

**Accountability, Selection, and Experience:
Theory and Evidence from U.S. Term Limits**

James Alt
Harvard University
jalt@latte.harvard.edu

Ethan Bueno de Mesquita
Harris School of Public Policy Studies
University of Chicago
bdm@uchicago.edu

Shanna Rose
Wagner School of Public Service
New York University
shanna.rose@nyu.edu

October 10, 2007

Abstract

Previous theoretical and empirical research on term limits has focused on the problem of accountability—that is, the possibility that term-limited politicians exert less effort than those who are eligible to run for reelection. We present a model with both accountability and selection effects, in which term limits not only cause incumbents to shirk but also interfere with voters' ability to reelect high-quality candidates. We also show how this model can be extended to address the possibility that incumbent quality increases with experience in office. We evaluate our model by taking advantage of variation in gubernatorial term limit laws across states and over time. We find evidence suggesting that term limits not only cause politicians to shirk, but also reduce expected incumbent quality by limiting voters' ability to select competent politicians and by limiting incumbents' ability to gain experience in office.

We thank Tim Besley, John Gasper, Christina Gathmann, Rema Hanna, Ilyana Kuziemko, Rene Lindstadt, and seminar participants at the University of Chicago, Florida State University, the 2006 Annual Meeting of the Midwest Political Science Association, and the 2007 Annual Meeting of the American Political Science Association for helpful comments.

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 empirical literature on this topic is surprisingly small and focuses almost exclusively on the problem of accountability, or moral hazard. This emphasis on elections as a mechanism for punishing poorly-behaved incumbents ignores the possibility raised by many observers that elections are primarily a selection mechanism that enables voters to choose competent politicians (see, for example, Fearon 1999), which would suggest that term limits have the additional effect of removing voters' ability to reelect high-quality politicians.

In this paper, we present and test a model that allows for voters to select among incumbents on the basis of ability, as well as for elections to discipline incumbents. We use this model to generate predictions about incumbent performance under one-term limits, two-term limits, and no term limits. The expected ability of incumbents improves with each additional term in office, offsetting the effect of shirking among second-term lame ducks but not first-term lame ducks.¹ The model therefore predicts that performance will be worse under one-term limits than under two-term limits, and worse under two-term limits than under no term limits. We also show that these results hold up even if, in addition to the selection effect, incumbent quality improves with experience in office as a result of on-the-job learning.

We test our model using data from the American states, which have witnessed a major change in gubernatorial term limit laws over the last half century: the gradual shift from one-term limits to two-term limits in nearly one-third of the states. During this same period, other states had two-term limits or no term limits. This variation in term limit laws across states and time

¹ 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 an incumbent continues to serve despite not having been re-elected.

allows us to isolate the selection and accountability effects of elections using two sets of empirical tests. We first examine the experience of states that switched from one-term limits to two-term limits, comparing the performance of first-term lame ducks, first-term governors who were eligible for reelection, and second-term lame ducks. We then examine the states that have had either had two-term limits or no term limits over the past fifty years, comparing the performance of first-term governors who were eligible for reelection, second-term lame ducks, and second-term governors who were eligible for reelection.

Our results support the model's predictions. Using several measures of economic performance, we find evidence suggesting that term limits have both accountability and selection effects. All else equal, per capita taxes and spending are higher under first-term governors than under second-term governors, indicating a selection effect, but are also higher under lame ducks than under governors who can run again, indicating an accountability effect. Similarly, state borrowing costs (which are inversely related to the skill with which state finances are managed) are higher under first-term governors than under second-term governors, and are also higher under term-limited governors than under governors who are eligible for reelection. State economic growth (relative to the national average) is higher under governors who win reelection than under those who are weeded out through the electoral process, and also higher under reelection-eligible governors than under lame ducks. We also find that, in addition to these selection and accountability effects, economic performance appears to improve as incumbents gain experience in office. Together, these findings suggest that term limits not only cause politicians to shirk, but also reduce expected incumbent quality both by limiting voters' ability to select competent politicians and by limiting incumbents' ability to gain experience in office.²

² Throughout the paper, we estimate the *marginal* effects of term limits. That is, we do not assume that term-limited governors have no career ambitions beyond the governor's office or that they do not care about their reputation *per se*

These findings have policy implications for term limits at other levels of government within the United States, such as the presidency and state legislatures. There are also many governments in other parts of the world that have one-term limits, such as Mexico and Turkey, or two-term limits, such as Portugal and Sri Lanka. A major advantage of studying term limits within the United States rather than across countries, however, is that the states share a broadly similar institutional and constitutional structure, minimizing the possibility that unobserved heterogeneity will bias the results.

1. Literature Review

The theoretical literature on elections discusses two ways in which elections serve the interests of voters. Barro (1973) and Ferejohn (1986) model elections as a sanctioning device in which the desire to be reelected in the future leads politicians to exert effort in the present. In these pure moral hazard models, because all candidates are identical, voters are indifferent between the incumbent and the challenger at the election. As such, voters are free to choose a retrospective voting rule that maximizes their *ex ante* expected utility by threatening not to reelect unless the outcome is good enough, which induces effort from the incumbent.

Other theorists have discussed the idea that elections can serve as a device for selecting candidates who have some intrinsic characteristic that is of value to voters, such as quality, honesty, or preference congruence. Some such models treat elections as pure selection mechanisms, having no effect on behavior (e.g., Zaller 1998, Gowrisankaran, Mitchell, and Moro 2006, and Ashworth and Bueno de Mesquita 2007), while others allow for electoral decisions to be based on both candidate characteristics and candidate behavior (e.g., Banks and Sundaram

(i.e. their “legacy”). Instead, we assume that since non-term-limited governors share these concerns, we can compare the two types of governors to determine whether term limits affect governors’ behavior at the margin.

1998, Persson and Tabellini 2000, Ashworth 2005, Ashworth and Bueno de Mesquita 2006, Besley 2006). We follow in this latter tradition, examining how term limits affect both politician behavior and voter capacity for selection of good types.

Our theoretical model differs from the existing literature in a few ways. We, like much of the literature, focus on an application of Holmström's (1999) "career concerns" framework in which there is moral hazard and symmetric uncertainty about candidate ability. Earlier papers in this tradition either considered games in finite time (Persson and Tabellini 2000, Ashworth 2005, Ashworth and Bueno de Mesquita 2006) or restricted attention to two-term limits (Banks and Sundaram 1993). In order to consider the effect of term limits on performance, we allow for politicians who are not term-limited in an infinitely repeated game. It is worth noting that in order to do so, we must simplify the environment from that studied in these earlier papers.

Besley (2006) also considers an infinitely-repeated model under different term-limit regimes. The key difference between our approach and Besley's is that he assumes that politicians have private information about their type and focuses on separating equilibria of a signaling game, while we follow Holmström in assuming symmetric uncertainty, so that politicians have no information advantage to exploit through signaling.³ Instead, both the politician and the voters learn about the politician's ability over time.

³ In our view, there are interesting political economy settings in which both symmetric and asymmetric uncertainty over candidate type are reasonable assumptions. Besley's asymmetric information environment is particularly natural when the uncertainty is over candidate characteristics that the politician can learn from introspection (e.g., primitive policy preferences). We think the symmetric uncertainty assumption we (and many others) make is natural when the uncertainty is over ability, in the sense of how effective the politician is likely to be at governance or at getting things done. Here, as long as the voters have access to the politician's past record, much of the uncertainty comes from how idiosyncratic facts about the candidate end up interacting with a variety of elements of the institutional and structural environment the politician ends up facing. It seems likely that both candidates and voters will face substantial uncertainty about such things.

Empirical Literature

The empirical literature on the role of elections in representative democracy has focused primarily on accountability rather than selection. To test the hypothesis that elections give incumbents an incentive to exert effort, the literature examines whether outcomes are any different in the absence of a reelection motive. For instance, a sizeable literature investigates whether members of Congress “shirk” in their final terms prior to retirement, where shirking is defined either in terms of reduced effort (i.e. voting less frequently) or ideological congruence with voters (i.e. voting more “sincerely”). Vanbeek (1991) and Lott and Bronars (1993) find that retiring members vote less frequently, but do not find statistically significant last-period effects in the content of members’ votes, while Figlio (1995), Tien (2001), and Snyder and Ting (2003) find evidence that members vote more sincerely in the final term. Similarly, McArthur and Marks (1988) find that, in post-election sessions, members who have not been reelected vote systematically differently than those who have been reelected.

A related literature looks for last-period effects in state legislatures. The 1990s saw one-third of states impose term limits on state legislators, although several states have since repealed these laws. A few recent studies examine the impact of the introduction of legislative term limits on the composition of state legislatures and the behavior of term-limited legislators and find mixed results (Kousser 2005; Carey et al. 2006; Kurtz, Cain, and Niemi 2007). However, the fact that most legislative term limits first took effect between 2000 and 2002 makes it difficult to conduct a systematic analysis of shirking or selection effects.

Gubernatorial term limits, by contrast, date from as early as 1787 (Delaware), and therefore offer greater scope for empirical analysis. The most influential paper on gubernatorial term limits is by Besley and Case (1995). The authors develop an accountability model in which

the prospect of reelection gives office-seeking politicians an incentive to exert effort, where effort is defined as “keeping taxes and expenditures down.” The model therefore predicts that term-limited governors tax and spend more than those who are eligible to run again. Besley and Case find evidence that, in the 48 contiguous states between 1950 and 1986, per capita spending was \$14 higher and per capita taxes were \$7 higher (in 1982 dollars) when the governor could not run for reelection. However, in a later (2003) paper, the authors repeat their analysis on a sample that extends through the mid-1990s and find that the effect of term limits has disappeared over time. The authors conclude that “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).

One potential explanation for this puzzling result has nothing to do with governors’ behavior: over the past fifty years, many states have switched from one-term limits to two-term limits, as shown in Table 1.⁴ In the early postwar years, the majority of term-limited incumbents served under “no succession” laws limiting them to a single term; today Virginia alone retains the practice. 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.⁵

[Table 1 about here]

If voters use elections to weed out low-quality incumbents, then as average tenure increases, the lame duck variable will increasingly pick up the offsetting effect of higher

⁴ In addition, many states that did not previously have term limits have adopted two-term limits; if these states are systematically different from those that already had term limits in 1950 for reasons that have nothing to do with term limits, 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. We return to this point in the empirical section, when designing our own regression specifications.

⁵ This specification has generated considerable confusion in the literature. 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.... Besley and Case do not analyze the behavior of single-term limited governors, a relatively rare institution in the United States” (p. 75).

incumbent quality over time. Indeed, in his 2006 book on government accountability, Besley acknowledges in a footnote that when he separates one-term limits from two-term limits, most of the impact on spending and taxes appears to be generated by governors serving under one-term limits (Ch. 3. n. 23).

The conjecture that one- and two-term limits might have different effects as a result of a selection effect remains largely unexamined. Only one previous paper tests this conjecture empirically: based on a panel of countries from 1972 to 1990, Johnson and Crain (2004) find that, as expected, the size of government has expanded more rapidly in countries with one-term limits than in countries with two-term limits. However, as mentioned earlier, studying the effects of term limits across countries is complicated by institutional and constitutional heterogeneity. Thus, in the remainder of the paper we develop a formal model of accountability, selection, and term limits and test its predictions using data from the American states.

2. Model

Consider a game in which there are two kinds of players: politicians and a voter. The order of play is as follows. At the beginning of each period, the incumbent politician chooses a level of effort to exert in providing public goods $a \in \{\underline{a}, \bar{a}\}$. 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. A politician is *term limited* or a *lame duck* if she can only serve one more period. If a politician is not term limited she runs against a randomly selected challenger (described later). If she is term limited, there is an open seat election with two randomly selected challengers.

Each politician can be of two types $\{\theta_L, \theta_H\}$. The probability that a randomly drawn politician is of type θ_H is π . There is symmetric uncertainty about the politicians' types (i.e., no one has private information, including the politician herself).⁶ Moreover, there is some chance in any round that a politician of high type becomes low type.⁷ This transition probability is ε , where $\pi < 1 - \varepsilon$ (i.e., it is more likely that a politician who was a high type last period is a high type this period than it is that a new politician is a high type).

The outcome space is $\{L, H\}$ where the function $f : \{\theta_L, \theta_H\} \times \{\underline{a}, \bar{a}\} \rightarrow \Delta(\{L, H\})$, mapping actions and types into lotteries over outcomes is given by:

$$f(\bar{a}, \theta_H) = \begin{cases} H & \text{with probability } \gamma \\ L & \text{with probability } 1 - \gamma \end{cases}$$

$$f(\underline{a}, \theta_H) = \begin{cases} H & \text{with probability } \lambda \\ L & \text{with probability } 1 - \lambda \end{cases}$$

$$f(\bar{a}, \theta_L) = L$$

$$f(\underline{a}, \theta_L) = L,$$

with $\gamma > \lambda$. We interpret the outcome H as the politician providing a high level of public goods and the outcome L as the politician providing a low level of public goods. To keep the model simple, we assume that low ability types are not able to provide a high level of public goods.⁸ We also strengthen our assumption on the priors, assuming that $(1 - \varepsilon)\lambda/\gamma > \pi$, which is a necessary (but not sufficient) condition for the voter to consider reelecting an incumbent.

Denote by $\mu(O | a, \mu_0)$ the voter's belief that the incumbent is a high type, given that the outcome was O , the voter believes the politician took action a , and the prior belief was μ_0 .

⁶For other models of electoral accountability that assume symmetric uncertainty of this sort see, for example, Ashworth (2005), Ashworth and Bueno de Mesquita (2006), and Persson and Tabellini (2000).

⁷This is a simple way to insure continuing uncertainty. In the absence of uncertainty, incentive effects disappear.

⁸This assumption simplifies the analysis without sacrificing any intuition. Similar results would follow in a model where the degenerate probability was on the high type yielding the good outcome and the low type with high effort could achieve that outcome probabilistically.

A strategy for the voter ($r(\mu)$) maps beliefs (and terms remaining, not notated) into a decision over whether or not to reelect (where $r(\mu) = 1$ is interpreted as reelect for certain and $r(\mu) = 0$ is interpreted as do not reelect). A strategy for the incumbent maps beliefs (and terms remaining) into a choice of effort.

The cost of high effort is c (low effort's cost is normalized to zero). The benefit of reelection is B . So, the politician's payoff from any period in which she is in office is:

$$B - c(a)$$

The politicians discount the future by δ .

We assume that voters are myopic (they care only about next period). The voter's payoff from an outcome O is given by V_O , where we normalize $V_H = 1$ and $V_L = 0$.

Given this set-up, we can characterize the voter's posteriors on the incumbent's type ($\mu(O | a, \mu_0)$), given a belief about actions, an outcome, and a prior:

$$\begin{aligned}\mu(H | \bar{a}, \mu_0) &= 1 - \varepsilon \\ \mu(H | \underline{a}, \mu_0) &= 1 - \varepsilon \\ \mu(L | \bar{a}, \mu_0) &= \frac{(1 - \gamma)\mu_0(1 - \varepsilon)}{(1 - \gamma)\mu_0 + 1 - \mu_0} \\ \mu(L | \underline{a}, \mu_0) &= \frac{(1 - \lambda)\mu_0(1 - \varepsilon)}{(1 - \lambda)\mu_0 + 1 - \mu_0}.\end{aligned}$$

We focus on pure strategy Markov perfect equilibria (MPE). This restricts us to equilibria that are stationary in the payoff relevant state variables—beliefs about ability and terms remaining prior to being term limited.

One-Term Limit

In the absence of any chance of reelection, effort imposes costs without offering any benefits for the politician. As such, a politician who can only serve one term will choose \underline{a} in that term. The following result is immediate and stated without proof.

Proposition 1. *With a one-term-limited politician, expected public goods provision is $\pi\lambda$ in every period.*

Two-Term Limit

Now suppose a politician can serve at most two terms (but that, if not reelected, can never run again). Just as in the one-term-limit case, it is clear that when she is a lame duck in her second term, an incumbent will choose \underline{a} . The question is, can \bar{a} be sustained in the first term through electoral incentives?

The answer turns out to depend on parameter values. The basic idea is that the voter will only reelect incumbents whom he believes are likely to be high ability. The voter forms his beliefs by observing outcomes. Thus, the benefit of effort is that it makes it more likely that the incumbent will provide a high level of public goods, convince the voter she is of high ability, and thereby achieve reelection. If, however, the costs of effort are too high, the benefits of office are too low, or the increase in probability of providing a high level of public goods that is achieved through high effort is too low, then the politician will choose not to invest in a high level of effort, even in her first term. These intuitions are formalized in the following result.⁹

Proposition 2.

1. *If $c \leq \delta B\pi(\gamma - \lambda)$, then there is a pure strategy MPE in which expected public goods provision from a first-term politician is $\pi\gamma$ and from a second-term politician is*

⁹ The proofs of this and all subsequent results appear in the appendix.

$(1 - \varepsilon)\lambda$.

2. If $c \geq \delta B\pi(\gamma - \lambda)$, then there is a pure-strategy Markov-perfect equilibrium in which expected public goods provision from a first-term politician is $\pi\lambda$ and from a second-term politician is $(1 - \varepsilon)\lambda$.

If the inequality is strict (i.e., $c \neq \delta B\pi(\gamma - \lambda)$), then the equilibria are unique in their respective parts of the parameter space.

This result implies that first-term outcomes are better for the voter under two-term limits than under one-term limits due to the incentive effect—politicians who can be reelected exert positive effort if the cost is not too high. Moreover, more public goods are also provided by a second-term lame duck than by a first-term lame duck. This is because, although no lame duck has an incentive to exert high effort, there is a selection effect. Only high-ability types survive to be lame ducks in the two-term-limit system. Consequently, second-term lame ducks are more likely to be of high ability than are first-term lame ducks.

No Term Limits

The absence of term limits enhances the incentive effect, since it provides the incumbent with a longer time horizon. In particular, unlike in the case of a two-term limit, the voter is able to provide second- (and later-) term politicians with electoral incentives. Moreover, selection effects continue to operate. Thus, we have the following result:

Proposition 3. *There exists a $\eta > 0$ such that, when $c \leq \delta B\pi(\gamma - \lambda) + \eta$, there exists a MPE in which politicians exert high effort in their first two terms. In this equilibrium, the expected output from first-term politicians is $\pi\gamma$ and the expected output from second-term politicians is $(1 - \varepsilon)\gamma$.*

This result demonstrates that there is an equilibrium in which the conditions for sustaining effort in the first period are weaker under no term limits than under two-term limits. We focus on this equilibrium.¹⁰ The reason is that, when politicians can serve more terms in office, the benefit of winning reelection includes not only the immediate payoff, but also the possibility of future terms. As such, politicians are willing to exert more effort to attempt to win reelection. Moreover, the voter can also induce high effort in the second term, since there are still electoral incentives. Finally, selection continues to operate.

Empirical Implications: Comparisons across Systems

Given these results, we can make comparisons across systems in order to derive testable implications of our model that can be evaluated using data from the states. First, the model yields the prediction that the two-term-limit system is better for the voter than the one-term-limit system. In the first term, the two-term-limit politician works harder than the one-term-limit politician, because she has electoral incentives. In the second term, the two-term-limit politician exerts low effort. However, the previous election selected for a high-ability politician. So, despite the fact that she exerts low effort, the second-term incumbent is nonetheless more likely to yield a good outcome because she is more likely to be of high quality.

Proposition 4. *Politicians in one-term-limit systems produce fewer public goods in expectation than do politicians in either term of a two-term limit system.*

A second testable implication of the model is that, in equilibrium, the no-term-limit system yields better outcomes for the voter than the two-term-limit system in both the first and second terms of a politician's time in office.

¹⁰ Restricting attention to a particular equilibrium which serves voter interests is standard when there are multiple equilibria (e.g., Besley 2006).

Proposition 5. *For all parameter values, there exists a MPE of the game with no term limits such that the expected level of public goods in a politician's first and second terms is weakly higher than the expected level of public goods in a politician's first and second terms in the pure strategy MPE of the game with a two-term limit. Moreover, there exist parameter values such that this relationship is strict rather than weak.*

The model also yields predictions about how changes in behavior from term to term differ in two-term-limit systems and no-term-limit systems. To see this, restrict attention only to those politicians who successfully achieve reelection to a second term. In both systems, these politicians had to be of the high type to achieve reelection. This means that, in the model, the probability that a second-term politician is the high type is $1-\varepsilon$, regardless of whether or not that politician is term-limited.

Now consider the comparison of that politician's expected level of public goods provision in her second versus her first term. In both two-term-limit and no-term-limit cases, conditioning on the fact of reelection, the politician is certain to have provided a high level of public goods in her first term (since only high producers are reelected). In the second period, in the two-term-limit system, the expected level of public goods is lower, since she is expected to be of high ability, but expends low effort. In the no-term-limit system, the expected level of public goods is also lower but less so, since the politician is still expected to be of high ability and continues to work hard, which makes her likely (but not certain) to produce the good outcome. This intuition is formalized below, and reported in the following table.

Proposition 6. *For small ε , the difference between second- and first-term performance among politicians who are reelected to a second term is expected to be negative in both a two-term-limit and a no term limit system. However, the difference will be larger in a two-term-limit system.*

[Table 2 about here]

Thus, the model's third testable implication is that the degree to which performance is lower in the second term should be more pronounced among politicians who face a term limit than among politicians who do not face a term limit. This prediction can be tested through a difference-in-differences design comparing first- and second-term performance among politicians who are reelected to a second term in states with and without two-term limits.

Finally, the model also has implications for the likelihood of reelection in the first term under different term-limit rules. In particular, reelection in the first term is more likely, the less stringent the term limits. The intuition is as follows. In both two-term-limit and no-term-limit systems, if the parameter values are such that first-term politicians exert a high level of effort in equilibrium, then high-ability types are reelected with probability γ and low-ability types are not reelected. If the parameter values are such that first-term politicians do not exert a high level of effort in equilibrium, then high-ability types are reelected with probability λ (i.e., if they happen to get the good outcome) and low ability types are still never reelected. Since $\gamma > \lambda$, reelection is more likely when the equilibrium calls for a high level of effort, since this better facilitates separation. Since first-term candidates exert a high level of effort for a larger set of parameter values with no term limits than with two-term limits, the absence of term limits increases the probability of reelection.

Proposition 7. *For all parameter values, there exists a MPE of the game with no term limits such that the probability of reelection of a first-term politician is weakly higher than the probability of reelection of a first-term politician under the two-term limit. Moreover, there exist parameter values such that this relationship is strict rather than weak.*

Incumbent Experience

Up to this point, we have ignored a potentially important aspect of differences between first- and second-term incumbents: politicians may learn on the job, becoming better at providing public goods over time. That is, second-term incumbents might be of higher expected quality than first-term incumbents not only because voters have weeded out the low types, but also because second-term incumbents are more experienced.

There are two natural ways to formalize incumbent learning in our environment. First, we might let λ and γ increase with tenure, so that more experienced politicians are more likely than less experienced politicians to achieve good outcomes, regardless of whether they exert low or high effort. Second, we might think that the difference between the good outcome and the bad outcome increases with tenure, so that politicians can provide better outcomes as experience increases.

Notice that neither extension changes Propositions 1-5 or 7 (or their proofs) in any qualitative way, because experience operates in the same direction as selection. However, when we test empirically for the presence of a selection effect, we will need to distinguish it from an experience effect. We return to this point in the next section.

The experience extension does suggest that our model might underestimate the expected output of second-term incumbents discussed in Proposition 6. However, since this would be the case for both types of term-limit systems, the difference-in-differences would remain negative. That is, in our model, relative to first-term performance, second-term incumbents without term limits are expected to outperform second term incumbents with term limits, even if incumbents learn from experience. In sum, the difference-in-differences relationship is not overturned by

adding incumbent learning to the model, nor does such a prediction arise from a model with learning but without the choice of effort included in our model.

3. Data and Methodology

To evaluate our model, we use panel data from the American states.¹¹ Unless otherwise noted, the sample period is 1950-1996.¹² All monetary values are reported in constant 1982 dollars. We adopt the specifications of Besley and Case (2003) and Besley (2006) to show how our results build on and modify theirs.

Dependent Variables

In the model presented in the previous section, the dependent variable that voters value and incumbents provide is 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. However, measuring the provision of government services is notoriously difficult.

Many political agency models, such as those of Besley and Case (1995) and Persson and Tabellini (2000), suggest that low taxes and spending are public goods. They interpret a small budget as reflecting effort, which assumes that incumbents who exert effort must provide relatively more goods at lower cost, as measured by the levels of spending and taxes. Although these measures of public goods are not perfect, we want to respond to and build on the existing literature, so we include these variables in our empirical analysis.¹³

¹¹ We thank Tim Besley for generously sharing the data from Besley and Case (2003). The data set has been amended to include some variables from additional sources, as noted below.

¹² This sample period avoids confounding the effect of gubernatorial term limits with that of legislative term limits, which first began to take effect in 1998.

¹³ Throughout our empirical analysis, we control for per capita state income and the partisan composition of government—two of the factors most frequently associated with the deliberate expansion of public spending and taxes—as well as state and year fixed effects.

When comparing performance under one- and two-term limits, our task is complicated by the upward trend in spending and taxes. Observations in which governors served under a two-term limit come later in the sample than those serving under a one-term limit; thus, even if term limits have no effect, using levels of spending and taxes as the dependent variable might generate the result that two-term limits are associated with higher spending and taxes than one-term limits. Therefore, rather than using levels, we use the percentage change in per capita spending and taxes in the regressions that compare one- and two-term limits.

We also use two additional dependent variables.¹⁴ The first is the state's financial condition, as measured by its borrowing costs. Comparable panel data on state bond interest rates are not available, so we use relative yield spreads on 20-year general obligation debt.¹⁵ A higher yield spread increases the cost of borrowing, which can reflect rising debt, poor economic performance, mismanagement, corruption, and other factors that bear on the state's ability to pay (Lowry and Alt 2001). Since higher interest costs translate into higher general fund expenditures that provide no public benefits, they represent poor performance that could affect re-election prospects.

Our final dependent variable is the state's economic growth rate. Studies of gubernatorial elections and job approval ratings suggest that voters value and respond to growth (Lowry, Alt, and Ferree 1998). However, because growth is clearly not completely within the control of governors, we use the difference between the state's economic growth rate and the national

¹⁴ In addition to these two measures, we examined another dependent variable, the log of the number of corruption convictions per 1000 public employees, as reported by the DoJ Public Integrity Section. The results (not shown) had expected signs, but no estimate substantially exceeded its standard error. We also considered the possibility of using approval ratings as a measure of incumbent performance, but data on approval (e.g. Beyle et al. 2002) do not go back far enough to encompass changes from one-term to two-term limits.

¹⁵ This Chubb Relative Value Survey generates biennial expert opinions of the difference in yields between each other state and, arbitrarily, New Jersey. We thank Jim Poterba and Kim Rueben for supplying the data. The data are only available for states that issue 10-year general obligation debt, and only go back to 1973, so the sample period begins in that year for the borrowing cost regressions. This enables us to include in the regression a few additional states that adopted two-term limits between 1950 and 1973 (see Table 1 for a list of these states).

average. In addition to removing the national trend, this approach is also consistent with previous findings that voters evaluate their governors' economic performance relative to the performance of the national economy (see for example Wolfers 2002).¹⁶ Data on state economic growth are from the U.S. Department of Commerce's Bureau of Economic Analysis. Descriptive statistics for all four dependent variables appear in Table 3.

[Table 3 about here]

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 two different subsets of states. The first set of tests examines the states that changed from one- to two-term limits. The second set of tests examines the states that either had two-term limits or no term limits during the sample period.¹⁷

Our first specification examines the relative performance of governors operating under one- and two-term limits. Recall that Proposition 4 predicts that a first-term governor who is eligible for reelection should perform better than a first-term lame duck because both have the same expected ability but the former exerts a higher level of effort. Proposition 4 also predicts that a second-term lame duck should perform better than a first-term lame duck because both exert the same level of effort but the former is, in expectation, of higher ability.

¹⁶ Since we include year effects in all regressions, we are already controlling for average effects across all states in each year; thus, using the state's growth rate or the difference between a state's growth rate and the national average yields the same results.

¹⁷ We do not compare states before and after the switch from no term limits to two-term limits because most states without term limits had two-year terms prior to the adoption of term limits, making it difficult to compare them to states with four-year terms.

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:¹⁸

$$performance_{it} = \alpha_1 first-term-eligible_{it} + \alpha_2 second-term-lame-duck_{it} + \alpha_3 X_{it} + \theta_i + \delta_t + v_{it}$$

where performance is as defined above (alternately: spending growth, tax growth, borrowing costs, and economic growth), X_{it} are control variables (described below), θ_i are state fixed effects, δ_t are year effects, and v_{it} is an error term. The omitted category is first-term lame ducks. It follows that α_1 , the coefficient on *first-term-eligible*, will provide an estimate of the accountability effect, since it compares first-term governors who are eligible for reelection to first-term lame ducks, while α_2 , the coefficient on *second-term-lame-duck*, will provide an estimate of the selection effect, since it compares second-term lame ducks to first-term lame ducks. Since performance is expected to be better under both of these types of governors than under first-term lame ducks, we expect both of these coefficients to be negative for spending, taxes, and borrowing costs, which are all “bads,” and positive for economic growth, which is a “good.”

We use the same controls as Besley and Case (1995, 2003); these are variables that might be expected to affect economic policy outcomes, namely, population size, state income, the proportions of aged and young in the population, and indicators relating to party control of government (Democratic governor, Democratic house, Democratic senate, and divided government).¹⁹ In addition, we control for the governor’s years of political experience prior to becoming governor.

¹⁸ Although a total of fourteen states abandoned one-term limits, we exclude Kentucky and Mississippi because they abandoned one-term limits so close to the end of the sample period (in 1994).

¹⁹ 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 therefore retain.

As mentioned in the previous section, it is possible that second-term lame ducks perform better than first-term lame ducks as a result of not only selection but also experience on the job. Ideally, we would be able isolate these two effects by separating first-term lame ducks according to whether or not they would have won reelection to a second term had they not been term limited. Evidence that second-term lame ducks perform better than not only first-term losers but also first-term winners would indicate a learning effect. However, we cannot make this comparison because first-term lame ducks do not have a chance to run for reelection so we do not observe which would have won and which would have lost.²⁰ However, we can isolate the two effects when comparing two-term limits to no term limits, as explained below.

Our second regression model evaluates Proposition 5, which compares outcomes under two-term limits and no term limits. This specification essentially modifies and extends the one used by Besley and Case (1995; 2003). Recall that their specification potentially attributes both selection effects (resulting from the switch from one- to two-term limits in many states) and changes in the sample (resulting from the adoption of two-term limits by states that previously did not have any term limits) to incumbent shirking.

To get around the latter problem, we include in our sample only those states that had either no term limits or two-term limits during the sample period.²¹ For the sake of direct comparison, we also exclude governors serving two- rather than four-year terms and governors serving their third (or higher) terms in office.²² That is, if a governor in a state without term

²⁰ It is possible to separate first-term *eligible* governors into winners and losers, but our model does not yield any predictions about how these governors compare to second-term lame ducks, even in the absence of incumbent learning.

²¹ As shown in Table 1, many states adopted two-term limits in the late 1950s and early 1960s. We experimented with several start dates and subsets of states and found similar results. We report results for the start date of 1959. The states in the sample are listed in the second and third columns of Table 2.

²² Currently New Hampshire is the only state with two-year terms, but many other states had two-year terms earlier in the sample period.

limits served three terms, we include in our sample his first and second terms but not his third term.

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 selection from accountability. Proposition 5 predicts that second-term governors who were eligible for reelection will perform better than first-term governors who were eligible for reelection (the selection effect) and that second-term governors who were eligible for reelection will perform better than second-term lame ducks (the accountability effect).

To test these hypotheses, we run the following regression:

$$performance_{it} = \beta_1 first-term-eligible_{it} + \beta_2 second-term-lame-duck_{it} + \beta_3 X_{it} + \theta_i + \delta_t + v_{it}$$

The omitted category is second-term governors who are eligible for reelection. Because the omitted category is different than in the previous specification, the interpretation of coefficients is also different. The coefficient on *first-term-eligible* (β_1) now provides an estimate of the selection effect, since it compares first- to second-term governors who are eligible for reelection, while the coefficient on *second-term-lame-duck* (β_2) now provides an estimate of the accountability effect, since it compares second-term lame ducks to second-term governors who are eligible for reelection. Since performance is expected to be worse under both of these types of governors than under second-term eligibles, we expect the coefficients will be of the opposite signs as in the previous regression: positive for spending, taxes, and borrowing costs, which are all “bads,” and negative for economic growth, which is a “good.” The control variables are the same as in the previous specification and, as before, we control for state and year effects.

As before, the selection effect is potentially confounded by an experience effect. We can modify the regression model to isolate the two effects by separating first-term governors who

went on to win reelection from those who did not. Evidence that second-term governors perform better than first-term winners would indicate an experience effect that has nothing to do with the weeding-out process. Evidence that second-term governors outperform first-term losers by more than they outperform first-term winners (that is, that losers are significantly worse than winners) would suggest a selection effect.

We can also modify this specification to test Proposition 6, which predicts that, among governors who were reelected to a second term, second-term minus first-term performance is lower in states with term limits than in states without term limits as a result of the accountability effect. Testing this proposition entails separating the first-term winners according to whether or not they served in a state with term limits. Since the omitted category is second-term governors in states without term limits, the coefficient on *first-term-winner,0TL* will provide an estimate of the difference between first- and second-term performance in states without term limits. An F-test comparing this coefficient to the difference between the coefficients on *first-term-winner,2TL* and *second-term-lame-duck* will reveal whether the latter difference is less than the former difference, as predicted by Proposition 6.²³

4. Results

One-Term Limits vs. Two-Term Limits

Table 4 tests Proposition 4 by comparing first-term lame ducks to first-term governors who can run for reelection and second-term lame ducks. Heteroskedasticity-consistent robust

²³ Due to the fairly small number of governors who lost or did not run, we do not distinguish them according to the term limit regime in order to avoid perfect multicollinearity. A difference-in-differences test would drop the losers from the sample and simply compare the first and second terms of two-term governors. However, throwing out the losers would result in the loss of nearly one-third of our observations, in large part because several states have identical patterns over time among two-term governors. It would also create an unbalanced panel. Nonetheless, compared to the results reported in the paper, the results are qualitatively similar (but in some cases lose statistical significance) when we drop the losers from the sample.

standard errors are reported in parentheses. The first column shows that, as predicted, per capita spending growth is about 0.2 percentage points lower under both first-term governors who are eligible for reelection and second-term lame ducks than under first-term lame ducks, supporting the accountability and selection hypotheses, respectively. While term limits do not appear to affect per capita tax growth (the coefficients are close to zero and statistically insignificant), borrowing costs are about 10 basis points lower under both first-term governors who are eligible for reelection and second-term lame ducks, compared to first-term lame ducks. As with spending growth, the selection and accountability effects appear to be of approximately the same magnitude.

[Table 4 about here]

As shown in the final column, the economic growth rate is 0.7 percentage points higher under first-term governors who are eligible for reelection than under first-term lame ducks, supporting the accountability hypothesis. The positive coefficient on “second-term lame duck” (the selection effect) goes in the correct direction but falls short of statistical significance at conventional levels.

Institutional Endogeneity

Any analysis of the policy effects of institutions raises inevitable concerns about institutional endogeneity. In the case of the switch from one- to two-term limits, the direction of the potential bias is not clear, *a priori*. On the one hand, a state might abandon one-term limits as a result of disillusionment with especially poor performance under the current lame-duck incumbent, which would bias in favor of our hypothesis that performance is worse under one-term limits than under two-term limits. On the other hand, the relaxation of term limit laws could be motivated by the desire to allow a highly competent politician to run for reelection, which would bias against our hypothesis. The fact that the majority of states that switched from one- to

two-term limits designed their laws to allow the current incumbent to run for reelection suggests that the bias is most likely working against our hypothesis, meaning that our results provide a lower bound on the true magnitude of the effect of one-term limits.

To address the potential bias resulting from institutional endogeneity, we conducted a robustness test in which we omitted from the sample the incumbent who was in office at the time of the policy change in each state. In the spending and borrowing-cost regressions, the coefficients and standard errors get slightly larger, leaving the significance levels virtually unchanged. In the growth regressions, the coefficient on first-term-eligible remains the same but the standard error gets larger, so that the result is only statistically significant at about the 20 percent level.²⁴

Two-Term Limits vs. No Term Limits

Table 5 tests Propositions 5 and 6 by comparing second-term governors who are eligible for reelection to first-term governors who are eligible for reelection and second-term lame ducks. Recall that because the omitted category is different than in the previous table, the expected signs are different, and the coefficient on first-term eligible now gives the selection effect while the coefficient on second-term lame duck gives the accountability effect.

[Table 5 about here]

Column (1) reveals that per capita spending in states without term limits is approximately \$41 higher under first-term governors than under second-term governors, which is consistent with the selection hypothesis. In addition, per capita spending is approximately \$45 higher under

²⁴ In case of policy lags, we also ran a robustness test in which we omitted not only the current incumbent at the time of the policy change but also the previous incumbent; this robustness test yields similar results, although in the spending regressions the standard errors rise enough to make the results statistically insignificant, which is perhaps not surprising since omitting eight years per state results in a much smaller sample size and, in particular, cuts the number of one-term-limit observations by about half.

second-term lame ducks than under second-term governors who are eligible for reelection, consistent with the accountability hypothesis. Similarly, per capita taxes are \$25 higher under first-term governors than under second-term governors within states without term limits, and \$17 higher under second-term lame ducks than under second-term governors who are eligible for reelection.

These results shed light on Besley and Case's (2003) puzzling finding that the effect of binding term limits on spending and taxes has disappeared over time. Like Besley and Case, we find an accountability effect: namely, that term-limited governors spend and tax more than governors who are eligible for reelection. However, we also find that second-term governors spend and tax less than first-term governors. Moreover, these effects are large enough to offset the accountability effect. Thus, our results suggest that as the states gradually switched from one- to two-term limits over the past fifty years, the fact that term-limited governors were increasingly second- rather than first-term incumbents made it appear as though the impact of term limits was declining to zero.

Column (1) also shows that borrowing costs are approximately 9 basis points higher under first-term eligible governors than under second-term eligible governors, supporting the selection hypothesis, and are about 11 basis points higher under second-term lame ducks than under second-term eligible governors, supporting the accountability hypothesis. In addition, the negative signs on the coefficients in the growth regression are consistent with the selection and accountability hypotheses, but both fall short of statistical significance at conventional levels.

Column (2) attempts to disentangle selection from experience by separating the first-term eligible governors who went on to win reelection from those who did not. The results are consistent with the presence of incumbent learning: second-term governors perform better than

first-term winners for spending, taxes, and borrowing costs, but not for growth. As predicted by the selection hypothesis, however, first-term losers are generally “more worse” than first-term winners when compared to second-term governors, and F-tests reveal that these differences are statistically significant for taxes and growth. However, for borrowing costs the opposite appears to be true.

Column (3) separates first-term winners in states with and without term limits in order to test Proposition 6, which predicts that, among governors who won reelection to a second term, the difference between second- and first-term performance in states with term limits is less than the difference between second- and first-term performance in states without term limits, provided that ε is small enough. The difference between the coefficients on *second-term lame duck* and *first-term winner, 2TL* is smaller than the coefficient on *first-term winner, 0TL* in the spending, tax, and borrowing-cost regressions; however, in the growth regression, both differences are close to zero. The F-test at the bottom of the table indicates that the difference in differences is statistically significant for borrowing costs but falls short of significance for the other variables.

Column (3) also provides additional support for the experience hypothesis. In states without term limits, second-term governors appear to perform significantly better than first-term winners when it comes to taxes and borrowing costs, as indicated by the positive and statistically significant coefficients on *first-term winner, 0TL*. Moreover, in states with two-term limits, first-term winners are statistically indistinguishable from second-term lame ducks. These findings are consistent with the hypothesis that incumbents learn from experience in office, and that term limits offset this effect by removing the incentive to exert effort in the final term.

Probability of Reelection

Our final set of tests relates to Proposition 7, which predicts that the first-term probability of reelection is higher under no term limits than under two-term limits. A simple difference of means reveals that the first-term probability of reelection is 0.67 in states without term limits, compared to 0.61 in states with two-term limits, but this difference is not statistically significant. However, the switch from no term limits to two-term limits does appear to be associated with a statistically significant reduction in the first-term probability of winning reelection. Limiting the sample to states that adopted two-term limits during the sample period, and excluding observations in which governors served two-year terms, the first-term probability of winning reelection was 0.74 before the adoption of term limits and 0.60 after the adoption of two-term limits. This difference is statistically significant at the five percent level, and is robust to the inclusion of year effects to account for changes in the probability of reelection over time.

5. Conclusions

The question of whether elections serve as a mechanism of accountability or selection is central to understanding the nature of representative democracy. This paper has attempted to shed light on this question by developing and testing a model of term limits that produces conjectures about the effects of accountability and selection in elections. We have found evidence of both effects.

First, we have found that, controlling for tenure in office, performance is better under governors who can run for reelection than under term-limited governors, suggesting that term limits remove the disciplining effect of elections. Specifically, state economic growth (relative to the national average) is higher, the growth of per capita government spending is lower, and state borrowing costs are lower under first-term governors who are eligible for reelection than under

first-term lame ducks. Similarly, taxes, spending, and borrowing costs are lower under second-term governors who are eligible for reelection than under second-term lame ducks.

Second, we have found that performance is better under governors who have survived reelection than under first-term governors, suggesting that elections allow voters to weed out low-quality incumbents—and that term limits interfere with this mechanism. Specifically, among governors who are eligible to run for reelection, per capita taxes, per capita spending, and borrowing costs are lower under second-term governors than under first-term governors, and growth is higher among governors who survive reelection than among those who do not. Similarly, among lame ducks, borrowing costs are lower and per capita spending grows less rapidly under second-term governors than under first-term governors.

In addition, we have found evidence consistent with incumbent learning. A comparison of the first and second terms of governors who won reelection to a second term in office indicates that economic performance improves with the incumbent's experience in office.

Previous research on term limits has focused primarily on the problem of accountability, but in addition to concerns about shirking, a common argument in the normative debate on term limits is that term limits interfere with voters' ability to choose the best-qualified representatives. Indeed, "the argument against term limits that seems to be the most compelling to the U.S. public is definitely not that term limits would make politicians unaccountable, but rather that term limits arbitrarily restrict the free choice of who can be one's representative" (Fearon 1999, 62). Our results show that concerns about selection are relevant, and that the length, not just the existence, of term limits matters. In short, the potential advantages of term limits addressed elsewhere in the literature—such as limiting careerism and the incumbency advantage—must be weighed against

the disadvantages, which include not only reduced accountability but also reduced scope for both incumbent learning and voter selection of competent politicians.

Appendix

Proof of Proposition 2. The argument in the text establishes that second-term effort is always \underline{a} .

Thus, second term expected output is $\mu\lambda$.

Next, consider the voter's behavior, characterized in the following lemma.

Lemma 1. *In a two-term limit system, the voter reelects an incumbent following an outcome H and does not reelect following an outcome of L .*

Proof of Lemma 1. Suppose the voter expects high effort from a newly elected politician. Then, the voter wants to reelect the incumbent when his beliefs are μ if

$$\mu\lambda \geq \pi\gamma.$$

Since, the voter's beliefs after H are $1-\varepsilon$, and $\frac{(1-\varepsilon)\lambda}{\gamma} > \pi$, the voter will reelect following H .

Since the voter's beliefs after L are $\frac{(1-\gamma)\pi(1-\varepsilon)}{(1-\gamma)\pi+1-\pi} < \pi$, the voter does not reelect following L .

Now suppose the voter expects low effort from a newly elected politician. Then, the voter wants to reelect the incumbent when his beliefs are μ if

$$\mu\lambda \geq \pi\lambda.$$

Clearly, the voter will reelect following H , since his beliefs are $1-\varepsilon > \pi$. Following L , the voter's beliefs are

$$\frac{(1-\lambda)\pi(1-\varepsilon)}{(1-\lambda)\pi+1-\pi} < \pi,$$

so the voter will not reelect.

Now consider first term effort.

Case 1: $c \geq \delta B\pi(\gamma - \lambda)$

Suppose the voter expects the politician to choose \bar{a} in the first term. Then the payoff to the politician of doing so is

$$B - c + \delta B(\pi\gamma r(1 - \varepsilon) + (1 - \pi)r(0)).$$

The payoff to the politician of choosing \underline{a} is

$$B + \delta B(\pi(\lambda r(1 - \varepsilon) + (1 - \lambda)r(0)) + (1 - \pi)r(0)).$$

According to Lemma 1, we have $r(1 - \varepsilon) = 1$ and $r(0) = 0$. Substituting, the politician will choose high effort in the first period if

$$\delta B\pi(\gamma - \lambda) \geq c$$

which is confirmed by the hypothesis of this case. Thus, expected first term output is $\pi\gamma$. Since only types who succeed will be reelected, beliefs about a second term incumbent are always $\mu = 1 - \varepsilon$. Since second-term incumbents will not exert effort, expected output is $(1 - \varepsilon)\lambda$.

This establishes existence, to see uniqueness for the strict inequality, consider the alternative. Suppose the voter expects \underline{a} . The payoff to the politician of choosing \underline{a} is:

$$B + \delta B(\pi\lambda r(1 - \varepsilon) + (\pi(1 - \lambda) + (1 - \pi))r(\mu(L | \underline{a}, \pi))).$$

The payoff of deviating to \bar{a} is

$$B - c + \delta B(\pi\gamma r(1 - \varepsilon) + (\pi(1 - \gamma) + (1 - \pi))r(\mu(L | \underline{a}, \pi))).$$

From Lemma 1, we have $r(1 - \varepsilon) = 1$ and $r(\mu(L | \underline{a}, \pi)) = 0$. This implies that the politician will choose low effort in the first period if

$$\delta B\pi(\gamma - \lambda) \leq c.$$

Thus, there is no equilibrium with low effort in this case when the inequality defining the case is strict.

Case 2: $c \leq \delta B\pi(\gamma - \lambda)$

An argument identical to the one directly above establishes that, in this case, there is an equilibrium where first period effort is \underline{a} and that there is no equilibrium where first period effort is \bar{a} , when the inequality is strict. Hence, in this case, first term expected output is $\pi\lambda$. Again, by Lemma 1, only high types are reelected, so second term expected output is still $(1-\varepsilon)\lambda$.

Proof of Proposition 3. The proof is by construction. Consider the following strategy profile:

$$r(\mu) = \begin{cases} 1 & \text{if } \mu \geq \pi \\ 0 & \text{else.} \end{cases}$$

$$a^*(\mu) = \begin{cases} \bar{a} & \text{if } \mu \geq \hat{\mu} \\ \underline{a} & \text{else.} \end{cases}$$

A politician who exerts high effort in a period where the beliefs about his type being high are μ and the voter expects high effort receives the following payoff:

$$B - c + \delta[\mu\gamma r(1-\varepsilon)V_{1-\varepsilon} + (\mu(1-\gamma) + (1-\mu))r(0)V_0],$$

where V_μ represents the continuation value given beliefs μ . If the politician chooses low effort instead, her payoff is:

$$B + \delta[\mu\lambda r(1-\varepsilon)V_{1-\varepsilon} + (\mu(1-\lambda) + (1-\mu))r(0)V_0].$$

Rearranging and substituting in the proposed voting rule shows that a politician of type μ will exert high effort when it is expected by the voter if and only if:

$$\mu\delta(\gamma - \lambda)V_{1-\varepsilon} \geq c. \tag{1}$$

Define $\hat{\mu} \equiv \frac{c}{\delta(\gamma-\lambda)V_{1-\varepsilon}}$. To see that $\hat{\mu} < \pi$, notice that we have $\delta B\pi(\gamma-\lambda) + \eta \geq c$ for some positive η , by hypothesis. Since $V_{1-\varepsilon}$ is strictly greater than B , this hypothesis implies that $\pi\delta(\gamma-\lambda)V_{1-\varepsilon} > c$. Thus, there is a $\mu < \pi$ that satisfies the inequality in (1), and the left hand side is clearly increasing in μ . Since the $\hat{\mu}$ we have chosen just satisfies inequality (1), it must be that $\hat{\mu} < \pi$.

Next consider a politician of type $\mu_0 < \hat{\mu} < \pi$ who is expected to exert low effort. If she exerts low effort, her payoff is:

$$B + \delta[\mu_0\lambda r(1-\varepsilon)V_{1-\varepsilon} + \mu_0(1-\lambda)r(\mu(L|\underline{a}, \mu_0))V_{\mu(L|\underline{a}, \mu_0)} + (1-\mu_0)r(\mu(L|\underline{a}, \mu_0))V_{\mu(L|\underline{a}, \mu_0)}].$$

If the politician chooses high effort, her expected payoff is:

$$B - c + \delta[\mu_0\gamma r(1-\varepsilon)V_{1-\varepsilon} + (1-\mu_0)r(\mu(L|\underline{a}, \mu_0))V_{\mu(L|\underline{a}, \mu_0)}].$$

Rearranging and substituting the reelection rule, the politician will choose \underline{a} if

$$c \geq \delta\mu_0(\gamma-\lambda)V_{1-\varepsilon}.$$

We have chosen $\hat{\mu}$ such that this condition holds for all $\mu < \hat{\mu}$. Thus, all politicians of type $\mu < \hat{\mu}$ choose \underline{a} .

Finally we need to show that the voting rule is a best response. The expected payoff from not reelecting a politician is simply $\pi\gamma$. The payoff from reelecting a politician of type $\mu \geq \pi$ is $\mu\gamma \geq \pi\gamma$. The supremum of the payoff from reelecting a politician of type $\mu < \pi$ is $\mu\gamma < \pi\gamma$. Thus, the voting rule is a best reply.

Given this, effort is high in the first term, yielding an expected output of $\pi\gamma$. Only high types are reelected, so effort is also high in the second term, yielding an expected output of $(1-\varepsilon)\gamma$.

Proof of Proposition 4. The proof is immediate from a comparison of Propositions 1 and 2.

Proof of Proposition 5. Define $\hat{\eta} \equiv \pi\delta^2(\gamma - \lambda)V_{1-\varepsilon}$, where $V_{1-\varepsilon}$ is the continuation value from the equilibrium in Proposition 3. Now there are three cases to consider.

Case 1: $c < \delta B\pi(\gamma - \lambda)$

In this case, according to Proposition 2, expected output in the first term of a two-term system is $\pi\gamma$ and in the second term of a two-term system is $(1 - \varepsilon)\lambda$. According to Proposition 3, there is an MPE of the game with no term limits where expected output in the first term is $\pi\gamma$ and expected output in the second term is $(1 - \varepsilon)\gamma$. Thus, expected output is the same in the first term and strictly higher in the second term with no term limits.

Case 2: $c \in [\delta B\pi(\gamma - \lambda), \delta B\pi(\gamma - \lambda) + \hat{\eta}]$

According to Proposition 2, expected output in the first term of a two-term system is $\pi\lambda$ and in the second term is $(1 - \varepsilon)\lambda$. Notice that $\hat{\eta}$ is chosen to just satisfy inequality (1) evaluated at $\mu = \pi$. Proposition 3 then implies that expected output in the first term with no term limits is $\pi\gamma$ and expected output in the second term is $(1 - \varepsilon)\gamma$. Thus, expected output is strictly higher in both the first and second terms with no term limits.

Case 3: $c > \delta B\pi(\gamma - \lambda) + \hat{\eta}$

According to Proposition 2 there is low effort in both terms. Thus, expected output in any MPE of the no term limit game must be at least as high.

Proof of Proposition 6. In both two term limit and no term limit system, the expected first-term output of a politician, conditioning on reelection, is γ . The expected second-term output of a politician reelected in a two term limit system is $(1-\varepsilon)\lambda$, whereas in a no term limit system it is $(1-\varepsilon)\gamma$. Thus, for any ε the difference is expected to be negative in a two term limit system. And, as ε goes to 0, the expected difference in a no term limit system goes to 0 as well.

Proof of Proposition 7. The probability of reelection, given that the equilibrium calls for a high level of effort, is simply π . The probability of reelection, given that the equilibrium calls for a low level of effort is $\pi\lambda$. As shown in Proposition 2, there is high effort in the first term with two term limits only if $c \leq \delta B\pi(\gamma - \lambda)$. With no term limits, the restriction for high effort in the first term is weaker. Thus, when $c \leq \delta B\pi(\gamma - \lambda)$, the probability of reelection is the same. However, there is an open set of parameter values such that there is high effort with no term limits but not with two term limits, and in these cases the probability of reelection is higher with no term limits. For cases where high effort cannot be sustained in either case, the probability of reelection must again be weakly higher with no term limits, since it cannot be lower than the probability with low effort.

References

- 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. 2007. "Electoral Selection, Strategic Challenger Entry, 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.
- Banks, Jeffrey S., and Rangarajan K. Sundaram. 1993. "Moral Hazard and Adverse Selection in a Model of Repeated Elections," in Barnett William A., Melvin J. Hinich, and Norman J. Schofield, eds. *Political Economy: Institutions, Information, Competition, and Representation*. Cambridge: Cambridge University Press.
- Barro, Robert. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(1): 19-42.
- 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. 2006. "Term Limits in the State Legislatures: Results from a New Survey of the 50 States." *Legislative Studies Quarterly*, 31: 105-136.
- 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.
- Figlio, David N. 1995. "The Effect of Retirement on Political Shirking: Evidence from Congressional Voting." *Public Finance Review* 23(2): 226-241.
- Gowrisankaran, Gautam, Matthew F. Mitchell, and Andrea Moro. 2006. "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: 169-82.
- 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.
- Kurtz, Karl T, Bruce Cain, and Richard G. Niemi, eds. 2007. *Institutional Change in American Politics: The Case of Term Limits*. Ann Arbor: University of Michigan Press.

- Lott, John R. and Stephen G. Bronars. 1993. "Time Series Evidence on Shirking in the U.S. House of Representatives." *Public Choice* 76: 125-149.
- Lowry, Robert C., and James Alt. 2001. "A Visible Hand? Bond Markets, Political Parties, Balanced Budget Laws, and State Government Debt." *Economics and Politics* 13:1 (March): 49-72.
- 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.
- McArthur, John, and Steven V. Marks. 1988. "Constituent Interest vs. Legislator Ideology: The Role of Political Opportunity Cost." *Economic Inquiry* 26(3): 461-470.
- Persson, Torsten, and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge: MIT Press.
- Snyder, James M., and Michael M. Ting. 2003. "Roll Calls, Party Labels and Elections." *Political Analysis* 11(4): 419-444.
- Tien, Charles. 2001. "Representatives, Voluntary Retirement, and Shirking in the Last Term." *Public Choice* 106:117-130.
- Vanbeek, James R. 1991. "Does the Decision to Retire Increase the Amount of Political Shirking?" *Public Finance Review* 19: 444-456.
- Wolfers, Justin. 2002. "Are Voters Rational? Evidence from Gubernatorial Elections." Stanford GSB Working Paper #1730.
- 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 Since 1950

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 1-term limit to 2-term limit (year of switch)	AL (1968), FL (1968), GA (1976), IN ^c (1972), KY (1994), LA (1966), MS (1994), MO (1965), NM, ^d NC ^c (1977), OK (1966), PA (1967), SC (1980), TN (1978), ^e WV (1970)
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), OH (1958), RI (1994), SD (1956), WY ^f (1992)

^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 IN and NC allow a third term after a four-year hiatus.

^d NM had a 2-term limit during the sample period except for 1972-1990, when it had a 1-term limit.

^e TN adopted the 1-term limit in 1955.

^f MT and WY restrict Governors to two terms or eight years in 16.

Table 2. First-Term Winners vs. Second-Term Incumbents under Different Term Limit Regimes

	First-Term Output (Winners)	Expected Second- Term Output	Difference between Second and First Terms
Two Term Limit	1	$(1 - \varepsilon)\lambda$	$(1 - \varepsilon)\lambda - 1 < 0$
No Term Limit	1	$(1 - \varepsilon)\gamma$	$(1 - \varepsilon)\gamma - 1 < 0$
Difference in Differences			$(1 - \varepsilon)(\lambda - \gamma) < 0$

Table 3. Summary Statistics

	States that switched from 1- to 2-term limits (1)	States with no term limits (2)	States with 2-term limits (3)
Dependent Variables:			
Per capita spending	896 (404)	1071 (515)	1061 (543)
Per capita spending growth	0.038 (0.084)	0.038 (0.071)	0.040 (0.080)
Per capita taxes	448 (195)	555 (268)	497 (258)
Per capita tax growth	0.037 (0.075)	0.039 (0.075)	0.042 (0.095)
Borrowing costs	4.74 (24.36)	16.29 (28.10)	12.51 (22.21)
Economic growth	0.024 (0.032)	0.020 (0.031)	0.020 (0.032)
Independent Variables:			
First-term lame duck	0.44 (0.50)	- -	- -
First-term eligible	0.36 (0.48)	0.51 (0.50)	0.64 (0.48)
Second-term lame duck	0.20 (0.40)	- -	0.36 (0.48)
Second-term eligible	- -	0.49 (0.50)	- -
Control Variables:			
State income per capita	8665 (2791)	10632 (3310)	11075 (2993)
Population ('000)	4917 (2797)	6147 (6084)	4764 (3525)
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)
Gov's yrs of prior experience	9.3 (7.5)	11.1 (7.4)	11.5 (7.5)
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 4. One-Term Limits vs. Two-Term Limits

Dependent variable:	Per capita spending growth	Per capita tax growth	Borrowing costs	Economic growth
Expected signs:	-	-	-	+
	(1)	(2)	(3)	(4)
First-term eligible (Accountability effect)	-0.021** (0.011)	0.005 (0.005)	-9.91** (4.35)	0.007* (0.004)
Second-term lame duck (Selection/experience effect)	-0.024** (0.011)	0.005 (0.014)	-10.35** (4.36)	0.004 (0.004)
Observations	520	520	242	520
R-squared	0.20	0.37	0.86	0.26

Note: The omitted category is first-term lame ducks. Controls include state income, income squared, income cubed, state population, population squared, population cubed, percent elderly, percent school-aged, dummies for Democratic governor, Democratic house, Democratic senate, and divided government, and the governor's years of prior political experience. All regressions include state and year effects. Robust standard errors in parentheses.

* Significant at 10% level.

** Significant at 5% level.

Table 5. Two-Term Limits vs. No Term Limits

Dependent variable:	Per capita spending			Per capita taxes			Borrowing costs			Economic growth		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Expected signs:	+			+			+			-		
Second-term lame duck (Accountability effect)	45.57** (21.29)	44.94** (21.26)	102.90** (28.24)	17.36* (10.56)	16.10 (10.50)	23.79 (17.64)	10.75** (3.89)	11.53** (3.92)	16.43** (5.44)	-0.003 (0.004)	-0.003 (0.004)	-0.007 (0.006)
First-term eligible (Selection/experience effect)	41.53** (16.32)			25.00** (8.06)			9.07** (3.07)			-0.004 (0.003)		
First-term loser (Selection effect)	49.55** (19.78)		81.98** (21.44)	41.27** (11.81)		45.57** (14.66)	5.49 (3.67)		7.40* (4.05)	-0.008* (0.004)		-0.010** (0.005)
First-term winner (Experience effect)	38.05** (16.77)			17.96** (7.96)			11.05** (3.29)			-0.002 (0.003)		
First-term winner, 0TL (Proposition 6)			26.14 (17.22)			16.38* (8.40)			10.23** (3.32)			-0.001 (0.003)
First-term winner, 2TL (Proposition 6)			107.72** (27.68)			27.21 (18.07)			17.05** (5.61)			-0.007 (0.006)
Observations	366	366	366	366	366	366	325	325	325	366	366	366
R-squared	0.97	0.97	0.97	0.96	0.96	0.96	0.66	0.66	0.67	0.34	0.35	0.35
F-test: First-term loser = First-term winner	0.56			5.01**			3.43*			2.73*		
F-test: First-term winner, 0TL = First-term winner, 2TL - Second-term lame duck			0.88			1.27			5.36**			0.05

Note: The omitted category is second-term governors who are eligible for reelection. Controls include state income, income squared, income cubed, state population, population squared, population cubed, percent elderly, percent school-aged, dummies for Democratic governor, Democratic house, Democratic senate, divided government, and the governor's years of prior political experience. All regressions include state and year effects.

Robust standard errors in parentheses.

* Significant at 10% level.

** Significant at 5% level.