Accountability, Selection, and Term Limits:

Theory and Evidence from U.S. State Elections

James Alt
Harvard University
jalt@latte.harvard.edu

Ethan Bueno de Mesquita
Washington University
ebuenode@artsci.wustl.edu

Shanna Rose
New York University
shanna.rose@nyu.edu

August 23, 2006

An earlier version of this paper was presented to the Annual Meeting of the Midwest Political Science Association, Chicago, April 2006. We thank Tim Besley, John Gasper and Christina Gathmann for useful comments.

Abstract

Previous theoretical and empirical work on term limits has focused on the problem of moral hazard—that is, the possibility that term-limited politicians exert less effort than those who are eligible to run for reelection. We present a model that includes both moral hazard and selection effects, in which term limits interfere with voters' ability to weed out low-quality incumbents. We evaluate our model by taking advantage of a natural experiment in the American states: variation across states and time in gubernatorial term limit laws. We find evidence suggesting that term limits have both selection and moral hazard effects.

Conventional wisdom holds that the desire to win reelection keeps politicians accountable for their policy choices. It is often suggested that term limits interfere with this mechanism by removing politicians' incentive to maintain a favorable reputation. However, the shirking hypothesis remains largely untested. Moreover, the empirical literature focuses exclusively on the problem of moral hazard; this emphasis on elections as a mechanism for punishing poorly-behaved incumbents ignores the possibility raised by many observers that elections are instead primarily a selection mechanism that helps voters choose competent politicians (see for example Fearon 1999).[1] But do such selection effects exist in practice? This paper presents a model and looks for empirical evidence of selection effects in elections.

The empirical literature on term limits focuses mostly on governors in the American states, and finds evidence suggesting that term-limited governors shirk. As we show in this paper, however, these findings are largely unreliable because the literature neglects two widespread changes in gubernatorial term limit laws that have taken place over the last half century.[2] During this period, many states that previously did not limit the governor's ability to seek reelection adopted two-term limits, while others switched from one-term limits to two-term limits. In fact; in the early postwar years, the majority of term-limited incumbents served under "no succession" laws limiting them to a single term, even though today Virginia alone retains the practice.

Because reelection is never possible under a one-term limit, voters have neither the opportunity to weed out low-quality incumbents nor the opportunity to discipline those who appear to have shirked. As a result, it is not unreasonable to expect that observable performance

[1] Models of elections with selection effects include those of Ashworth (2005), Ashworth and Bueno de Mesquita (2005), Gowrisankaran, Mitchell, and Moro (2006), and Zaller (1998).
[2] Gubernatorial term limits date from 1787 (Delaware). The 1990s saw over a third of states impose term limits on state legislators (Kousser 2005; Carey et al., forthcoming), but these are too recent for the sort of analysis we undertake.

by first-term "lame ducks" will be worse than performance by either second-term governors operating under a binding two-term limit (who have survived selection but who have an incentive to shirk) and second- (or higher-) term governors who are eligible for reelection (and therefore feel the effects of both selection and sanctioning).[3]

Thus, we consider variation in term limit laws across states and time to be an opportunity rather than a nuisance. To take advantage of this "natural experiment," we begin by developing a formal model that allows us to make several predictions about how performance should differ under different term limits. The key feature of our model is that it allows for voters to select among incumbents on the basis of ability, as well as for elections to discipline incumbents.[4]

To see why selection effects might matter, consider the following. In a pure moral hazard model like Ferejohn's (1986), in which retrospective voters choose a voting rule to induce effort, and no learning about the incumbent's type takes place, the pool of candidates has the same expected level of ability in all periods. Thus, incumbents have the same expected performance in the period in which they are term limited , regardless of whether the limit occurs in the first or second term. When both moral hazard and selection effects are present, on the other hand, the expected ability of incumbents improves with each additional term, potentially offsetting the effect of shirking among second-term lame ducks but not first-term lame ducks.[5]

[3] We use the phrase "lame duck" to refer to the entire term of an incumbent who is ineligible to run for re-election, rather than to the brief period after an election in which some incumbents continue to serve despite not having been re-elected.

[4] We abstract away from the possibility that incumbents might want to run for higher office, care about their place in history, or care about having a successor from the same party, which would give them an incentive to perform well even in a final term. We further assume that however strong those possible last-period incentives are, they are of the same strength in the final period no matter how many terms the governor has served, and therefore that differences in last-period incentives are not the cause of any differences we may observe between first- and second-term lame ducks.

[5] However, as the model shows, their higher ability might allow more competent politicians to reduce effort even in the absence of a term limit, so the implications for observed performance are not obvious.

To evaluate our model's predictions, we undertake two sets of empirical tests, both designed to avoid the problems that characterize the existing literature. The first set of tests examines the experience of states that switched from one-term limits to two-term limits. We compare performance in these states under different types of governors: first-term lame ducks, first-term governors who were eligible for reelection, and second-term lame ducks. Our second set of empirical tests focuses on a different sub-sample of states: those that either had two-term limits or no term limits during the sample period. In this case we compare performance under first-term governors who were eligible for reelection, second-term lame ducks, and second-term governors who were eligible for reelection.

Our results appear to support the model's predictions. In particular, the switch from one- to two-term limits appears to be associated with improved performance. Economic growth is lower under second-term lame ducks and (especially) first-term lame ducks, compared to governors who are eligible for reelection. And taxes and spending decline with a governor's seniority in office, indicating a selection effect, but are higher under second-term lame ducks than under second-term governors who can run again, suggesting that term-limited incumbents do indeed shirk.

The paper proceeds as follows. In the next section, we discuss problems with the empirical literature on term limits.[6] We show that previous work on term limits suffers from omitted variable bias, and explain why existing evidence about selection effects is ambiguous. The following section analyzes a repeated game between politicians and voters in which there is symmetric uncertainty about ability and politicians can be term-limited after either one or two

[6] We do not evaluate the normative literature on whether term limits are a good thing. However, the concluding section does place our results in the context of the normative debate.

periods. We then summarize our data and specifications and present the empirical results. The final section offers concluding remarks and directions for future research.

2. Problems with the Existing Literature

The most widely-cited paper in the small literature on term limits and shirking is by Besley and Case (1995). The authors develop a simple reputation-building model in which voters are uncertain about politicians' competence, and politicians' actions, in turn, affect voters' utility through policy outcomes. Voters reelect incumbents who deliver high pay-offs in their first term, since such politicians are likely to be competent and continue to deliver high pay-offs in the future. Thus, the prospect of reelection gives office-seeking politicians an incentive to exert effort. The model yields the highly intuitive prediction that term-limited governors do not have this incentive and thus expend less effort than governors who can stand for reelection, all else equal.

The authors point out that the model's policy implications are not obvious; the predicted effect of term limits on effort "may show up in all manner of policy choices. The theoretical model would, however, have to be enriched to handle the details of each" (p. 773). Nonetheless, the authors proceed to test their general model by studying the effects of term limits on a specific policy choice: the size of government. They define effort as "keeping taxes and expenditures down" without providing further explanation. However, it is not clear that all voters agree that small budgets are desirable.

Nonetheless, the authors find evidence that appears to support their theory. Specifically, in the 48 contiguous states between 1950 and 1986, per capita spending was $14 higher and

taxes were $7 higher (in 1982 dollars) when the governor could not run for reelection.[7] However, in a later (2003) paper, the authors repeat this analysis on a sample that extends through the mid-1990s and find that the effect of term limits on the size of government has changed dramatically over time. The authors conclude that:

> ...governors in the first half of our time period spent and taxed more when they could not stand for reelection but that, over time, this phenomenon has changed. It seems likely that some omitted variable is responsible for the *change in behavior* observed for governors working under a term limit. This is an area ripe for future research (p. 55; italics added).

But there is another possible explanation for this puzzling result that has nothing to do with governors' behavior. Over the past fifty years, there have been two widespread changes in term limit laws in the American states. First, as shown in Table 1, many states have abandoned one-term limits in favor of two-term limits, and second, other states without term limits have adopted two-term limits. These two trends have changed what it means to be a term-limited governor in potentially important ways.

[Table 1 about here]

First, consider the adoption of term limits by states that previously did not restrict the governor's ability to run for reelection. If these states are systematically different from those that already had term limits in 1950, then the lame-duck variable will, over time, begin to pick up these inter-state differences. Indeed, the states that adopted term limits during the sample period have comparatively small governments, on average, controlling for income and demographic characteristics—which could help explain the apparent decline in the effect of term limits. Thus, Besley and Case may be picking up changes in the composition of states with term limits that

[7] Although the effect of binding term limits on total taxes falls just short of significance at the 10 percent level, the authors find significant effects for two specific types of taxes: sales taxes and income taxes (see Besley and Case 1995 for details).

have little to do with term limits themselves. We return to this point in the empirical section, when designing our own regression specifications.

Second, and perhaps more importantly, the abandonment of one-term limits in favor of two-term limits has increased the average tenure of lame-duck governors. At the beginning of the sample period, 70 percent of lame ducks were first-term governors; by the end of the sample, this fraction had declined to 5 percent. However, Besley and Case do not distinguish between first- and second-term lame ducks; their regressions simply include a dummy variable for governors who cannot run for reelection. This specification has generated considerable confusion in the interpretation of their results. For example, Johnson and Crain (2004) state that:

> Besley and Case (1995a) find evidence of cyclical activity in policy variables for two-term limit states. For example, taxes and spending are lower in the first term than in the second. Besley and Case do not analyze the behavior of single-term limited governors, a relatively rare institution in the United States (p. 75).

In fact, more than *two-thirds* of the lame ducks in Besley and Case's original sample are first-term governors.

If a governor's seniority affects the fiscal policies he chooses, then Besley and Case's model suffers from omitted variable bias. That is, if there is a "selection effect," whereby voters use elections to weed out low-quality incumbents, then the lame duck variable will increasingly pick up this effect over time, as average tenure increases. Indeed, in his most recent work on this subject, Besley acknowledges that:

> there were some states that switched from a one-term limit ….If these … are separated out … we find that most of the term limit effects are coming from them. This is in line with existence of selection effects and suggests that the selection effect for incumbents who survive beyond a single term is strong (2006, Ch. 3. n. 23).

In order for selection effects to explain the diminishing effect of term limits over time, it must be that incumbent quality improves with tenure in office, and thus that governors constrained by a

two-term limit perform better than their one-term-limited counterparts. Besley does not provide this evidence; he merely offers a conjecture about the possibility of selection effects. This conjecture motivates both our use of the change from one- to two-term limits as a natural experiment and our development of an explicit model that allows for selection effects.

3. Model

Consider a game in which there are two players: a politician and a voter. The order of play is as follows. At the beginning of each period, the politician chooses a level of effort to exert in providing public goods. At the end of each period—after observing the level of public goods provided, but not the incumbent's level of effort—the voter selects between two candidates (one of whom may be the incumbent). The game is infinitely repeated.

We will say that there is a t*-period term limit* if a politician can only serve in office for t periods. We will say a politician is *term-limited* if there is a t-period term limit and the politician has served $t-1$ periods in office. If a politician is not term limited he runs for reelection against a randomly selected challenger (to be described later). If he is term limited, there is an open seat election with two randomly selected challengers.

The incumbent's action in period i is labeled $a_i \in A = [\underline{a}, \overline{a}] \subset \mathbb{R}$. The incumbent politician in period i has true ability θ_i. Politicians' abilities are distributed according to a mean 0, normal distribution with variance σ_θ^2. There is symmetric uncertainty about the incumbent's ability (i.e., neither the politicians nor the voters know the true abilities).

In each period the voter observes the level of public goods provided, $s_i = \theta_i + a_i + \varepsilon_i$, where ε is a mean zero, normal random variable, with variance σ_ε^2. Thus, public goods are a function of incumbent effort, incumbent ability, and random noise. This basic "career concerns"

set-up (Holmström 1999, Dewatripont, Jewitt and Tirole 1999) has been used in a variety of models of electoral politics.[8] The set-up here most closely resembles that used in Ashworth (2005).

In each period, the voter has a payoff of s_i. The incumbent bears a cost of effort $c(a_i)$ and gains benefit B from holding office. The voter discounts the future according to δ_v and the politicians according to δ_p.

We solve for Markov Perfect equilibria (MPE).[9]

One-term limits

Suppose the politician can only serve a single term in office. In this case, since there is no possibility of reelecting the politician, there is no payoff-relevant state variable. Consequently, the MPE correspond to the equilibria in a one-shot game.

The incumbent faces no electoral threat whatsoever. A period i incumbent, then, has expected payoff from exerting effort a_i given by:

$$B - c(a_i).$$

Clearly, the optimal choice in each period is to choose $a_i^* = \underline{a}$.

Since all candidates are *ex ante* identical, the voter in each period is indifferent between the two challengers and so any reelection rule is optimal.

[8] As Besley (2006, p. 103) notes, in the case of ability, it is plausible, indeed natural, to assume that politicians do not know their capacities better than voters, let alone completely. When personal preferences are part of the politician's type, however, this assumption of symmetric uncertainty is much less plausible. For other applications of the career concerns framework to models of politics see, for example, Alesina and Tabellini (2003), Ashworth (2005), Ashworth and Bueno de Mesquita (2006), Lohmann (1998), and Persson and Tabellini (2000).

[9] This solution concept allows voters to make use of payoff relevant information about the state of the world (i.e., term limits and beliefs about incumbent ability), but excludes complicated punishment strategies requiring coordination among the players.

Proposition 1. The expected level of public goods provided in each period in the one-term limit system is $\underline{a}$.

Proof. As shown in the text, regardless of the voter's rule, the incumbent exerts effort $\underline{a}$. Thus, the expected level of public goods in a period is $E[\underline{a}+\theta_i+\varepsilon]=\underline{a}$. □

Two-term limits

Now consider a model where politicians can serve for two terms. In each period the world is described by a vector of two payoff relevant state variables. The first indicates whether or not the incumbent is term limited. The second indicates the expected ability of the incumbent.

In his first term in office, an incumbent's expected payoff for effort (a_1) is:

$$B-c(a_1)+\delta_p E\left[\Pr(\text{Reelect}\mid a_1,\theta_i)(B-c(a_2))\right].$$

If reelected, the incumbent's payoff in his second term in office is simply:

$$\delta_p(B-c(a_2)).$$

Clearly, when term limited in the second period, the incumbent will choose effort $a_2^*=\underline{a}$. To determine the incumbent's first-period action, we need to know how the voter will behave, since this will pin down the probability of reelection in the incumbent's objective function.

As stated above, at the time of the election either there is an incumbent or there is an open seat election. Let v_O be the voter's expected net present value from an open seat election and v_I the voter's expected net present value from an election with an incumbent. Further, let $\tilde{a}_1$ be the voter's beliefs about the level of effort by a period 1 incumbent. Then, we have that $v_O=\tilde{a}_1+\delta_v v_I$.

Using standard facts about normal learning (DeGroot 1970), the posterior beliefs about the incumbent's ability after observing a level of public goods (s_1), are distributed normally with mean:

$$\theta' = \lambda(s_1 - \tilde{a}_1) = \lambda(\theta_1 + a_1 + \varepsilon - \tilde{a}_1),$$

where $\lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}$.

The expected payoff to the voter of electing the incumbent is:

$$\underline{a} + \theta_i + \delta_v v_O.$$

The expected payoff of electing the challenger is:

$$\tilde{a}_1 + \delta_v v_I.$$

Thus, the voter reelects the incumbent only if:

$$\underline{a} + \theta_i + \delta_v v_O \geq \tilde{a}_1 + \delta_v v_I, \tag{1}$$

which, after substituting in for v_O and θ_i, is true if an only if:

$$\varepsilon \geq \frac{(1-\delta_v)(\tilde{a}_1 + \delta_v v_I) - \underline{a}}{\lambda} + \tilde{a}_1 - \theta_i - a_1.$$

The probability that an incumbent is reelected is:

$$\Pr(\text{Reelect} \mid a_1) = 1 - \Phi\left(\frac{\frac{(1-\delta_v)(\tilde{a}_1 + \delta_v v_I) - \underline{a}}{\lambda} + \tilde{a}_1 - \theta_i - a_1}{\sigma_\varepsilon^2}\right),$$

where Φ is the cdf of the standard normal distribution.

Remark 1. Candidates who perform better (i.e., have a higher s_i) are more likely to be reelected.

Taking expectations, a first-term office holder's problem is the following:

$$\max_{a_1} B - c(a_1) + \delta_p\left(1 - \Phi\left(\frac{\frac{(1-\delta_v)(\tilde{a}_1 + \delta_v v_I) - \underline{a}}{\lambda} + \tilde{a}_1 - a_1}{\sigma_\varepsilon^2}\right)\right)(B - c(\underline{a})).$$

At an interior solution, the following first-order condition characterizes the optimal level of effort:

$$\phi\left(\frac{\frac{(1-\delta_v)(\tilde{a}_1 + \delta_v v_I) - \underline{a}}{\lambda} + \tilde{a}_1 - a_1^*}{\sigma_\varepsilon^2}\right)\frac{\delta_p(B - c(\underline{a}))}{\sigma_\varepsilon^2} = c'(a_1^*).$$

In equilibrium, $\tilde{a}_1 = a_1^*$ (i.e., the voter's beliefs about actions are correct), so the first-order condition reduces to:

$$\phi\left(\frac{\frac{(1-\delta_v)\left(a_1^*+\delta_v v_I\right)-\underline{a}}{\lambda}}{\sigma_\varepsilon^2}\right)\frac{\delta_p(B-c(\underline{a}))}{\sigma_\varepsilon^2} = c'(a_1^*). \tag{2}$$

There may be multiple solutions to this first-order condition. However, any interior optimum must satisfy the first-order condition.[10]

Intuitively, an incumbent who faces another election has incentives to exert effort that a term limited incumbent does not. As long as the benefits of reelection are sufficiently large, relative to the marginal cost of effort, this will induce an incumbent who is not term limited to exert a greater level of effort than a term limited incumbent.

Proposition 2. An MPE equilibrium exists in both the game with one- and two-term limits. The expected level of public goods provided by a first-term incumbent who faces a two-term term limit is weakly higher than the expected level of public goods provided by a first-term incumbent who is term limited in the first term. Moreover, if the rewards of reelection (B) are sufficiently large, then the relationship is strict.

Proof. Existence follows from the argument in the text and the fact that A is compact. The weak condition is trivial given that a term limited incumbent chooses effort at the lower bound.

The left-hand side of equation (2) is increasing to infinity as B goes to infinity and the right-hand side is finite. So, there exists a B' such that, for all $B \geq B'$, the left-hand side evaluated at the lower bound is greater than the right-hand side evaluated at the lower bound. For such a B, either there is a corner solution at the upper bound or the first-order condition characterizes the optimum. If there is a corner solution at the upper bound, the strict result is

[10] See Ashworth (2005) for a detailed discussion of second-order conditions in this sort of model.

established. If the first-order condition characterizes the optimum, the marginal benefit of effort (the left-hand side of equation (2)) to a first term incumbent with a two-term limit is strictly greater than 0, whereas it is 0 for a term-limited first-term incumbent. Moreover, the marginal cost of effort is the same for both. Thus, the relationship is strict whenever the first-order condition characterizes the optimum.□

Notice that, in deciding whether or not to reelect the incumbent, the voter is selecting for quality. As a result, incumbents who have survived an electoral challenge are, in expectation, higher quality than politicians who have just been elected. This gives rise to the following result.

Proposition 3. The expected level of public goods provided is greater when there is a term limited incumbent who faced a two-term limit than when there is a term limited incumbent who faces a one-term limit.

Proof. Both types of incumbent choose a level of effort $\underline{a}$. Thus, it is sufficient to show that $E[\theta' \mid reelection] > 0$. Applying Bayes' Rule, the posterior distribution of abilities, condition reelection is given by:

$$f\left(\theta_i \mid reelection\right) = \frac{\Pr\left(reelect \mid \theta_i\right)\Pr\left(\theta_i\right)}{\Pr\left(reelect\right)} = \frac{\left(1-\Phi\left(\frac{\frac{a_1^* + \delta_v(v_I - v_O)}{\lambda} - \theta_i}{\sigma_\varepsilon^2}\right)\right)\phi\left(\frac{\theta_i}{\sigma_\theta^2}\right)}{\left(1-\Phi\left(\frac{\frac{a_1^* + \delta_v(v_I - v_O)}{\lambda}}{\sigma_\varepsilon^2+\sigma_\theta^2}\right)\right)}.$$

Fix $\theta' > \theta$. Then we have

$$\frac{f\left(\theta_i \mid reelection\right)}{f\left(\theta_i^{'} \mid reelection\right)} = \frac{\left(1-\Phi\left(\frac{\frac{a_1^* + \delta_v(v_I - v_O)}{\lambda} - \theta_i}{\sigma_\varepsilon^2}\right)\right)\phi\left(\frac{\theta_i}{\sigma_\theta^2}\right)}{\left(1-\Phi\left(\frac{\frac{a_1^* + \delta_v(v_I - v_O)}{\lambda} - \theta_i^{'}}{\sigma_\varepsilon^2}\right)\right)\phi\left(\frac{\theta_i^{'}}{\sigma_\theta^2}\right)} > \frac{\phi\left(\frac{\theta_i}{\sigma_\theta^2}\right)}{\phi\left(\frac{\theta_i^{'}}{\sigma_\theta^2}\right)},$$

where the inequality follows from the monotonicity of the normal cdf. Thus, the distribution of posteriors conditional on reelection monotone likelihood ratio (MLRP) dominates the prior distribution. MLRP dominance implies first-order stochastic dominance, which implies that the expectation under the conditional posterior is higher than the expectation under the prior. Thus $E[\theta' \mid reelection] > 0$, which establishes the result.□

Costly reelection campaigns

Another possible specification of the model would assume that there is some small cost associated with seeking reelection. In this case, only incumbents who expect to be reelected (i.e., those with $\underline{a} + \theta_i + \delta_v v_O \geq \tilde{a}_1 + \delta_v v_I$) would actually seek reelection. Under this specification, we get the following result.

Corollary 1. If seeking reelection is costly, then the performance of those politicians who do not seek reelection, *ex post*, will be lower than the performance of those politicians who seek reelection. Moreover, politicians who have sought reelection and are then term limited will, in expectation, provide more public goods than politicians who face a one-term limit.[11]

Empirical implications and longer horizons

To summarize, Proposition 1 derives from the fact that under a one-term limit the marginal benefit of effort is zero, and gives a baseline expectation for effort in the term-limited case. Proposition 2 derives from the fact that the voters' reelection rule means that in equilibrium under a two-term limit the marginal benefit of effort is not zero for a first-term incumbent who values reelection sufficiently. It characterizes empirical conjectures with respect to effort when

[11] Note this result is similar in spirit to Gordon, Huber, and Landa's (2006) study of costly entry by *challengers*. In that model, only challengers who believe that the incumbent is of sufficiently low quality are willing to enter because mounting a challenge is costly.

reelection to a second term is possible. Finally, Proposition 3 derives from the fact that the conditional expectation of ability of reelected incumbents is higher than the overall average level of ability. It characterizes the empirical conjectures with respect to selection.

Two slightly broader interpretations of these results would take them to imply, first, that politicians with longer term limits work harder earlier in their careers than politicians with shorter term limits, and second, that for a fixed number of terms remaining, a politician who has stood for reelection more times will provide a greater level of goods than a politician who has stood for reelection a shorter number of times.

Unfortunately, technical issues in characterizing equilibrium preclude presenting results like ours comparing politicians who are not term limited with those having arbitrarily long term limits. To get some intuition for the problems that arise, imagine a model with a politician who has a longer term limit. It is useful to consider in turn several effects associated with an incumbent having held office for a longer length of time.

The first effect is the intuitive one described in Remark 1 and Proposition 3. With each electoral success the politician's expected ability increases. Since ability enters positively into the production of public goods, this makes the voter like the incumbent more with each electoral success.[12] We refer to this as the "screening effect."

Second, effort and ability are substitutes in the production function for public goods. Moreover, as incumbent ability increases, it becomes less likely that the voter will conclude the politician is of low enough ability to not reelect. Thus, for a fixed electoral rule, incumbent effort is declining in incumbent ability. We refer to this as the "discipline effect."

Over time, the screening effect increases expected public goods provision from a given

[12] This result is in the spirit of the results on screening and the incumbency advantage in Ashworth (2005), Ashworth and Bueno de Mesquita (2005), Gowrisankaran, Mitchell, and Moro (2005), and Zaller (1998).

incumbent, while the discipline effect decreases expected public goods provision. It is not clear, *a priori*, which effect is larger. That is, if the decrease in effort more than offsets the increase in expected ability, total production of public goods could actually be *decreasing* in effort in a model with a longer time horizon.[13]

Finally, the longer an incumbent has been in office, the less uncertainty there is about her ability. Perhaps counter-intuitively, this decrease in uncertainty makes the incumbent *less* attractive to a voter who likes high ability incumbents. The reason is as follows. An increase in uncertainty about incumbent ability means that, from the voter's perspective, the incumbent is more likely to be either much better or much worse than her expected ability. But a voter who concludes that a politician is of low quality can decline to reelect that politician. Since the voter throws out all politicians below some quality level, the voter does not care about the increased probability of very low quality politicians that increased variance implies. But the voter keeps high quality politicians, and the higher quality the better. So the increased probability of a very high quality politician associated with an higher variance is good for the voter. Thus, when there are more than two electoral rounds, the decrease in uncertainty about incumbent ability makes the incumbent less attractive over time. Put differently, with each term in office, the challenger's option value is growing relative to the incumbent's. We refer to this as the "option value" effect.

The combination of these three effects make it difficult to characterize equilibrium in a model with a longer time horizon. In particular, if the "discipline" and "option value" effects dominate the "screening effect", then, perversely, the voter might not elect the candidate with the highest expected ability. If that were the case, then the incumbent's incentives to invest in effort

[13] Ashworth (2005) provides sufficient conditions for the "screening effect" to dominate the "effort effect" in a model with three periods, which is different than an infinitely repeated model with three-term term limits, which would be the three-term limit analogue to the model we study here.

in order to convince the voter that she is of high ability would be confounded. While we believe substantively that the broader interpretations of results from our two period model are likely to hold in the real world, for the reasons outlined above, characterizing sufficient conditions for this to be the case in the more general model presents significant technical problems that we leave for future research.

4. Data and methodology

In the model just described, the dependent variable that voters value and incumbents provide is "public goods." Public goods are notoriously difficult to measure; as proxies, we consider economic growth, fiscal balance, spending, and taxes, as explained below. We briefly discuss the possibilities of directly measuring incumbent experience and analyzing congruence between incumbent ideology and public opinion, noting why both of these are less than ideal fits for our model. We then turn to a description of our regression specifications.

Public goods

Ideally, in order to test our model, we would be able to measure the amount of (widely-valued, non-targeted) public goods provided per tax dollar collected. However, measuring the provision of government services is quite difficult. Thus, as one indicator of a public good we employ economic growth, which is widely valued and not narrowly targeted. Studies of gubernatorial elections (such as Lowry, Alt, and Ferree 1998) and job approval ratings suggest that voters value and respond to growth even though it is not completely within the control of governors. Fiscal balance (revenues minus spending) is another example of a positive policy outcome—it is not targeted and voters are known to like it.[14]

[14] Lowry, Alt, and Ferree (1998) show that voters punish incumbents for deviations from fiscal balance.

We also examine aggregate real per capita spending and taxes. Some models (such as those of Besley and Case and Persson and Tabellini) suggest that low spending and taxes are public goods.[15] They interpret a small budget as reflecting effort, which assumes (since effort *per se* is not valuable to voters) that incumbents who exert effort must provide relatively more goods at lower cost, as measured by the levels of spending and taxes.[16] Although it is debatable whether small budgets are public goods, since voters might prefer larger or smaller budgets depending on their ideology and the extent to which spending is targeted to them, we want to build on the existing literature, so we include these variables in our empirical analysis.[17]

Experience

We also consider a direct measure of the incumbent's ability, namely years of prior political experience. Examining experience is awkward in our model: experience is clearly not simply a function of effort or ability, but might be an indicator of having been hard-working or able enough to remain in office over time. Thus, for example, politicians who produce more public goods at lower unit cost might be overrepresented among the "survivors," who have served more years in politics. However, inferring quality from experience is not so simple. For one thing, experience is an input whereas we are interested in the output: performance. Moreover, effort and ability are substitutes in the public goods production function in our model. In any round (not just the final term in office), if we fix the expected ability of challengers, the higher the expected ability of the incumbent, the less effort the incumbent will exert. Hence,

[15] They get that result by making debatable public choice assumptions about the nature of government and voter preferences. As mentioned earlier, we find it unlikely that voters all agree that small budgets are good or indicative of highly competent government, unless we could hold constant the level of services provided, which we cannot do.
[16] Since the states have balanced budget rules, taxes are simply a cost of spending.
[17] We considered the possibility of using voters' job approval ratings as a measure of incumbent performance, but systematic data collections (Beyle et al. 2002) do not go back far enough to encompass changes from one-term to two-term limits.

more experienced politicians may well reduce effort, making their relative performance uncertain.

A further concern about experience is that it is perfectly observable by everyone. This means that voters should select on it, but it also means that politicians with experience, knowing that they have this valence advantage, can adjust their level of effort down and the voters cannot commit to punishing them for doing so. So, if you are a voter, do you want to have a highly experienced leader? Only as long as the screening effect dominates the disciplining effect (that is, only if the fact that they are higher ability more than compensates for the fact that they will exert less effort).[18]

Because of the observability problem, we provide some intuition for the underlying process by which the change from a one-term to a two-term limit could boost the average experience of candidates. Assume a pool of people in politics, amassing experience in town councils, the state legislature, etc. Some have more experience (because they are able, have worked hard, or were lucky) and these individuals will run for higher office such as Congress or the governorship. The switch from one-to two-term limits, by adding the option ultimately to serve a second term, makes the governor's job relatively more attractive. Consequently, relatively more of the more experienced individuals should run for that office, since even though they do not know their own ability, they value the governorship more than when it is only possible to serve a single term.

Ideological Congruence

Besley (2006) proposes another way to measure a politician's quality: the congruence between the politician's ideology and voters' opinions. He employs a measure of the

[18] A publicly observable measure of ability like experience may also raise a variety of strategic issues that don't emerge with unobserved ability.

resemblance between elite and mass opinion from Berry et al (1998), which indicates whether the governor's ideology tracks the center of public opinion.[19] This definition of congruence is also problematic in our model. For one thing, it is targeted (to the median voter), whereas public goods are not. Moreover, congruence with public opinion depends on whether the governor's policy preferences match those of voters. If voters select a "good" representative in the sense that she has policy preferences similar to theirs, then reduced effort under a term limit would mean the politician's preferences subsequently become more divergent. Representatives could display ability in implementing voters' preferences, and choose a level of effort to put into that implementation, in which divergence would accompany reduced effort. But while effort has a natural interpretation in an ability model, it is less clear that ability has an equally natural interpretation in a policy preference model, in which private information is more important.[20] Nonetheless, we examine this measure of quality in order to respond to the existing literature.

Regression specifications

We take advantage of variation in gubernatorial term limits across states and time to examine the conjectures of the model. We employ two sets of empirical tests, focusing on different subsets of states, to avoid the aforementioned problems that afflict the existing literature. First, we compare performance under one-term limits and two-term limits. We examine performance in the twelve states that changed from one- to two-term limits, performing a detailed comparison of first-term lame ducks, first-term governors who were eligible for

[19] These variables are annual measures of citizen and government ideology for the American states from 1960 through 2002, and are based on roll call voting scores. The roll call scores are provided by the Americans for Democratic Action (ADA) and the AFL-CIO Committee on Political Education (COPE), of state congressional delegations, outcomes of congressional elections, the partisan division of state legislatures, and the party of the governor, based on assumptions regarding voters and state political elites. The data are available at the ICPSR Publication-Related Archive, study # 1208. Because both measures yield similar results, we only report results for one measure (ADA).

[20] Our model is continuous and Besley's is binary, which may also affect the divergence of results. It is possible to combine congruence with ability in a model: see Austen-Smith and Banks (1989) for an example..

reelection, and second-term lame ducks in these states. Second, we look at states with two-term limits and states without term limits, comparing performance under first-term governors who were eligible for reelection, second-term governors who were eligible for reelection, and second-term lame ducks. For all regressions, we use Besley and Case's data, amended to include some variables from Alt and Lowry (2000). We adapt the specifications and use the controls from Besley and Case (2003) and Besley (2006) to show how our results build on and modify theirs. Unless otherwise noted, the sample period is 1950-1996. Descriptive statistics for the variables appear in Table 2.

[Table 2 about here]

One-term limits vs. two-term limits

One might think that performance would simply improve *on average* within a state after a switch from one- to two-term limits, as a result of increased effort or ability or both, but it is not so simple. State personal income growth (relative to the U.S. average) and fiscal balance appeared to improve in many states after they abandoned one-term limits, but both results fall short of statistical significance. Governors had on average three years more political experience, but lower ideological congruence, after the switch from one- to two-term limits. The latter result is entirely driven by southern states, undoubtedly reflecting the evolution of a two-party system in the South. Results for spending and taxes fall short of statistical significance, but time series estimates for these variables are not reliable as both spending and taxes clearly trend upward. Thus, more sophisticated panel-data (rather than time series) analysis is necessary.

Our first specification examines what happens when one "regime" gives way to another—that is, when a state abandons one-term limits in favor of two-term limits.[21] Consider the three different types of governors within the subset of states that switched from one- to two-term limits: (1) a first-term lame duck, (2) a first-term governor who was eligible for reelection, and (3) a second-term lame duck. Let b_i be the expected performance of politician i. Then from the propositions in Section 3 we have the following:

> $b_1 < b_3$ because types (1) and (3) exert the same level of effort but type (3) is, in expectation, of higher ability due to the selection effect.
>
> $b_1 < b_2$ because types (1) and (2) have the same expected ability but type (2) works harder than type (1).
>
> We cannot order b_2 and b_3 because type (2) works harder but type (3) has higher expected ability.

To test these hypotheses, we run the following regression on a sample of twelve states that changed from a one- to a two-term limit during the sample period: [22]

> performance = α_1 first-term-eligible + α_2 first-term-lame-duck + α_3 second-term-lame-duck + α_4 controls + α_5 fixed effects + α_6 year effects

[21] We treat the rule change as an exogenous shock, even though some first-term lame ducks may have played a role in bringing about reform. Some certainly did try to evade the limit, as the case of Alabama reveals. There, George Wallace became governor in 1962, but failed to get the legislature to amend the "no succession" rule. He even stood down in 1966 so that his wife Lurleen could run in his place, even though it was clear that George would effectively still be governor. She ran and won, but died two years later, so the lieutenant governor took her place. Meanwhile, the term limit law was changed to permit a second term. The lieutenant governor subsequently lost to George Wallace in the next primary, and Wallace was then elected unopposed and served two terms. One can interpret this as endogenous reform, but note that this lack of exogeneity would lead us to underestimate rather than overestimate any effect of the change in limits on performance, as we do not correct for the presence in office before the reform of incumbents willing to work especially hard for the chance of re-election.

[22] Although fourteen states abandoned one-term limits, we exclude Kentucky and Mississippi because they abandoned one-term limits too close to the end of the sample period (1994). Thus, the sample includes twelve states.

where performance is as defined above (alternately: growth, surplus, spending, taxes, political experience, and congruence). The variable "first-term-eligible" drops out due to multicollinearity and is thus the omitted category to which other types of governors are compared.

Furthermore, candidates who perform better (that is, have higher ability and/or exert more effort) are more likely to be reelected. So, among first-term eligibles, those who either choose not to run ("withdraw") or run and lose should show worse performance than those who run and win:

$$b_{2W}, b_{2L} < b_{2R}$$

where b_{2W} , $b_{2L,}$ and b_{2R} are first-term governors who are eligible for reelection and who withdraw, run and lose, and run and win, respectively.

These differences among the various type (2) politicians raise another issue. Consider b_{2R}, the performance of first-termers who go on to get reelected. In their second term, they become lame ducks and their performance is b_3. The model does not yield predictions about the relative sizes of b_2 and b_3, but it does imply that $b_{2R} - b_3$ is positive in expectation, since ability is the same and effort is higher. If we think of ability and effort as making weighted contributions to public goods provision, then the bigger is $b_{2R} - b_3$, the more important is effort relative to ability.

Thus, we regress measures of performance on dummy variables for first-term lame ducks and second-term lame ducks (the omitted category is first-term governors who are eligible for reelection). Our specification uses the same control variables as Besley and Case (2003) and Besley (2006). These are variables that might be expected to affect economic policy outcomes,

like population size, income,[23] the proportion of aged and young and indicators relating to party control of government. State and year fixed effects also are included.

Finally, the model does predict a connection between experience (as an indicator of ability) and reelection, consequent on the switch from one-term to two-term limits. However, we should point out that while it allows for the average level of experience to rise after the switch, the model does not explicitly predict this.[24] Second-term lame ducks should also be more experienced than their one-term-limited predecessors. By force (they have served another term) second-term lame ducks are more experienced than first term eligibles ($b_3 - b_2 \geq 4$). Any extent to which the difference is greater than four years indicates a potential selection effect.

Two-term limits vs. no term limits

Our second set of empirical tests returns to Besley and Case's original specification—which, as explained in Section 2, has been widely but mistakenly interpreted as a test of the effect of a binding two-term limit—and attempts to resolve its shortcomings. Recall that their specification potentially attributes both selection effects (resulting from the transition from one- to two-term limits in many states) and compositional effects (resulting from the adoption of two-term limits by states that previously did not have term limits) to incumbent shirking.

To avoid the compositional effect problem, we include in our sample only those states that had either no term limits or two-term limits during the sample period.[25] For the sake of

[23] Besley and Case (2003) include the square and cube of population and income to allow for nonlinearities in these relationships. Besley (2006) substitutes the logarithms of income and population, but this alternative specification is never statistically superior to the polynomial coding, which we retain below.

[24] Suppose the office became more valuable and more competitive after the switch from one- to two-term limits. Then, if candidates knew their ability, we would expect the average ability of the pool of candidates. But in our career concerns model, they do not know their ability, and so it is not obvious that the pool will improve, independent of the more experienced being more likely to win reelection.

[25] As shown in Table 1, many states adopted two-term limits in 1959 or 1967. We experiment with both of these start dates and find similar results in both sets of regressions. We report results for the former. The states are: Delaware, Maryland, New Jersey, Ohio, and Oregon (all of which had two-term limits between 1959 and 1996) and

direct comparison, we also exclude governors serving two- rather than four-year terms[26] and governors serving their third (or higher) terms in office.[27] We then conduct a test similar to the one described above, in which we compare performance under different types of governors, in an attempt to isolate the selection and moral hazard effects.

Consider the three different types of governors within this new sub-sample: (4) a first-term governor who was eligible for reelection, (5) a second-term governor who was eligible for reelection, and (6) a second-term lame duck. As before, let b_i be the expected performance of politician i. Then the model is consistent with the following hypotheses:[28]

> $b_4 < b_5$ because types (4) and (5) exert the same level of effort but type (5) is, in expectation, of higher ability due to the selection effect.
>
> $b_6 < b_5$ because types (5) and (6) have the same expected ability but type (5) works harder than type (6).
>
> We cannot order b_4 and b_6 because type (4) works harder but type (6) has higher expected ability.

To test these hypotheses, we run the following regression on a sample of eleven states that had either no term limits or two-term limits during the sample period:

> performance = β_1 first-term-eligible + β_2 second-term-eligible + β_3 second-term-lame-duck + β_4 controls + β_5 fixed effects + β_6 year effects

A slight rearrangement of this regression equation reveals that it is just a difference-in-differences specification in which we control for the effect of being in a two-term limit state (the

---

Connecticut, Idaho, Illinois, New York, Utah, and Washington (none of which had two-term limits between 1959 and 1996).

[26] Currently, New Hampshire and Vermont are the only states with two-year terms, but other states had two-year terms earlier in the sample period.

[27] That is, if a governor in a state without term limits served three terms, we include in our sample his first and second terms but not his third term.

[28] Because of the difficulty of extending our model to a third or subsequent term limit, these regressions to not relate directly to conjectures from the model, but are consistent with the spirit of the model.

treatment group, in the language of program evaluation), the effect of being a second-term governor (*after*), and the interactive effect of being a second-term governor in a two-term limit state—i.e. the effect of being subject to a binding two-term limit (*treatment group*after*):

$$\text{performance} = \gamma_1 \text{ term-limit state} + \gamma_2 \text{ second-term governor} + \gamma_3 \text{ term-limit state} * \text{second-term-governor} + \gamma_4 \text{ controls} + \gamma_5 \text{ fixed effects} + \gamma_6 \text{ year effects}$$

The variable "term-limit state" does not change over time and therefore is absorbed by the state fixed effects. The coefficient γ_2 is the effect of being a second-term governor who can run for reelection, relative to a first-term governor who can run for reelection (β_2 in the previous regression), while the sum of the coefficients γ_2 and γ_3 is the effect of being a second-term lame duck, relative to a first-term governor who can run for reelection (β_3 in the previous regression).

We expect performance to be better under second-term governors than under first-term governors, but that this effect will be partially or wholly offset by the presence of a binding term limit. Thus, a significant positive (in the sense of improving performance) coefficient on "second-term governor" would suggest a selection effect, while a significant (and opposite-signed) coefficient on the interaction between "second-term governor" and "term-limit state" would suggest a countervailing moral hazard effect. Besley and Case's specification effectively includes the interaction term but not the main effects, making it impossible to isolate moral hazard from selection.

5. Results

Public goods: one-term limits vs. two-term limits

The first and third columns of Table 3 show the results for surplus and growth, respectively, when we compare performance under first-term lame ducks, second-term lame

ducks, and first-term governors who are eligible for reelection (the omitted category). We do not find significant results, although the signs of the coefficients in the growth regressions go in the expected directions—that is, growth is lower under first-term lame ducks than under first-term governors who can run again, and also slightly lower than under second-term lame ducks.

[Table 3 about here]

This specification has the disadvantage of lumping together three very different types of first-term governors who are eligible for reelection: those who did not run for reelection, those who lost their reelection bids, and those who went on to win reelection to a second term. When we separate out these three types of first-term governors (columns two and four), we see that growth performance is exactly as predicted by the model: growth is lower under second-term lame ducks and (especially) first-term lame ducks than under first-term governors who are later "selected" into a second term (the omitted category). Not surprisingly, growth is also substantially lower under first-term governors who are defeated. The regressions for fiscal surplus appear to isolate the presence of deficits to those cases in which incumbents chose not to run for reelection. This makes intuitive sense, but the results do not achieve statistical significance.

Public goods: two-term limits vs. no term limits

Table 4 shows the results of regressions comparing performance in states with and without two-term limits under three different types of governors: first-term governors who were eligible for reelection, second-term lame ducks, and second-term governors who were eligible for reelection. Because this regression specification is not based on changes in term limit laws

over time, we are able to run these regressions on spending and taxes (as well as surplus and growth) without concern about their upward trends.

[Table 4 about here]

Although we do not get significant results for surplus or growth, the coefficients have the expected signs; that is, surpluses and growth appear to be slightly higher under second-term governors than under first-term governors, suggesting a selection effect, but lower under second-term lame ducks than under second-term governors who can run again, suggesting a moral hazard effect.

We do get significant results for spending and taxes, however. The negative coefficients on "second-term governor" in both regressions suggest that spending and taxes are lower under second-term governors than under first-term governors within states without term limits. This result is consistent with a selection effect if we follow Besley and Case in interpreting small budgets as a public good. Similarly, the positive coefficients on the interaction between "second-term governor" and "term-limit state" reveal that this positive effect of tenure in office on performance is offset by the removal of the reelection incentive—that is, by moral hazard. In the case of spending, this effect more than offsets the selection effect, so that spending is slightly higher under second-term lame ducks, relative to first-term governors. In the case of taxes, the moral hazard effect only partially offsets the selection effect, so that taxes are slightly lower under second-term lame ducks, compared to first-term governors. In summary, our findings for spending and taxes support the presence of both selection and moral hazard effects.

Experience

A comparison of the political experience of first-term lame ducks, second-term lame ducks, and first-term governors who are eligible for reelection (see Table 5) presents an interesting picture, showing what a powerful filter elections can be. First-term lame ducks have eight years of political experience, on average, compared to second-term lame ducks, who have approximately fourteen years of experience. The regression also reveals a striking contrast regarding first-term governors eligible for reelection: those who subsequently lost or withdrew have about six years less political experience than those who won reelection.[29] For the reasons mentioned above, ability and performance are not the same thing; however, if experience is at all connected with ability, then these differences in experience bear witness to the importance of reelection motives in eliciting better performance, and of term limits in removing that motivation.

[Table 5 about here]

Ideological Congruence

A governor's congruence depends on how far her ideological position is from that of the average voter in the state. Besley (2006: 122-3) finds (in a full cross-section analysis not controlling for regime changes) that term-limited governors are more congruent than those who can run for re-election. That result seems inconsistent with the usual argument that we get worse performance from the last-period types because of moral hazard.[30] Besley himself regards his result as "suggestive of a strong selection effect into last periods in office where voters get to

[29] We believe the reason that second-term lame ducks do not have a full four years more experience than first-term governors who went on to get reelected is because of censoring at the end of the sample.
[30] Indeed, in Table 3.5 Besley shows, with the same specification, that spending and taxes are higher among governors who cannot run, indicating shirking.

find incumbents who are more in tune with their preferences," because voters, anticipating a last-period problem, try hard to find congruent types.

Table 5 shows that first-term lame ducks were more than two points[31] less congruent, and subsequent first-termers who lost re-election bids were more than three points less congruent, than those who withdrew, won re-election, or became second-term lame ducks. If congruence reflects effort, then these results are quite intuitive: first-term lame ducks put less effort into carrying out popular preferences, as do those first-termers under a two-term limit who fail to be re-elected. Since second-term lame ducks outperform first-term lame ducks, this is consistent with a selection effect.[32]

The first column of Table 6 compares governors in states with and without two-term limits and reveals that second-term governors are about 3.7 points less congruent, while second-term lame ducks are 4.6 points more congruent, than first-term governors, apparently confirming Besley's result.

[Table 6 about here]

However, in states with two differentiated parties, divergence can also be interpreted to mean that the governor drifts away from her own party's position toward the other side. In terms of our model, effort and ability reflect the governor implementing policies reflecting her party's position, and divergence is what we might expect of a last-term incumbent who shirks. Thus, the second column of Table 6 replaces the dependent variable with the difference between

[31] The underlying "point" is a unit on an ADA scale score, which is an index ranging from zero to 100. Since the average governor is twenty-plus points away from the center of opinion (see Table 2), the difference attributable to first-term lame duck status is about ten per cent of the mean value.

[32] The congruence data begins in 1960, so first-term lame ducks are less frequent than in Table 3. On the other hand, the fact that in the early period many of the one-term-limit states were southern, and effectively had only one party, makes interpreting the result as we do more straightforward.

(Democratic) governors' ideological scores and the corresponding scores for the state Democratic Party. Now we see that those governors who have been re-elected are approximately four points more congruent with their party's ideology than were first-termers, indicating a selection effect but not toward the center of public opinion. Second-term lame ducks, indeed, are 2.6 points more divergent than those who can be re-elected; this finding is consistent with moral hazard in a model in which effort means that governors implement party positions, though the estimate is barely larger than its standard error. Results for Republican governors are statistically insignificant. Perhaps the safest conclusion is that while there is some tantalizing evidence of selection effects, our model is not ideal for generating conjectures about congruence. This is partly because policy preferences do not naturally fit an ability model, but also because if implementing party positions is important, the parties need to be in the model from the outset.

6. Conclusions

We have developed a model and employed a natural experiment to examine a variety of conjectures about the effects of discipline and selection in elections. We have found evidence of both effects, although the results are somewhat mixed. Evidence of selection effects includes the finding that growth is higher under second-term lame ducks than first-term lame ducks. Moreover, our results for spending and taxes are consistent with the existence of both selection and moral hazard effects if one assumes, as the previous literature does, that lower taxes and spending mean that the incumbent is providing relatively more goods at lower cost.

One important caveat about our analysis is that none of our dependent variables are perfect measures of gubernatorial performance. Indeed, some of our measures, like growth, are probably largely determined by factors outside the governor's control. We would like to have

used job approval ratings, but unfortunately the data do not extend back far enough to study one-term-limited governors, possibly because, with no prospect of re-election, there was little reason to conduct such a survey.

Even though we cannot settle the normative debate on term limits, our results do make a contribution to this discussion. The debate contrasts incumbent advantage in elections with incumbent experience in office, careerism with professionalism, and the desirability of bringing in new blood with voters' ability to choose the best-qualified representative. Our results show that concerns about experience are relevant, and that voters do seem to try to select for ability as well as punish shirking. More importantly, we show that the length, not just the existence, of term limits matters: two-term limits allow entry by more experienced candidates than do one-term limits, and experience appears to affect performance. Unfortunately, as a result of the difficulties we encountered in generating equilibria for multiple-term settings – and the lack of data for terms of more than four years – we can only suggest that further inquiry into optimal term limits would be rewarding.[33]

The question of whether elections are a selection mechanism as well as a mechanism with which to discipline incumbent performance remains an important one. We believe this paper offers sufficient theoretical and empirical grounds to redirect scholarly attention toward exploiting variation caused by institutional change in order to understand the nature of representative democracy. We look forward to exploring this research agenda further in the future.

[33] See Chappell and Keech, 1983, for an example of such an evaluation.

Works Cited

Alesina, Alberto and Guido Tabellini. 2003. "Bureaucrats or Politicians?" NBER Working Paper no. 9809.

Alt, James E. and Robert C. Lowry. 2000. "A Dynamic Model of State Budget Outcomes Under Divided Partisan Government." *Journal of Politics* 62: 1035-69.

Ashworth, Scott. 2005. "Reputational Dynamics and Political Careers." *Journal of Law, Economics and Organization* 21: 441-446.

Ashworth, Scott and Ethan Bueno de Mesquita. 2005. "Electoral Selection and the Incumbency Advantage." Washington University typescript.

Ashworth, Scott and Ethan Bueno de Mesquita. 2006. "Delivering the Goods: Legislative Particularism in Different Electoral and Institutional Settings." *Journal of Politics* 68:168-179.

Austen-Smith, David and Jeffrey Banks. 1989. "Electoral Accountability and Incumbency." In Peter Ordeshook (ed.), *Models of Strategic Choice in Politics*. Ann Arbor: University of Michigan Press.

Berry, William, Evan Ringquist, Richard Fording and Russell Hanson. "Measuring Citizen and Government Ideology in the American States, 1960-1993." *American Journal of Political Science* 42 (January 1998): 327-48.

Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government*. Oxford: Oxford University Press.

Besley, Timothy and Anne Case. 1995. "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits." *Quarterly Journal of Economics* 110 (3): 769-798.

Besley, Timothy and Anne Case. 2003. "Political Institutions and Policy Choices: Evidence from the United States." *Journal of Economic Literature* 41(1): 7-73.

Beyle, Thad, Richard G. Niemi, and Lee Sigelman. 2002. "Gubernatorial, Senatorial, and State-level Presidential Job Approval: The U.S. Officials Job Approval Ratings (JAR) Collection." *State Politics & Policy Quarterly* 2(3), 215-29.

Carey, John, Gary F. Moncrief, Richard G. Niemi, Lynda W. Powell, "Term Limits in the State Legislatures: Results from a New Survey of the 50 States," *Legislative Studies Quarterly*, forthcoming.

Chappell, Jr., Henry W. and William R. Keech. 1983. "Welfare Consequences of the Six Year Presidential Term Evaluated in the Context of a Model of the U.S. Economy. *American Political Science Review*, 77: 75-91.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *Review of Economic Studies* 66(1):199-217.

Fearon, James. 1999. "Electoral Accountability and the Control of Politicians." In Adam Przeworski, Susan C. Stokes**,** and Bernard Manin (eds.)**,** *Democracy, Accountability and Representation.* New York: Cambridge University Press, pp. 55-97.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5-26.

Gordon, Sanford, Gregory Huber, and Dimitri Landa. 2006. "Challenger Entry and Voter Learning." NYU typescript.

Gowrisankaran, Gautam, Matthew F. Mitchell, and Andrea Moro. 2005. "Why Do Incumbent Senators Win? Evidence From a Dynamic Selection Model." Washington University typescript.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66(1):169-182.

Johnson, Joseph M., and W. Mark Crain. 2004. "Effects of Term Limits on Fiscal Performance: Evidence from Democratic Nations." *Public Choice* 119: 73-90.

Kousser, Thad. 2005. *Term Limits and the Dismantling of State Legislative Professionalism*. New York: Cambridge University Press.

Lohmann, Susanne. 1998. "An Informational Rationale for the Power of Special Interest." *American Political Science Review* 92(4):809-827.

Lowry, Robert C., James E. Alt, and Karen Ferree. 1998. "Fiscal Policy and Elections in American States." *American Political Science Review* 92(4):759-74.

Persson, Torsten and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge: MIT Press.

Zaller, John. 1998. Politicians as Prize Fighters: Electoral Selection and the Incumbency Advantage. In *Politicians and Party Politics,* ed. John G. Geer. Baltimore, MD: Johns Hopkins University Press.

Table 1. Gubernatorial Term Limit Laws in the American States, 1950-2000

Type of term limit law	States
No term limit	CT, ID,[a] IL, IA, MA, MN, NH, NY, ND, TX, UT,[b] VT, WA, WI
1-term limit	VA
2-term limit	DE, NJ, OR
Switched from no term limit to 2-term limit (year of switch)	AZ (1992), AR (1992), CA (1990), CO (1990), KS (1972), ME (1966), MD (1954), MI (1992), MT[f] (1992), NE (1966), NV (1970), NM,[c] OH (1958), RI (1994), SD (1956), WY[f] (1992)
Switched from 1-term limit to 2-term limit (year of switch)	AL (1968), FL (1968), GA (1976), IN[e] (1972), KY (1994), LA (1966), MS (1994), MO (1965), NC[e] (1977), OK (1966), PA (1967), SC (1980), TN (1978),[d] WV (1970)

[a] ID had a 1-term limit until 1955.
[b] UT adopted a 3-term limit in 1994, but it was never binding during the sample period.
[c] NM had a 2-term limit during the sample period except for 1972-1990, when it had a 1-term limit.
[d] TN adopted the 1-term limit in 1955.
[e] IN and NC allow a third term after a four-year hiatus.
[f] MT and WY restrict Governors to two terms or eight years in 16.

Table 2. Summary Statistics

	States that switched from 1- to 2-term limits		States with no term limits		States with 2-term limits	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Surplus	9.87	51.11	4.50	73.91	10.50	108.8
Growth	0.024	0.032	0.020	0.031	0.020	0.032
Congruence	-21.33	10.85	-29.33	10.83	-23.74	9.92
Experience	9.52	7.45	10.57	6.88	10.66	6.77
Spending (per capita)	896.41	404.18	1071.01	514.8	1061.06	542.56
Taxes (per capita)	448.08	195.33	554.72	268.05	497.38	257.63
First-term lame duck	0.44	0.50				
Second-term lame duck	0.20	0.40				
First-term eligible - lost	0.05	0.22				
First-term eligible - withdrew	0.08	0.27				
First-term eligible - reelected	0.23	0.42				
Term-limit state			0.00	0.00	1.00	0.00
Second-term governor			0.49	0.50	0.36	0.48
Term-limit state * second-term governor			0.00	0.00	0.36	0.48
State income	8,664.55	2,790.55	10,631.70	3,310.28	11,074.66	2,992.84
Population	4,917,404	2,797,157	6,147,243	6,084,351	4,764,112	3,525,412
Percent elderly	0.11	0.03	0.10	0.02	0.10	0.02
Percent school-aged	0.23	0.03	0.23	0.04	0.22	0.03
Democratic governor	0.71	0.45	0.50	0.50	0.52	0.50
Democratic House	0.89	0.31	0.45	0.50	0.60	0.49
Democratic Senate	0.86	0.35	0.40	0.49	0.66	0.47
Divided government	0.26	0.44	0.49	0.50	0.42	0.49
States in sub-sample	AL, FL, GA, IN, LA, MO, NC, OK, PA, SC, TN, WV		CT, ID, IL, NY, UT, WA		DE, MD, NJ, OH, OR	

Table 3. One- vs. Two-Term Limits

Dependent variable:	Surplus		Growth	
First-term lame duck	3.64	3.32	-0.004	-0.006*
	(7.90)	(8.19)	(0.003)	(0.004)
Second-term lame duck	-1.63	-3.93	-0.003	-0.005**
	(6.29)	(6.19)	(0.002)	(0.002)
First-term eligible - lost		0.67		-0.009**
		(9.32)		(0.004)
First-term eligible - withdrew		-13.28		-0.004
		(12.32)		(0.006)
Constant	74.18	-0.85	0.081	0.077
	(174.85)	(2.04)	(0.066)	(0.067)
Number of observations	531	531	520	520
R-squared	0.40	0.40	0.67	0.67

Notes: The constant term captures the omitted category (first-term governors eligible for reelection in columns one and three; first-term governors who went on to win reelection in columns two and four). Control variables in the growth regression are: state population, population squared and cubed, percent elderly, percent school-aged, Democratic governor, Democratic House, Democratic Senate, divided government, state and year effects. Surplus regressions also include state income, income squared and income cubed. Robust standard errors in parentheses.

* Significant at 10% level.

** Significant at 5% level.

Table 4. Two-Term Limits vs. No Term Limits

Dependent variable:	Surplus	Growth	Spending	Taxes
Second-term governor	2.26	0.004	-39.23**	-25.16**
	(9.99)	(0.003)	(15.02)	(7.80)
Second-term governor * Term-limit state	-15.45	-0.004	44.55**	17.43*
	(14.08)	(0.004)	(20.85)	(10.55)
Constant	1149.36**	-0.064	1357.08**	-1203.52**
	(383.72)	(0.070)	(557.16)	(298.88)
Number of observations	366	366	366	366
R-squared	0.60	0.60	0.97	0.96

Notes: "Term limit state" drops out because it is perfectly collinear with the state fixed effects. Control variables are: state population, population squared and cubed, percent elderly, percent school-aged, Democratic governor, Democratic House, Democratic Senate, divided government, state and year effects. Surplus, spending and tax regressions also include state income, income squared and income cubed. Robust standard errors in parentheses.
* Significant at 10% level.
** Significant at 5% level.

Table 5. One- vs. Two-Term Limits: Experience and Ideological Congruence

Dependent variable:	Experience		Congruence	
First-term lame duck	-2.48	-4.63*	-1.86**	-2.63**
	(2.63)	(2.62)	(0.92)	(0.95)
Second-term lame duck	3.25	1.59	0.19	0.26
	(2.14)	(2.53)	(0.58)	(0.60)
First-term eligible - lost		-6.11		-3.42**
		(4.05)		(1.43)
First-term eligible - withdrew		-5.97		1.82*
		(3.91)		(1.08)
Constant	10.25	12.76*	87.83**	83.38
	(6.57)	(6.52)	(29.84)	(30.55)
Number of observations	136	136	408	408
R-squared	0.43	0.46	0.83	0.84

Notes: The constant term captures the omitted category (first-term governors eligible for reelection in columns one and three; first-term governors who went on to win reelection in columns two and four). Control variables are: state population, population squared and cubed, percent elderly, percent school-aged, Democratic governor, Democratic House, Democratic Senate, divided government, state and year effects. Robust standard errors in parentheses.
* Significant at 10% level.
** Significant at 5% level.

Table 6. Two-Term Limits vs. No Term Limits: Ideological Congruence and Political Parties

Dependent variable:	Congruence	Democratic Congruence	Republican Congruence
Second-term governor	-3.67**	4.06**	-0.24
	(1.13)	(1.96)	(0.84)
Second-term governor * Term-limit state	4.56**	-2.61	0.02
	(1.58)	(2.49)	(0.95)
Constant	-40.58	-73.60	39.72
	(46.22)	(68.99)	(43.16)
Number of observations	344	195	147
R-squared	0.70	0.76	0.98

Notes: "Term limit state" drops out because it is perfectly collinear with the state fixed effects. Control variables are: state population, population squared and cubed, percent elderly, percent school-aged, Democratic governor, Democratic House, Democratic Senate, divided government, state and year effects. Robust standard errors in parentheses.
* Significant at 10% level.
** Significant at 5% level.