}$$

$$\frac{\partial T_J}{\partial N_J} = \frac{T_J}{N_J} > 0 \tag{A.16}$$

From (A.12) and $V_{tt} < 0$ we have $\frac{\partial t^*(\gamma, \mu_J)}{\partial \mu_J} < 0$, which combined with $\frac{\partial \gamma^*}{\partial \mu_J} > 0$ implies:

$$\frac{\partial T_J(\mu_J, d_J, N_J)}{\partial \mu_J} < 0, \tag{A.17}$$

with the equilibrium value of time given by the solution to::

$$T_J(\mu_J^*(T_J, d_J, N_J), d_J, N_J) = T_J. \tag{A.18}$$

Finally, notice that the monotonic transformation of the preferences V_J to $(V_J - V_0)/N_j$ yields an objective function that is independent of N_J. Dividing the constraint by N_J we see that:

$$\mu_J^*\left(\frac{T_J}{N_J}, d_J, 1\right) = \mu_J^*(T_J, d_J, N_J).$$

Using (A.15-A.18) we can readily verify that:

$$\frac{\partial \mu_J^*}{\partial d_J}, \frac{\partial \mu_J^*}{\partial T_J} < 0, \frac{\partial \mu_J^*}{\partial N_J} > 0. \tag{A.19}$$

A.2.2 Comparative Static Results

An increase in V_0 increases U_J without affecting the marginal value of time in judging. Since V_0 does not affect the marginal utility of time, we have that

$$\frac{\partial \epsilon_J(T_J, d_J, N_J)}{\partial V_0} < 0.$$

which along with Proposition 1 implies the following:

Proposition 7. *An increase in the prestige from being a judge (V_0) reduces time allocated to judging.*

The reason for this result is that an increase in prestige for judging increases the marginal value of other activities, while having no effect on the marginal return from judging.

50

At the optimum the utility is equal to the Lagrangian:

$$U_J\left(T_J, d_J, N_J\right) = L\left(t^*\left(\gamma, \mu_J^*(\frac{T_J}{N_J}, d_J)\right), \gamma^*(\mu_J^*(\frac{T_J}{N_J}, d_J)), \mu_J^*(\frac{T_J}{N_J}, d_J), T_J, d_J, N_J\right),$$

and we have:

$$\frac{\partial U_J\left(T_J, d_J, N_J\right)}{\partial T_J} = \mu_J^*\left(\frac{T_J}{N_J}, d_J\right)$$

and hence:

$$\beta_J\frac{\partial U_J\left(T_J, d_J, N_J\right)}{\partial T_J} = \beta_J\mu_J^*\left(\frac{T_J}{N_J}, d_J\right) = \mu^* U_J > 0. \tag{A.20}$$

Namely, the multiplier gives us the marginal utility of time allocated to judging, which multiplied by elasticity of judging β_J is equal to the overall price of time. As time becomes more expensive, then this multiplier is larger, which will have some clear predictions regarding the allocation of time to cases.

$$\frac{\partial U_J}{\partial d_J} = -\int_0^{\gamma^*} \Delta\left(\gamma, \mu_J\right) f\left(\gamma\right) d\gamma > 0. \tag{A.21}$$

The latter inequality follows from the fact that at the optimum $\Delta\left(\gamma, \mu_J\right) < 0$ for $\gamma < \gamma^*$. Since $\frac{\partial \mu_J^*\left(\frac{T_J}{N_J}, d_J\right)}{\partial d_J} < 0$, this implies:

$$\frac{\partial^2 U_J}{\partial T_J \partial d_J} < 0.$$

This result with (A.21) and (A.20) implies

$$\frac{\partial \epsilon_J\left(T_J, d_J\right)}{\partial d_J} = \frac{T_J}{U_J}\frac{\partial^2 U_J}{\partial T_J \partial d_J} - \frac{T_J}{U_J^2}\frac{\partial U_J}{\partial T_J}\frac{\partial U_J}{\partial d_J} < 0,$$

which with proposition 1 implies that T_J falls. In particular, this implies that the cost of time μ^* falls, however the effect upon the time allocated to hard cases is ambiguous since:

$$\frac{d\mu_J^*}{dd_J} = \frac{d\mu^*}{dd_J}U_J + \mu^*\frac{dU_J}{dd_J}.$$

The first term on the left is negative, while the second term is positive. This implies proposition (5).

Next observe that if we let

$$V\left(T_J, d_J\right) = U\left(T_J, d_J, 1\right) - V_0,$$

we can write the indirect utility from judging as:

$$U_J\left(T_J, d_J, N_J\right) = V_0 + N_J V\left(\frac{T_J}{N_J}, d_J\right).$$

Thus we have:

$$
\begin{aligned}
\frac{\partial U_J}{\partial T_J} &= V_{T_J}\left(\frac{T_J}{N_J}, d_J\right) = \mu_J^*\left(\frac{T_J}{N_J}, d_J\right) > 0, \\
\frac{\partial U_J}{\partial N_J} &= V\left(\frac{T_J}{N_J}, d_J\right) - V_{T_J}\left(\frac{T_J}{N_J}, d_J\right) T_J/N_J > 0,
\end{aligned}
$$

with the second inequality following from the fact that $V\left(0, d_J\right) \geq 0$ and V is concave in T_J.[25]

We also have:

$$\frac{\partial^2 U_J}{\partial T_J \partial N_J} = -\frac{V_{T_J T_J}\left(\frac{T_J}{N_J}, d_J\right)}{N_J^2} > 0,$$

Given these we have:

$$\frac{\partial^2 u_J\left(T_J, d_J, N_j\right)}{\partial T_J \partial N_J} = \frac{1}{U_J}\left(\frac{\partial^2 U_J}{\partial T_J \partial N_J} - \frac{1}{U_J}\frac{\partial U_J}{\partial T_J}\frac{\partial U_J}{\partial N_J}\right).$$

Given that the marginal returns do not vary with V_0, then for large V_0 this term is positive, but negative for small V_0. From this we have that

$$\frac{\partial \epsilon_J\left(T_J, d_J, N_J\right)}{\partial N_J} > 0,$$

if and only if V_0 is sufficiently large. This implies proposition (6).

B   Empirical Appendix: Construction of the Data Set

The judge performance data were collected from Bloomberg Law (bloomberglaw.com), which has the data presented in a standard HTML format that is amenable to automated parsing. Because scraping the data were disallowed by the Terms of Use, we had to hire undergrad research help to manually download the cases and related metadata. We downloaded all cases published at state courts of last resort between January 1, 1947 and December 31,

---

[25]Let $f\left(x\right) = V\left(x\right) - xV_x\left(x\right)$, then $f_x = V_x - V_x - xV_{xx} \geq 0$, while $f\left(0\right) \geq 0$ and hence $f\left(x\right) \geq 0$ for $x \geq 0$.

1994. Collection of the case data began in December 2011 and concluded in November 2013.

The case data were matched to states and employment conditions from the name of the court and from the date the case was published. The cases were matched to judges by the name of the judge, using the list of judges that were working on the court at the time. When all judges have different last names, this was done by last name. When two or more judges had the same last name, we matched by first and last name. In the one case where there were two judges with the same first and last name (Thomas M. Kavanagh and Thomas G. Kavanagh of the Wisconsin Supreme Court), we matched on middle initial. To allow for misspellings, we implemented a fuzzy string matching algorithm that would match to the correct judge if one or two characters were misspelled.

The case text files had fields for the opinion text as well as metadata attached by Bloomberg staff attorneys. The opinion text was already divided between majority opinions and supplemental opinions, which could be matched to judges the same way as majority opinions. The opinion length measures included the number of words in the opinions after having removed all HTML markup. Citations to previous case were already annotated, so we could count those citations to construct the Table of Cases length measure.

The citation counts were constructed from Bloomberg Law's citation analysis feature, which gives the full record of subsequent cases that have cited each decision. This includes all state and federal court cases in the United States, up through December 2011 (when data collection began). To generate out-of-state cites, we used the state where the citing court was located (either state or federal).

We loaded the case text and institutional variables into a PostgreSQL relational database. A series of Python scripts interface with the Postgres database to parse the text, compute performance measures, and merge with the institutional variables. We exported to CSV files for statistical analysis in Stata.

TABLE 1.1

State Supreme Court Employment Conditions Changes

Rule Change	Description	Treatment Events
Intermediate Appellate Court	Before, state supreme court judges reviewed a case directly from trial, with mandatory review. After, an intermediate court reviewed the case first, and the court exercised discretionary review.	FL (1956), MI (1963), AZ (1964), NM (1965), MD (1966), NC (1967), OR (1969), WA (1969), CO (1970), MA (1972),IA (1976), KS (1976), WI (1977), AR (1978), HI (1979), AK (1980), ID (1981), CT (1982), MN (1983), VA (1984), ND (1987), NE (1990)
Term Length Change	Increase or decrease in the length of term of office before facing re-election.	HI (1968), MT (1972), SD (1972), LA (1974), VT (1974)
Compensation Increase	Increases in judge annual salary due to legislation.	All states
Electoral Cycle	In states where judges are elected, the elections are staggered with a subset of judges up for election in any given year.	States with elections (see Table 1.2).

Notes. This table summarizes the changes in state supreme court employment conditions that are used as treatments in our empirical analysis. Column 1 names the rule change, Column 2 gives a brief description, and Column 3 lists the states that enacted the rule change and the year that they did so. Note that these reforms occurred in other states but are not listed here because they co-occurred with other reforms.

FIGURE 1.1
State Supreme Court Judge Term Lengths
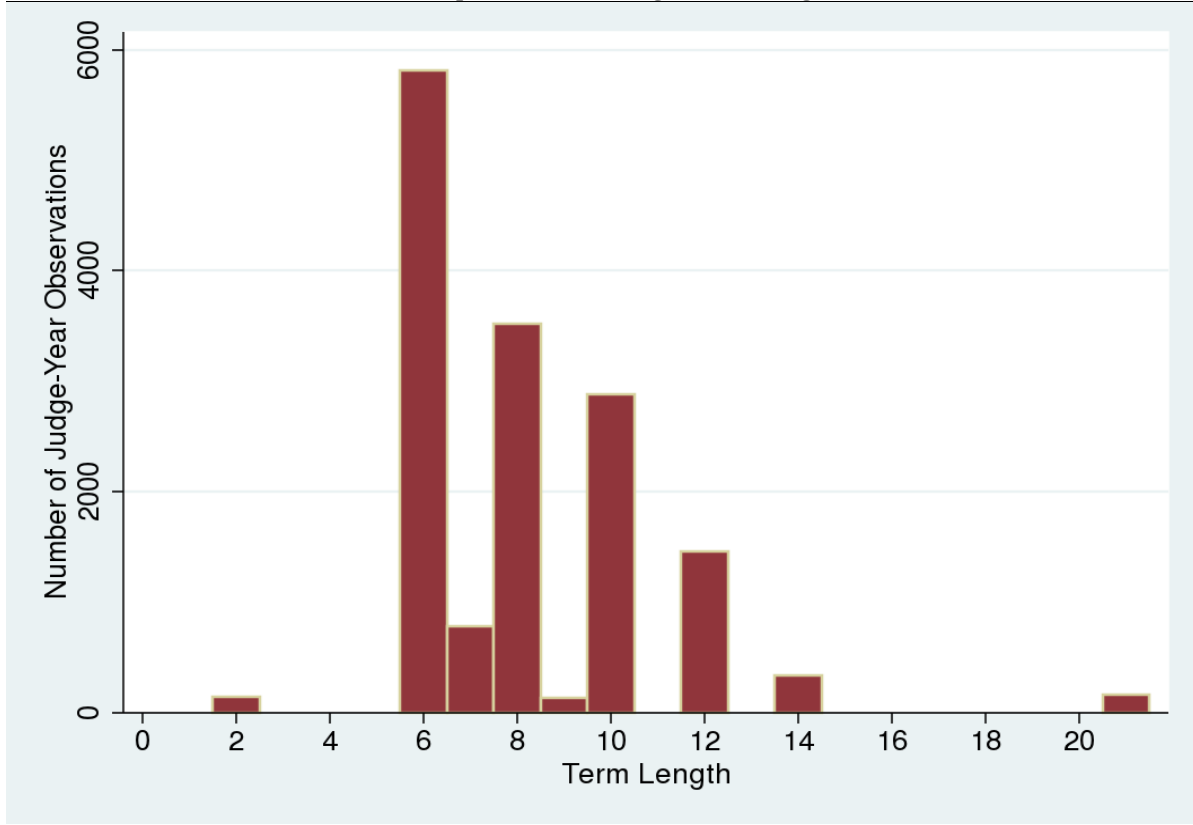
Notes. Frequency distribution over the operative term length for each judge. "Term length" is the number of years in between elections or governor reappointment.

Figure 1.2
History of State Supreme Court Reforms
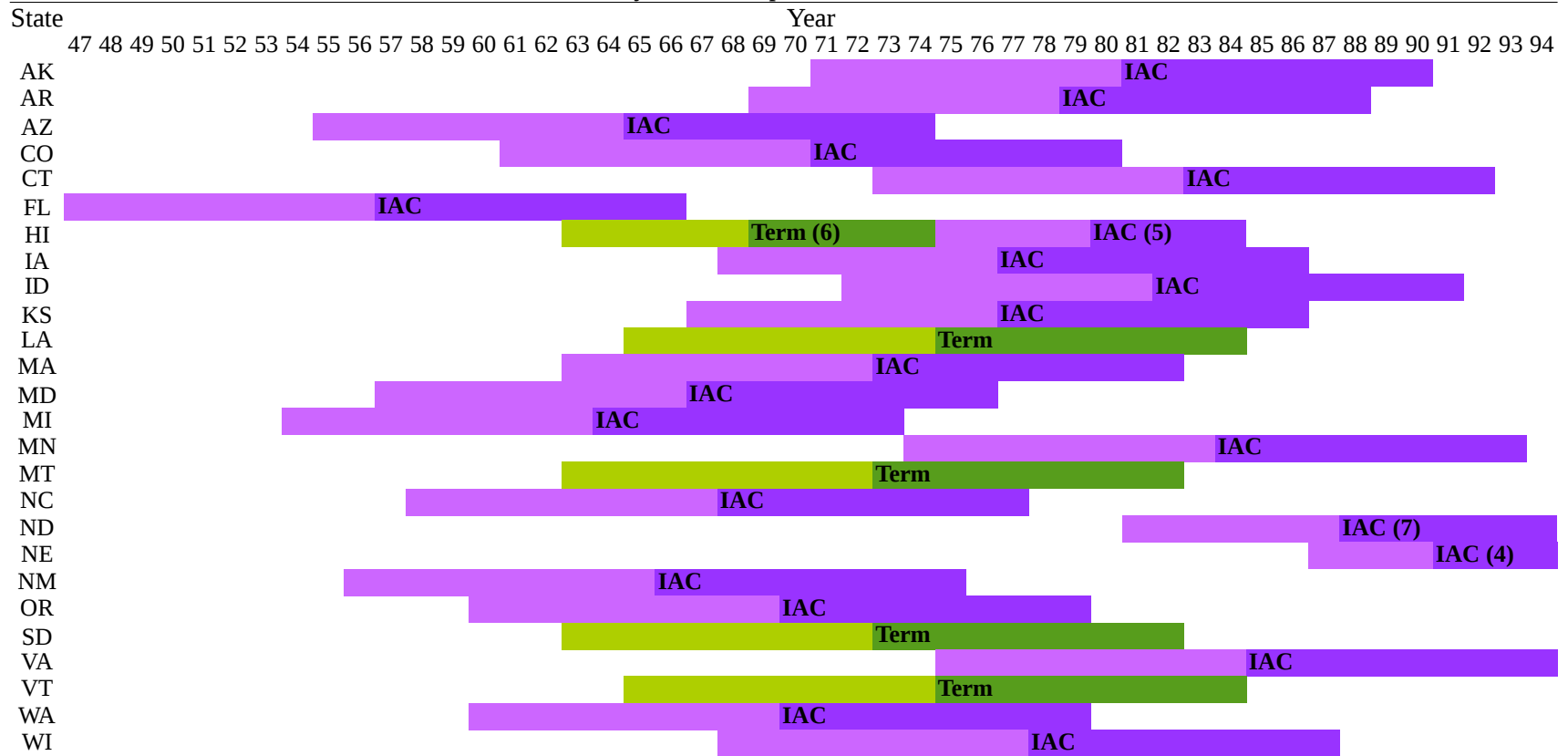
Notes. History of institutional treatments: establishment of intermediate courts (IAC) and term length changes (Term). Shaded areas indicate years included in treatment dummies. Numbers in parentheses are included when the time window used is less than ten years.

FIGURE 1.3
Mean Real Salary for State Supreme Court Judges, 1976-1994
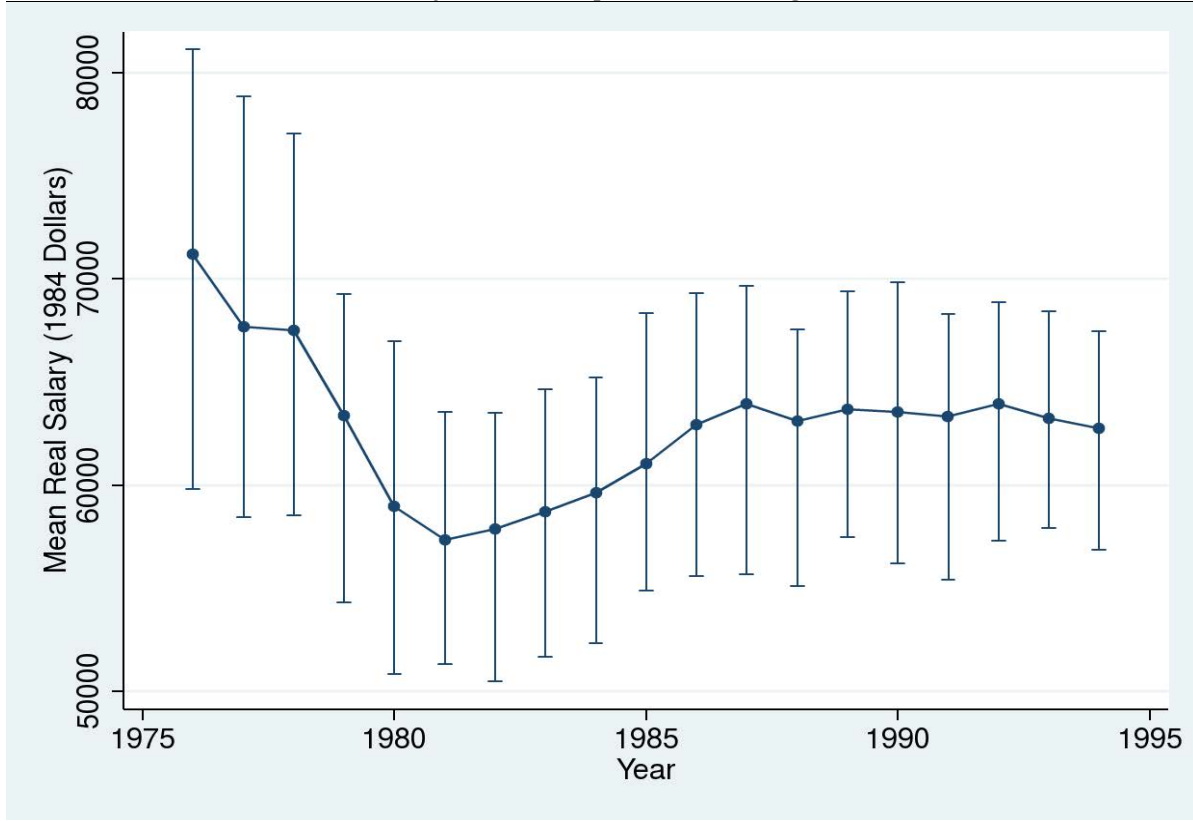
Notes. Average real salary of state supreme court judges for the years 1976 through 1994, adjusted to 1984 dollars using the BLS's CPI. Error spikes depict 25th and 75th quantiles. 1974 and 1975 omitted due to missing data for some states.

FIGURE 1.4
Number of State Supreme Court Salary Changes By Year, 1976-1994
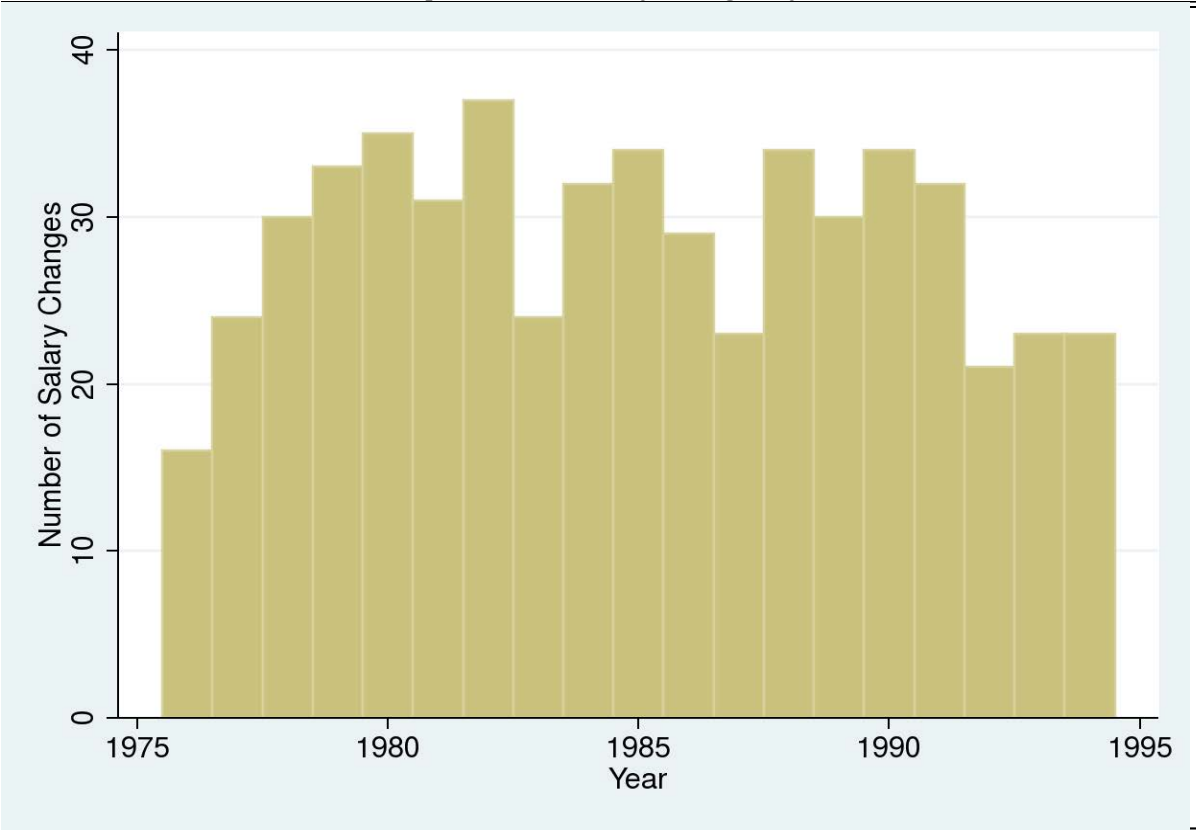
Notes. Histogram of discrete salary increases for state supreme court judges for the years 1976 through 1994. 1974 and 1975 omitted due to missing data for some states.

FIGURE 1.5
Average Percent Change in Salary Due to Discrete Changes
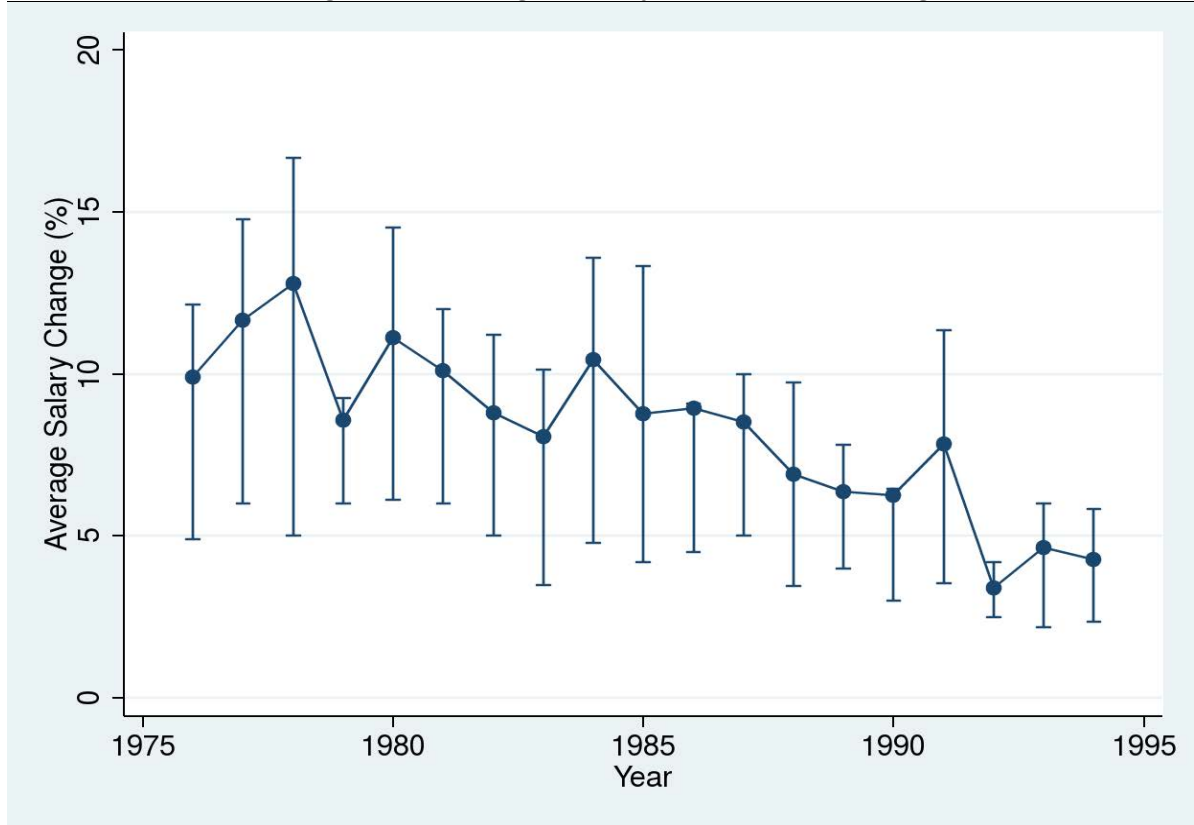
Notes. Average percent change of salary increases for state supreme court judges for the years 1976 through 1994, conditional on having an increase. Error spikes depict 25th and 75th quantiles. 1974 and 1975 omitted due to missing data for some states.

TABLE 1.2
State Supreme Court Retention Systems

Retention System	Contested or Uncontested	List of States
Partisan Elections	Contested	AL, AR, CO (-1966), FL (-1972), GA (-1982), IA (-1962), IL (-1962), IN (-1970), KS (-1958), KY (-1975), LA, MI, MS, NC, NE (-1962), NM (-1988), NY (-1977), OH, OK (-1967), TN (-1971, 1975-), TX, UT (-1951), WV
Nonpartisan Elections	Contested	AZ (-1972), FL (1973-1976), GA (1983-), ID, KY (1976-), MD (-1976), MN, MS, MT, ND, NE, NV, OR, SD (-1980), UT (1952-1984), WA, WI, WY (-1972)
Uncontested Elections	Uncontested	AK, AZ (1973-), CA, CO (1967-) FL (1977-), IA (1963-), IL (1963-), IN (1971-), KS (1968-), MD (1977-), MO, NE (1963-), NM (1989-), OK (1968-), PA (1969-), SD (1981-), TN (1972-1974), UT (1985-), WY (1973-)
Governor Retention	Contested	CT, DE, HI, ME, NY (1978-)
Legislative Retention	Contested	SC, VA, VT
Life Tenure	Uncontested	MA, NH, NJ, RI

Notes. This table lists the retention systems for state supreme court judges observed in our data. The middle column notes whether the system involved a contested retention process, as used in Table 5.2 and Table 6.Pennsylvania had single terms without retention before 1969.

TABLE 1.3
State Supreme Court Rules for Case Selection

Procedural Rule	Description	List of States
Full Discretion	State supreme court judges have full discretion in whether to review most non-death-penalty cases.	CA, CT, FL, HI, IA, ID, IL, IN, LA, MI, MT, NC, NH, NY, OK, TN, TX, VA, WA, WI, WV
Partial Discretion	State supreme court judges have discretionary review for some types of cases, but a significant proportion of cases require re-	AK, AL, AR, CO, DE, GA, KS, KY, MA, MD, MN, MO, NE, NJ, NM, OR, PA, RI, SC
Mandatory Review	State supreme court judges have no discretion whether to accept for review; some review is always mandatory.	AZ, ME, MS, ND, NV, SD, UT, VT, WY

Notes. This table summarizes rules for case selection at state courts that play a role in our empirical analysis. Column 1 names the rule, Column 2 describes it, and Column 3 lists the states that follow that rule in their supreme court.

TABLE 2.1

Summary Statistics on Judge Characteristics

	Mean	S.D.	Min	Max
Start Age	53.26	8.44	32.00	80.00
Career Length	12.69	8.07	1.00	42.00
Female	0.05	0.21	0.00	1.00
Top School	0.15	0.35	0.00	1.00
Promoted	0.04	0.19	0.00	1.00
Retire/Resign	0.73	0.44	0.00	1.00
Died In Office	0.16	0.36	0.00	1.00
Lost Election	0.06	0.23	0.00	1.00

Notes. Biographical information for N=1,636 state supreme court judges. *Start Age* is judge age upon joining the court. *Career Length* is number of years working on the court, conditional on having left the court before 2014. *Female* is a dummy for being female. *Top School* means the judge attended law school at Yale, Harvard, Columbia, Stanford, or Chicago. *Promoted* means they became a judge on a federal court. *Retire/Resign* means they left the court voluntarily. *Died in Office* means they died while sitting on the court. *Lost Election* means they were removed because they lost election.

TABLE 2.2
Summary Statistics on Institutional Variables

	Mean	Std. Dev.	Min	Max
Intermediate Appellate Court	0.72	0.45	0.00	1.00
Real Salary	621.61	109.96	399.97	1177.73
Number of Judges	6.49	1.27	3.00	9.00
Term Length	8.29	2.60	2.00	21.00
Non-Partisan Elections	0.25	0.43	0.00	1.00
Partisan Elections	0.31	0.46	0.00	1.00
Uncontested Election	0.20	0.40	0.00	1.00
Governor Retention	0.09	0.28	0.00	1.00
Life Tenure	0.08	0.27	0.00	1.00
Legislative Retention	0.06	0.24	0.00	1.00
Min Age Requirement	0.32	0.47	0.00	1.00
Max Age Requirement	0.57	0.49	0.00	1.00
Partial Discretion	0.38	0.48	0.00	1.00
Mandatory Review	0.18	0.39	0.00	1.00

Notes. Observation is a state-year. *Intermediate Appellate Court* is a dummy for state-years where the court had an IAC operating. *Real Salary* is deflated by 1984 dollars. *Number of Judges* is the number of judge positions on the court (rather than the number actually sitting due to vacancies). *Term Length* is the number of years in between elections or reappointment. *Nonpartisan Election, Partisan Election, Uncontested Election*, *Governor Retention, Life Tenure*, and *Legislative Retention* are dummies for the methods of judge retention described in Section 4. *Min Age Requirement* is a dummy equaling one in state-years that judges have to be above a certain age to join the court. *Max Age Requirement* is a dummy for state-years that judges have a mandatory retirement age. *Partial Discretion* is a dummy for having partial discretion in selecting cases for review. *Mandatory Review* is a dummy for having to review all cases. This leaves 44.4% of the state-years left over for discretionary review.

TABLE 2.3
Summary Statistics on Case Characteristics

Outcome Variable	Full Sample		Authored Majority Opinions	
	Mean	Std. Dev.	Mean	Std. Dev.
Majority Opinion Length (Words)	984.38	1663.96	2228.50	1894.34
Table of Cases Length	9.01	18.58	20.59	24.04
Positive Cites	5.19	19.35	11.74	27.81
Distinguishing Cites	0.77	3.21	1.82	4.82
Negative Cites	0.20	1.05	0.47	1.59
Discuss Cites	1.15	3.39	2.64	4.92
Quoted Cites	1.23	4.88	2.83	7.20
Out-of-State Cites	0.76	3.94	1.73	5.94
Affirm	0.24	0.41	0.50	0.48
Reverse	0.12	0.31	0.25	0.42
Remand	0.10	0.29	0.20	0.39
Modify	0.03	0.16	0.06	0.23
Civil Law Case	0.28	0.45	0.50	0.50
Criminal Law Case	0.22	0.42	0.30	0.46
Administrative Law Case	0.04	0.20	0.08	0.26
Constitutional Law Case	0.05	0.22	0.10	0.29
One Dissent	0.05	0.21	0.10	0.30
Two Dissents	0.01	0.07	0.01	0.11
Three Dissents	0.00	0.03	0.00	0.04
One Concurrence	0.02	0.15	0.05	0.22
Two Concurrences	0.00	0.05	0.01	0.07
Three Concurrences	0.00	0.02	0.00	0.03
One Concurring/Dissenting Opinion	0.01	0.09	0.02	0.13
Two Concurring/Dissenting Opinions	0.00	0.03	0.00	0.04
Number of Observations	1025461		387905	

Notes. Means and standard deviations of case data. The first pair of columns are from the full sample of cases, while the second pair of columns are from the set of authored majority opinions, at least 7 sentences long, that are used in the empirical analysis. See accompanying text for definitions of variables.

TABLE 2.4
Summary Correlations of Case Characteristics

	Maj. Op. Length	TOC Length	Pos Cites	Dist Cites	Neg Cites	Discuss Cites	Quoted Cites	Out-State Cites	Dissents	Concurs
Maj. Op. Length	1.000									
TOC Length	0.744	1.000								
Pos Cites	0.218	0.232	1.000							
Dist Cites	0.329	0.350	0.531	1.000						
Neg Cites	0.276	0.309	0.462	0.717	1.000					
Discuss Cites	0.308	0.314	0.658	0.537	0.412	1.000				
Quoted Cites	0.276	0.293	0.752	0.483	0.401	0.682	1.000			
Out-State Cites	0.156	0.168	0.353	0.277	0.173	0.398	0.447	1.000		
Dissents	0.092	0.081	0.075	0.114	0.114	0.125	0.080	0.039	1.000	
Concurrences	0.094	0.088	0.070	0.092	0.095	0.116	0.078	0.040	0.123	1.000

Notes. Observation is a case, N=387,905. Pairwise correlations between case-level variables. All correlations are statistically significant with p<.01, using Bonferroni-adjusted significance levels. Includes majority opinions that are authored by a state supreme court judge and at least seven sentences long. See accompanying text for variable definitions.

FIGURE 2.1

Distribution of Time Between Case Publication and Citation
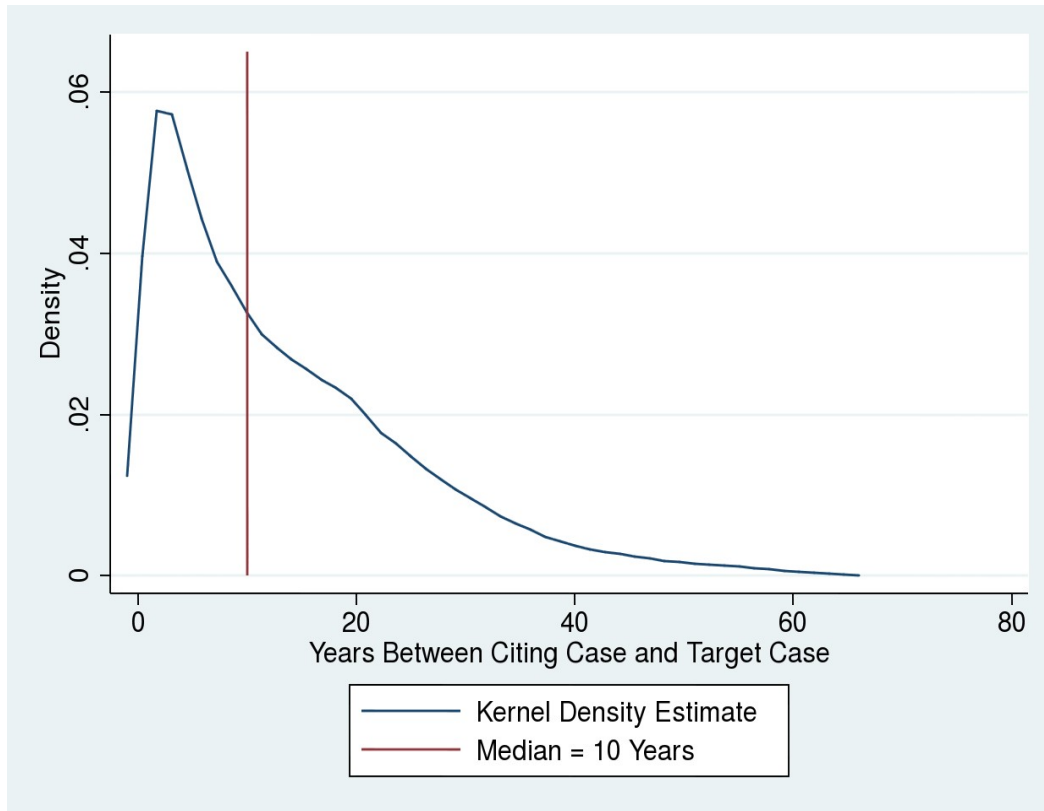
Notes. Kernel density estimate of case citations by the number of years between the original case and the citing case, using the Epanechnikov kernel with bandwidth=1. Includes all cases. Vertical line indicates median delay of 10 years. The kink in the distribution at 18-19 years reflects that our case data ends in 1994 and citations go through 2012.

TABLE 2.5
Summary Statistics on Judge-Year Performance Variables

Outcome Variable	Mean	Std. Dev.	Min	Max
Number of Majority Opinions Written	24.98	16.11	1.00	330.00
Number of Dissents Written	3.76	5.70	0.00	153.00
Number of Concurrences Written	1.84	3.48	0.00	60.00
Total Words Written	63831.30	40462.83	298.00	429770.00
Average Length of Majority Opinion	2485.19	1355.36	47.00	38240.00
Average Length of Table of Cases	22.95	17.07	0.00	557.50
Positive Cites Per Opinion	13.26	13.14	0.00	373.00
Distinguishing Cites Per Opinion	2.18	2.78	0.00	51.59
Negative Cites Per Opinion	0.59	0.91	0.00	20.63
Discuss Cites Per Opinion	3.01	2.78	0.00	105.50
Quoted Cites Per Opinion	3.36	4.28	0.00	196.50
Out-of-State Cites Per Opinion	1.86	2.54	0.00	82.43
Total Positive Cites	293.74	277.46	0.00	6134.00
Total Distinguishing Cites	45.63	56.26	0.00	877.00
Total Negative Cites	11.82	16.12	0.00	246.00
Total Discuss Cites	66.04	54.87	0.00	902.00
Total Quoted Cites	70.86	73.27	0.00	1527.00
Total Out-of-State Cites	43.21	77.26	0.00	2308.00
Total Federal Circuit Cites	16.94	107.92	0.00	3503.00

Notes. Observation is a judge-year, N=15,486. These statistics are constructed from each judge's yearly output of cases. "Per Opinion" measures are divided by the number of majority opinions written that year. See variable definitions in the accompanying text.

TABLE 2.6
Summary Correlations by Judge-Year Output Measure

	# Maj Ops	# Dis Ops	# Con Ops	Total Words	Maj Op Length	TOC Length	Pos Cites	Dist Cites	Neg Cites	Discuss Cites	Quoted Cites	Out-State Cites
# Maj Ops	1.000											
# Dis Ops	0.2148*	1.000										
# Con Ops	0.1961*	0.6213*	1.000									
Total Words	0.5745*	0.4352*	0.4032*	1.000								
Maj Op Length	-0.2921*	-0.016	0.0351*	0.3692*	1.000							
TOC Length	-0.2139*	0.007	0.0595*	0.3397*	0.8113*	1.000						
Pos Cites	-0.1783*	0.1354*	0.1289*	0.1261*	0.3942*	0.4170*	1.000					
Dist Cites	-0.1985*	0.0998*	0.0838*	0.1769*	0.4295*	0.4653*	0.6061*	1.000				
Neg Cites	-0.1970*	0.0986*	0.0847*	0.1158*	0.3547*	0.4003*	0.5661*	0.8659*	1.000			
Discuss Cites	-0.2051*	0.1249*	0.1580*	0.1580*	0.5072*	0.5120*	0.7919*	0.4939*	0.4256*	1.000		
Quoted Cites	-0.1903*	0.0942*	0.0998*	0.1597*	0.5109*	0.5552*	0.8609*	0.5142*	0.4443*	0.8544*	1.000	
Out-State Cites	-0.0825*	0.011	0.0354*	0.0889*	0.2347*	0.2568*	0.4027*	0.2382*	0.1820*	0.4668*	0.4491*	1.000

Notes. Observation is a judge-year, N=15,486. These statistics are constructed from each judge's yearly output of cases. A star * indicates that the correlation is statistically significant with p<.01, using Bonferroni-adjusted significance levels. Citation measures are per opinion. See vari-able definitions in the accompanying text.

TABLE 3.1
Effect of Reform: Establishment of Intermediate Appellate Court

Outcome	Aggregate Effects			Baseline Effect	Limited Discretion
	(1)	(2)	(3)	(4a)	(4b)
Log Number of Majority Opinions Written	-0.225**	-0.166*	-0.162**	-0.283**	0.256**
	(0.0792)	(0.0664)	(0.0516)	(0.0628)	(0.0752)
Log Number of Discretionary Opinions Written	-0.146+	-0.0569	-0.0322	-0.0160	-0.0431
	(0.0837)	(0.0613)	(0.0495)	(0.0670)	(0.0724)
Log Total Words Written	-0.117*	-0.124*	-0.0619	-0.117*	0.0991
	(0.0570)	(0.0486)	(0.0381)	(0.0493)	(0.0667)
Log Average Length of Majority Opinion	0.115*	0.0429	0.100**	0.158**	-0.134*
	(0.0541)	(0.0387)	(0.0293)	(0.0340)	(0.0534)
Log Average Length of Table of Cases	0.123*	0.0494	0.111**	0.142**	-0.0791
	(0.0515)	(0.0374)	(0.0252)	(0.0350)	(0.0487)
Log Positive Cites Per Opinion	0.101+	0.0640+	0.0869**	0.115**	-0.0842+
	(0.0578)	(0.0367)	(0.0281)	(0.0356)	(0.0430)
Log Distinguishing Cites Per Opinion	0.198**	0.112*	0.106**	0.160**	-0.136*
	(0.0692)	(0.0455)	(0.0362)	(0.0515)	(0.0588)
Log Negative Cites Per Opinion	0.128**	0.0960**	0.0821**	0.134**	-0.126**
	(0.0435)	(0.0330)	(0.0269)	(0.0401)	(0.0450)
Log Discuss Cites Per Opinion	0.0722+	0.0397	0.0620**	0.0793**	-0.0551+
	(0.0408)	(0.0281)	(0.0216)	(0.0233)	(0.0292)
Log Quoted Cites Per Opinion	0.0899	0.0387	0.0669**	0.0867**	-0.0609+
	(0.0549)	(0.0328)	(0.0239)	(0.0276)	(0.0304)
Log Out-of-State Cites Per Opinion	-0.0157	0.0308+	0.0417*	0.0478*	-0.0210
	(0.0637)	(0.0167)	(0.0189)	(0.0221)	(0.0329)
Fixed Effects	None	State	Judge	Judge	
Trends	None	State	State	State	
Treated States	22	22	22	10	12
Treated Judges	139	139	139	66	73

N= 15,277 judge-years. Treatment states listed in Table 1.1. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Columns 1 through 3 report aggregate effects. Columns 4a and 4b report the baseline treatment effect (4a) interacted with a dummy for limited discretion over case selection (4b).

FIGURE 3.1

Dynamic Judge Responses to Establishment of Intermediate Appellate Courts
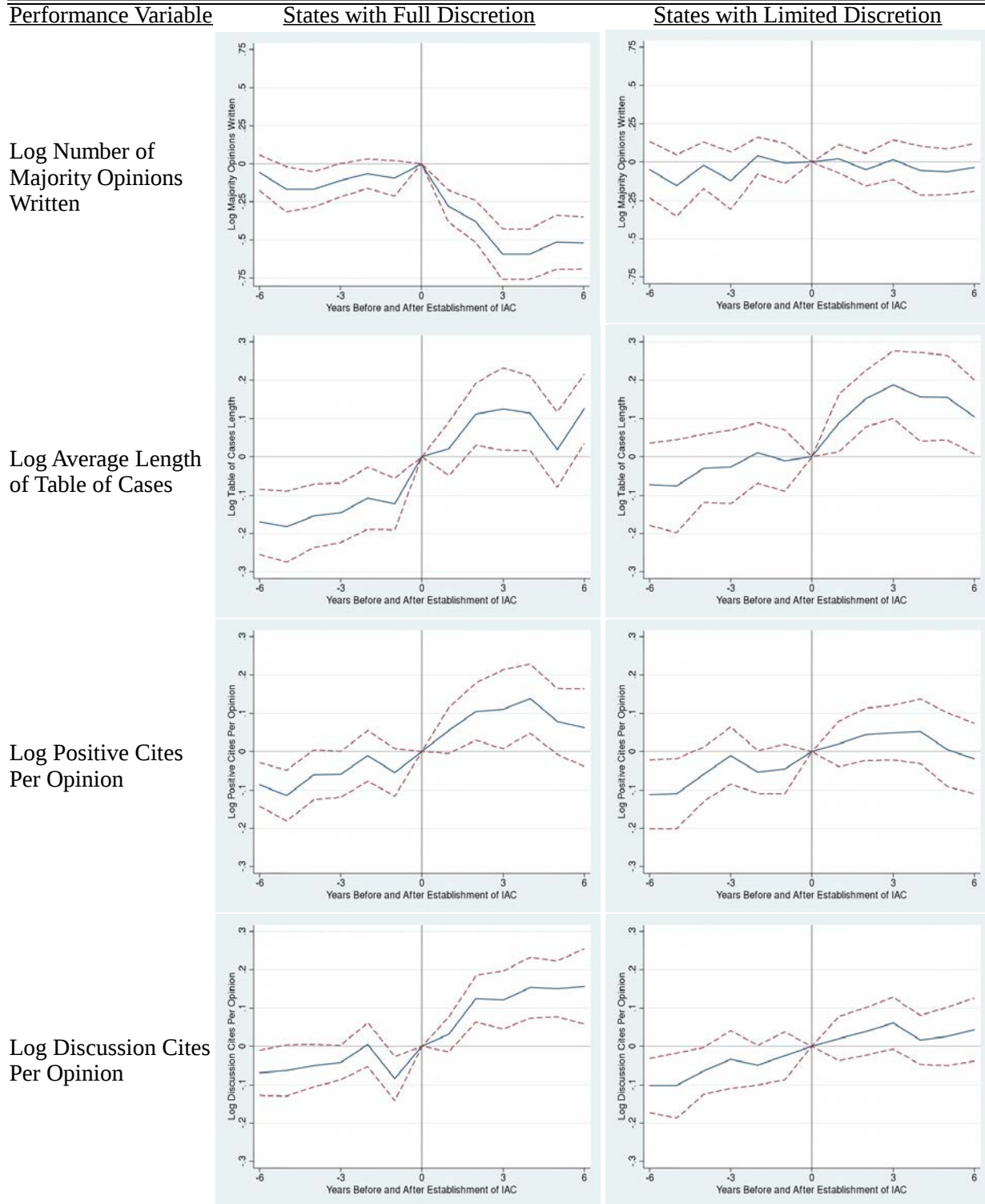
Performance Variable	States with Full Discretion	States with Limited Discretion

Notes. Plotted coefficients obtained from regression of the performance variable on indicators for six years before and after IAC reform, relative to the year the reform was enacted. Regressions include judge fixed effects. Error bands indicate 90% confidence intervals using standard errors clustered by judge. Lists of states with full discretion and limited discretion are listed in Table 1.3.

TABLE 3.2
Effect of Establishing an Intermediate Appellate Court on Bottom-Quality and Top-Quality Cases

Outcome	Bottom 5 Cases		Top 5 Cases	
	Baseline Effect	Limited Discretion	Baseline Effect	Limited Discretion
	(1a)	(1b)	(2a)	(2b)
Log Average Length of Majority Opinion	0.152**	-0.079	0.0815**	-0.0912*
	(0.0362)	(0.0592)	(0.0276)	(0.0444)
Log Average Length of Table of Cases	0.178**	-0.0412	0.0715+	-0.0483
	(0.0566)	(0.0812)	(0.0364)	(0.0401)
Log Positive Cites Per Opinion	0.256**	-0.188*	0.0049	-0.141
	(0.0733)	(0.0784)	(0.0286)	(0.0925)
Log Distinguishing Cites Per Opinion	0.0627*	-0.0471+	0.0289	-0.0118
	(0.0242)	(0.0266)	(0.0594)	(0.0707)
Log Negative Cites Per Opinion	0.0577**	-0.0592**	0.0886+	-0.0884
	(0.0190)	(0.0213)	(0.0454)	(0.0579)
Log Discuss Cites Per Opinion	0.124**	-0.0711	0.00247	-0.109
	(0.0386)	(0.0463)	(0.0311)	(0.0655)
Log Quoted Cites Per Opinion	0.114*	-0.0799+	-0.0225	-0.0959
	(0.0433)	(0.0420)	(0.0241)	(0.0747)
Log Out-of-State Cites Per Opinion	0.0552	-0.04	-0.033	-0.109
	(0.0377)	(0.0419)	(0.0413)	(0.120)
Treated States	10	13	10	13
Treated Judges	66	82	66	82

N= 13,438 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Sample restricted to judge-years with at least 10 published cases. Columns 1a and 1b include the bottom 5 cases, and Columns 2a and 2b include the top 5 cases, for each judge-year as ranked by the number of positive citations. Each regression includes a year fixed effect, judge fixed effect, and state trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline including the ten years before and after the policy change. Columns 1a and 2a report the baseline treatment effect, which is interacted with a dummy for limited discretion over case selection (1b and 2b).

TABLE 4
Effect of Reform: Term Length Increase

Outcome	Aggregate Effects			Baseline Effect	Limited Discretion
	(1)	(2)	(3)	(4a)	(4b)
Log Number of Majority Opinions Written	-0.0989	-0.0995	-0.116	-0.212	0.247
	(0.237)	(0.202)	(0.217)	(0.337)	(0.341)
Log Number of Discretionary Opinions Written	-0.286	-0.206	-0.0903	-0.103	0.00879
	(0.501)	(0.229)	(0.191)	(0.307)	(0.315)
Log Total Words Written	-0.293	-0.125	-0.125	-0.149	0.0648
	(0.182)	(0.169)	(0.182)	(0.299)	(0.311)
Log Average Length of Majority Opinion	-0.177	-0.0349	-0.0306	0.0157	-0.110
	(0.131)	(0.0659)	(0.0597)	(0.0752)	(0.0835)
Log Average Length of Table of Cases	-0.125	-0.00754	0.0242	0.102+	-0.221**
	(0.151)	(0.0678)	(0.0619)	(0.0543)	(0.0722)
Log Positive Cites Per Opinion	-0.492**	0.0321	0.0447	0.0741	-0.0935
	(0.182)	(0.0576)	(0.0361)	(0.0452)	(0.0726)
Log Distinguishing Cites Per Opinion	-0.306+	0.115+	0.115+	0.154+	-0.0987
	(0.173)	(0.0610)	(0.0602)	(0.0821)	(0.0984)
Log Negative Cites Per Opinion	-0.174+	0.0741+	0.0660	0.111*	-0.110+
	(0.0932)	(0.0389)	(0.0432)	(0.0541)	(0.0611)
Log Discuss Cites Per Opinion	-0.271*	0.0617	0.0727*	0.114**	-0.112+
	(0.117)	(0.0511)	(0.0298)	(0.0189)	(0.0575)
Log Quoted Cites Per Opinion	-0.253+	0.0720	0.0825*	0.132**	-0.134*
	(0.143)	(0.0540)	(0.0383)	(0.0369)	(0.0575)
Log Out-of-State Cites Per Opinion	-0.0549	0.0419	0.0248	0.0546	-0.0625
	(0.144)	(0.0498)	(0.0353)	(0.0456)	(0.0677)
Fixed Effects	None	State	Judge	Judge	
Trends	None	State	State	State	
Treated States	5	5	5	3	2
Treated Judges	35	35	35	25	10

N= 15,277 judge-years. Treatment states listed in Table 1.1. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Columns 1 through 3 report aggregate effects. Columns 4a and 4b report the baseline treatment effect (4a) interacted with a dummy for limited discretion over case selection (4b).

TABLE 5.1

Effect of Real Salary Changes

Outcome	Aggregate Effects			Baseline Effect	Limited Discretion
	(1)	(2)	(3)	(4a)	(4b)
Log Number of Majority Opinions Written	-0.799**	0.0345	-0.0438	-0.14	0.168
	(0.273)	(0.171)	(0.171)	(0.295)	(0.315)
Log Number of Discretionary Opinions Written	0.594	-0.418*	-0.2	-0.176	-0.0429
	(0.439)	(0.195)	(0.173)	(0.222)	(0.255)
Log Total Words Written	0.0304	-0.0402	-0.0318	-0.147	0.2
	(0.308)	(0.156)	(0.188)	(0.295)	(0.304)
Log Average Length of Majority Opinion	0.592*	-0.00848	0.0333	0.00996	0.0407
	(0.278)	(0.116)	(0.112)	(0.134)	(0.179)
Log Average Length of Table of Cases	0.53	0.0507	0.11	0.128	-0.0311
	(0.367)	(0.123)	(0.109)	(0.133)	(0.178)
Log Positive Cites Per Opinion	1.930**	-0.0289	0.0264	0.318+	-0.508*
	(0.363)	(0.146)	(0.136)	(0.169)	(0.244)
Log Distinguishing Cites Per Opinion	1.796**	0.0539	0.12	0.192	-0.126
	(0.381)	(0.130)	(0.127)	(0.185)	(0.205)
Log Negative Cites Per Opinion	1.071**	0.0221	0.0548	0.11	-0.0963
	(0.271)	(0.0844)	(0.0819)	(0.123)	(0.126)
Log Discuss Cites Per Opinion	1.232**	-0.00447	0.0334	0.280+	-0.430*
	(0.276)	(0.118)	(0.119)	(0.154)	(0.201)
Log Quoted Cites Per Opinion	1.660**	0.0126	0.0888	0.290+	-0.352
	(0.322)	(0.123)	(0.107)	(0.166)	(0.211)
Log Out-of-State Cites Per Opinion	0.591*	-0.0923	-0.0805	0.0326	-0.197
	(0.284)	(0.0972)	(0.0881)	(0.130)	(0.188)
Fixed Effects	None	State	Judge	Judge	
Trends	None	State	State	State	
Treated States	50	50	50	22	28
Treated Judges	868	868	868	413	455

N= 6,288 judge-years. Treatment states listed in Table 1.1. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is log real salary in state s at year t. Columns 1 through 3 report aggregate effects. Columns 4a and 4b report the baseline treatment effect (4a) interacted with a dummy for limited discretion over case selection (4b).

TABLE 5.2
Effect of Real Salary Changes by Contestedness of Retention Process

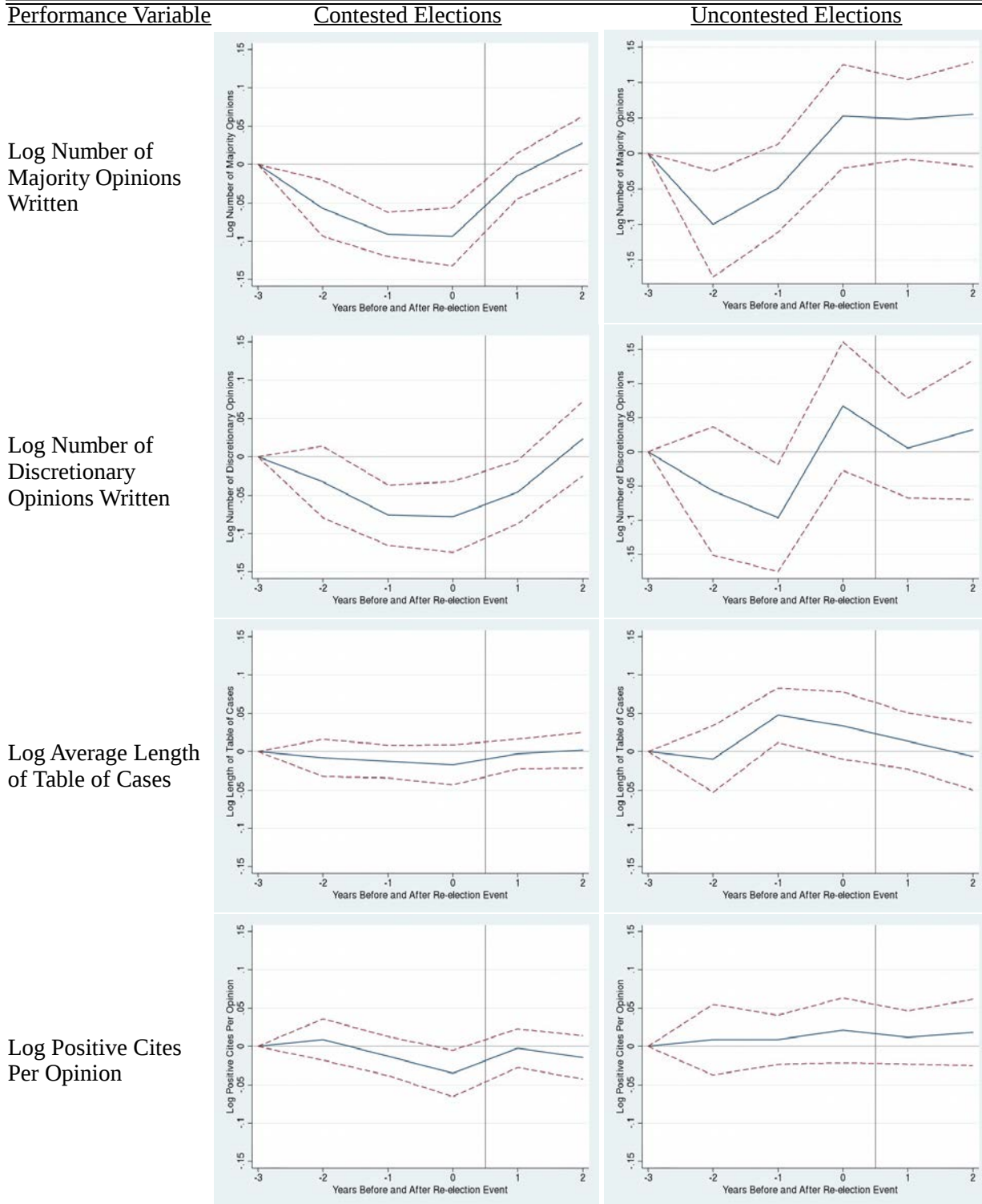| Outcome	Contested Retention		Uncontested Retention	
	Baseline Effect	Limited Discretion	Baseline Effect	Limited Discretion
	(1a)	(1b)	(2a)	(2b)
Log Number of Majority Opinions Written	-0.138	0.173	-0.189	0.23
	(0.289)	(0.312)	(0.298)	(0.322)
Log Number of Discretionary Opinions Wri	-0.171	-0.0404	-0.191	-0.0383
	(0.222)	(0.255)	(0.223)	(0.262)
Log Total Words Written	-0.144	0.206	-0.198	0.262
	(0.289)	(0.302)	(0.296)	(0.309)
Log Average Length of Majority Opinion	0.0107	0.0412	0.0061	0.0433
	(0.134)	(0.180)	(0.136)	(0.181)
Log Average Length of Table of Cases	0.125	-0.0332	0.143	-0.0441
	(0.135)	(0.179)	(0.137)	(0.180)
Log Positive Cites Per Opinion	0.314+	-0.512*	0.340+	-0.526*
	(0.169)	(0.244)	(0.172)	(0.246)
Log Distinguishing Cites Per Opinion	0.188	-0.129	0.214	-0.139
	(0.184)	(0.204)	(0.187)	(0.208)
Log Negative Cites Per Opinion	0.107	-0.0963	0.103	-0.0749
	(0.123)	(0.127)	(0.124)	(0.128)
Log Discuss Cites Per Opinion	0.277+	-0.433*	0.300+	-0.448*
	(0.153)	(0.201)	(0.156)	(0.204)
Log Quoted Cites Per Opinion	0.289+	-0.354+	0.304+	-0.365+
	(0.165)	(0.211)	(0.169)	(0.214)
Log Out-of-State Cites Per Opinion	0.029	-0.2	0.047	-0.204
	(0.130)	(0.186)	(0.131)	(0.188)
Treated States	16	17	7	15
Treated Judges	288	276	131	200

N= 6,288 judge-years. Treatment states listed in Table 1.1. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Each coefficient is from a separate regression that includes a year fixed effect, judge fixed effect, and state time trend. The treatment variable is log real salary in state s at year t. Columns 1a and 1b interact the salary effect with a dummy for states with contested elections, governor retention, or legislative retention. Columns 2a and 2b interact the salary effect with a dummy for states with uncontested elections or life tenure.

TABLE 6

Effect of Being Up For Election

Outcome	Contested (1a)	Uncontested (1b)	Contested (2a)	Uncontested (2b)	Contested (3a)	Uncontested (3b)
Log Number of Majority Opinions Written	-0.0650**	0.0520+	-0.0671**	0.0437	-0.0695*	0.0561
	(0.0221)	(0.0270)	(0.0245)	(0.0265)	(0.0294)	(0.0392)
Log Number of Discretionary Opinions Written	-0.0769**	0.0840*	-0.0534**	0.0591*	-0.0588**	0.0461
	(0.0248)	(0.0374)	(0.0198)	(0.0237)	(0.0185)	(0.0304)
Log Total Words Written	-0.0936**	0.0761*	-0.0917**	0.0667*	-0.0954**	0.0891*
	(0.0274)	(0.0307)	(0.0279)	(0.0292)	(0.0329)	(0.0434)
Log Average Length of Majority Opinion	-0.013	0.0154	-0.0160*	0.0157	-0.0184*	0.0218+
	(0.00899)	(0.0217)	(0.00718)	(0.0126)	(0.00826)	(0.0128)
Log Average Length of Table of Cases	-0.0121	0.0294	-0.0141	0.0176	-0.0133	0.0323*
	(0.0132)	(0.0218)	(0.0112)	(0.0132)	(0.0117)	(0.0157)
Log Positive Cites Per Opinion	-0.0270*	0.0216	-0.0356**	0.00837	-0.0282*	-0.0014
	(0.0111)	(0.0185)	(0.0110)	(0.0181)	(0.0131)	(0.0170)
Log Distinguishing Cites Per Opinion	-0.0292*	0.0369+	-0.0283*	0.0189	-0.0274+	0.0209
	(0.0136)	(0.0212)	(0.0124)	(0.0218)	(0.0147)	(0.0277)
Log Negative Cites Per Opinion	-0.0241**	0.0166	-0.0218**	0.013	-0.0238**	0.018
	(0.00755)	(0.0139)	(0.00710)	(0.0128)	(0.00885)	(0.0188)
Log Discuss Cites Per Opinion	-0.0248**	0.0191	-0.0272**	0.00857	-0.0208*	0.00842
	(0.00894)	(0.0128)	(0.00817)	(0.0119)	(0.0103)	(0.0123)
Log Quoted Cites Per Opinion	-0.0181*	0.0221	-0.0235**	0.0116	-0.0195*	0.000989
	(0.00878)	(0.0168)	(0.00713)	(0.0166)	(0.00865)	(0.0161)
Log Out-of-State Cites Per Opinion	-0.0173*	0.0266+	-0.0142+	0.0244+	-0.00829	0.0157
	(0.00821)	(0.0153)	(0.00804)	(0.0138)	(0.0112)	(0.0173)
Fixed Effects	State, Year		Judge, Year		Judge, State-Year	
Trends	State		State		None	
Treated States	35	18	35	18	35	18
Treated Judges	702	277	702	277	702	277
Election Events	1327	451	1327	451	1327	451

N= 15,277 judge-years. Retention systems described in Table 1.2. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01. Each coefficient is from a separate regression that includes a year fixed effect. Treatment variable is a dummy equaling one for years judge is facing reelection.

FIGURE 3.2
Dynamic Judge Responses to the Electoral Cycle

Performance Variable	Contested Elections	Uncontested Elections

Log Number of Majority Opinions Written

Log Number of Discretionary Opinions Written

Log Average Length of Table of Cases

Log Positive Cites Per Opinion

Notes. Plotted coefficients obtained from regression of the performance variable on indicators for two years before and after a re-election event, relative to three years before the re-election event. Regressions include judge fixed effects. Error bands indicate 90% confidence intervals using standard errors clustered by judge. Table 1.2 lists states by retention system.